MPLS Application, Services & Best Practices for Deployment

Monique Morrow (mmorrow@cisco.com)

Martin Winter (mwinter@cisco.com)

APRICOT 2009 MANILA

Manila, 26th February 2009

House Rules

- Please put your mobile phones into silent mode.
- Kindly do not take calls inside of this room while the session is going on.
- Your feedback on the session is extremely important!
- We assume that you will be awake and keep us awake as well ☺

Session Agenda

- MPLS Layer 3 VPN
- MPLS Traffic Engineering
- MPLS Layer 2 VPN
- Q&A

MPLS Layer 3 VPN

Agenda

- MPLS VPN Explained
- MPLS VPN Services
- Best Practices
- Conclusion

Prerequisites

- Must understand basic IP routing, especially BGP
- Must understand MPLS basics (push, pop, swap, label stacking)

Terminology

- LSR: Label switch router
- LSP: Label switched path

    The chain of labels that are swapped at each hop to get from one LSR to another
- VRF: VPN routing and forwarding

    Mechanism in Cisco IOS® used to build per-interface RIB and FIB
- MP-BGP: Multiprotocol BGP
- PE: Provider edge router interfaces with CE routers
- P: Provider (core) router, without knowledge of VPN
- VPNv4: Address family used in BGP to carry MPLS-VPN routes
- RD: Route distinguisher

    Distinguish same network/mask prefix in different VRFs
- RT: Route target

    Extended community attribute used to control import and export policies of VPN routes
- LFIB: Label forwarding information base
- FIB: Forwarding information base

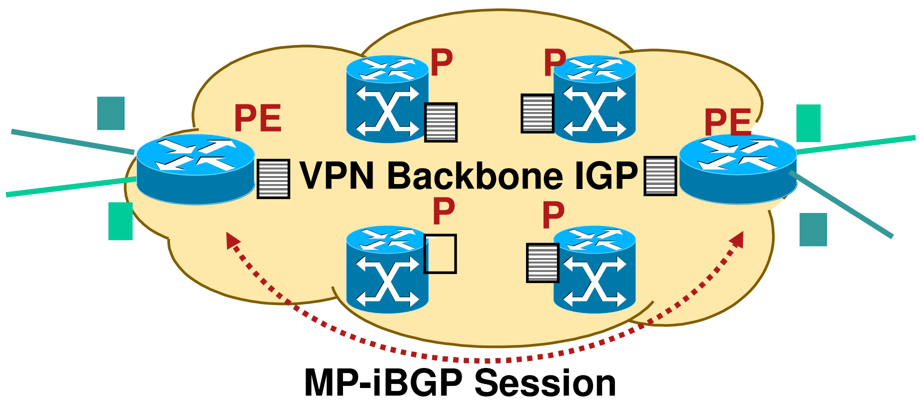
Agenda

- MPLS VPN Explained
    - Technology
- MPLS-VPN Services
- Best Practices
- Conclusion

MPLS-VPN Technology

- Control plane—VPN route propagation
- Data plane—VPN packet forwarding

MPLS-VPN Technology
MPLS VPN Connection Model



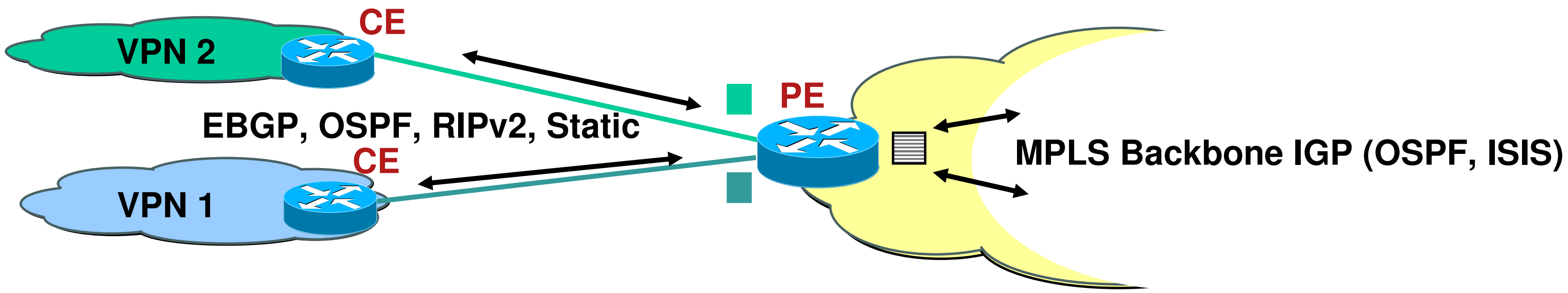
PE Routers

- Edge routers
- Use MPLS with P routers
- Uses IP with CE routers
- Connects to both CE and P routers Distribute VPN information through MP-BGP to other PE router with VPN-IPv4 addresses, extended community, label

P Routers

- P routers are in the core of the MPLS cloud
- P routers do not need to run BGP and doesn't need to have any VPN knowledge
- Forward packets by looking at labels
- P and PE routers share a common IGP

MPLS-VPN Technology
Separate Routing Tables at PE



VRF Routing Table

- Routing (RIB) and forwarding table (CEF) associated with one or more directly connected sites (CEs)
- The routes the PE receives from CE routers are installed in the appropriate VRF routing table(s)

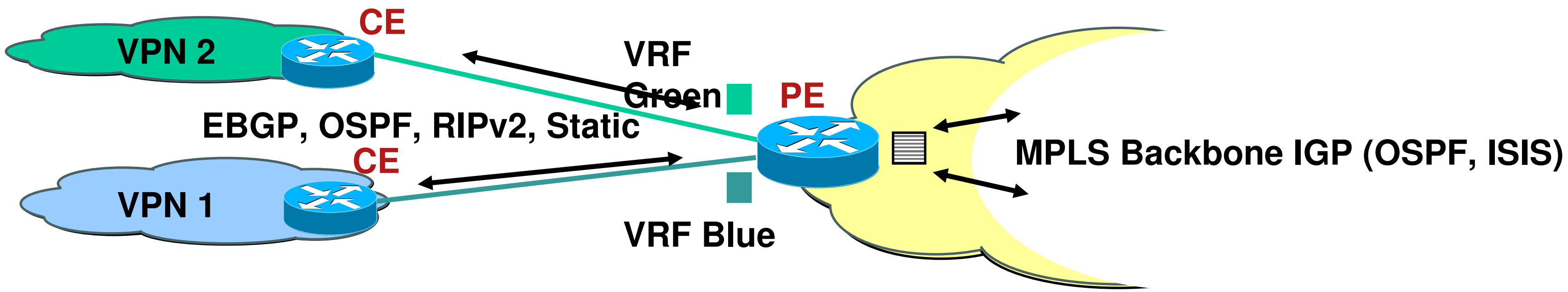
  blue VRF routing table or

  green VRF routing table

The Global Routing Table

- Populated by the IGP within MPLS backbone

MPLS-VPN Technology
Virtual Routing and Forwarding Instance (1)



- What's a VRF ?
- Associates to one or more interfaces on PE

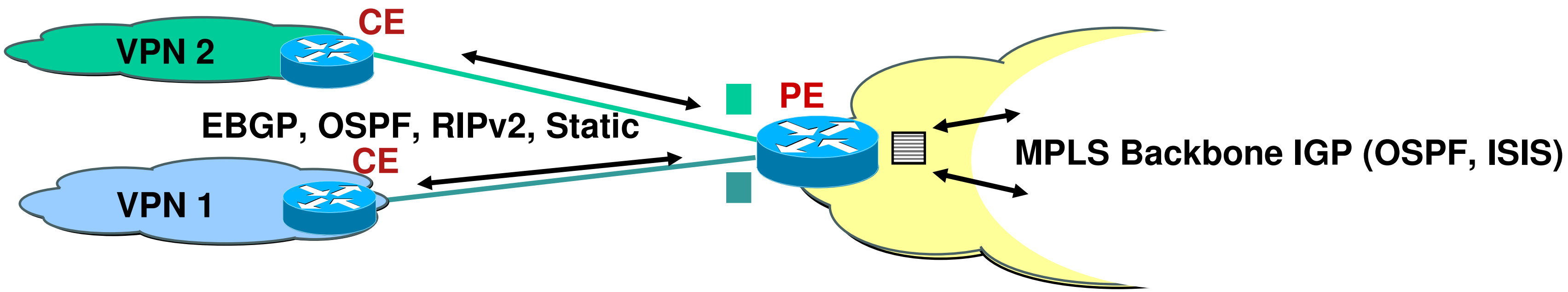
    Privatize an interface i.e., coloring of the interface

- Has its own routing table and forwarding table (CEF)
- VRF has its own instance for the routing protocol

    (static, RIP, BGP, EIGRP, OSPF)

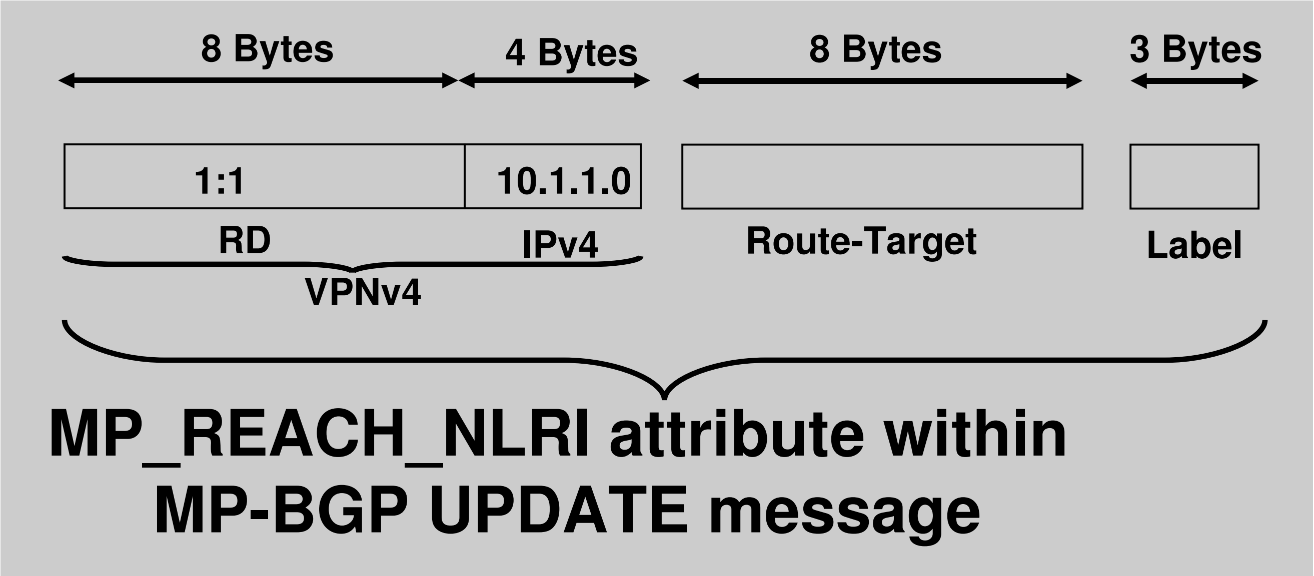
- CE router runs standard routing software

MPLS-VPN Technology
Virtual Routing and Forwarding Instance (2)



- PE installs the routes, learned from CE routers, in the appropriate VRF routing table(s)
- PE installs the IGP (backbone) routes in the global routing table
- VPN customers can use overlapping IP addresses

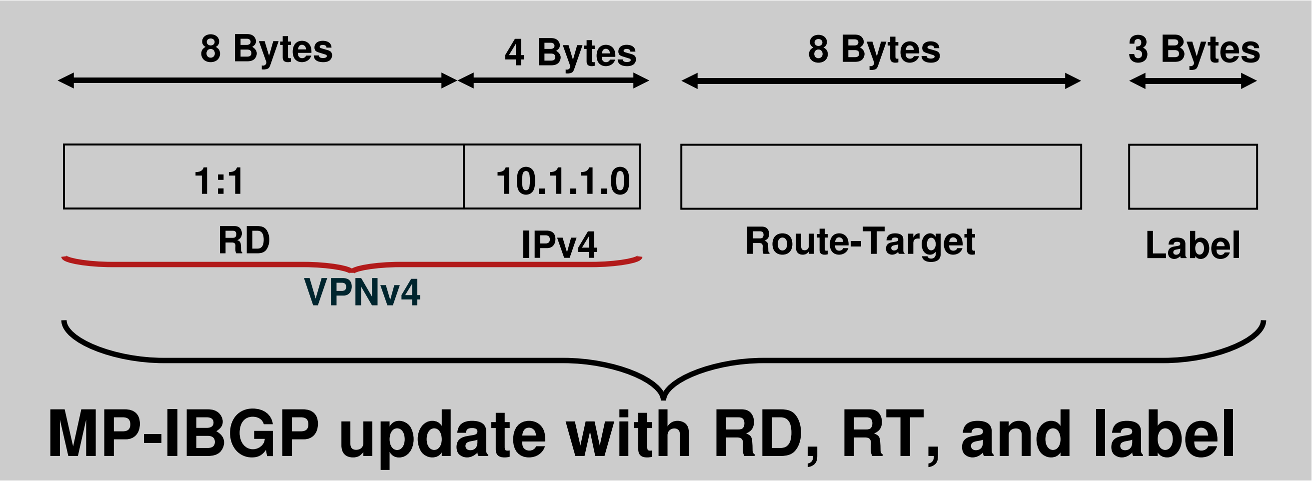
MPLS-VPN Technology:
Control Plane



Let's Discuss:

- Route Distinguisher (RD); VPNv4 route
- Route Target (RT)
- Label

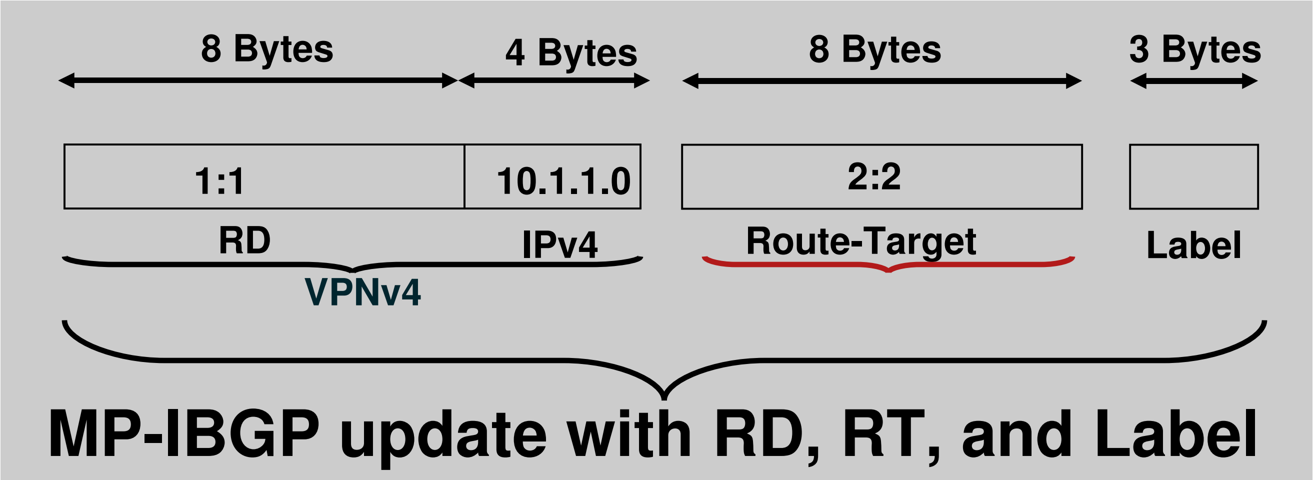
MPLS VPN Control Plane
MP-BGP Update Components: VPNv4 Address



- To convert an IPv4 address into a VPNv4 address, RD is appended to the IPv4 address i.e. 1:1:10.1.1.0
  - Makes the customer's IPv4 route globally unique
- Each VRF must be configured with an RD at the PE
  - RD is what that defines the VRF

```
!
ip vrf v1
 rd 1:1
!
```

MPLS VPN Control Plane
MP-BGP Update Components: Route-Target

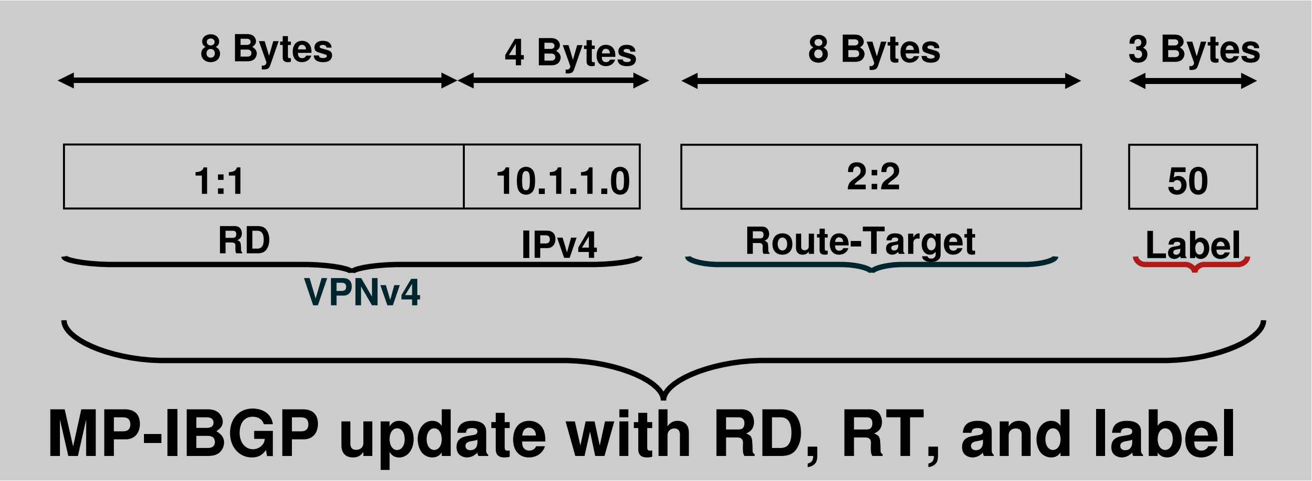


- Route-target (RT): Identifies the VRF for the received VPNv4 prefix. It is an 8-byte extended community (a BGP attribute)
- Each VRF is configured with RT(s) at the PE

    RT helps to color the prefix

```
!
ip vrf v1
 route-target import 1:1
 route-target export 1:2
!
```

MPLS VPN Control Plane
MP-BGP Update Components: Label



- The Label (for the VPNv4 prefix) is assigned only by the PE whose address is the next-hop attribute

  PE routers rewrite the next-hop with their own address (loopback)

  "Next-hop-self" towards MP-iBGP neighbors by default

- PE addresses used as BGP next-hop must be uniquely known in the backbone IGP

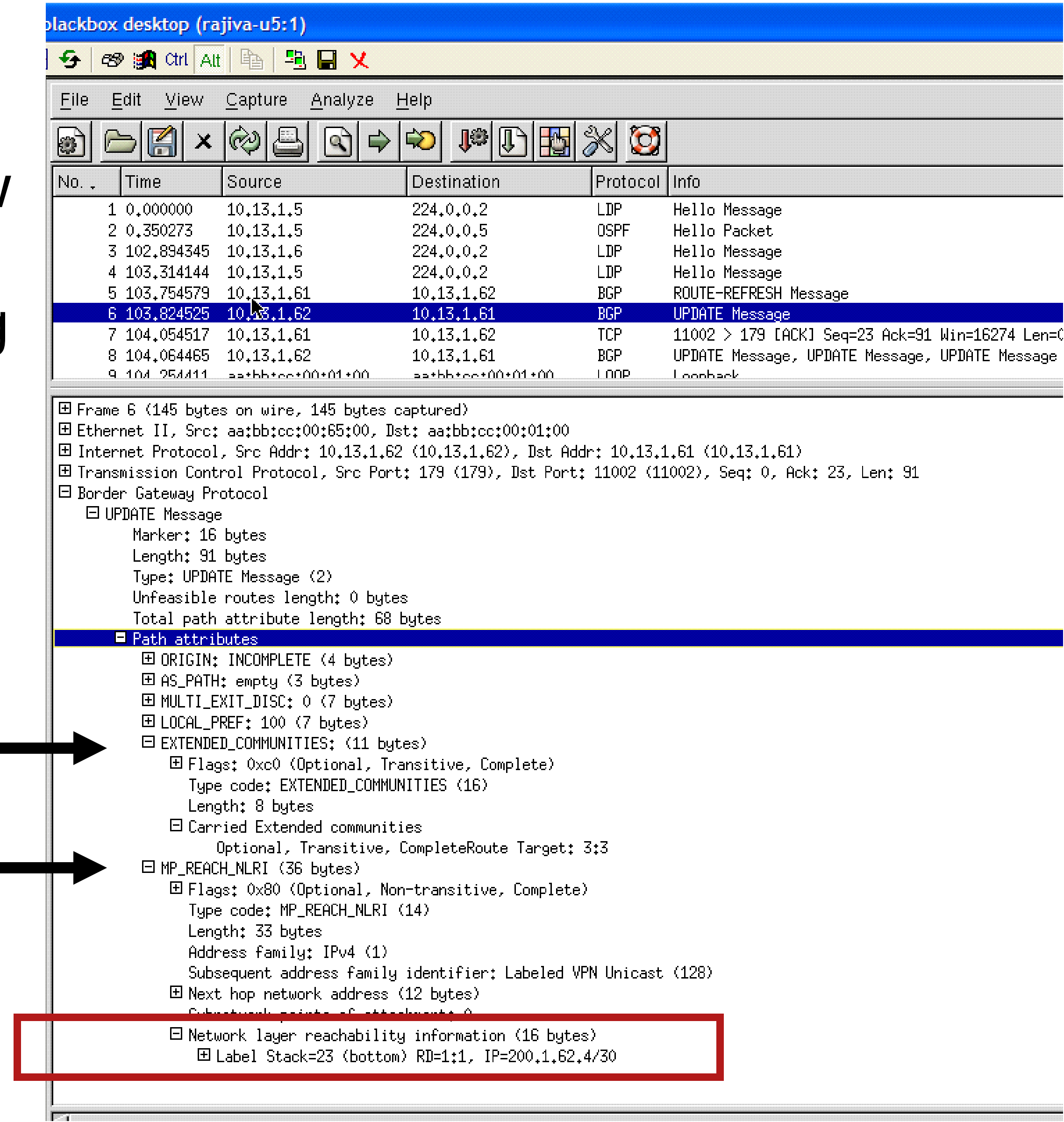
  Do Not Summarize the PE Loopback Addresses in the Core

MPLS-VPN Technology: Control Plane
MP-BGP UPDATE Message Capture

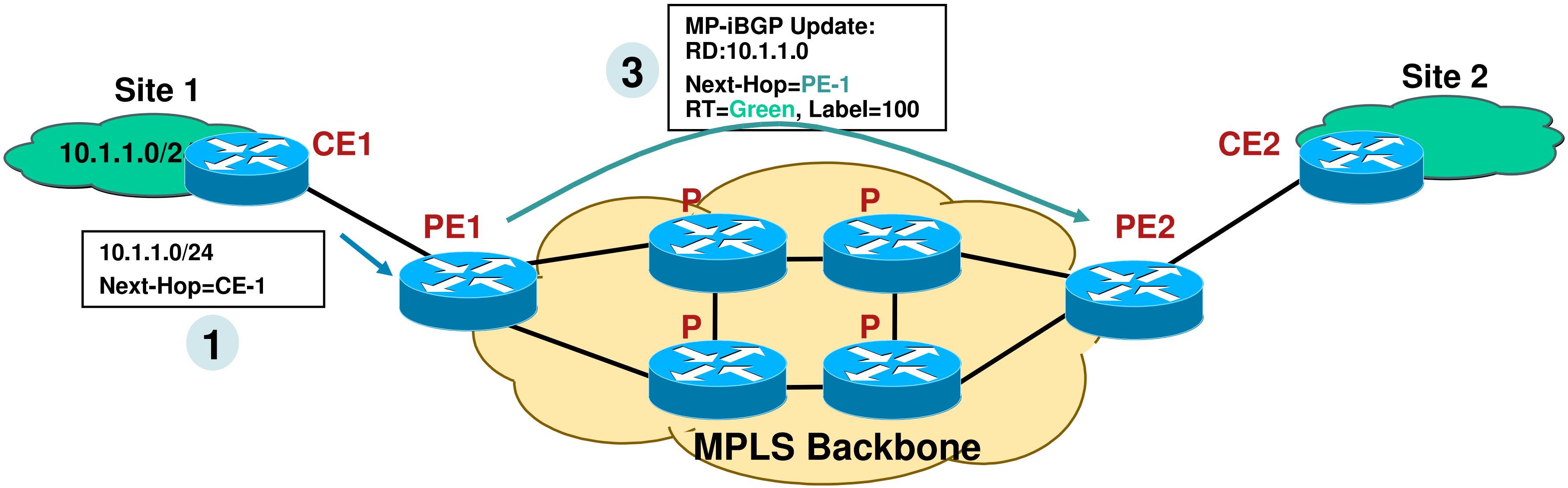
- This capture might help to visualize how the BGP UPDATE message advertising VPNv4 routes look like.
- Notice the Path Attributes.

Route Target 3:3

MP_REACH_NLRI 1:1:200.1.62.4/30

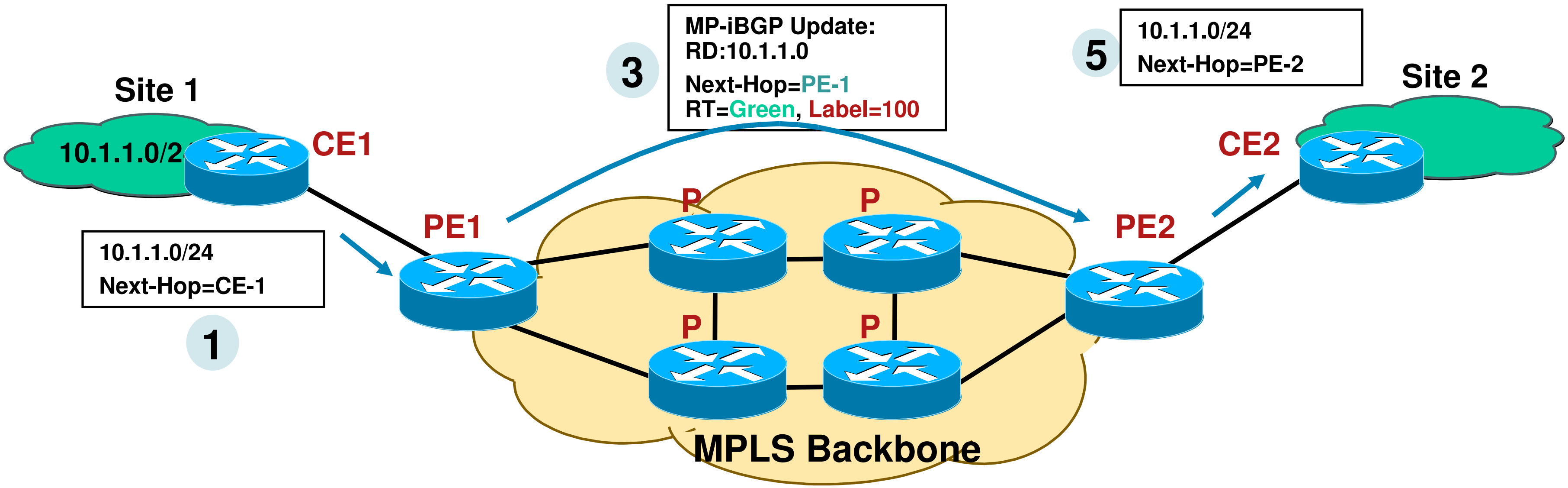


MPLS VPN Control Plane:
Putting It All Together



1. PE1 receives an IPv4 update (eBGP,OSPF,EIGRP)
2. PE1 translates it into VPNv4 address
   - Assigns an RT per VRF configuration
   - Rewrites next-hop attribute to itself
   - Assigns a label based on VRF and/or interface
3. PE1 sends MP-iBGP update to other PE routers

MPLS VPN Control Plane:
Putting It All Together



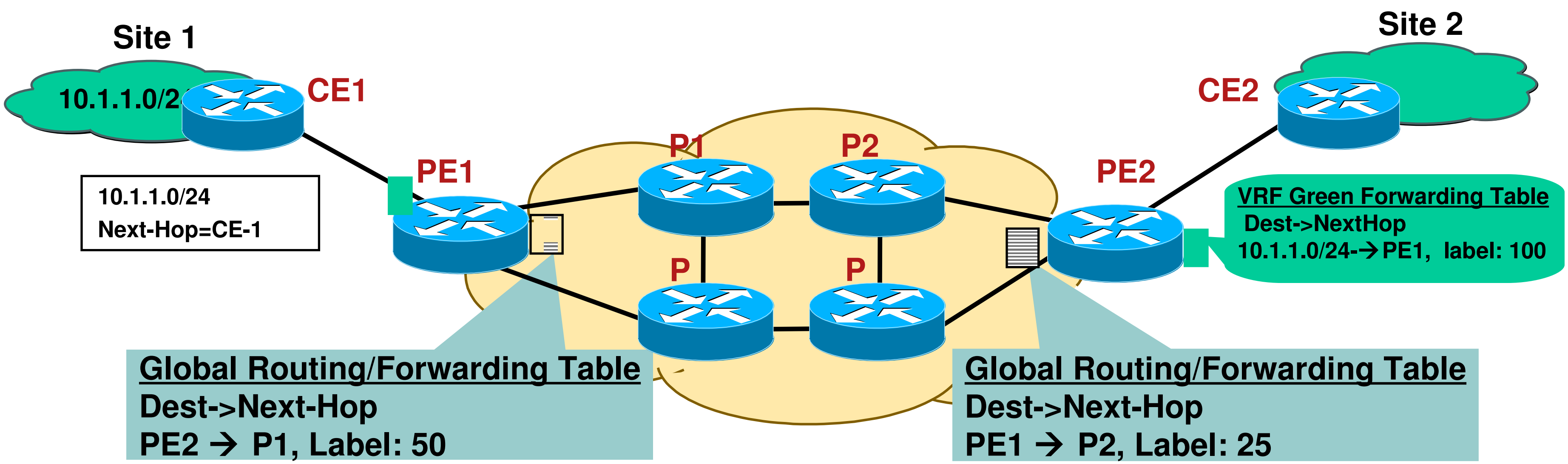
1. PE2 receives and checks whether the RT=green is locally configured within any VRF, if yes, then
2. PE2 translates VPNv4 prefix back into IPv4 prefix,

   Installs the prefix into the VRF routing table

   Updates the VRF CEF table with label=100 for 10.1.1.0/24

   Advertise this IPv4 prefix to CE2 (EBGP, OSPF, EIGRP)

MPLS-VPN Technology:
Forwarding Plane



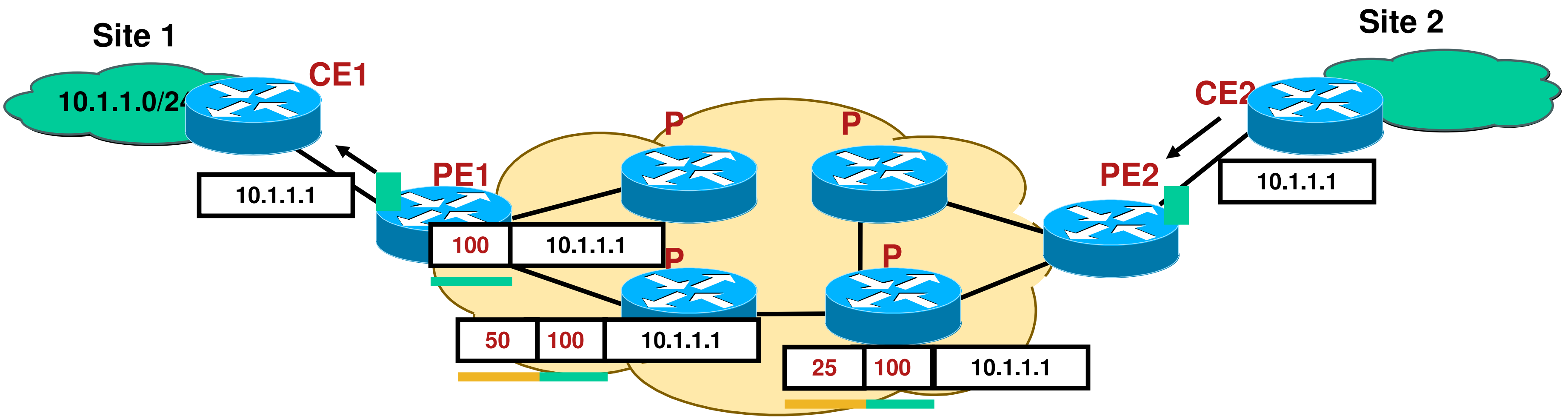
The Global Forwarding Table (show ip cef)

- PE routers store IGP routes
- Associated labels
- Label distributed through LDP/TDP

VRF Forwarding Table (show ip cef vrf <vrf>)

- PE routers store VPN routes
- Associated labels
- Labels distributed through MP-BGP

MPLS-VPN Technology:
Forwarding Plane



- PE2 imposes TWO labels for each packet going to the VPN destination 10.1.1.1
- The top label is LDP learned and derived from an IGP route
  - Represents LSP to PE address (exit point of a VPN route)
- The second label is learned via MP-BGP
  - Corresponds to the VPN address

MPLS-VPN Technology: Control Plane
MPLS Packet Capture

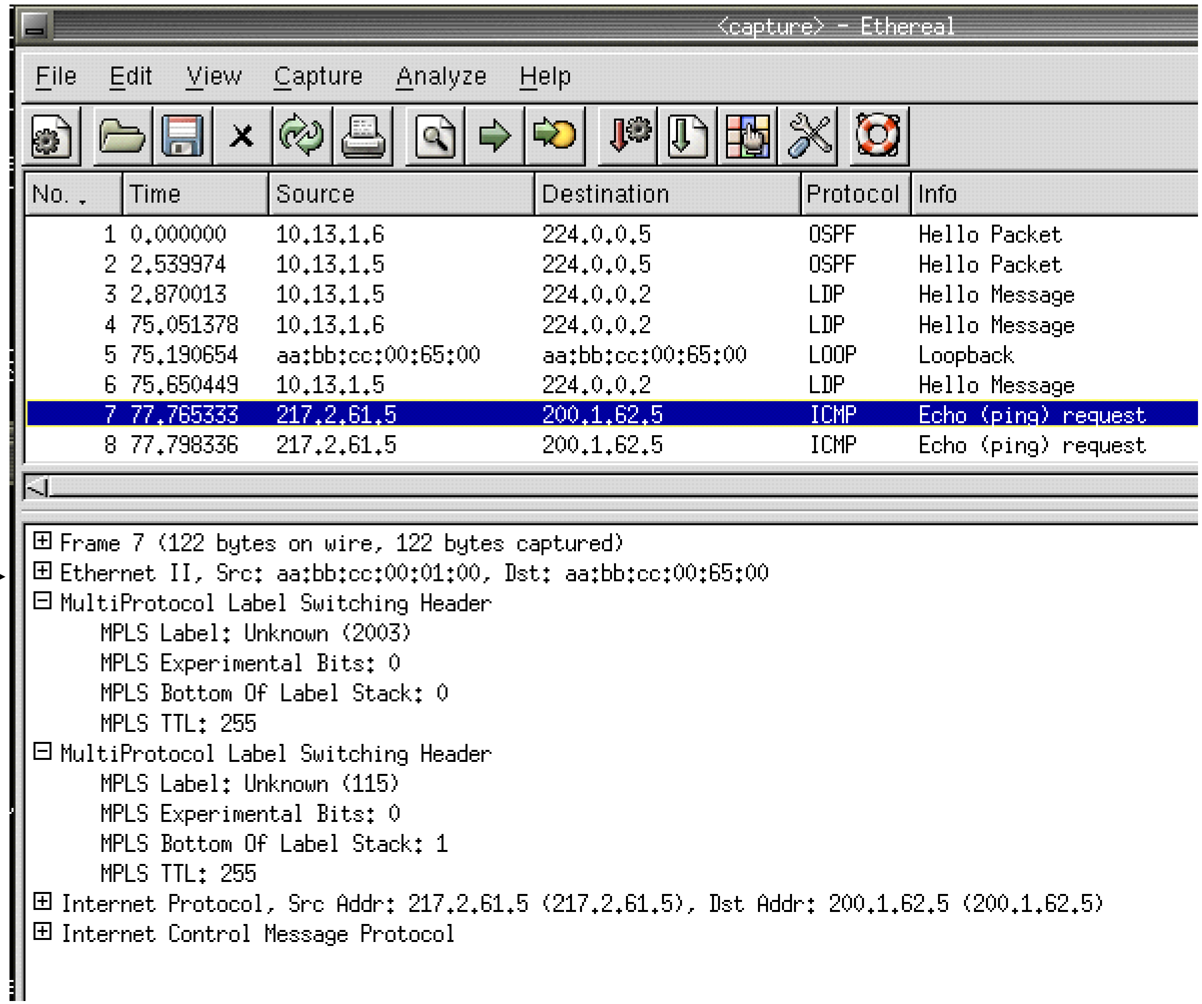
- This capture might be helpful if you never captured an MPLS packet before.

Ethernet Header

Outer Label

Inner Label

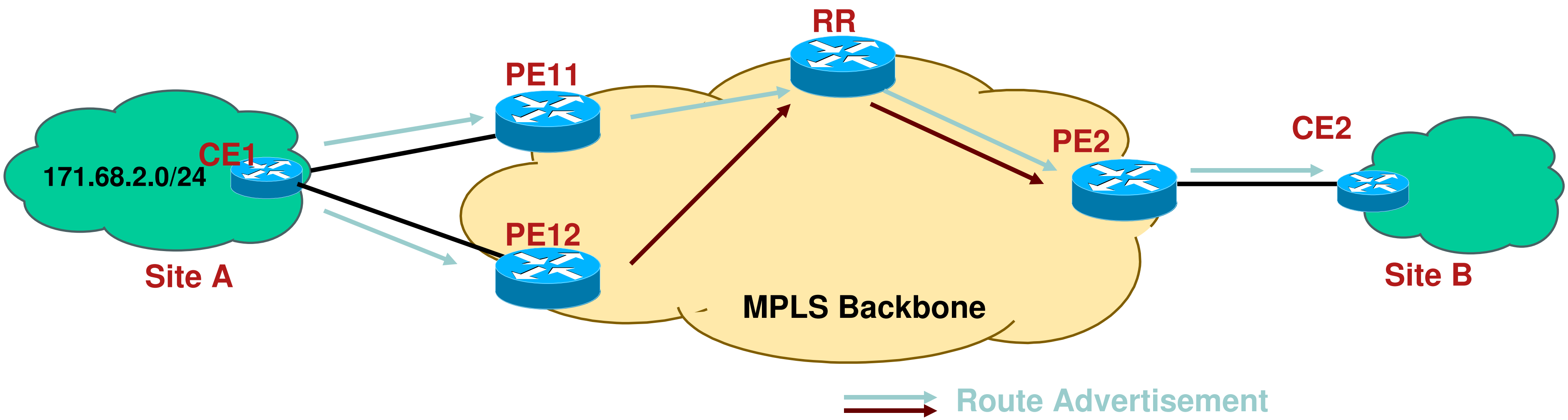
IP packet



Agenda

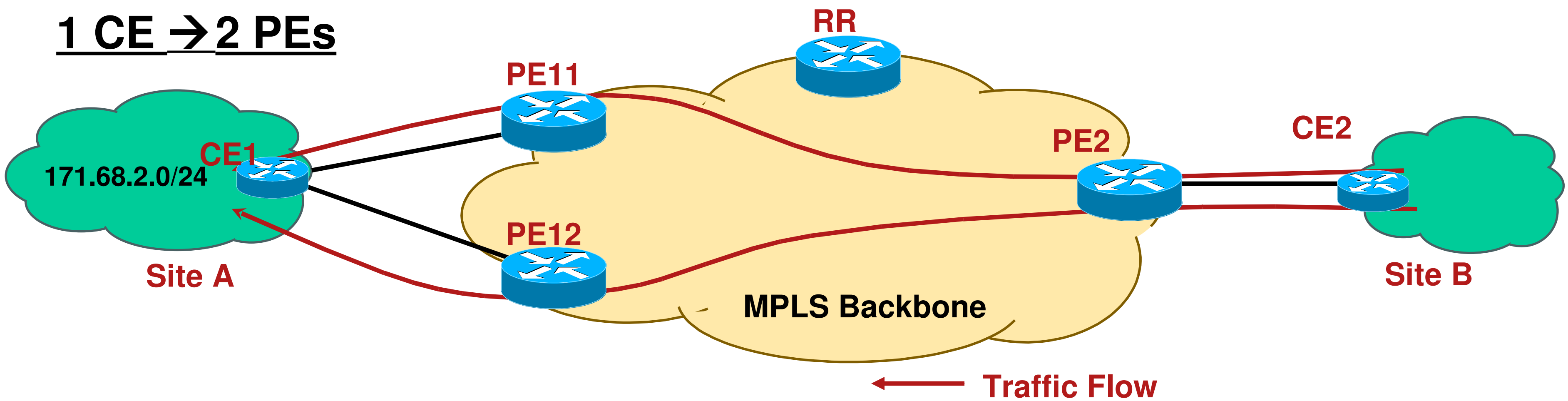
- MPLS VPN Explained
- MPLS-VPN Services
  1. Providing Load-Shared Traffic to the Multihomed VPN Sites
  2. Providing Hub and Spoke Service to the VPN Customers
  3. Providing Internet Access Service to VPN Customers
  4. Providing VRF-Selection Based Services
  5. Providing Remote Access MPLS VPN
  6. Providing VRF-Aware NAT Services
  7. Providing MPLS VPN over IP Transport & Multi-VRF CE Services
- Best Practices
- Conclusion

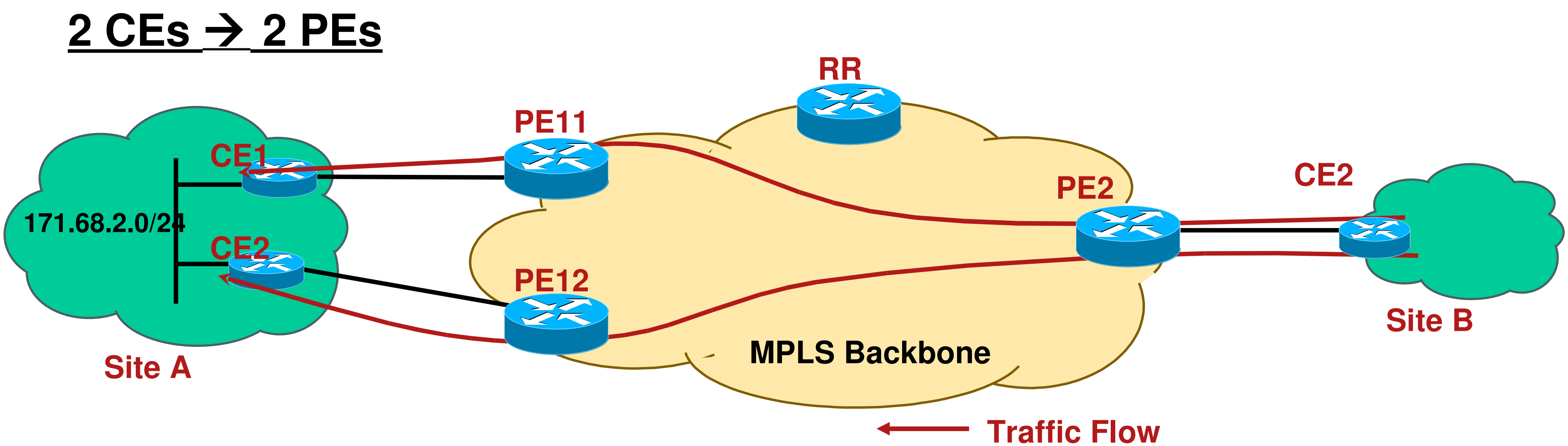
MPLS VPN Services:
1. Loadsharing for the VPN Traffic



- VPN sites (such as Site A) could be multihomed
- VPN customer may demand the traffic (to the multihomed site) be loadshared

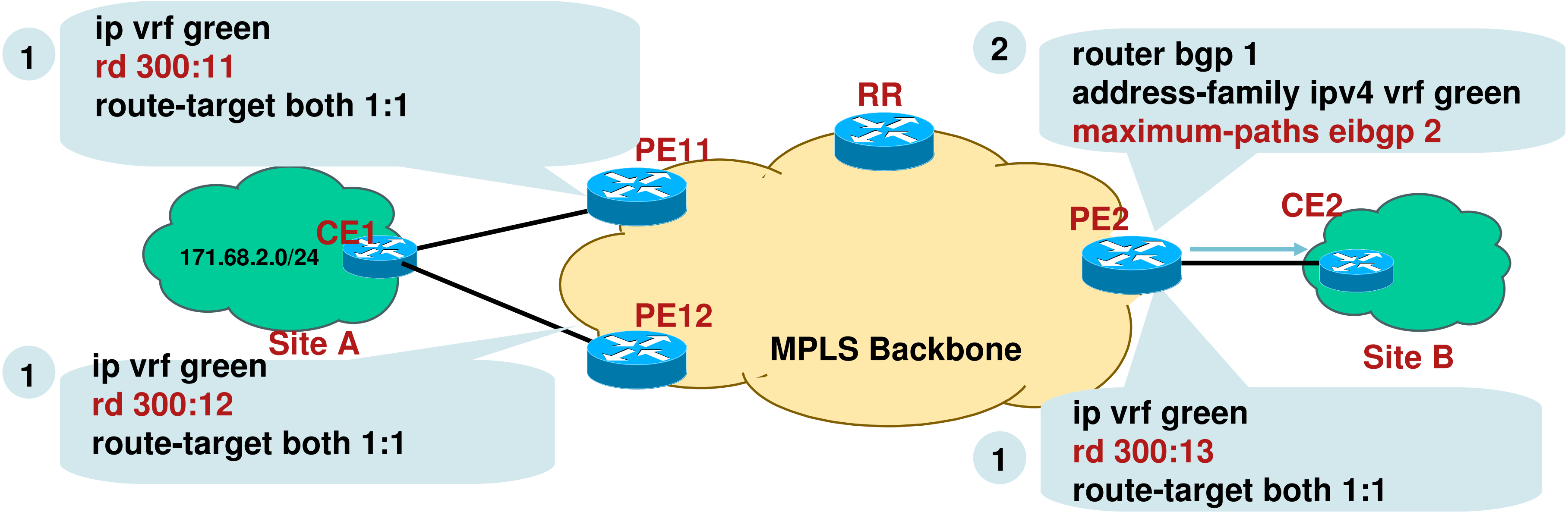
MPLS VPN Services:
1. Loadsharing for the VPN Traffic: Cases



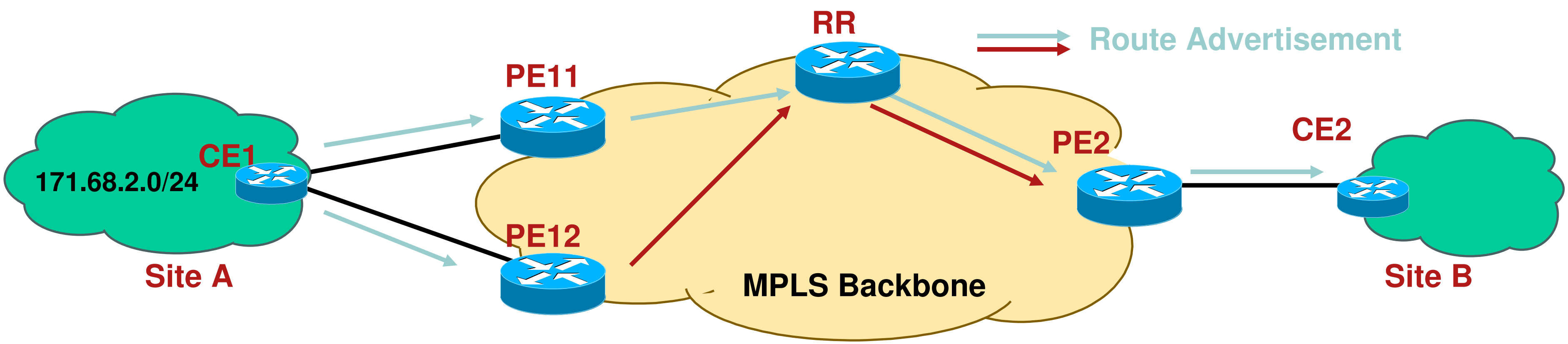


MPLS VPN Services:
1. Loadsharing for the VPN Traffic: Deployment

- How to deploy the loadsharing?
- Configure unique RD per VRF per PE for multihomed site/interfaces
- Enable BGP multipath within the relevant BGP VRF address-family at remote/receiving PE2 (why PE2?)



MPLS VPN Services:
1. Loadsharing for the VPN Traffic



- RR must advertise all the paths learned via PE11 and PE12 to the remote PE routers

Please note that without ‘unique RD per VRF per PE’, RR would advertise only one of the received paths for 171.68.2.0/24 to other PEs. ☹

- Watch out for the increased memory consumption (within BGP) due to multipaths at the PEs
- “eiBGP multipath” implicitly provides both eBGP and iBGP multipath for VPN paths

Agenda

- MPLS VPN Explained
- MPLS-VPN Services
  1. Providing Load-Shared Traffic to the Multihomed VPN Sites
  2. Providing Hub and Spoke Service to the VPN Customers
  3. Providing Internet Access Service to VPN Customers
  4. Providing VRF-Selection Based Services
  5. Providing Remote Access MPLS VPN
  6. Providing VRF-Aware NAT Services
  7. Providing MPLS VPN over IP Transport & Multi-VRF CE Services
- Best Practices
- Conclusion

MPLS-VPN Services:
2. Hub and Spoke Service to the VPN Customers

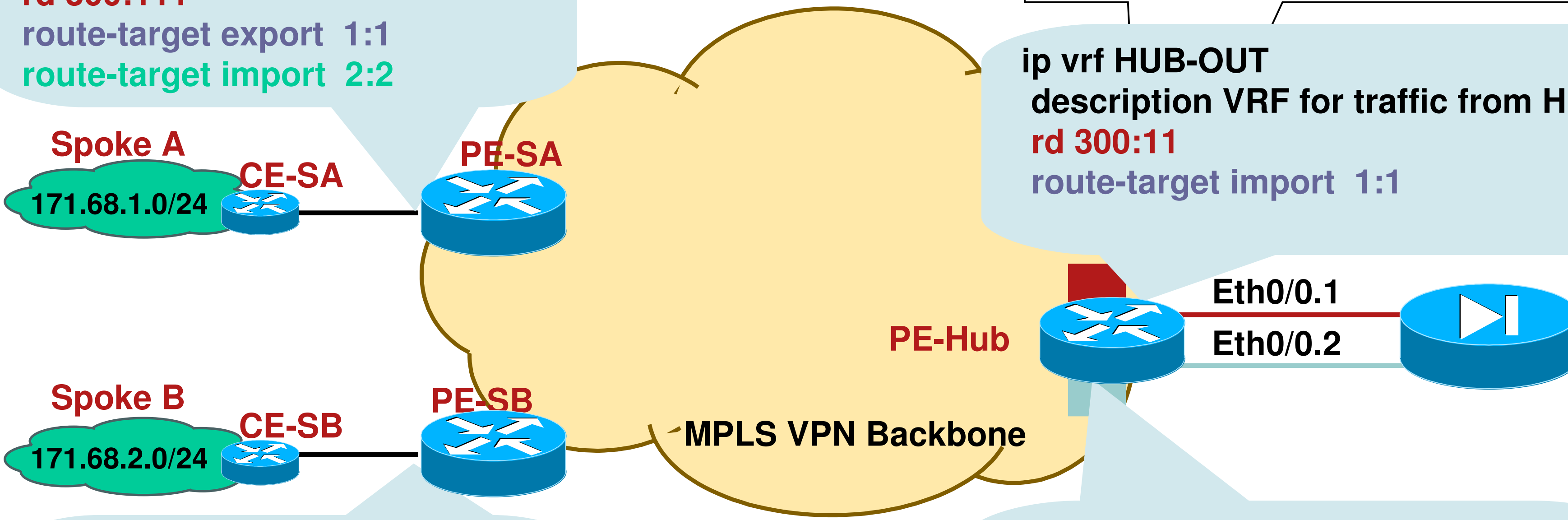
- Traditionally, VPN deployments are Hub and Spoke
    - Spoke to spoke communication is via Hub site only
- Despite MPLS VPN’s implicit any-to-any, i.e, full-mesh connectivity, Hub and Spoke service can easily be offered
    - Done with import and export of route-target (RT) values

MPLS-VPN Services:
2. Hub and Spoke Service: Configuration

```
ip vrf green-spoke1
 description VRF for SPOKE A
 rd 300:111
 route-target export  1:1
 route-target import  2:2
```

```
router bgp <ASN>
 address-family ipv4 vrf HUB-OUT
 neighbor <CE> as-override
```

```
ip vrf HUB-OUT
 description VRF for traffic from HUB
 rd 300:11
 route-target import  1:1
```




```
ip vrf green-spoke2
 description VRF for SPOKE B
 rd 300:112
 route-target export  1:1
 route-target import  2:2
```

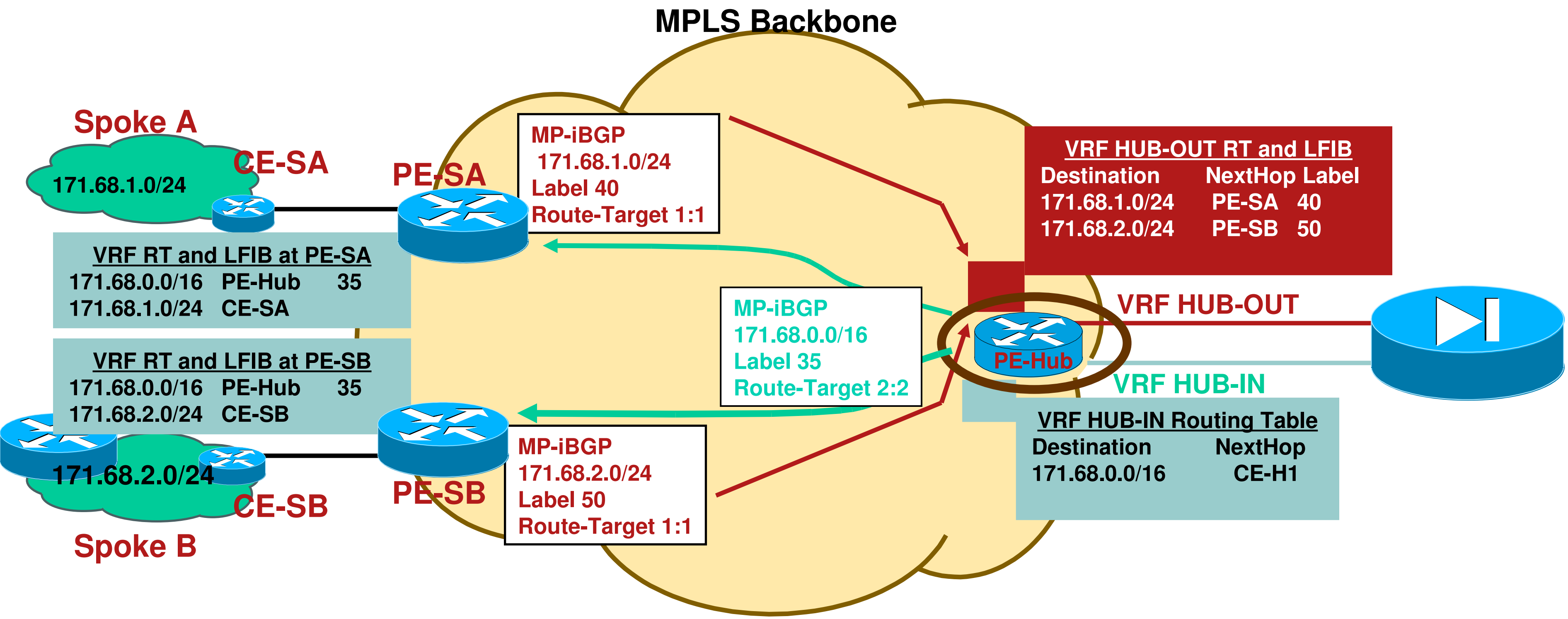
```
ip vrf HUB-IN
 description VRF for traffic to HUB
 rd 300:12
 route-target export  2:2
```

```
router bgp <ASN>
 address-family ipv4 vrf HUB-IN
 neighbor <CE> allowas-in 2
```

MPLS-VPN Services:
2. Hub and Spoke Service: Configuration

- If BGP is used end-to-end, then as-override and allowas-in knobs must be used at the PE_Hub
  - Otherwise AS_PATH looping will occur
- If the spoke sites only need the default route from the hub site, then it is possible to use a single interface between PE-hub and CE-hub (instead of two interfaces as shown on the previous slide)
  - Let CE-hub router advertise the default or aggregate
  - Avoid generating a BGP aggregate at the PE

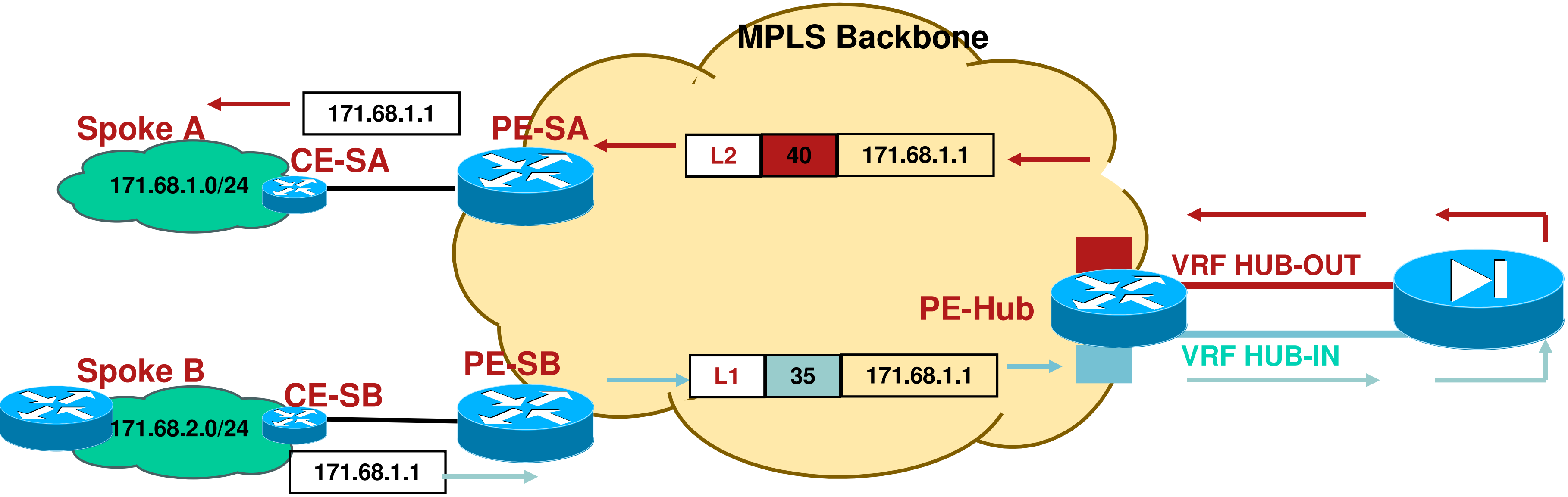
MPLS-VPN Services:
2. Hub and Spoke Service: Control Plane



- All traffic between spokes must pass through the hub/central site
  - Hub site could offer FireWall, NAT like applications
- Two VRF solutions at the PE-hub:
  - VRF HUB_OUT would have knowledge of every spoke routes
  - VRF HUB_IN only have a 171.68.0.0/16 route and advertise that to spoke PEs
- Import and export route-target within a VRF must be different

MPLS-VPN Services:
2. Hub and Spoke Service: Forwarding Plane

This is how the spoke-to-spoke traffic flows -



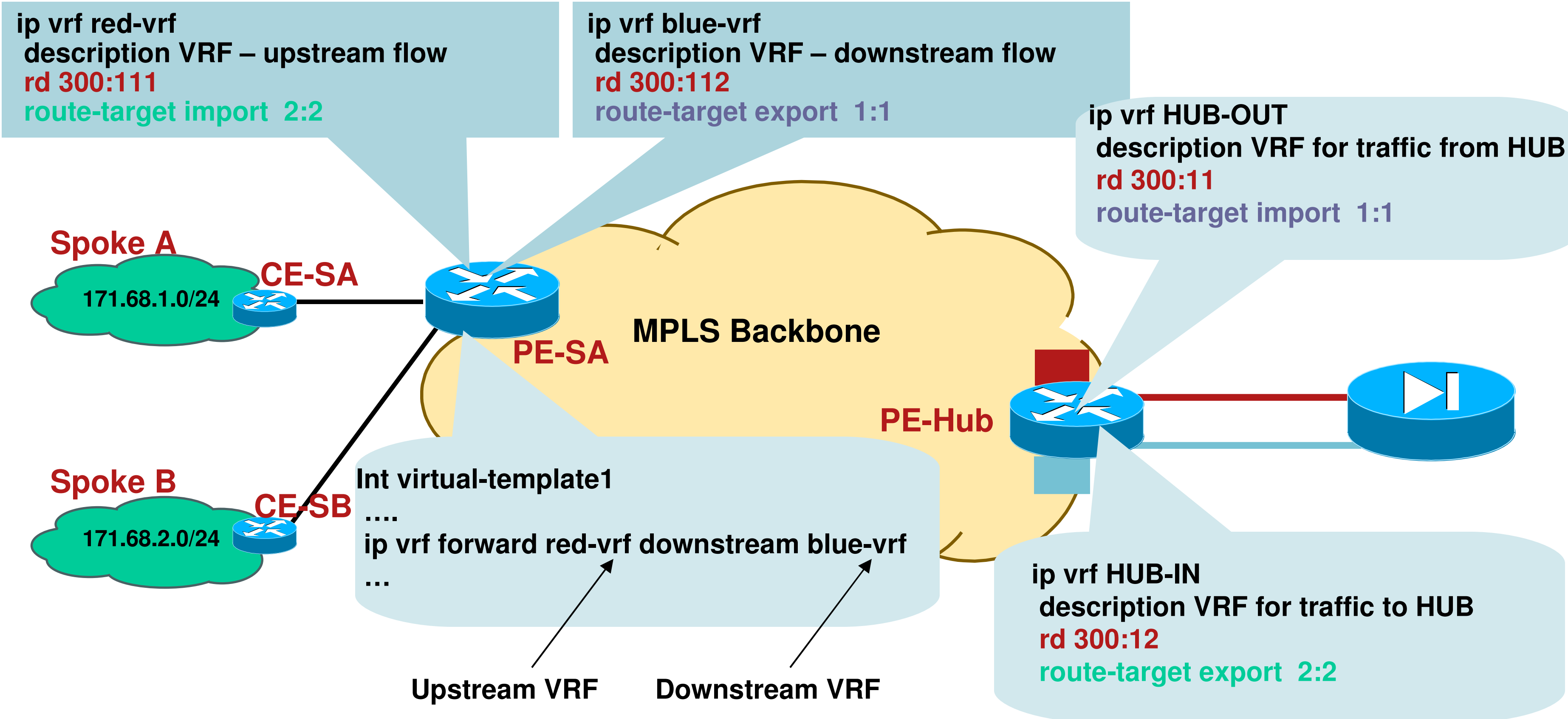
L1 is the label to get to PE-Hub

L2 is the label to get to PE-SA

MPLS-VPN Services:
2. Hub and Spoke Service: Half-Duplex VRF

- When do we need Half-duplex VRF?
- If more than one spoke router (CE) connects to the same PE router within the single VRF, then such spokes can reach other without needing the Hub
  - This defeats the purpose of doing Hub and Spoke
- Half-duplex VRF is the answer.
  - Half-duplex VRF is specific to dial-users i.e., virtual-template
- It requires two VRFs on the PE router
  - Upstream VRF for Spoke->Hub communication
  - Downstream VRF for Spoke<-Hub communication

MPLS-VPN Services:
2. Hub and Spoke Service: Half-Duplex VRF



PE-SA installs the spoke routes only in downstream VRF i.e. blue-VRF

PE-SA forwards the incoming IP traffic (from Spokes) using the upstream VRF i.e. red-vrf routing table

Agenda

- MPLS VPN Explained
- MPLS-VPN Services
    1. Providing Load-Shared Traffic to the Multihomed VPN Sites
    2. Providing Hub and Spoke Service to the VPN Customers
    3. Providing Internet Access Service to VPN Customers
    4. Providing VRF-Selection Based Services
    5. Providing Remote Access MPLS VPN
    6. Providing VRF-Aware NAT Services
    7. Providing MPLS VPN over IP Transport & Multi-VRF CE Services
- Best Practices
- Conclusion

MPLS-VPN Services
3. Internet Access Service to VPN Customers

- Internet access service could be provided as another value-added service to VPN customers
- Security mechanism must be in place at both provider network and customer network
    - To protect from the Internet vulnerabilities
- VPN customers benefit from the single point of contact for both Intranet and Internet connectivity

MPLS-VPN Services
3. Internet Access: Different Methods of Service

- Four ways to provide the Internet service
  1. VRF specific default route with "global" keyword
  2. Separate PE-CE sub-interface (non-VRF)
  3. Extranet with Internet-VRF
  4. VRF-aware NAT

MPLS-VPN Services

3. Internet Access: Different Methods of Service

1. VRF specific default route

   1.1 Static default route to move traffic from VRF to Internet (global routing table)

   1.2 Static routes for VPN customers to move traffic from Internet (global routing table) to VRF

2. Separate PE-CE sub-interface (non-VRF)

   May run BGP to propagate Internet routes between PE and CE

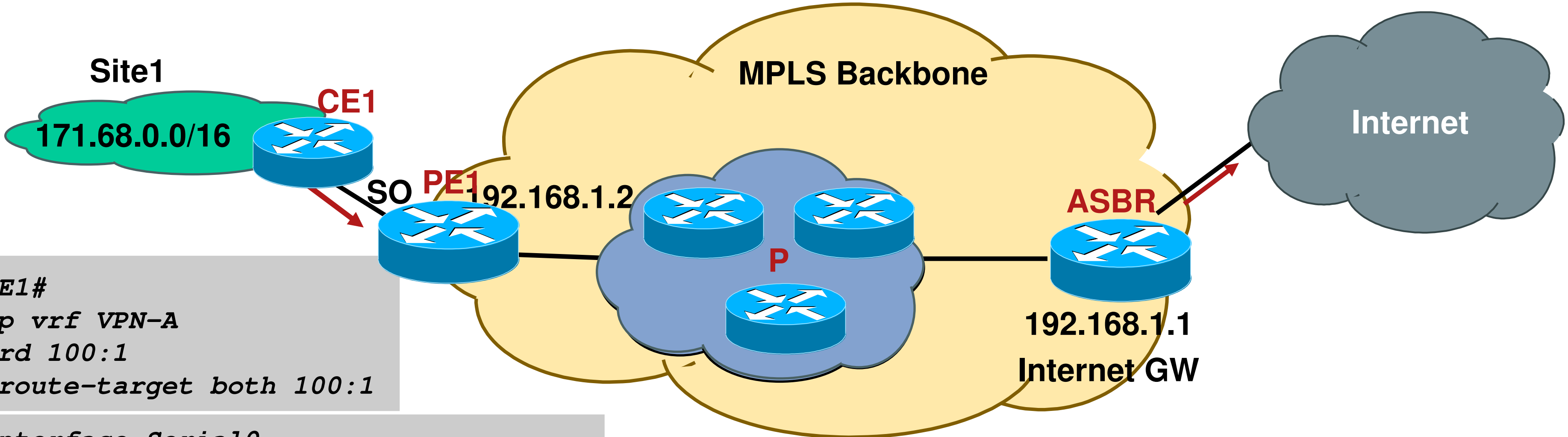
3. Extranet with Internet-VRF

   VPN packets never leave VRF context; issue with overlapping VPN address

4. Extranet with Internet-VRF along with VRF-aware NAT

   VPN packets never leave VRF context; works well with overlapping VPN address

MPLS-VPN Services:
3.1 Internet Access: VRF Specific Default Route (Config)




```
PE1#
ip vrf VPN-A
 rd 100:1
 route-target both 100:1
```

```
Interface Serial0
ip address 192.168.10.1 255.255.255.0
ip vrf forwarding VPN-A
```

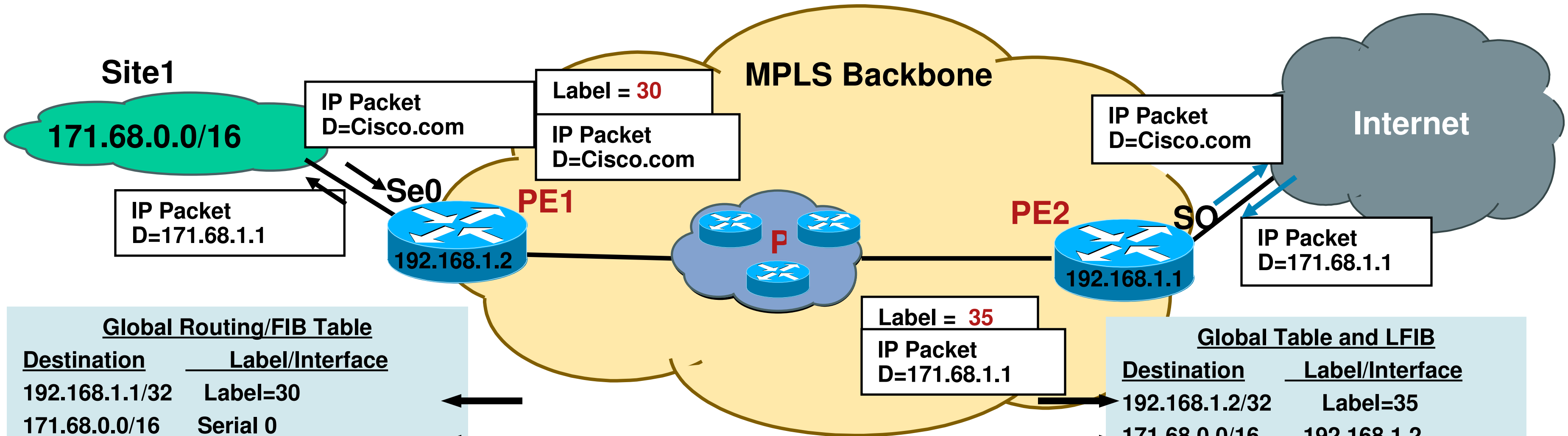
```
Router bgp 100
 no bgp default ipv4-unicast
 redistribute static
 neighbor 192.168.1.1 remote 100
 neighbor 192.168.1.1 activate
 neighbor 192.168.1.1 next-hop-self
 neighbor 192.168.1.1 update-source loopback0
```

```
ip route vrf VPN-A 0.0.0.0 0.0.0.0 192.168.1.1 global
ip route 171.68.0.0 255.255.0.0 Serial0
```

- A default route, pointing to the ASBR, is installed into the site VRF at each PE
- The static route, pointing to the VRF interface, is installed in the global routing table and redistributed into BGP

MPLS-VPN Services:

3.1 Internet Access: VRF Specific Default Route (Forwarding)



Global Routing/FIB Table

Destination	Label/Interface
192.168.1.1/32	Label=30
171.68.0.0/16	Serial 0

Global Table and LFIB

Destination	Label/Interface
192.168.1.2/32	Label=35
171.68.0.0/16	192.168.1.2
Internet	Serial 0

VRF Routing/FIB Table

Destination	Label/Interface
0.0.0.0/0	192.168.1.1 (global)
Site-1	Serial 0

Pros

- Different Internet gateways can be used for different VRFs
- PE routers need not to hold the Internet table
- Simple configuration

Cons

- Using default route for Internet routing does NOT allow any other default route for intra-VPN routing Increasing size of global routing table by leaking VPN routes
- Static configuration (possibility of traffic blackholing)

MPLS-VPN Services

3.2 Internet Access

1. VRF specific default route

   1.1 Static default route to move traffic from VRF to Internet (global routing table)

   1.2 Static routes for VPN customers to move traffic from Internet (global routing table) to VRF

2. Separate PE-CE sub-interface (non-VRF)

   May run BGP to propagate Internet routes between PE and CE

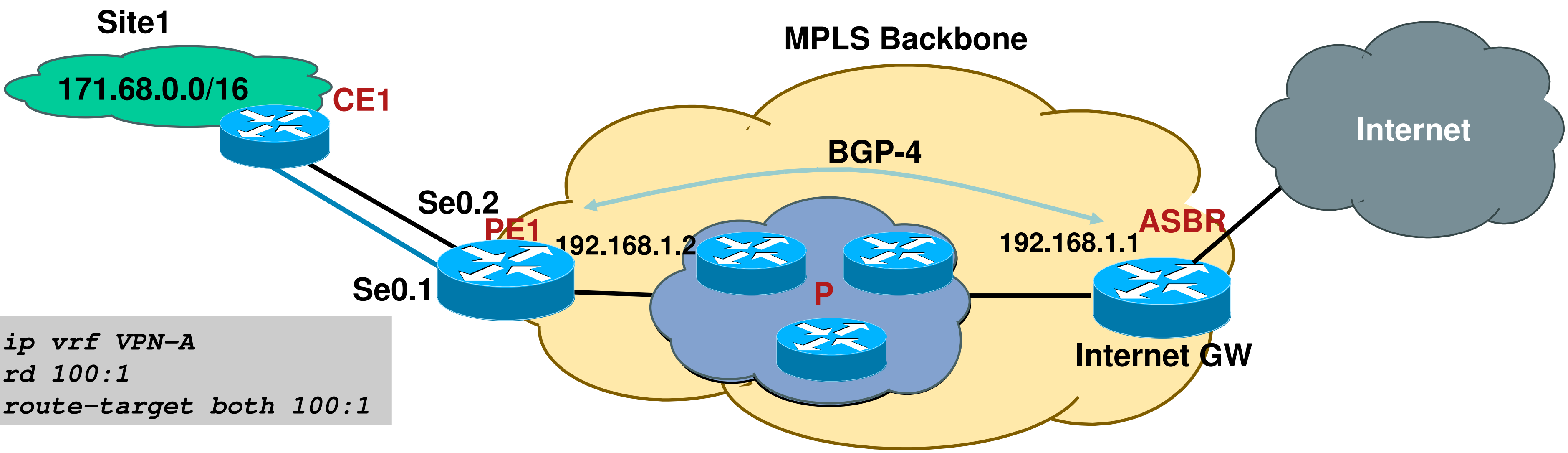
3. Extranet with Internet-VRF

   VPN packets never leave VRF context; overlapping VPN addresses could be a problem

4. Extranet with Internet-VRF along with VRF-aware NAT

   VPN packets never leave VRF context; works well with overlapping VPN addresses

3.2 Internet Access Service to VPN Customers
Using Separate Sub-Interface (Config)



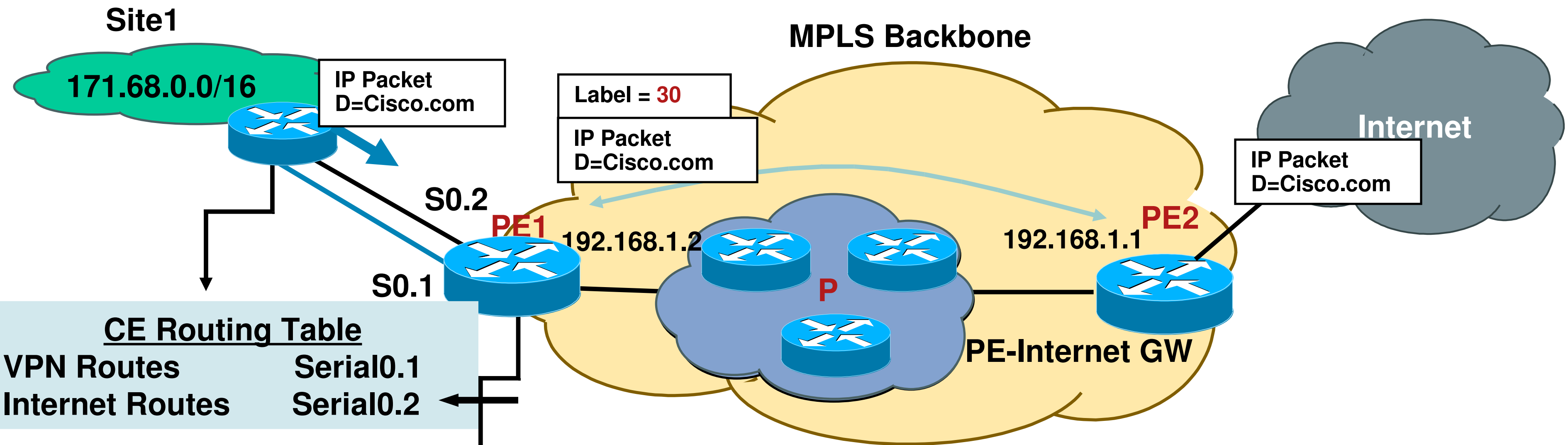

```
ip vrf VPN-A
rd 100:1
route-target both 100:1
```

```
Interface Serial0.1
 ip vrf forwarding VPN-A
 ip address 192.168.20.1 255.255.255.0
 frame-relay interface-dlci 100
!
Interface Serial0.2
 ip address 171.68.10.1 255.255.255.0
 frame-relay interface-dlci 200
!
```

```
Router bgp 100
no bgp default ipv4-unicast
neighbor 171.68.10.2 remote-as 502
```

- One sub-interface for VPN routing associated to a VRF
- Another sub-interface for Internet routing associated to the global routing table
- Could advertise full Internet routes or a default route to CE
- The PE will need to advertise VPN routes to the Internet (via global routing table)

Internet Access Service to VPN Customers
3.2 Using Separate Sub-Interface (Forwarding)



CE Routing Table

VPN Routes	Serial0.1
Internet Routes	Serial0.2

PE Global Table and FIB

Internet Routes	192.168.1.1
192.168.1.1	Label=30

Pros

CE could dual home and perform optimal routing

Traffic separation done by CE

Cons

PE to hold full Internet routes

BGP complexities introduced in CE; CE1 may need to aggregate to avoid AS_PATH looping

Internet Access Service

3.3 Extranet with Internet-VRF

- The Internet routes could be placed within the VRF at the Internet-GW i.e. ASBR
- VRFs for customers could ‘extranet’ with the Internet VRF and receive either default, partial or full Internet routes
- Be careful if multiple customer VRFs, at the same PE, are importing full Internet routes
- Works well only if the VPN customers don’t have overlapping addresses

Internet Access Service

3.4 Internet Access Using VRF-Aware NAT

- If the VPN customers need Internet access without Internet routes, then VRF-aware NAT can be used at the Internet-GW i.e. ASBR
- The Internet GW doesn’t need to have Internet routes either
- Overlapping VPN addresses is no longer a problem
- More in the “VRF-aware NAT” slides…

Agenda

- MPLS VPN Explained
- MPLS-VPN Services
    1. Providing Load-Shared Traffic to the Multihomed VPN Sites
    2. Providing Hub and Spoke Service to the VPN Customers
    3. Providing Internet Access Service to VPN Customers
    4. Providing VRF-Selection Based Services
    5. Providing Remote Access MPLS VPN
    6. Providing VRF-Aware NAT Services
    7. Providing MPLS VPN over IP Transport & Multi-VRF CE Services
- Best Practices
- Conclusion

MPLS VPN Service
4. VRF-Selection

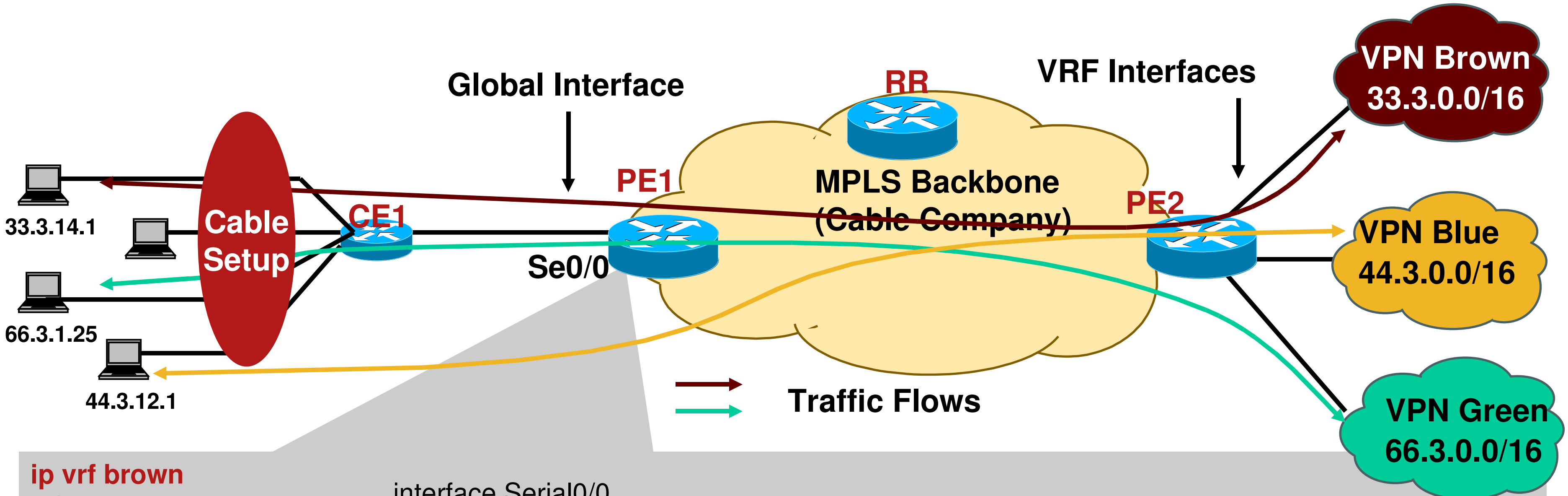
- The common notion is that the VRF must be associated to an interface
- "VRF-selection" breaks this association and associate multiple VRFs to an interface
- Each packet on the PE-CE interface could be handled (based on certain criteria) via different VRF routing tables

  Criteria such as source/dest IP address, ToS, TCP port, etc. specified via route-map

- Voice and data can be separated out into different VRFs at the PE; Service enabler

MPLS VPN Service

4. VRF-Selection: Based on Source IP Address




```
ip vrf brown
 rd 3000:111
 route-target export 3000:1
 route-target import 3000:1
!
ip vrf blue
 rd 3000:222
 route-target export 3000:2
 route-target import 3000:2
!
ip vrf green
 rd 3000:333
 route-target export 3000:3
 route-target import 3000:3
```

```
interface Serial0/0
 ip address 215.2.0.6 255.255.255.252
 ip policy route-map PBR-VRF-Selection
 ip receive brown
 ip receive blue
 ip receive green

access-list 40 permit 33.3.0.0 0.0.255.255
access-list 50 permit 44.3.0.0 0.0.255.255
access-list 60 permit 66.3.0.0 0.0.255.255
```

```
route-map PBR-VRF-Selection permit 10
 match ip address 40
 set vrf brown

route-map PBR-VRF-Selection permit 20
 match ip address 50
 set vrf blue

route-map PBR-VRF-Selection permit 30
 match ip address 60
 set vrf green
```

Agenda

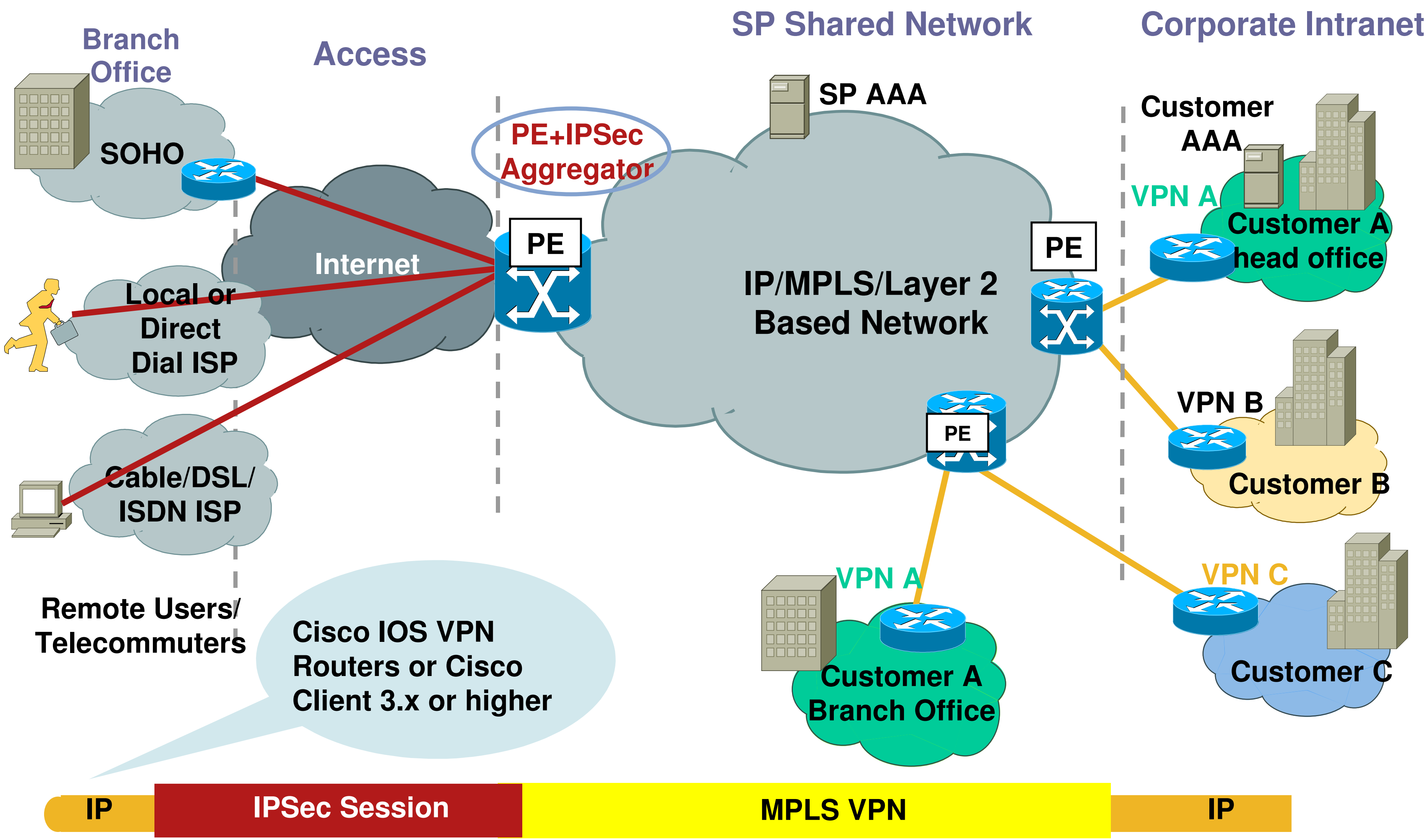
- MPLS VPN Explained
- MPLS-VPN Services
    1. Providing Load-Shared Traffic to the Multihomed VPN Sites
    2. Providing Hub and Spoke Service to the VPN Customers
    3. Providing Internet Access Service to VPN Customers
    4. Providing VRF-Selection Based Services
    5. Providing Remote Access MPLS VPN
    6. Providing VRF-Aware NAT Services
    7. Providing MPLS VPN over IP Transport & Multi-VRF CE Services
- Best Practices
- Conclusion

MPLS VPN Service

5. Remote Access Service

- Remote access users i.e. dial users, IPSec users could directly be terminated in VRF
  - PPP users can be terminated into VRFs
  - IPSec tunnels can be terminated into VRFs
- Remote access services integration with MPLS VPN opens up new opportunities for providers and VPN customers

MPLS VPN Service
5. Remote Access Service: IPSec to MPLS VPN



Agenda

- MPLS VPN Explained
- MPLS-VPN Services
  1. Providing Load-Shared Traffic to the Multihomed VPN Sites
  2. Providing Hub and Spoke Service to the VPN Customers
  3. Providing Internet Access Service to VPN Customers
  4. Providing VRF-Selection Based Services
  5. Providing Remote Access MPLS VPN
  6. Providing VRF-Aware NAT Services
  7. Providing MPLS VPN over IP Transport & Multi-VRF CE Services
- Best Practices
- Conclusion

MPLS-VPN Services
6. VRF-Aware NAT Services

- VPN customers could be using ‘overlapping’ IP address i.e. 10.0.0.0/8
- Such VPN customers must NAT their traffic before using either “Extranet” or “Internet” or any shared* services
- PE is capable of NATting the VPN packets (eliminating the need for an extra NAT device)

*** VoIP, Hosted Content, Management, etc.**

MPLS-VPN Services
6. VRF-Aware NAT Services

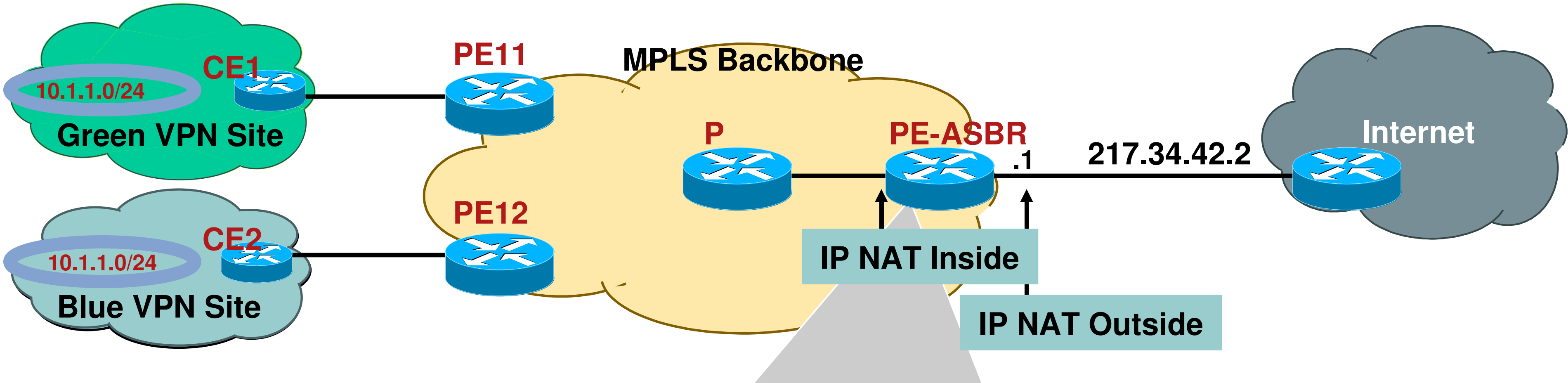
- Typically, inside interface(s) connect to private address space and outside interface(s) connect to global address space

  NAT occurs after routing for traffic from inside-to-outside interfaces

  NAT occurs before routing for traffic from outside-to-inside interfaces

- Each NAT entry is associated with the VRF
- Works on VPN packets in the following switch paths: IP->IP, IP->MPLS and MPLS->IP

MPLS-VPN Services:
6. VRF-Aware NAT Services: Internet Access



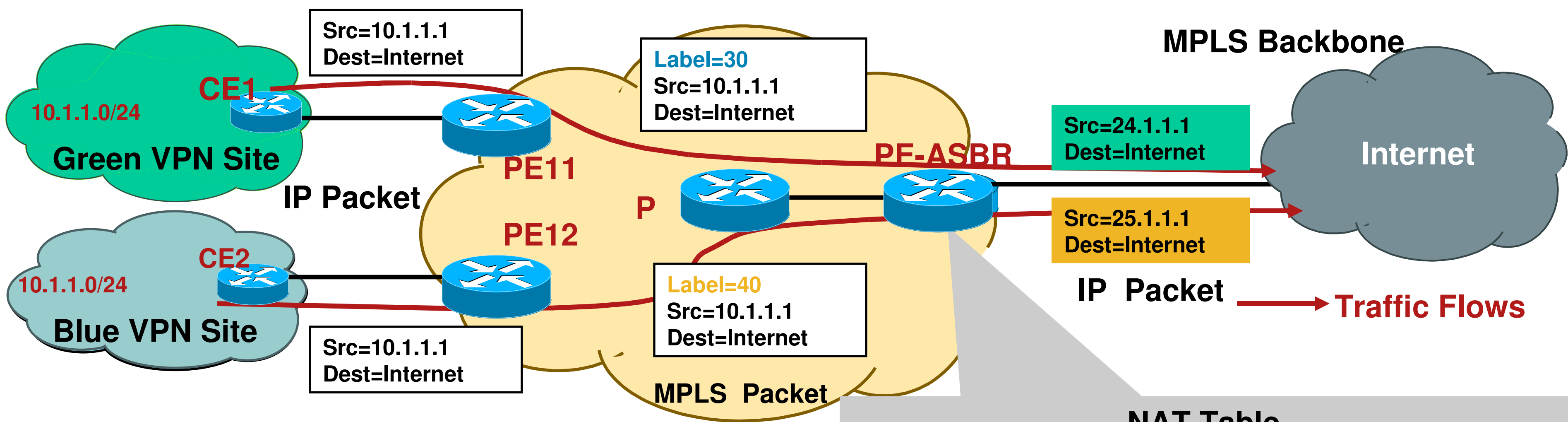

```
ip vrf green
 rd 3000:111
 route-target both 3000:1
ip vrf blue
 rd 3000:222
 route-target both 3000:2

router bgp 3000
 address-family ipv4 vrf green
 network 0.0.0.0
 address-family ipv4 vrf blue
 network 0.0.0.0
```

VRF Specific Config

```
ip nat pool pool-green 24.1.1.0 24.1.1.254 prefix-length 24

ip nat pool pool-blue 25.1.1.0 25.1.1.254 prefix-length 24

ip nat inside source list vpn-to-nat pool pool-green vrf green
ip nat inside source list vpn-to-nat pool pool-blue vrf blue

ip access-list standard vpn-to-nat
 permit 10.1.1.0 0.0.0.255

ip route vrf green 0.0.0.0 0.0.0.0 217.34.42.2 global
ip route vrf blue 0.0.0.0 0.0.0.0 217.34.42.2 global
```

VRF-Aware NAT Specific Config

MPLS-VPN Services:
6. VRF-Aware NAT Services: Internet Access



NAT Table

VRF IP Source	Global IP	VRF-Table-Id
10.1.1.1	24.1.1.1	green
10.1.1.1	25.1.1.1	blue

- PE-ASBR removes the label from the received MPLS packets per LFIB
- Performs NAT on the resulting IP packets
- Forwards the packet to the internet
- Returning packets are NATed and put back in the VRF context and then routed

- This is also one of the ways to provide Internet access to VPN customers with or without overlapping addresses

Agenda

- MPLS VPN Explained
- MPLS-VPN Services
  1. Providing Load-Shared Traffic to the Multihomed VPN Sites
  2. Providing Hub and Spoke Service to the VPN Customers
  3. Providing Internet Access Service to VPN Customers
  4. Providing VRF-Selection Based Services
  5. Providing Remote Access MPLS VPN
  6. Providing VRF-Aware NAT Services
  7. Providing MPLS VPN over IP Transport & Multi-VRF CE Services
- Best Practices
- Conclusion

MPLS-VPN Services:
7. Providing MPLS/VPN over IP Transport

- What if the core (P) routers are not capable of running MPLS
- MPLS/VPN (rfc2547) can be deployed using IP transport

  NO LDP anywhere

- Instead of using the MPLS label to reach the next-hop, an IP tunnel is used.

  IP tunnel could be L2TPv3, GRE etc.

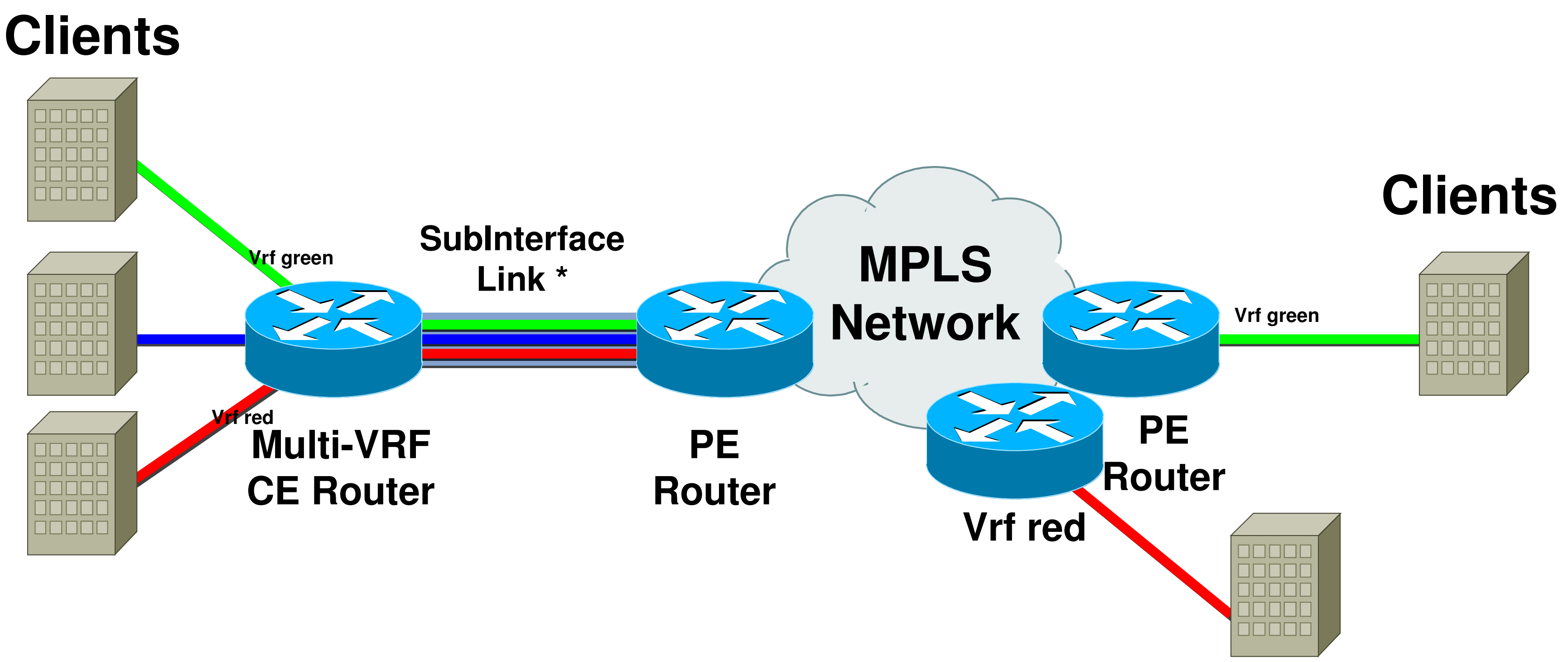
- MPLS labels are still allocated for the VPN prefix and used only by the PE routers

MPLS-VPN Services:
7. Providing Multi-VRF CE Service

- Is it possible for an IP router to keep multiple customer connections separated ?
  - Yes, “multi-VRF CE” aka vrf-lite is the answer.
- “Multi-VRF CE” provides multiple virtual routing tables (and forwarding tables) per customer at the CE router
  - Not a feature but an application based on VRF implementation
  - Any routing protocol that is supported by normal VRF can be used in a Multi-VRF CE implementation
- There is no MPLS functionality on the CE, no label exchange between the CE and any router (including PE)

MPLS-VPN Services:
7. Providing Multi-VRF CE Service

One Deployment Model—Extending MPLS/VPN



SubInterface Link – Any Interface type that supports Sub Interfaces, FE-Vlan, Frame Relay, ATM VC's

Agenda

- MPLS VPN Explained
- MPLS-VPN Services
- Best Practices
- Conclusion

L3 VPN Deployment Best Practices

1. Use RR to scale BGP; deploy RRs in pair for the redundancy
   - Keep RRs out of the forwarding paths and disable CEF (saves memory)
2. RT and RD should have ASN in them i.e. ASN: X
   - Reserve first few 100s of X for the internal purposes such as filtering
3. Consider unique RD per VRF per PE, if load sharing of VPN traffic is required
4. Don't use customer names as the VRF names; nightmare for the NOC. Use simple combination of numbers and characters in the VRF name
   - For example: v101, v102, v201, v202, etc. Use description.
5. PE-CE IP address should come out of SP's public address space to avoid overlapping
   - Use /31 subnetting on PE-CE interfaces
6. Define an upper limit at the PE on the number of prefixes received from the CE for each VRF or neighbor
   - Max-prefix within the VRF configuration
   - Max-prefix per neighbor within the BGP VRF af (if BGP on the PE-CE)

Agenda

- MPLS VPN Explained
- MPLS-VPN Services
- Best Practices
- Conclusion

Conclusion

- MPLS VPN is a cheaper alternative to traditional l2vpn
- MPLS-VPN paves the way for new revenue streams

    VPN customers could outsource their layer3 to the provider

- Straightforward to configure any-to-any VPN topology

    Partial-mesh, Hub and Spoke topologies can also be easily deployed

- VRF-aware services could be deployed to maximize the investment

MPLS Traffic Engineering

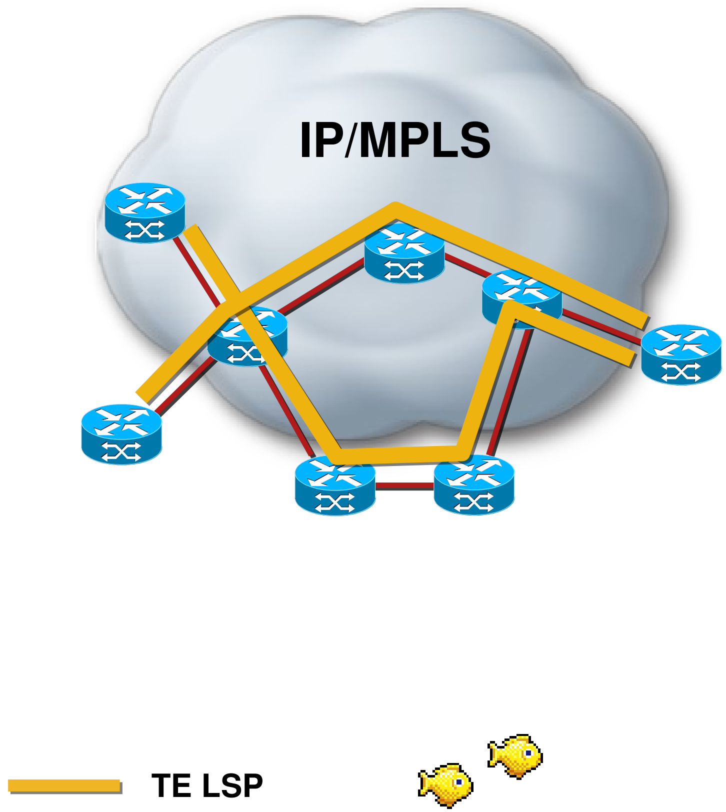
Agenda

- Technology Overview
- MPLS TE Deployment Models
    - Bandwidth optimization
    - Fast Re-route
    - TE for QoS
- Inter-Domain Traffic Engineering
- General Deployment Considerations

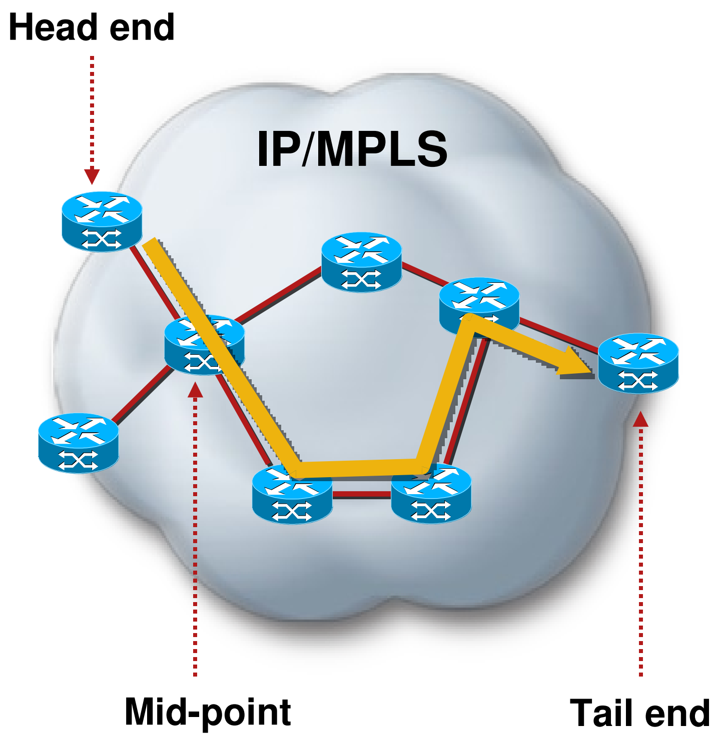
Technology Overview

MPLS TE Overview

- Introduces explicit routing
- Supports constrained-based routing
- Supports admission control
- Provides protection capabilities
- Uses RSVP-TE to establish LSPs
- Uses ISIS and OSPF extensions to advertise link attributes

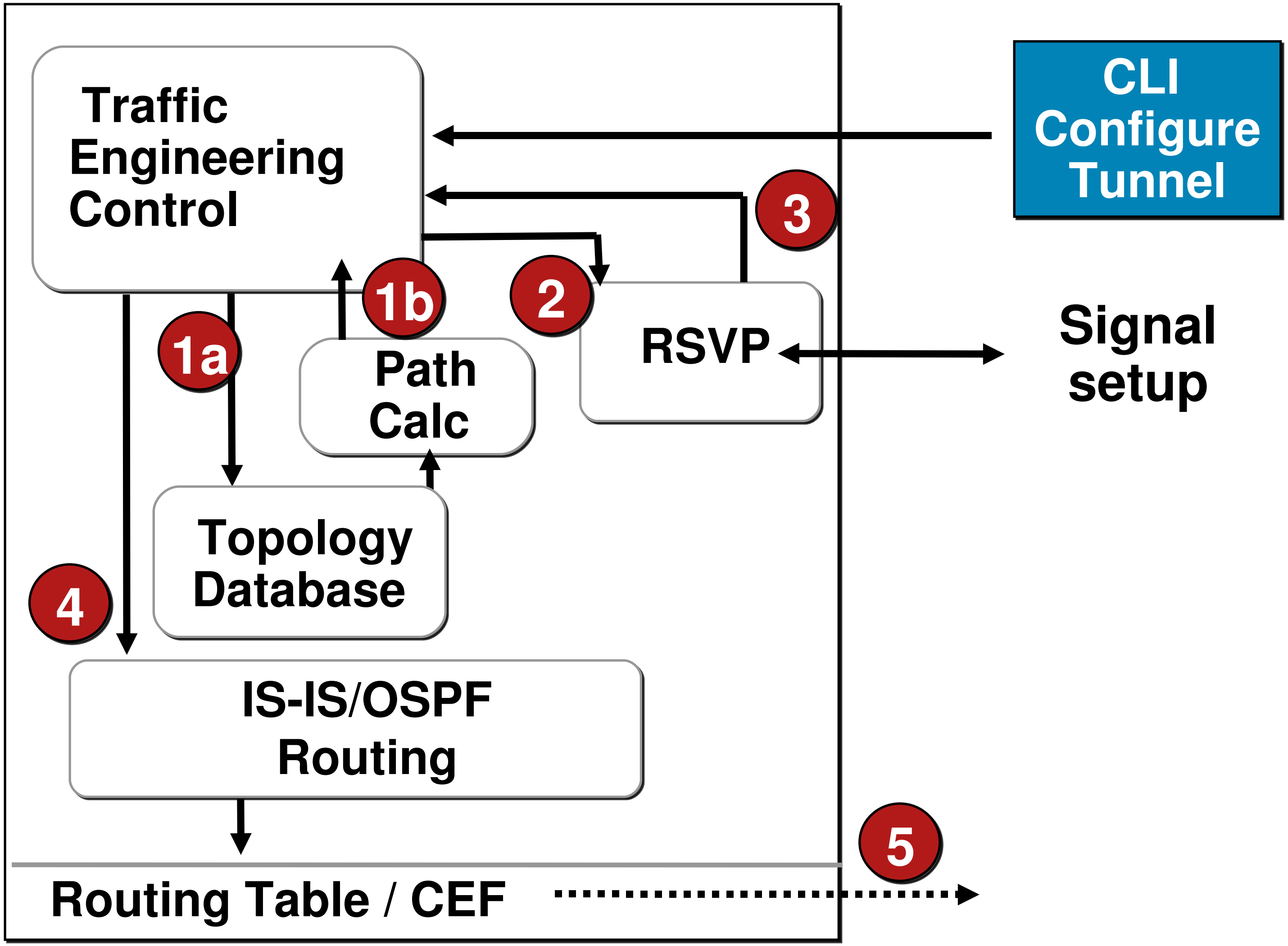


How MPLS TE Works



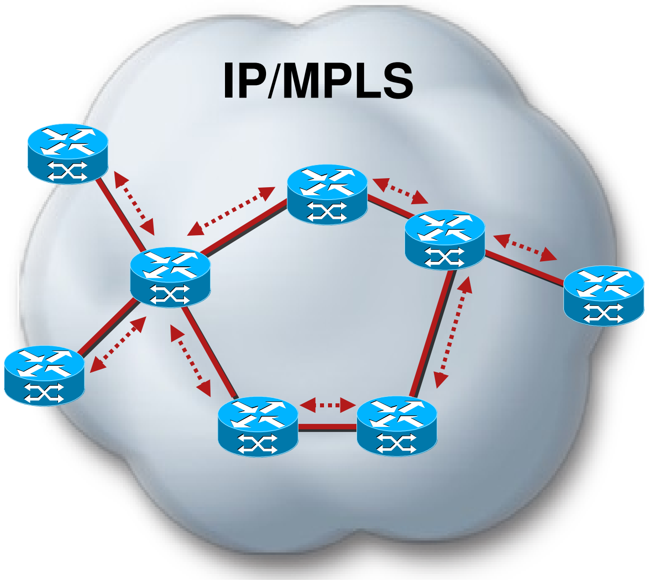
- Link information Distribution
  - ISIS-TE
  - OSPF-TE
- Path Calculation (CSPF)
- Path Setup (RSVP-TE)
- Forwarding Traffic down Tunnel
  - Auto-route
  - Static
  - PBR
  - CBTS
  - Forwarding Adjacency
  - Tunnel select

MPLS TE Router Operation

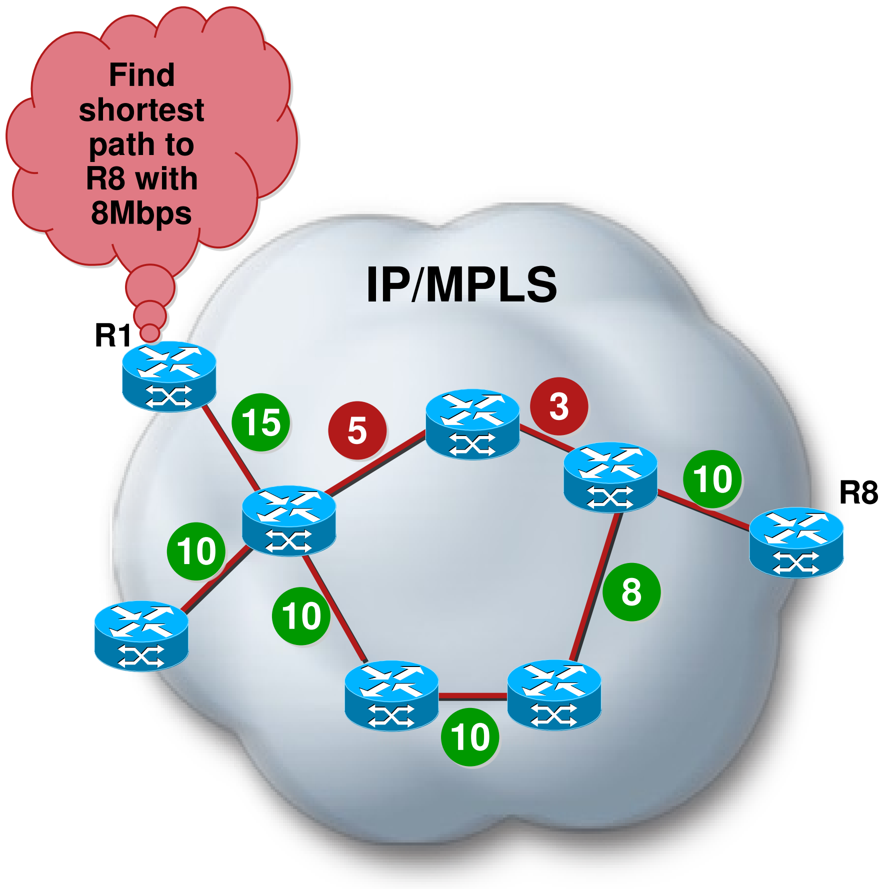


Link Information Distribution

- Additional link characteristics
  - Interface address
  - Neighbor address
  - Physical bandwidth
  - Maximum reservable bandwidth
  - Unreserved bandwidth (at eight priorities)
  - TE metric
  - Administrative group (attribute flags)
- IS-IS or OSPF flood link information
- TE nodes build a topology database



Path Calculation



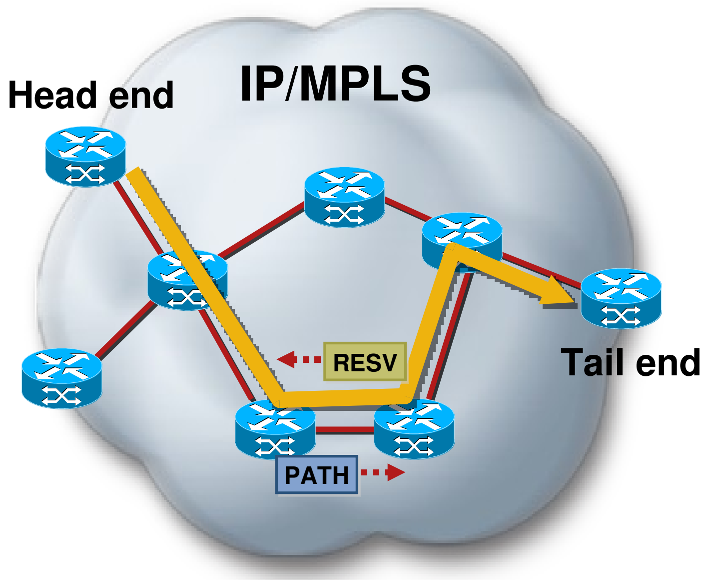
Link with insufficient bandwidth

Link with sufficient bandwidth

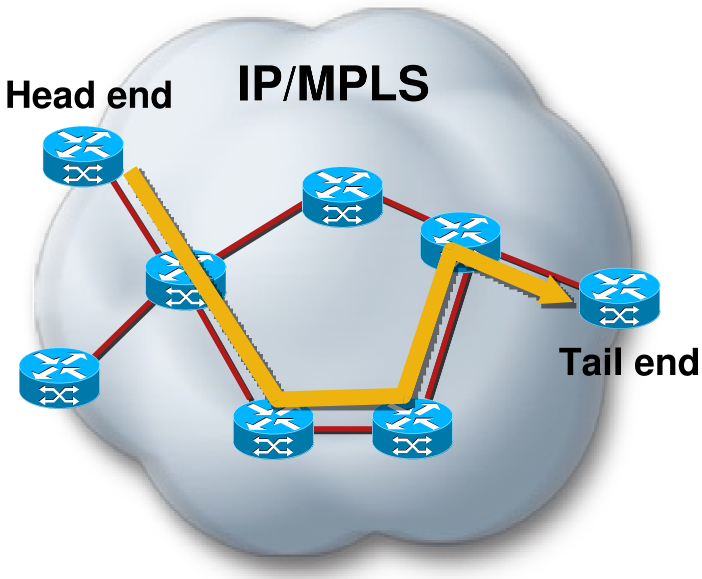
- TE nodes can perform constraint-based routing
- Constraints and topology database as input to path computation
- Shortest-path-first algorithm ignores links not meeting constraints
- Tunnel can be signaled once a path is found

TE LSP Signaling

- Tunnel signaled with TE extensions to RSVP
- Soft state maintained with downstream PATH messages
- Soft state maintained with upstream RESV messages
- New RSVP objects
    - LABEL_REQUEST (PATH)
    - LABEL (RESV)
    - EXPLICIT_ROUTE
    - RECORD_ROUTE (PATH/RESV)
    - SESSION_ATTRIBUTE (PATH)
- LFIB populated using RSVP labels



Traffic Selection



- Multiple traffic selection options
    - Auto-route
    - Static routes
    - Policy Based Routing
    - Forward Adjacency
    - Pseudowire Tunnel Selection
    - Class Based Tunnel Selection
- Tunnel path computation independent of routing decision injecting traffic into tunnel
- Traffic enters the tunnel at the head end

Configuring MPLS TE and Link Information Distribution Using IS-IS

```
mpls traffic-eng tunnels
!
interface POS0/1/0
 ip address 172.16.0.0 255.255.255.254
 ip router isis
 mpls traffic-eng tunnels
 mpls traffic-eng attribute-flags 0xF
 mpls traffic-eng administrative-weight 20
 ip rsvp bandwidth 100000
!
router isis
 net 49.0001.1720.1625.5001.00
 is-type level-2-only
 metric-style wide
 mpls traffic-eng router-id Loopback0
 mpls traffic-eng level-2
 passive-interface Loopback0
!
```

- **Enable MPLS TE on this node**
- **Enable MPLS TE on this interface**
- **Attribute flags**
- **TE metric**
- **Maximum reservable bandwidth**
- **Enable wide metric format and TE extensions (TE Id, router level)**

Configuring Tunnel at Head End

```
interface Tunnel1
 description FROM-ROUTER-TO-DST1
 ip unnumbered Loopback0
 tunnel destination 172.16.255.3
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng priority 5 5
 tunnel mpls traffic-eng bandwidth  10000
 tunnel mpls traffic-eng affinity 0x0 mask 0xF
 tunnel mpls traffic-eng path-option 5 explicit name
PATH1
 tunnel mpls traffic-eng path-option 10 dynamic
!
ip explicit-path name PATH1 enable
 next-address 172.16.0.1
 next-address 172.16.8.0
!
```

Destination (tunnel tail end)

TE tunnel (as opposed to GRE or others)

Setup/hold priorities

Signaled bandwidth

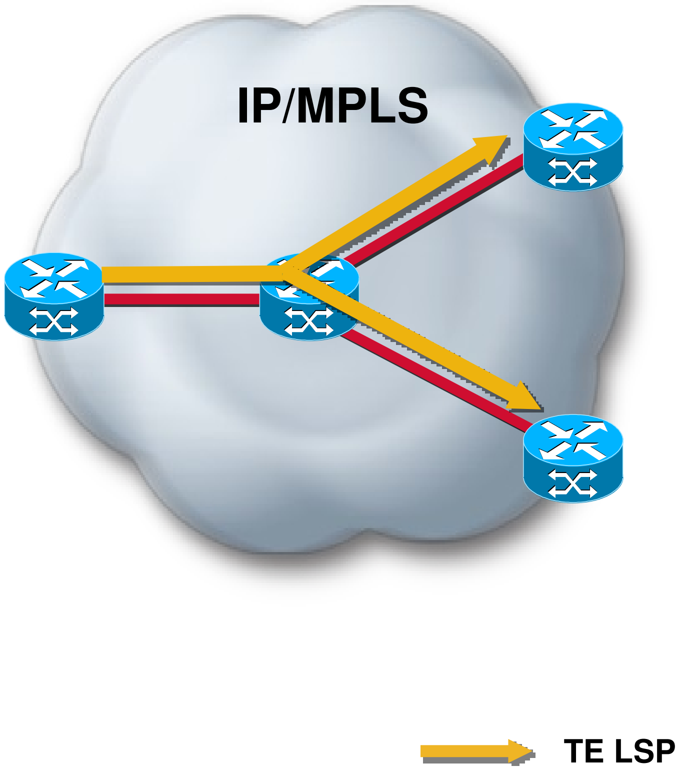
Consider links with 0x0/0xF as attribute flags

Tunnel path options (PATH1, otherwise dynamic)

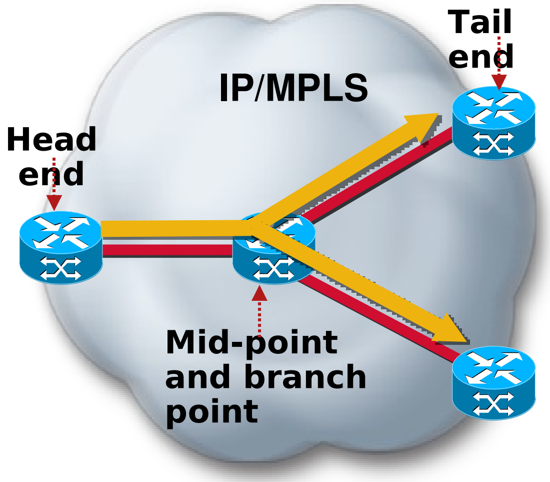
Explicit PATH1 definition

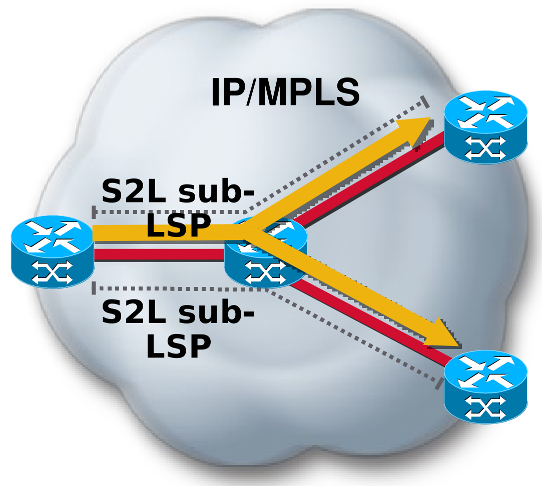
Characteristics of P2MP TE LSP

- Unidirectional
- Explicitly routed
- One head end, but one or more tail ends (destinations)
- Same characteristics (constraints, protection, etc.) for all destinations



P2MP TE LSP Terminology

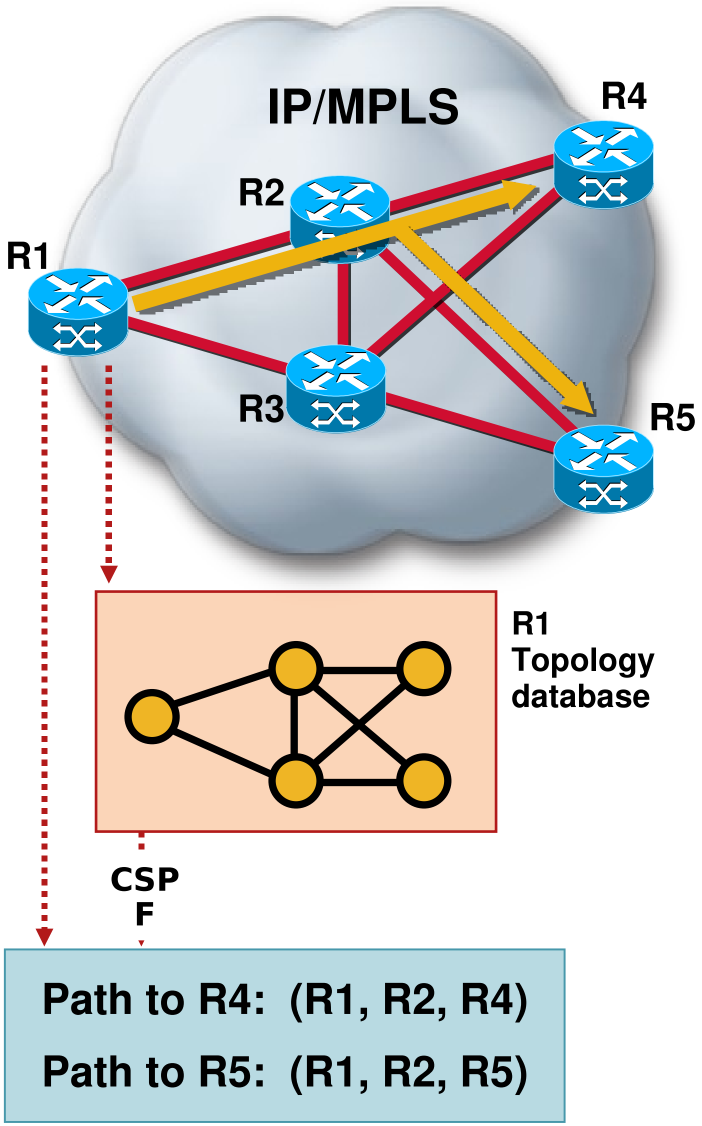




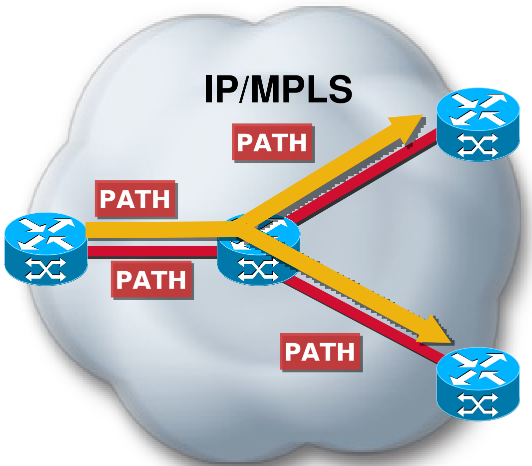
- Head-end/Source: Node where LSP signaling is initiated
- Mid-point: Transit node where LSP signaling is processed (not a head-end, not a tail-end)
- Tail-end/Leaf/destination: node where LSP signaling ends
- Branch point: Node where packet replication is performed
- Source-to-leaf (S2L) sub-LSP: P2MP TE LSP segment that runs from source to one leaf

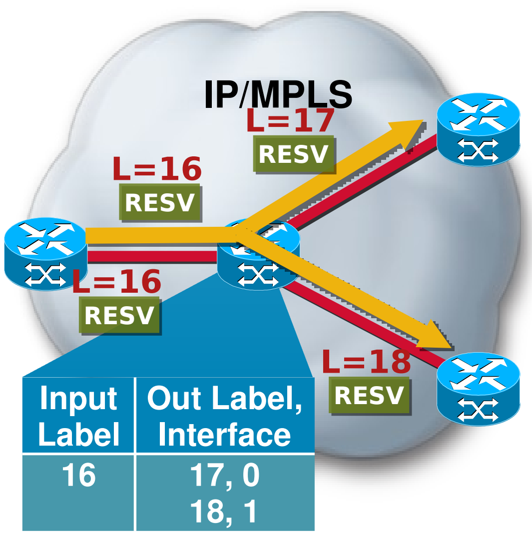
P2MP TE LSP Path Computation

- CSPF suitable to dynamically find an adequate tree
- CSPF executed per destination
- TE topology database and tunnel constraints as input for path computation
- Path constraints may include loose, included, excluded hops
- Same constraints for all destinations (bandwidth, affinities, priorities, etc.)
- Path computation yields explicit path to each destination
- No changes to OSPF/IS-IS TE extensions
- Static paths possible with offline path computation



P2MP TE LSP Signaling

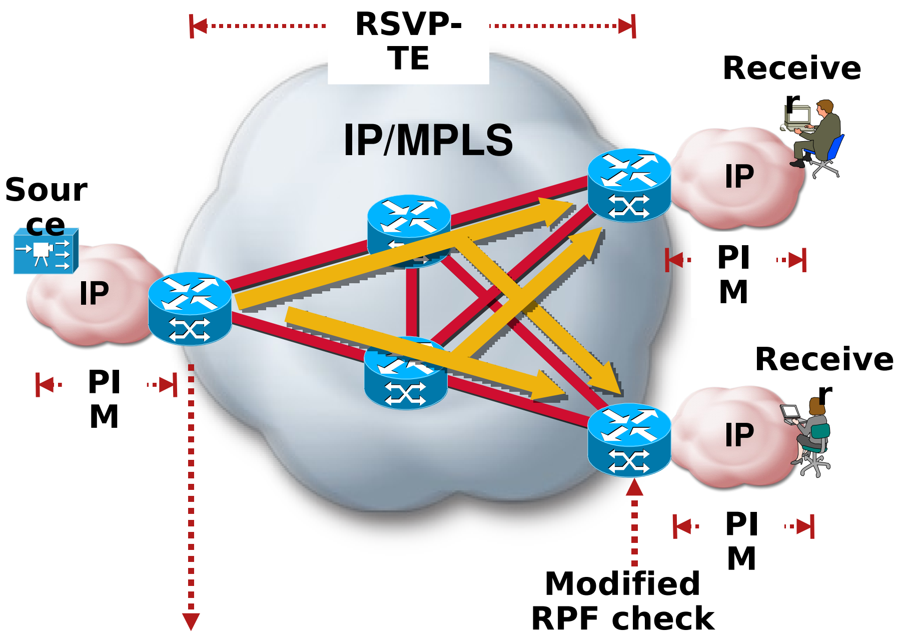




Input Label	Out Label, Interface
16	17, 0 18, 1

- Source sends unique PATH message per destination
- LFIB populated using RSVP labels allocated by RESV messages
- Multicast state built by reusing sub-LSP labels at branch points

P2MP TE LSP Traffic Selection



Static IGMP Joins

P2MP Tunnel	Multicast Group
Tunnel1	(192.168.5.1, 232.0.0.1) (192.168.5.1, 232.0.0.1)
Tunnel2	(192.168.5.1, 232.0.0.3)

- One or more IP multicast groups mapped to a Tunnel
- Groups mapped via static IGMP join
- PIM outside of MPLS network
- Modified egress RPF check against TE LSP and tunnel head end (source address)
- Egress node may abstract TE LSP as a virtual interface (LSPVIF) for RPF purposes

Configuring P2MP Tunnel at Head End (Cisco IOS)

```
mpls traffic-eng destination list name P2MP-LIST-DST1
 ip 172.16.255.1 path-option 10 explicit name PATH1
 ip 172.16.255.2 path-option 10 dynamic
 ip 172.16.255.3 path-option 10 dynamic
 ip 172.16.255.4 path-option 10 dynamic
!
interface Tunnel1
 description FROM-ROUTER-TO-LIST-DST1
 ip unnumbered Loopback0
 ip pim sparse-mode
 ip igmp static-group 232.0.0.1 source 192.168.5.1
 ip igmp static-group 232.0.0.2 source 192.168.5.1
 tunnel mode mpls traffic-eng point-to-multipoint
 tunnel destination list mpls traffic-eng name P2MP-LIST-DST1
 tunnel mpls traffic-eng priority 7 7
 tunnel mpls traffic-eng bandwidth 1000
!
```

- **Destination list with one `path-option` per destination**
- **Enable PIM-SM (historical)**
- **Multicast groups mapped to tunnel**
- **P2MP TE Tunnel**
- **Destination list**
- **Setup/hold priorities**
- **Signaled bandwidth**

Configuring RPF Check at P2MP Tunnel Tail End (Cisco IOS)

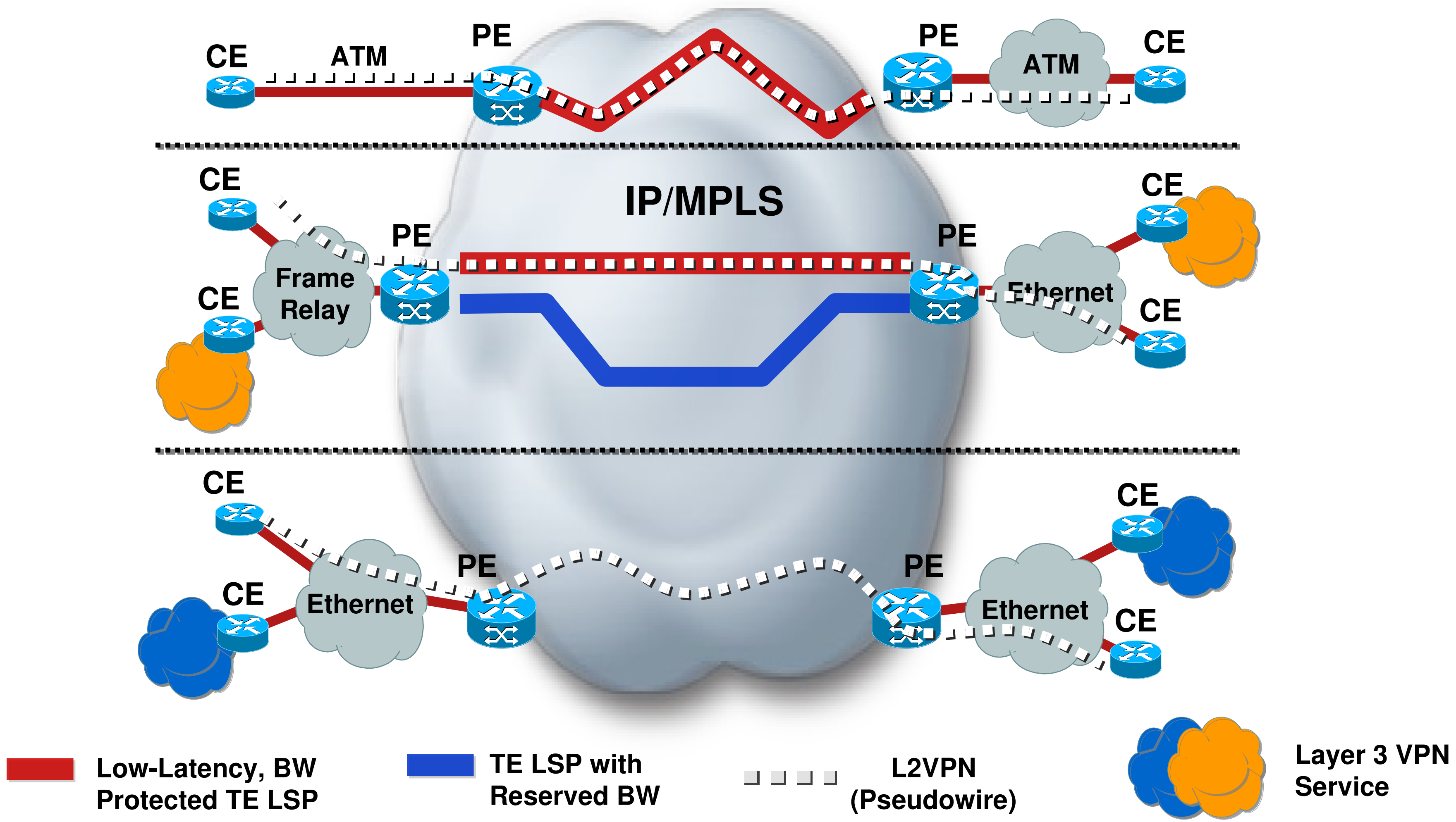
```
ip multicast mpls traffic-eng
ip multicast mpls source Loopback0
ip mroute 192.168.5.1 255.255.255.255 172.16.255.5
!
```

- `ip multicast mpls traffic-eng`: Enable IPv4 multicast over P2MP TE LSP
- `ip multicast mpls source Loopback0`: LSPVIF unnumbered (loopback0)
- `ip mroute 192.168.5.1 255.255.255.255 172.16.255.5`: Tunnel source (172.16.255.5) as next-hop for IP Multicast source (192.168.5.1) RPF check

MPLS TE Deployment Models

MPLS TE and L2/L3VPN

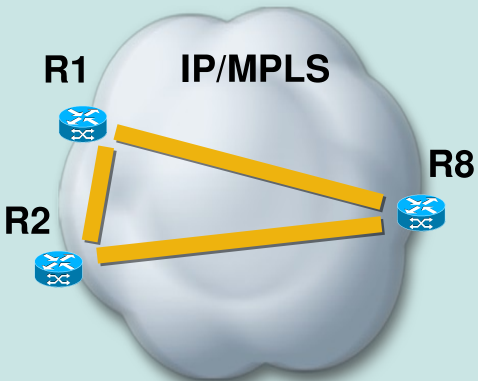
MPLS TE acts as transport for other application and services



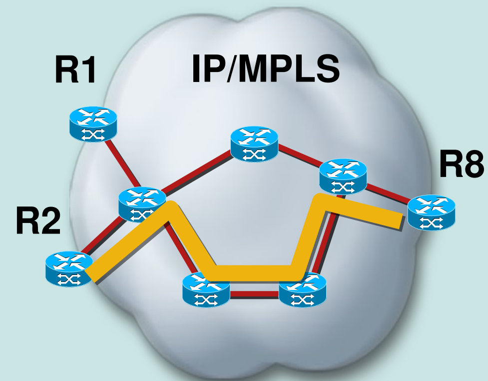
MPLS TE Deployment Models

Bandwidth Optimization

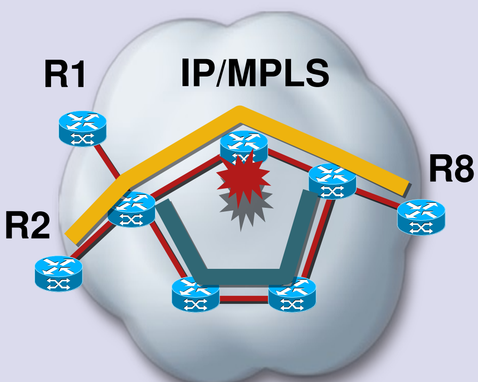
Strategic



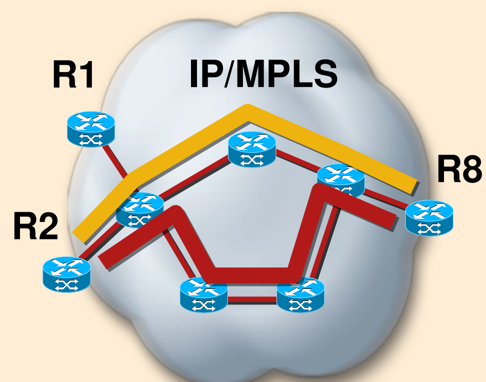
Tactical



Protection



Point-to-Point SLA



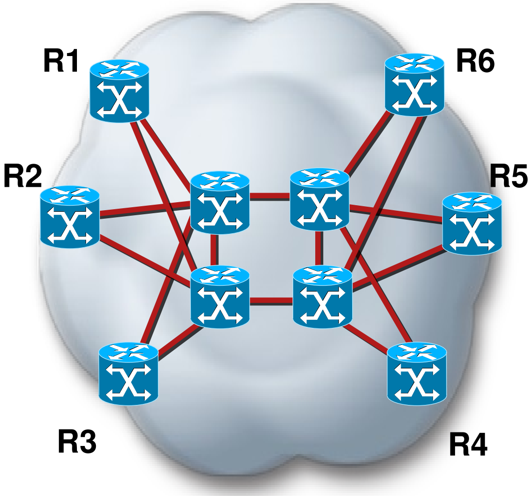
Bandwidth Optimization

Strategic Bandwidth Optimization

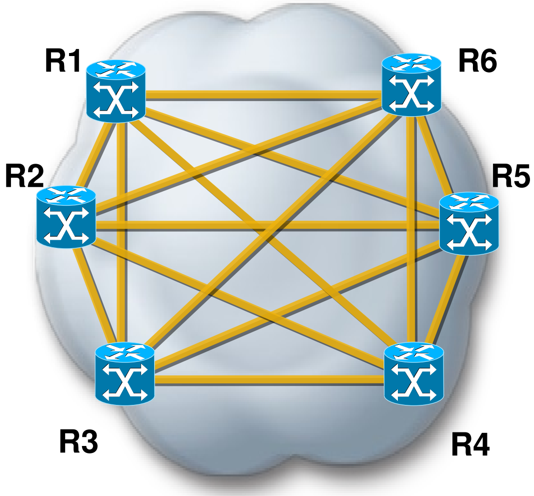
Traffic Matrix

	R1	R2	R3	R4	R5	R6
R1	4	7	1	5	4	5
R2	2	2	4	7	2	3
R3	1	2	9	5	5	5
R4	9	1	4	1	3	1
R5	3	7	9	2	7	7
R6	6	3	5	4	9	12

Physical Topology



Tunnel mesh to satisfy traffic matrix



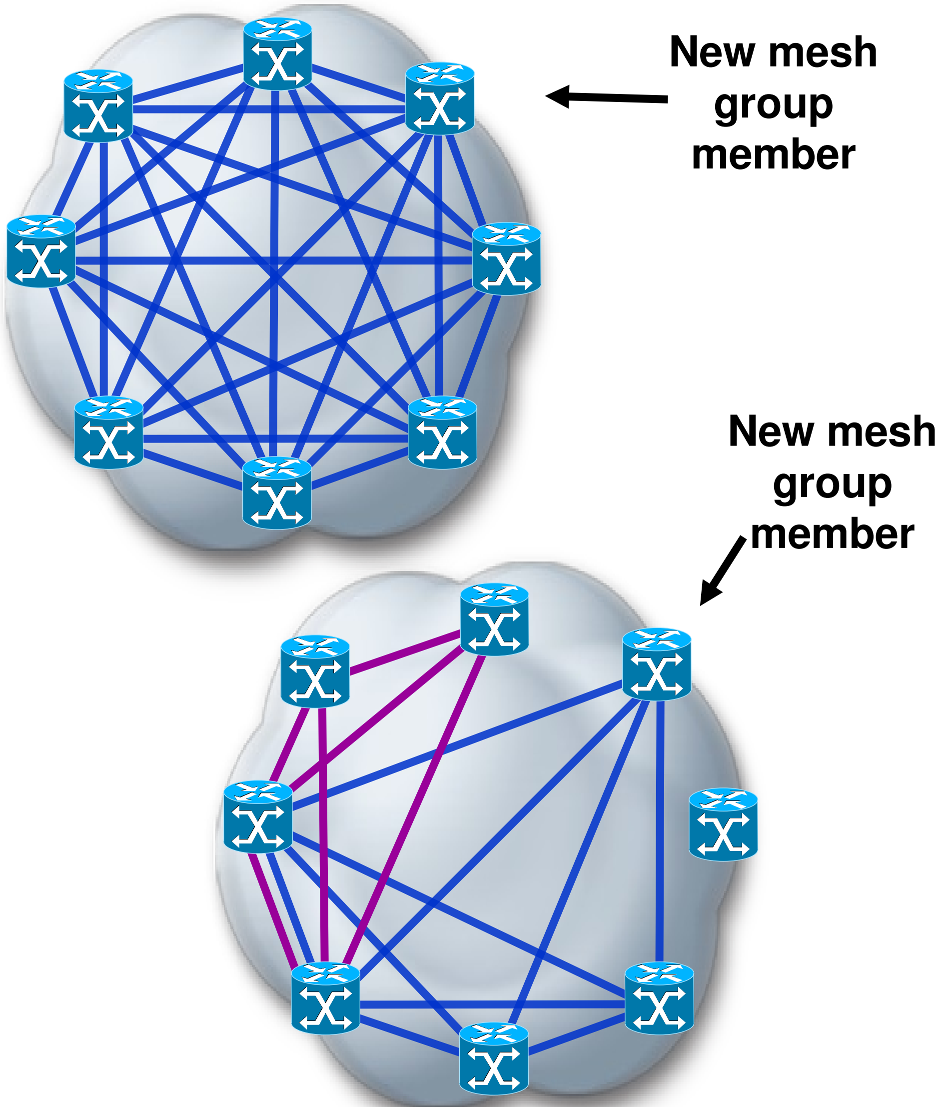
- Tries to optimize underlying physical topology based on traffic matrix
- Key goal is to avoid link over/under utilization
- On-line (CSPF) or off-line path computation
- May result in a significant number of tunnels
- Should not increase your routing adjacencies

AutoTunnel Mesh

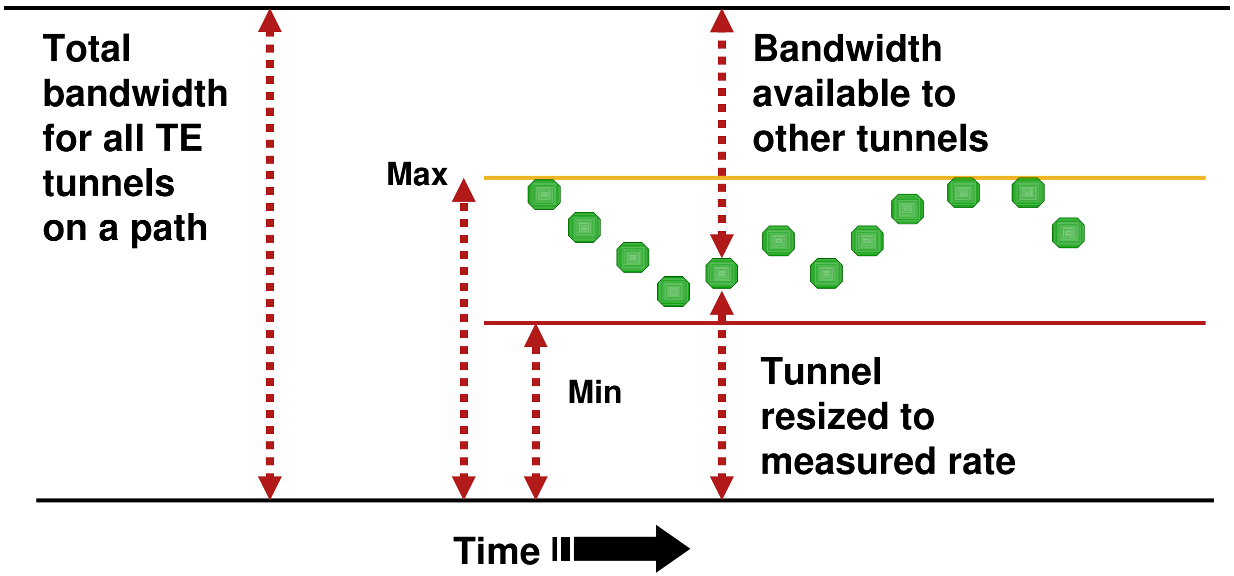
- Mesh group: LSRs to mesh automatically
- Membership identified by

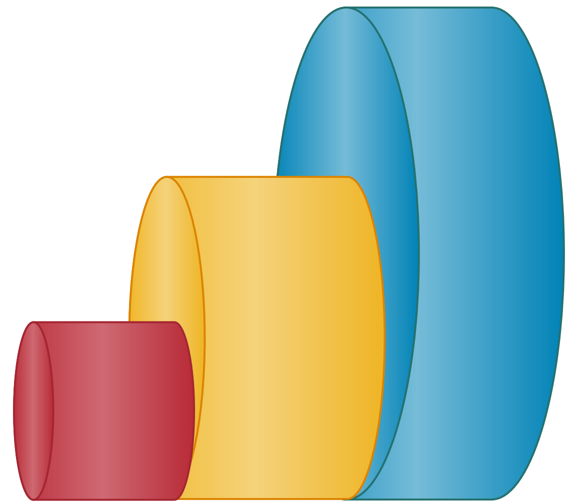
  Matching TE Router ID against ACL

  IGP mesh-group advertisement
- Each member automatically creates tunnel upon detection of a member
- Tunnels instantiated from template
- Individual tunnels not displayed in router configuration



Auto Bandwidth





- Dynamically adjust bandwidth reservation based on measured traffic
- Optional minimum and maximum limits
- Sampling and resizing timers
- Tunnel resized to largest sample since last adjustment

Configuring AutoTunnel Mesh

```
mpls traffic-eng tunnels
mpls traffic-eng auto-tunnel mesh
!
interface Auto-Template1
 ip unnumbered Loopback0
 tunnel destination mesh-group 10
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng autoroute announce
 tunnel mpls traffic-eng path-option 10 dynamic
 tunnel mpls traffic-eng auto-bw frequency 3600
!
router ospf 16
 log-adjacency-changes
 mpls traffic-eng router-id Loopback0
 mpls traffic-eng area 0
 mpls traffic-eng mesh-group 10 Loopback0 area 0
 passive-interface Loopback0
 network 172.16.0.0 0.0.255.255 area 0
!
```

Enable Auto-tunnel Mesh

Tunnel template

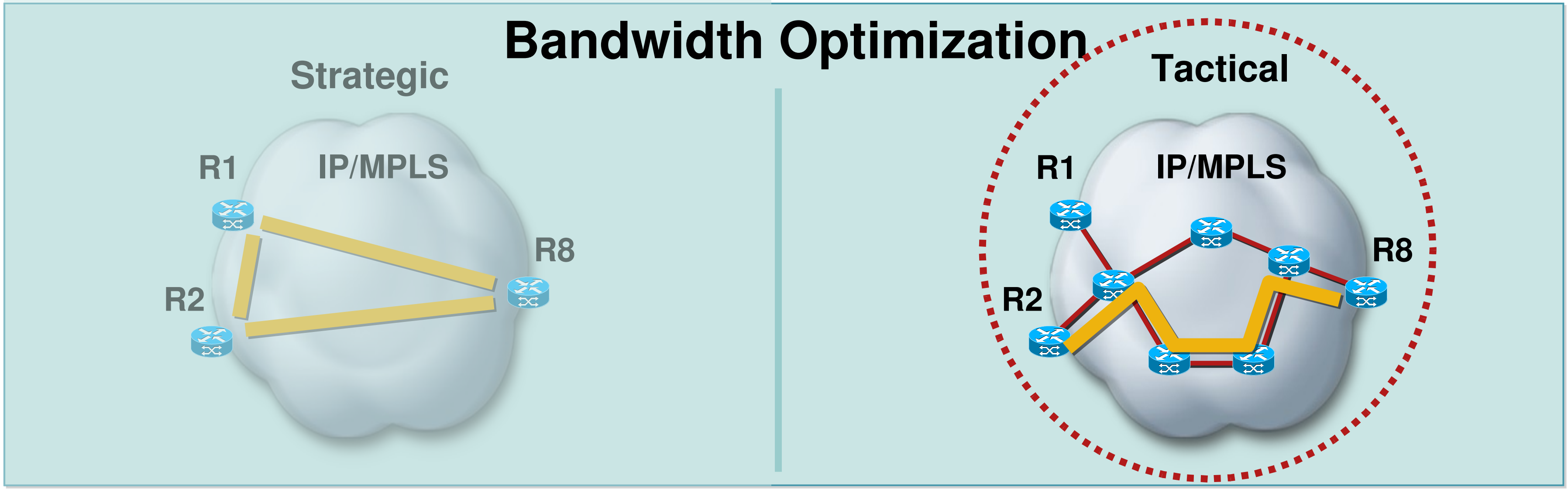
Template cloned for each member of mesh group 10

Dynamic (CSPF) path to each mesh group member

Tunnels will adjust bandwidth reservation automatically

Advertise mesh group 10 membership in area 0

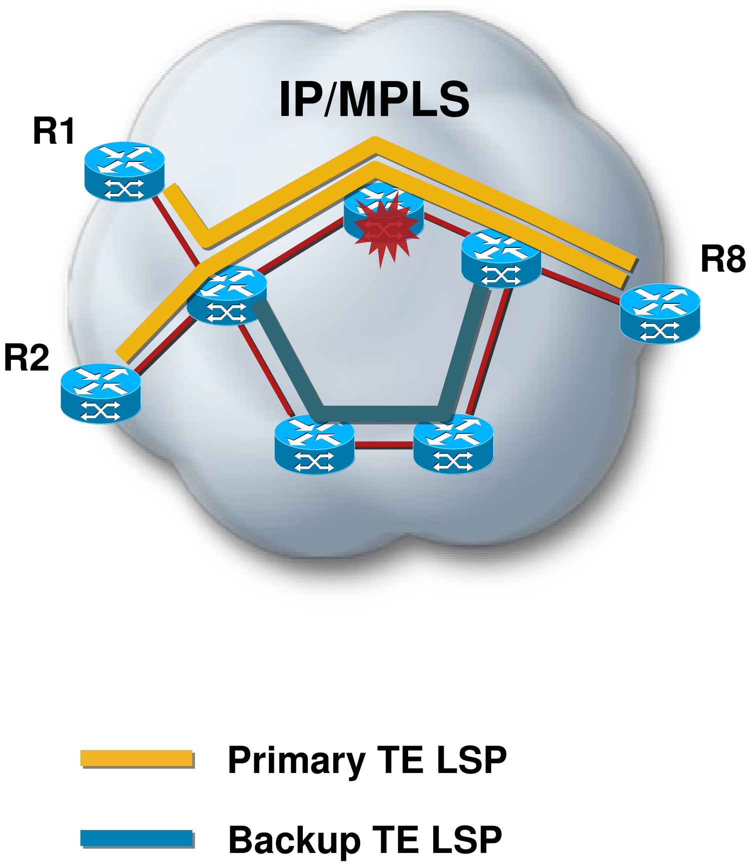
Tactical Bandwidth Optimization



- Selective deployment of tunnels when highly-utilized links are identified
- Generally, deployed until next upgrade cycle alleviates affected links

Fast Re-route

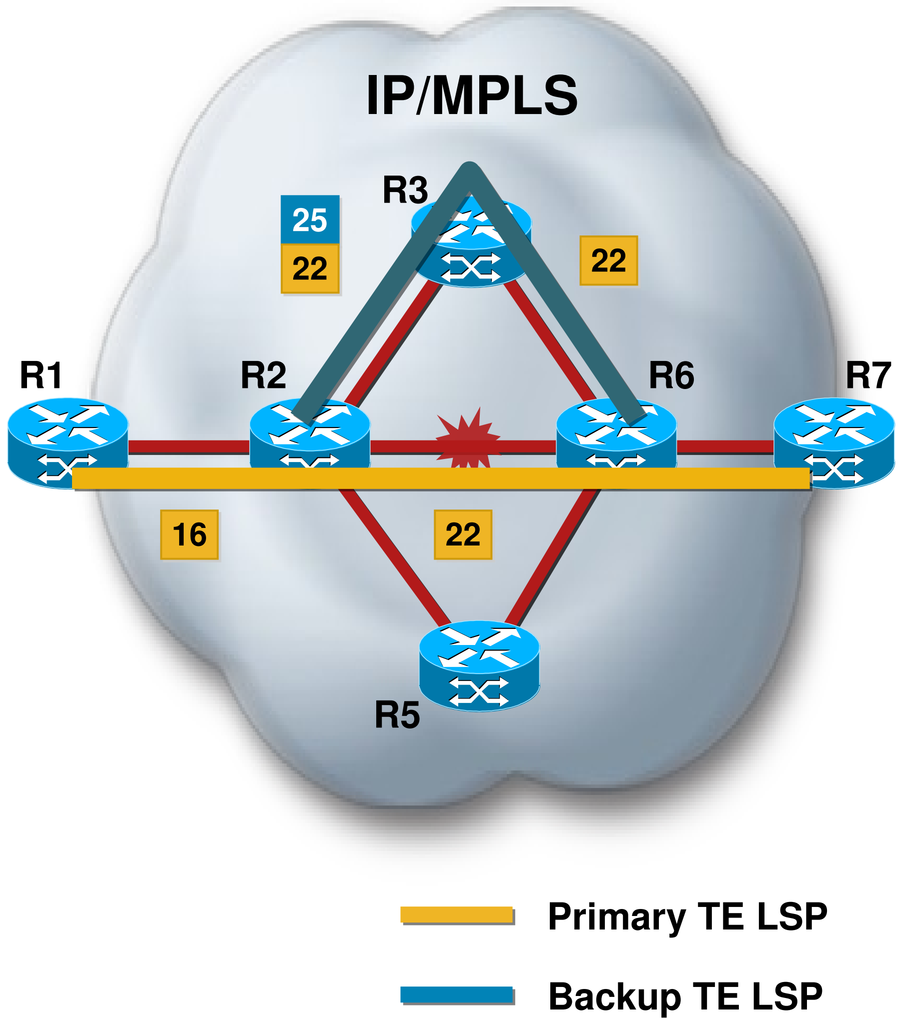
MPLS TE Fast Re-Route (FRR)



- Subsecond recovery against node/link failures
- Scalable 1:N protection
- Greater protection granularity
- Cost-effective alternative to optical protection
- Bandwidth protection

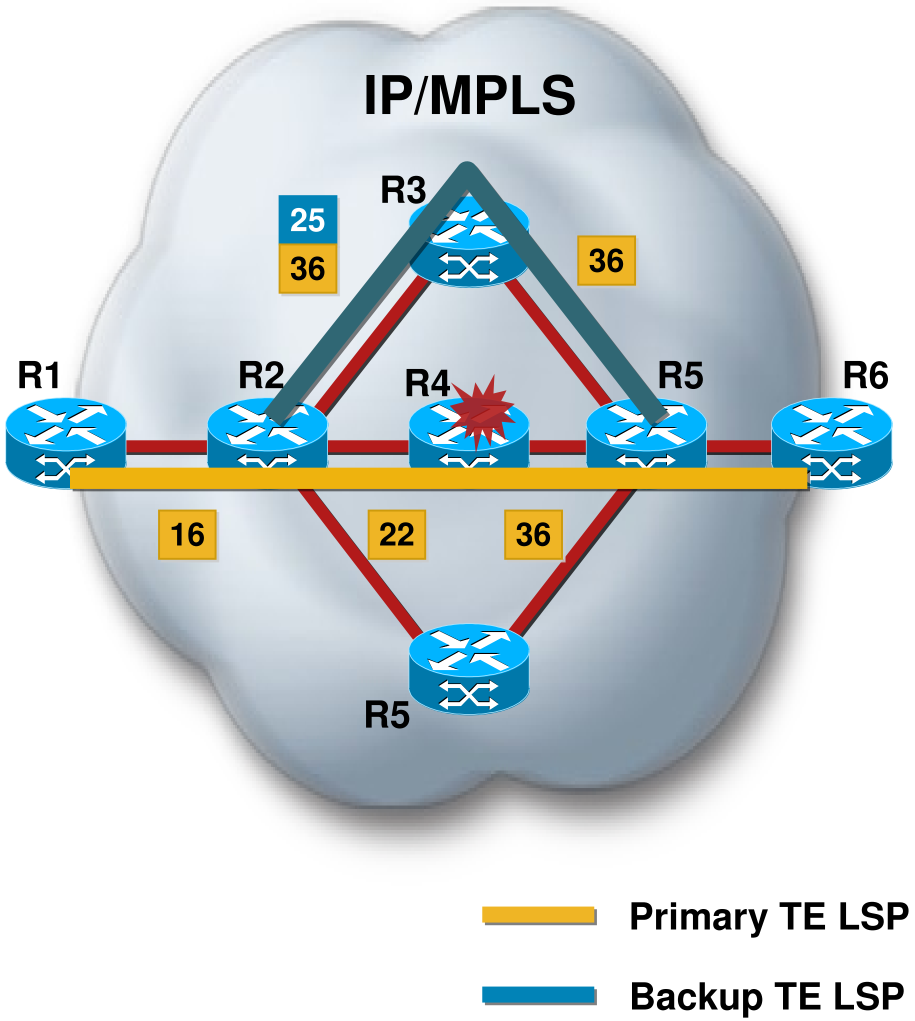
FRR Link Protection Operation

- Requires next-hop (NHOP) backup tunnel
- Point of Local Repair (PLR) swaps label and pushes backup label
- Backup terminates on Merge Point (MP) where traffic rejoins primary
- Restoration time expected under ~50 ms



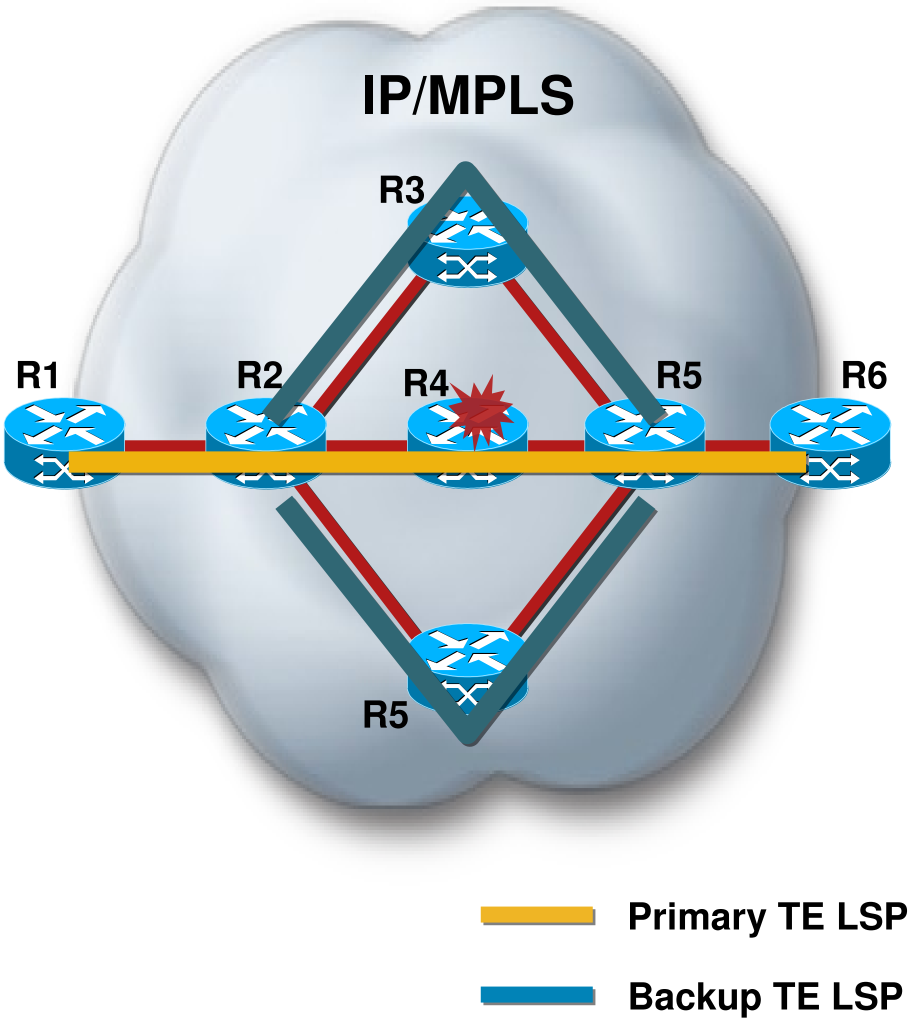
FRR Node Protection Operation

- Requires next-next-hop (NNHOP) backup tunnel
- Point of Local Repair (PLR) swaps next-hop label and pushes backup label
- Backup terminates on Merge Point (MP) where traffic rejoins primary
- Restoration time depends on failure detection time



Bandwidth Protection

- Backup tunnel with associated bandwidth capacity
- Backup tunnel may or may not actually signal bandwidth
- PLR will decide best backup to protect primary (nhop/nnhop, backup-bw, class-type, node-protection flag)



Configuring FRR

Primary Tunnel

```
interface Tunnel1
 description FROM-ROUTER-TO-DST1-FRR
 ip unnumbered Loopback0
 tunnel destination 172.16.255.2
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng bandwidth  20000
 tunnel mpls traffic-eng path-option 10 dynamic
 tunnel mpls traffic-eng fast-reroute
!
```

Indicate the desire for local protection during signaling

Backup Tunnel

```
interface Tunnel1
 description NNHOP-BACKUP
 ip unnumbered Loopback0
 tunnel destination 172.16.255.2
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng path-option 10 explicit name PATH1
!
interface POS1/0/0
 ip address 172.16.192.5 255.255.255.254
 mpls traffic-eng tunnels
 mpls traffic-eng backup-path Tunnel1
 ip rsvp bandwidth
!
```

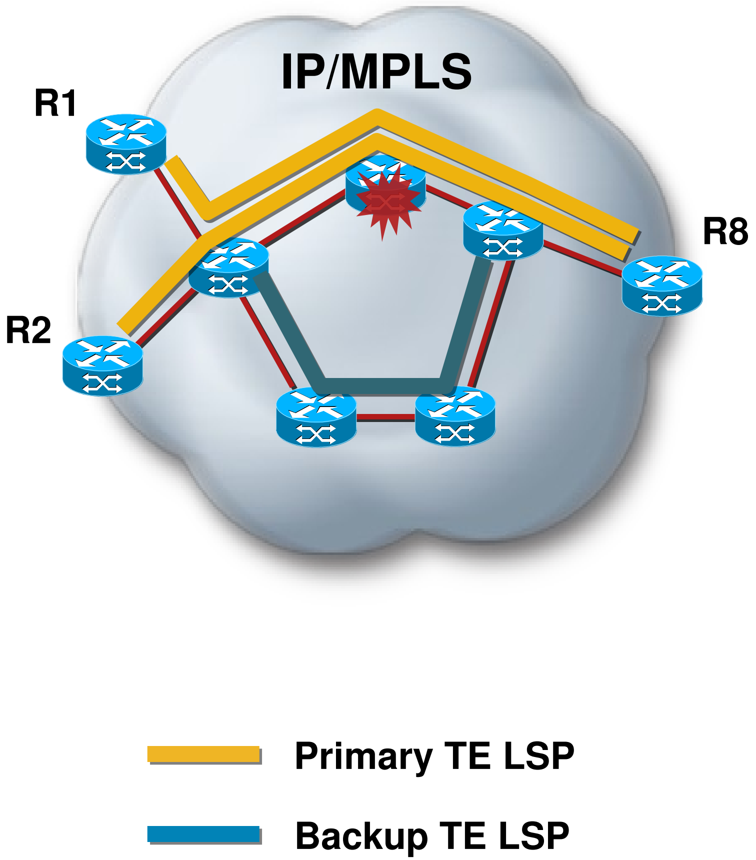
Explicitly routed backup to *172.16.255.2* with zero bandwidth

Use *Tunnel1* as backup for protected LSPs through *POS1/0/0*

AutoTunnel: Primary Tunnels

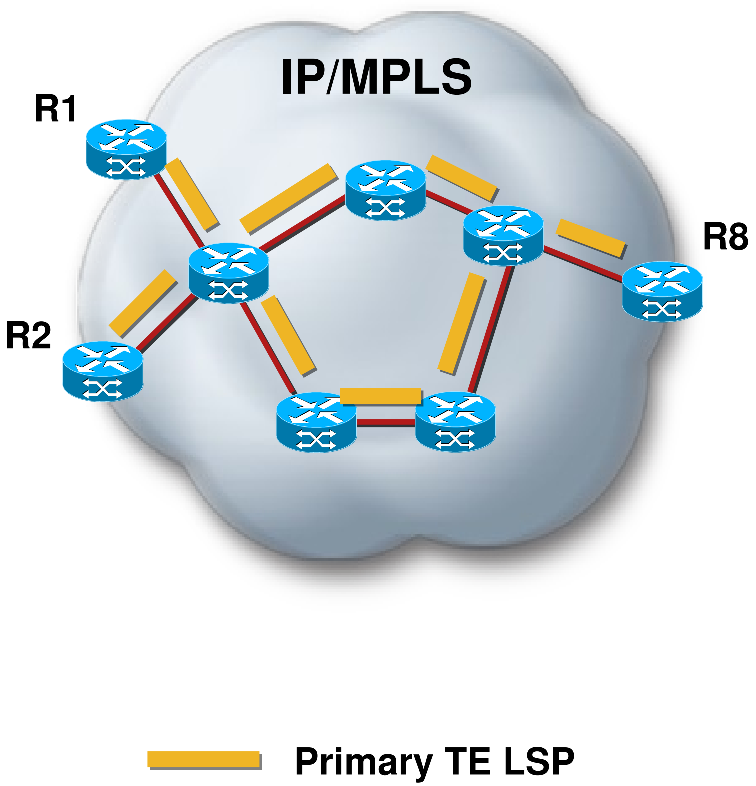
What's the Problem?

- FRR can protect TE Traffic
- No protection mechanism for IP or LDP traffic
- How to leverage FRR for all traffic?
- What if protection desired without traffic engineering?



AutoTunnel: Primary Tunnels

What's the Solution?



Forward all traffic through a one-hop protected primary TE tunnel

- Create protected one-hop tunnels on all TE links

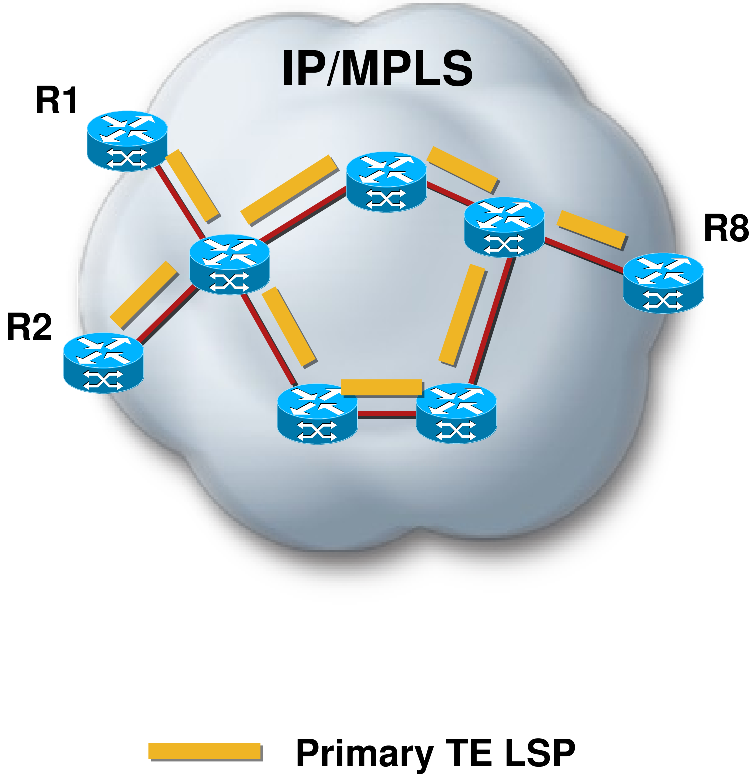
Priority	7/7
Bandwidth	0
Affinity	0x0/0xFFFF
Auto-BW	OFF
Auto-Route	ON
Fast-Reroute	ON
Forwarding-Adj	OFF
Load-Sharing	OFF

- Tunnel interfaces not shown on router configuration
- Configure desired backup tunnels (manually or automatically)

AutoTunnel: Primary Tunnels

Why One-Hop Tunnels?

- CSPF and SPF yield same results (absence of tunnel constrains)
- Auto-route forwards all traffic through one-hop tunnel
- Traffic logically mapped to tunnel but no label imposed (imp-null)
- traffic is forwarded as if no tunnel was in place



Configuring AutoTunnel Primary Tunnels

```
mpls traffic-eng tunnels
mpls traffic-eng auto-tunnel primary onehop
mpls traffic-eng auto-tunnel primary tunnel-num min 900 max 999
!
```

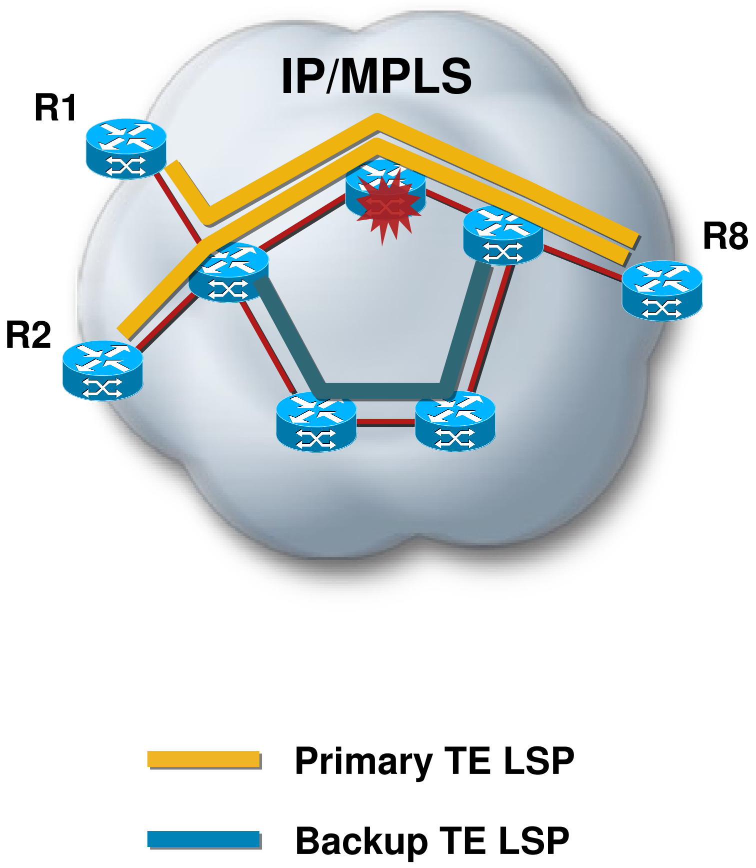
Enable auto-tunnel primary

Range for tunnel interfaces

AutoTunnel: Backup Tunnels

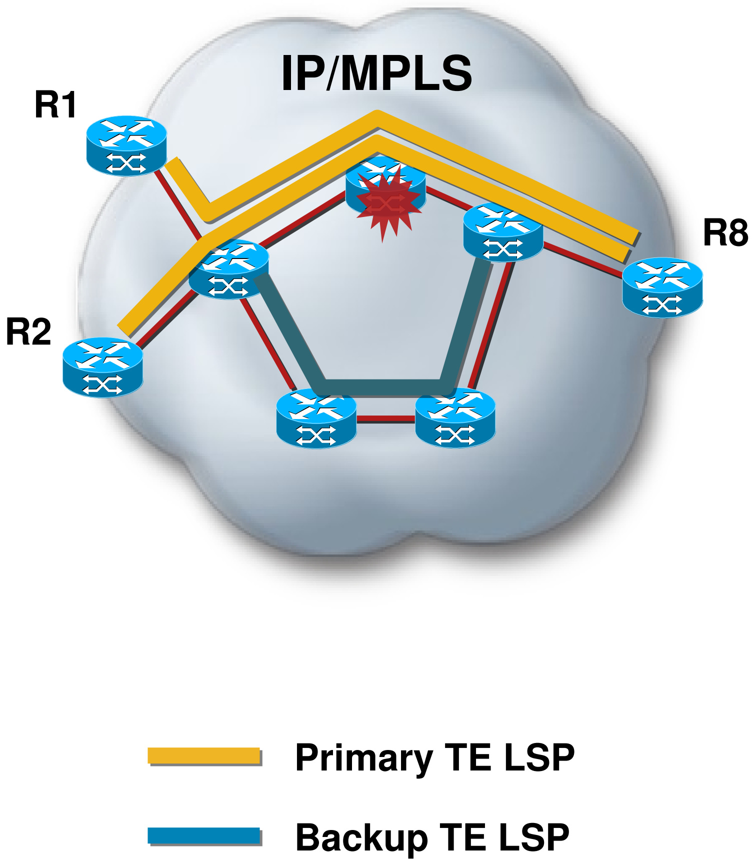
What's the Problem?

- MPLS FRR requires backup tunnels to be preconfigured
- Automation of backup tunnels is desirable



AutoTunnel: Backup Tunnels

What's the Solution?

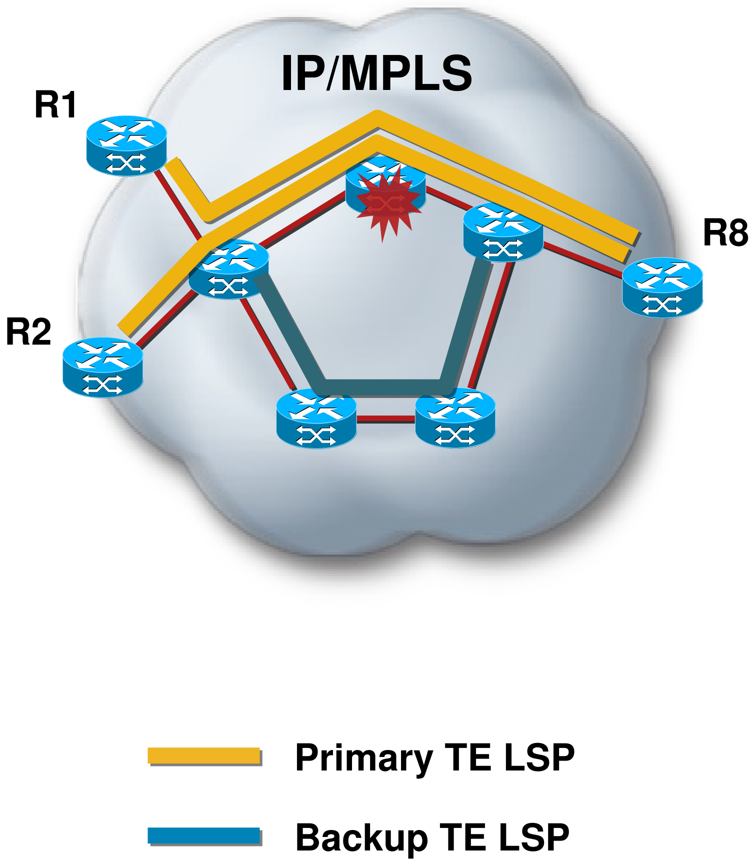


Create backup tunnels automatically as needed

- Detect if a primary tunnel requires protection and is not protected
- Verify that a backup tunnel doesn't already exist
- Compute a backup path to NHOP and NHOP excluding the protected facility
- Optionally, consider shared risk link groups during backup path computation
- Signal the backup tunnels

AutoTunnel: Backup Tunnels

What's the Solution? (Cont.)



- Backup tunnels are preconfigured

Priority	7/7
Bandwidth	0
Affinity	0x0/0xFFFF
Auto-BW	OFF
Auto-Route	OFF
Fast-Reroute	OFF
Forwarding-Adj	OFF
Load-Sharing	OFF

- Backup tunnel interfaces and paths not shown on router configuration

Configuring AutoTunnel Backup Tunnels

```
mpls traffic-eng tunnels
mpls traffic-eng auto-tunnel backup nhop-only
mpls traffic-eng auto-tunnel backup tunnel-num min 1900 max 1999
mpls traffic-eng auto-tunnel backup timers removal unused 7200
mpls traffic-eng auto-tunnel backup srlg exclude preferred
!
```

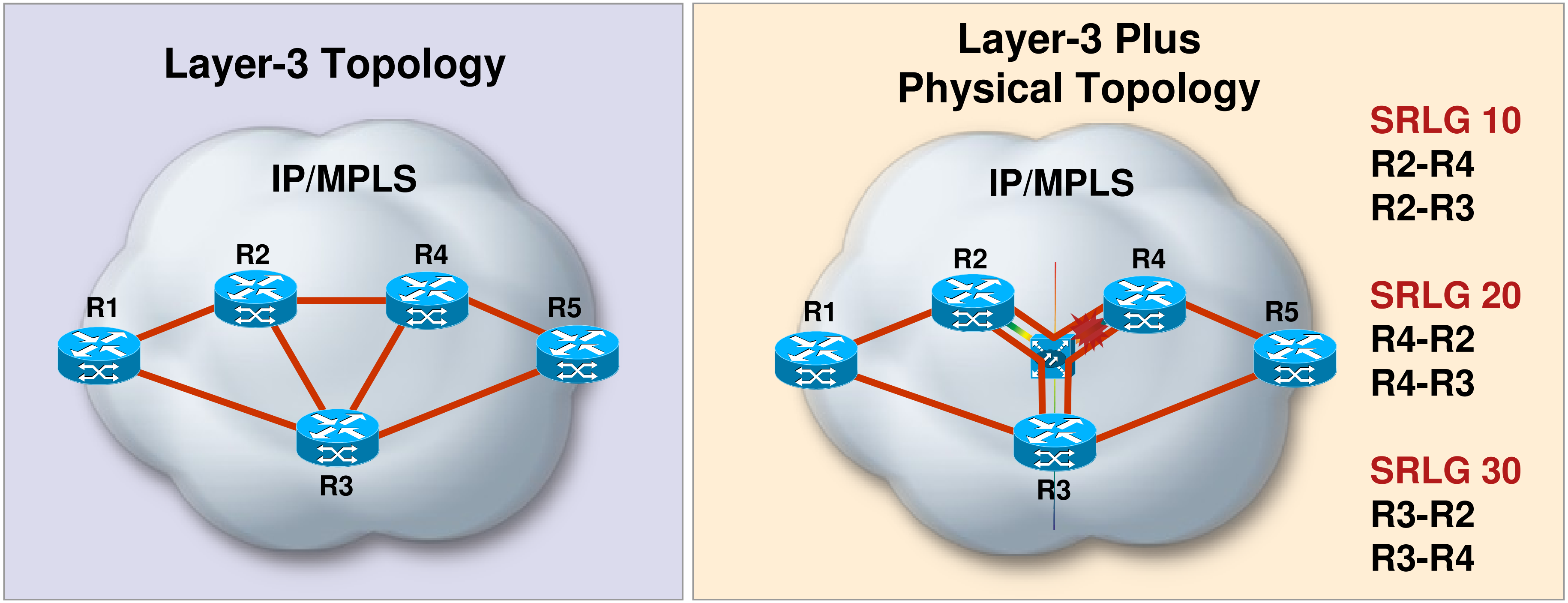
Enable auto-tunnel backup (NHOP tunnels only)

Range for tunnel interfaces

Tear down unused backup tunnels

Consider SRLGs preferably

Shared Risk Link Group (SRLG)



- Some links may share same physical resource (e.g. fiber, conduit)
- AutoTunnel Backup can force or prefer exclusion of SRLG to guarantee diversely routed backup tunnels
- IS-IS and OSPF flood SRLG membership as an additional link attribute

Configuring SRLG

```
mpls traffic-eng tunnels
mpls traffic-eng auto-tunnel backup nhop-only
mpls traffic-eng auto-tunnel backup srlg exclude force
!
interface POS0/1/0
 ip address 172.16.0.0 255.255.255.254
 mpls traffic-eng tunnels
 mpls traffic-eng srlg 15
 mpls traffic-eng srlg 25
 ip rsvp bandwidth
!
interface POS1/0/0
 ip address 172.16.0.2 255.255.255.254
 mpls traffic-eng tunnels
 mpls traffic-eng srlg 25
 ip rsvp bandwidth
!
```

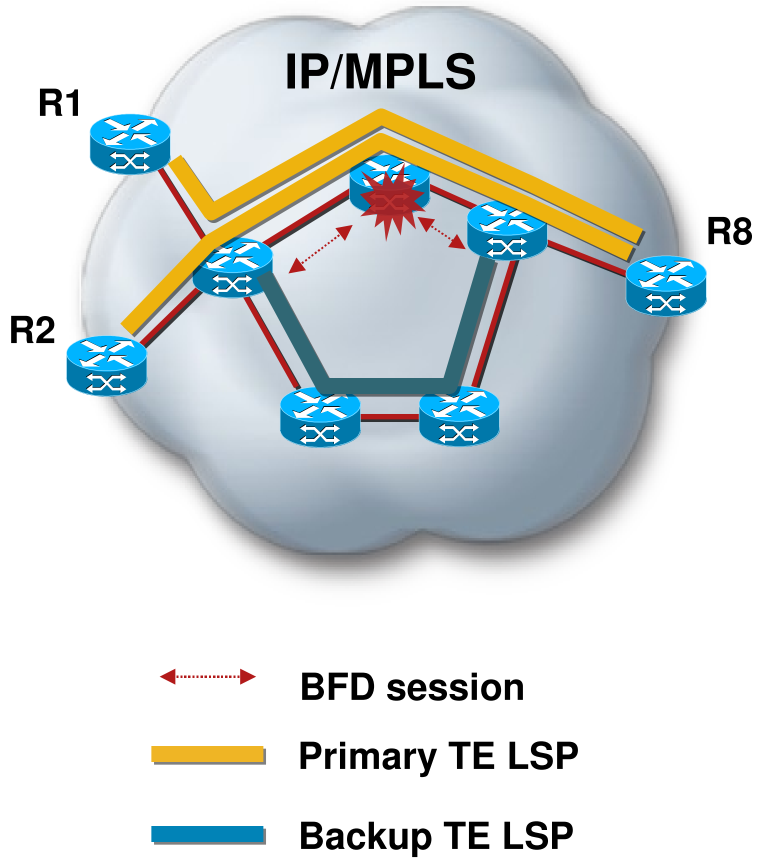
Force SRLG exclusion during backup path computation

Interface member of SRLG 15 and 25

Interface member of SRLG 25

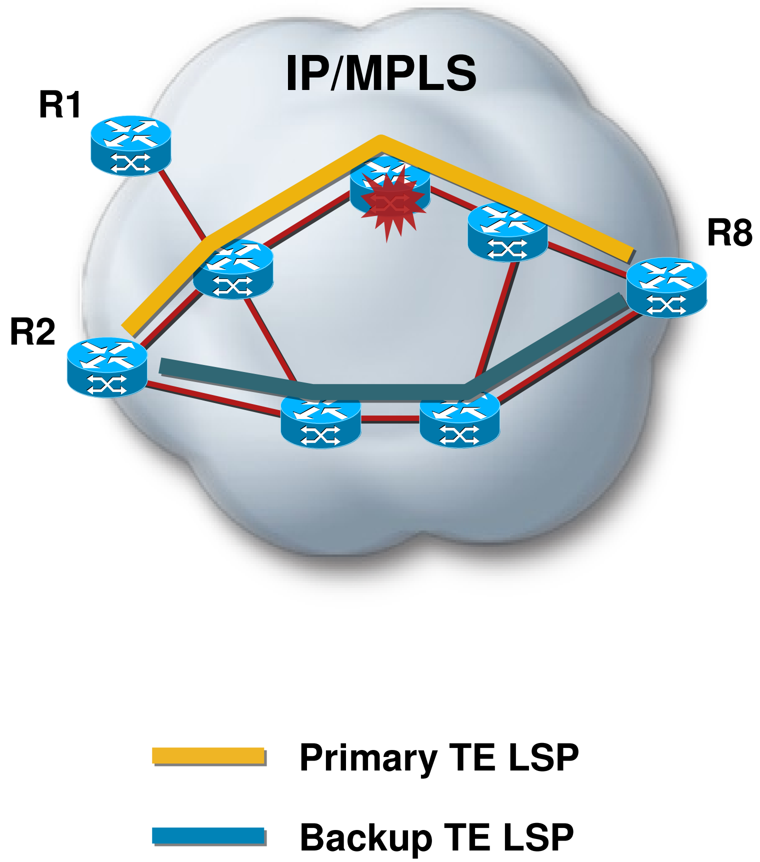
Bidirectional Forwarding Detection Trigger for FRR

- FRR relies on quick PLR failure detection
- Some failures may not produce loss of signal or alarms on a link
- BFD provides light-weight neighbor connectivity failure



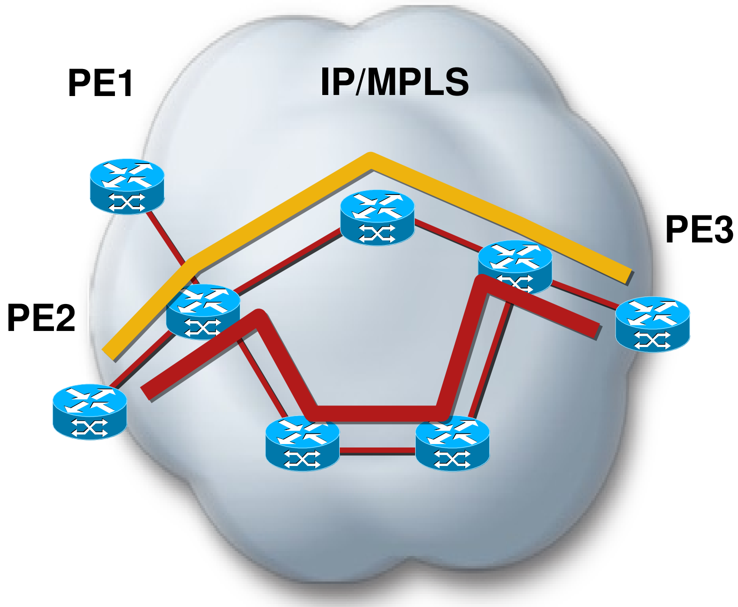
What About Path Protection?

- Primary and backup share head and tail, but diversely routed
- Expected to result in higher restoration times compared to local protection
- Doubles number of TE LSPs (1:1 protection)
- May be an acceptable solution for restricted topologies (e.g. rings)



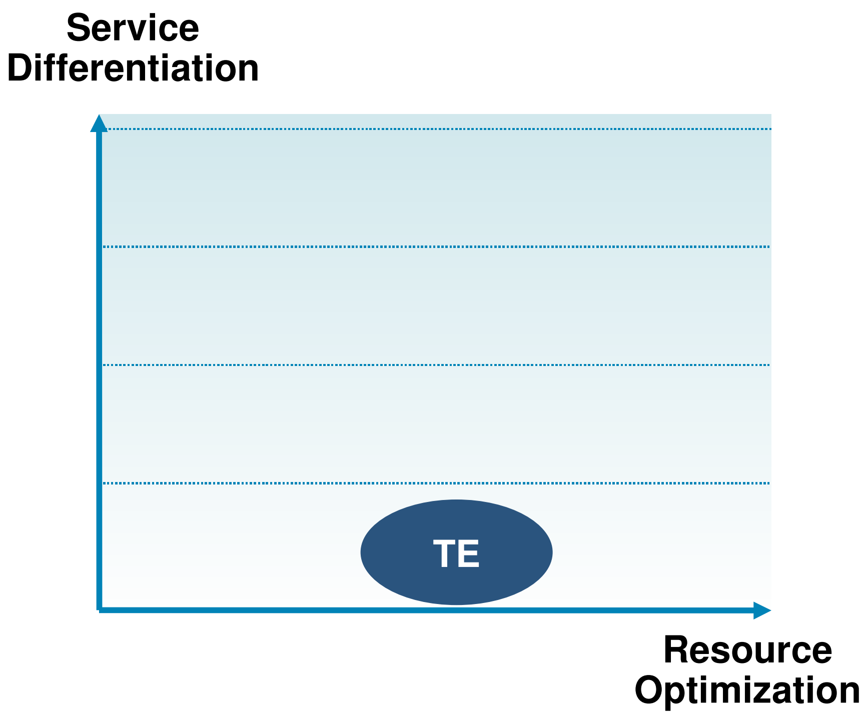
TE for QoS

Motivations

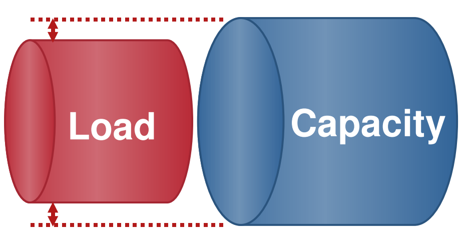


- Point-to-point SLAs
- Admission control
- Integration with DiffServ
- Increased routing control to improve network performance

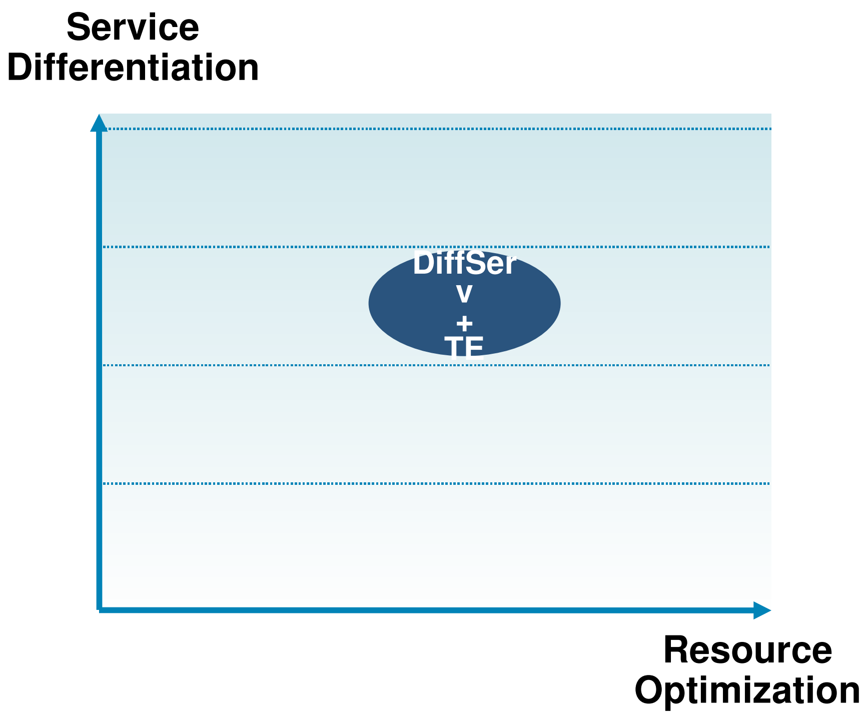
Network with MPLS TE

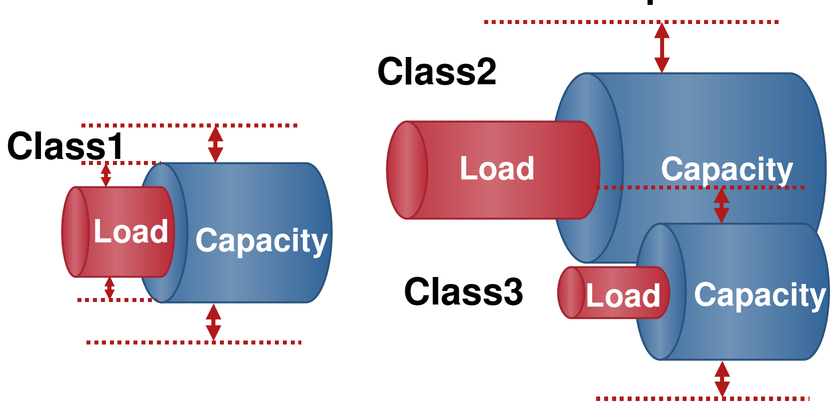


- A solution when:
  - No differentiation required
  - Optimization required
- Full mesh or selective deployment to avoid
  over-subscription
- Increased network utilization
- Adjust link load to actual
  link capacity



Network with MPLS DiffServ and MPLS TE



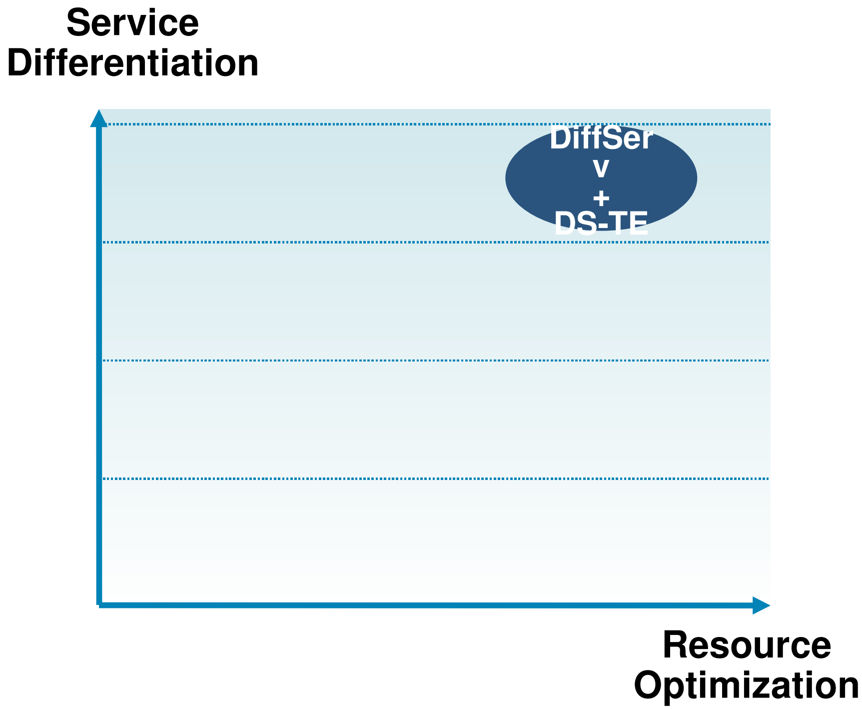


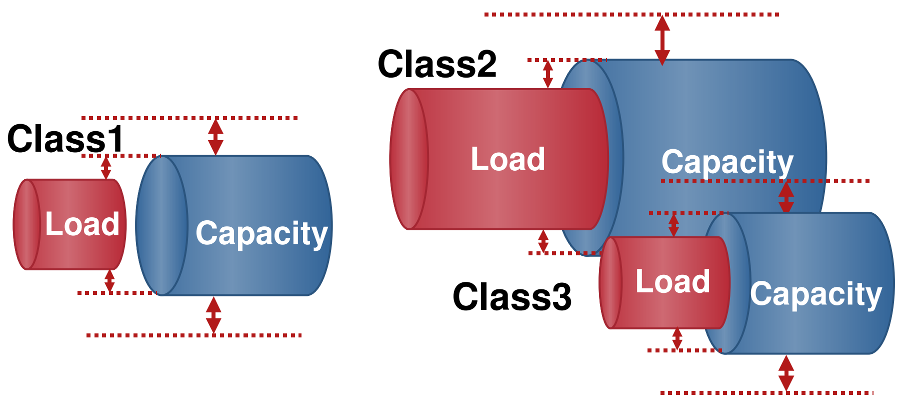
- A solution when:

  Differentiation required

  Optimization required
- Adjust class capacity to expected class load
- Adjust class load to actual class capacity for one class
- Alternatively, adjust link load to actual link capacity

Network with MPLS DiffServ and MPLS DS-TE



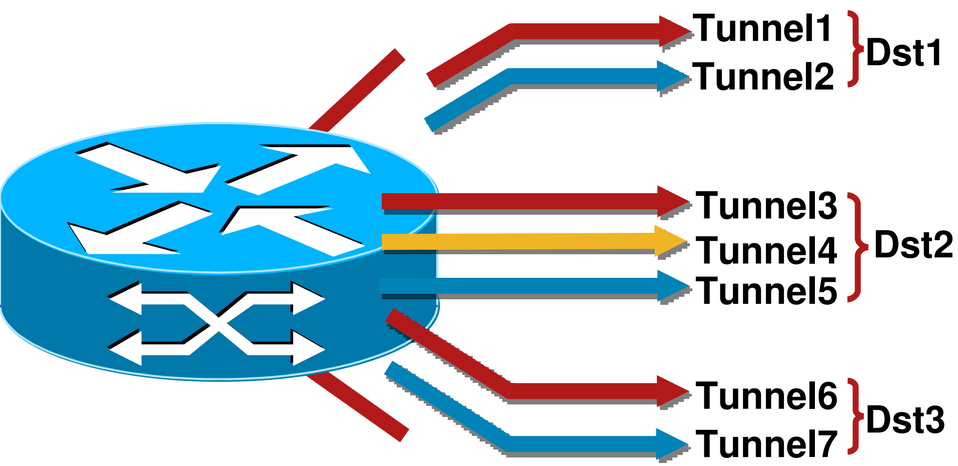


- A solution when:
  - Strong differentiation required
  - Fine optimization required
- Adjust class capacity to expected class load
- Adjust class load to actual class capacity

DiffServ-Aware TE

- Regular TE allows for one reservable bandwidth amount per link
- Regular (FIFO) queuing allows for one queue per link
- DiffServ queuing (e.g. LLQ) allows for more than one queue per link
- DS-TE allows for more than one reservable bandwidth amount per link
- Basic idea: connect PHB class bandwidth to DS-TE bandwidth sub-pool
- Still a control-plane reservation only

Class-Based Tunnel Selection: CBTS



FIB

Dst1, exp 5	Tunnel1
Dst1, *	Tunnel2
Dst2, exp 5	Tunnel3
Dst2, exp 2	Tunnel4
Dst2, *	Tunnel5
Dst3, exp 5	Tunnel6
Dst3, *	Tunnel7

***Wildcard EXP Value**

- EXP-based selection between multiple tunnels to same destination
- Local mechanism at head-end
- Tunnels configured with EXP values to carry
- Tunnels may be configured as default
- No IGP extensions
- Supports VRF traffic, IP-to-MPLS and MPLS-to-MPLS switching
- Simplifies use of DS-TE tunnels

Configuring CBTS

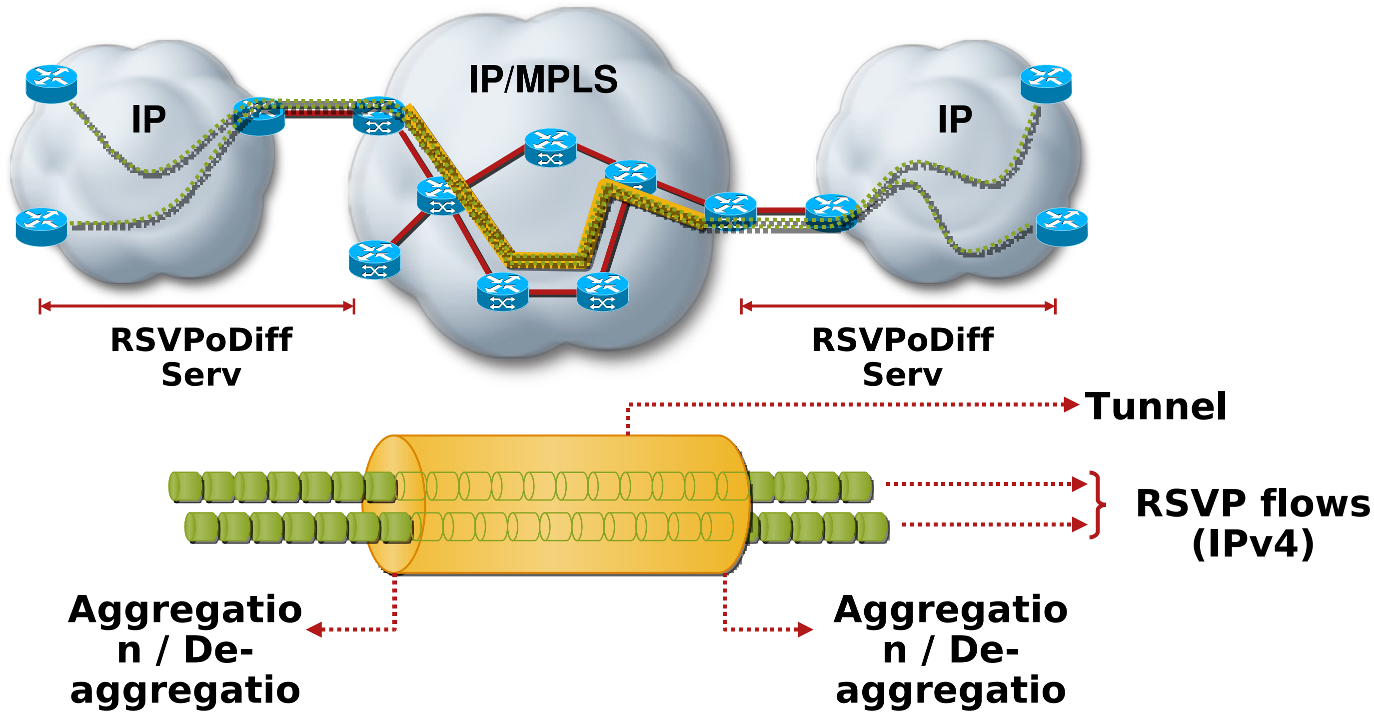
```
interface Tunnel1
 ip unnumbered Loopback0
 tunnel destination 172.16.255.3
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng priority 5 5
 tunnel mpls traffic-eng bandwidth  10000
 tunnel mpls traffic-eng path-option 10 dynamic
 tunnel mpls traffic-eng exp 5
!
interface Tunnel2
 ip unnumbered Loopback0
 tunnel destination 172.16.255.3
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng path-option 10 dynamic
 tunnel mpls traffic-eng exp default
!
ip route 192.168.0.0 255.255.255.0 Tunnel1
ip route 192.168.0.0 255.255.255.0 Tunnel2
!
```

Tunnel1 **will carry packets with MPLS EXP 5**

Tunnel2 **will carry packets with MPLS EXP other than 5**

CBTS performed on prefix *192.168.0.0/24* using *Tunnel1* and *Tunnel2*

Tunnel-based Admission Control



- Tunnel aggregates RSVP (IPv4) flows
- No per-flow state in forwarding plane (only DiffServ)
- No per-flow state in control plane within MPLS TE network
- RSVP enhancements enable end-to-end admission control solution (Receiver Proxy, Sender Notification, Fast Local Repair)

Configuring Tunnel-based Admission Control (Cisco IOS)

```
interface Tunnel1
 ip unnumbered Loopback0
 tunnel destination 172.16.255.2
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng autoroute announce
 tunnel mpls traffic-eng priority 7 7
 tunnel mpls traffic-eng bandwidth 100000
 tunnel mpls traffic-eng path-option 10 dynamic
 ip rsvp policy local default
  maximum senders 200
  maximum bandwidth single 1000
  forward all
 ip rsvp bandwidth 100000
!
interface GigabitEthernet3/3/0
 ip address 192.168.0.1 255.255.255.254
 service-policy output OUT-POLICY
 ip rsvp bandwidth percent 10
 ip rsvp listener outbound reply
 ip rsvp data-packet classification none
 ip rsvp resource-provider none
!
ip rsvp qos
!
```

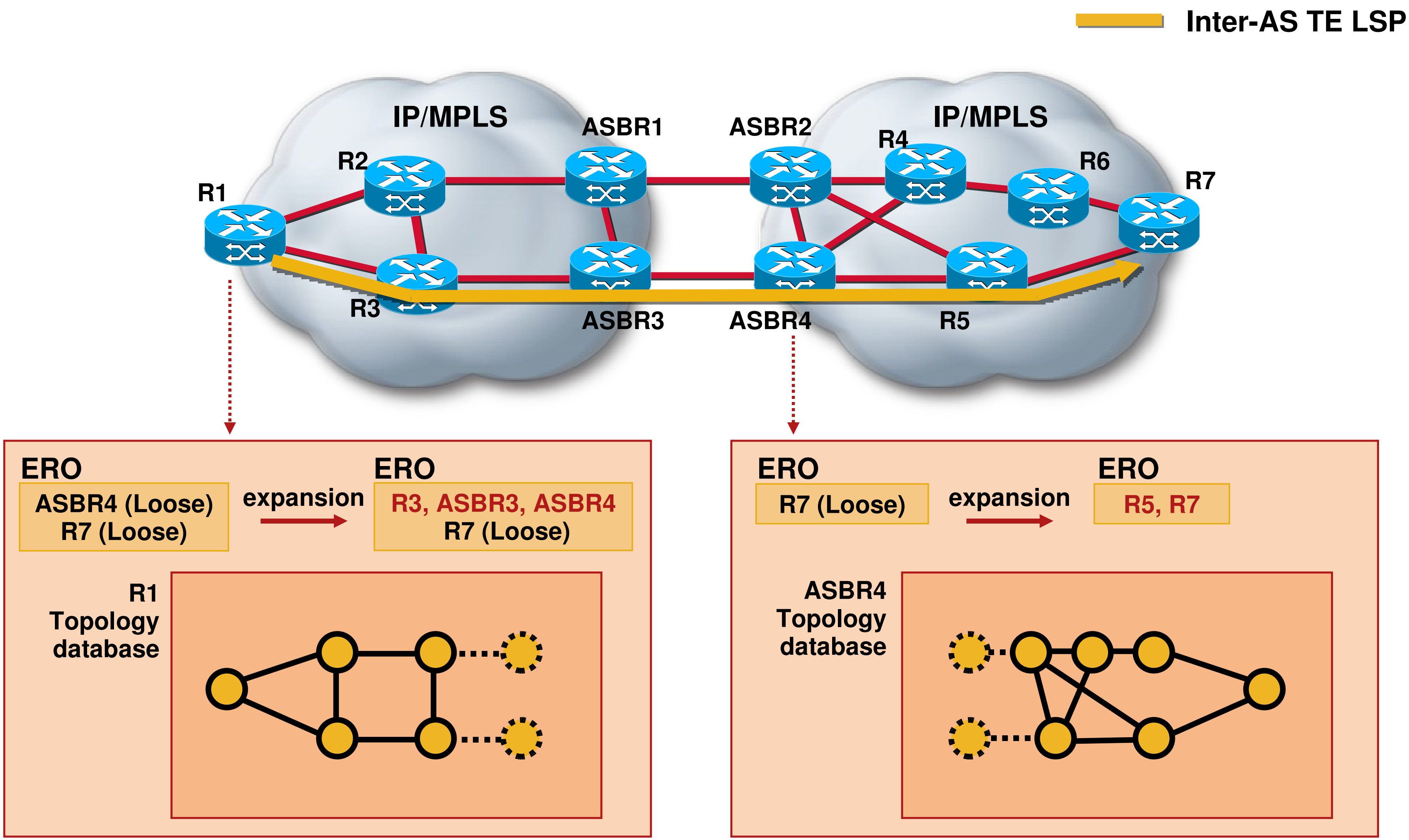
- Signaled bandwidth
- RSVP local policy (200 flows max, 1Mbps per flow max)
- Maximum reservable bandwidth
- Interface QoS policy (DiffServ)
- Maximum reservable bandwidth
- Act as RSVP receiver proxy on this interface
- No RSVP flow classification
- No RSVP flow queuing
- Enable per-flow RSVP

Inter-Domain Traffic Engineering

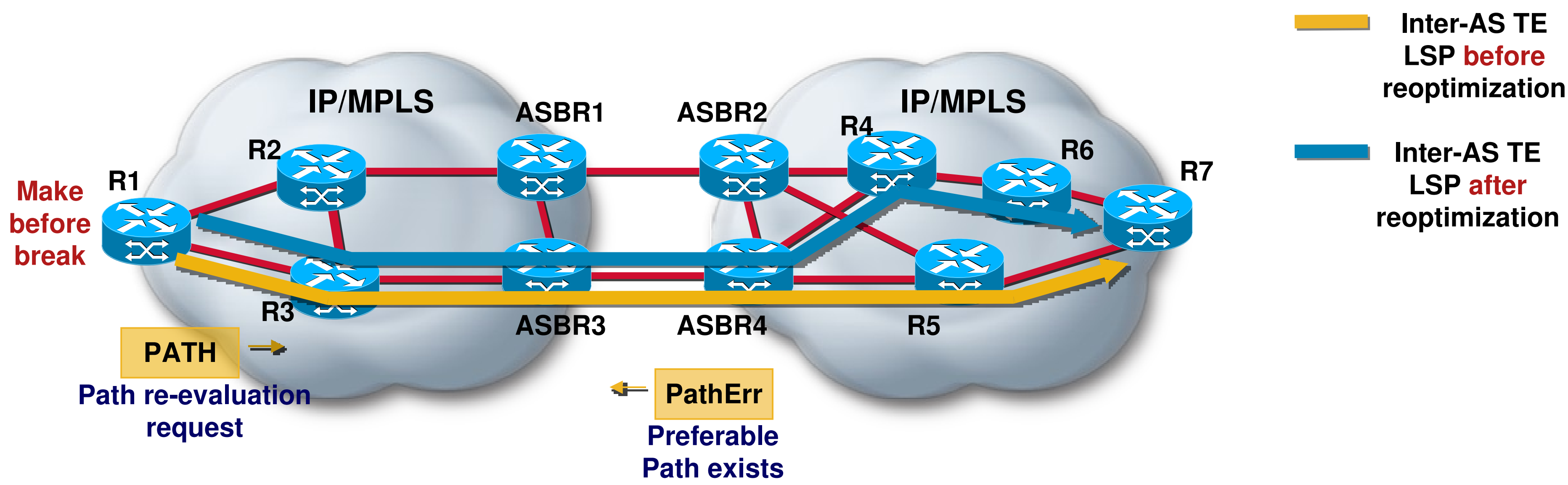
Inter-Domain Traffic Engineering: Introduction

- Domain defined as an IGP area or autonomous system
- Head end lacks complete network topology to perform path computation in both cases
- Two path computation approaches
  - Per-domain (ERO loose-hop expansion)
  - Distributed (Path Computation Element)

Per-Domain Path Computation Using ERO Loose-hop Expansion

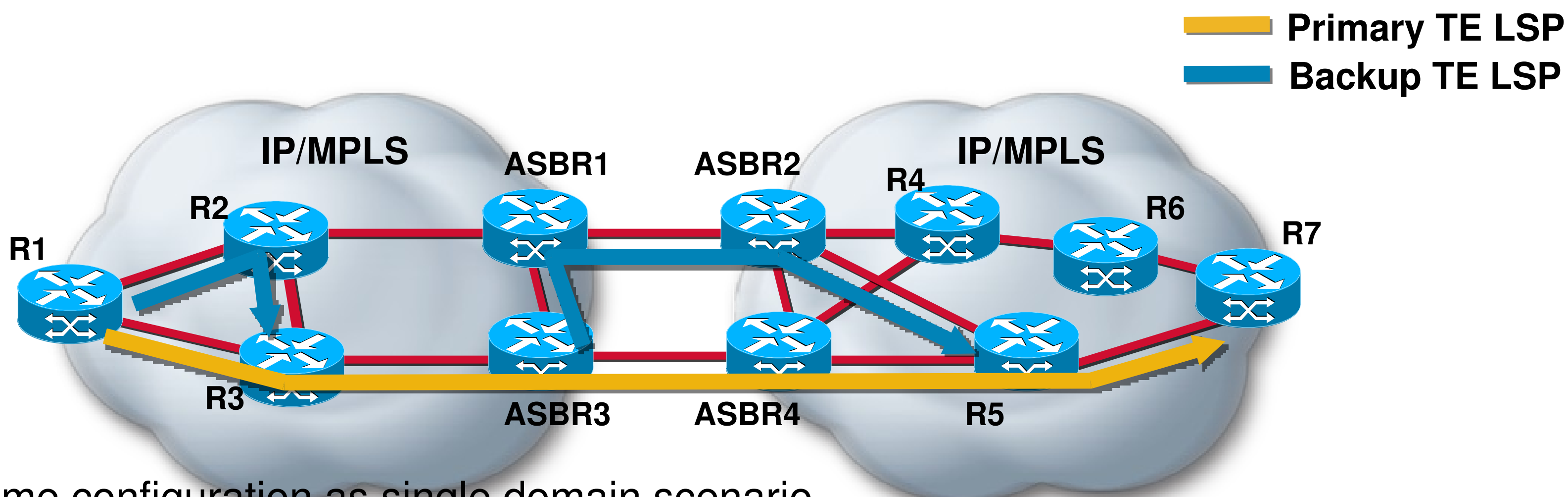


Inter-Domain TE – TE LSP Reoptimization



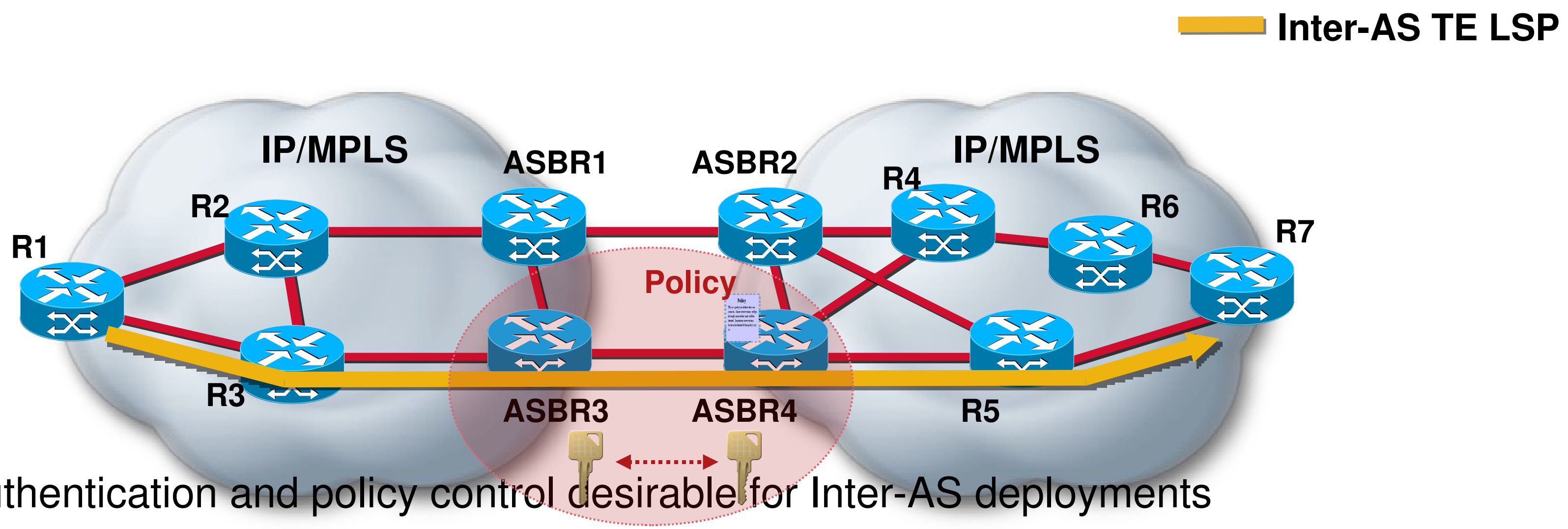
- Reoptimization can be timer/event/admin triggered
- Head end sets 'path re-evaluation request' flag (SESSION_ATTRIBUTE)
- Head end receives PathErr message notification from boundary router if a preferable path exists
- Make-before-break TE LSP setup can be initiated after PathErr notification

Inter-Domain TE – Fast Re-route



- Same configuration as single domain scenario
- Support for node-id sub-object required to implement ABR/ASBR node protection
- Node-id helps point of local repair (PLR) detect a merge point (MP)

Inter-Domain TE – Authentication and Policy Control



- Authentication and policy control desirable for Inter-AS deployments
- ASBR may perform RSVP authentication (MD5/SHA-1)
- ASBR may enforce a local policy for Inter-AS TE LSPs (e.g. limit bandwidth, message types, protection, etc.)

Configuring Inter-AS Tunnels

```
mpls traffic-eng tunnels
!
interface Tunnel1
 ip unnumbered Loopback0
 no ip directed-broadcast
 tunnel destination 172.31.255.5
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng priority 7 7
 tunnel mpls traffic-eng bandwidth  1000
 tunnel mpls traffic-eng path-option 10 explicit name LOOSE-PATH
!
ip route 172.31.255.5 255.255.255.255 Tunnel1
!
ip explicit-path name LOOSE-PATH enable
 next-address loose 172.24.255.1
 next-address loose 172.31.255.1
!
```

Loose-hop path

Static route mapping IP traffic to *Tunnel1*

List of ASBRs as loose hops

Configuring Inter-AS TE at ASBR

```
mpls traffic-eng tunnels
!
key chain A-ASBR1-key
 key 1
  key-string 7 151E0E18092F222A
!
interface Serial1/0
 ip address 192.168.0.1 255.255.255.252
 mpls traffic-eng tunnels
 mpls traffic-eng passive-interface nbr-te-id 172.16.255.4 nbr-igp-id ospf 172.16.255.4
 ip rsvp bandwidth
 ip rsvp authentication key-chain A-ASBR1-key
 ip rsvp authentication type sha-1
 ip rsvp authentication
!
router bgp 65024
 no synchronization
 bgp log-neighbor-changes
 neighbor 172.24.255.3 remote-as 65024
 neighbor 172.24.255.3 update-source Loopback0
 neighbor 192.168.0.2 remote-as 65016
 no auto-summary
!
ip rsvp policy local origin-as 65016
 no fast-reroute
 maximum bandwidth single 10000
 forward all
!
```

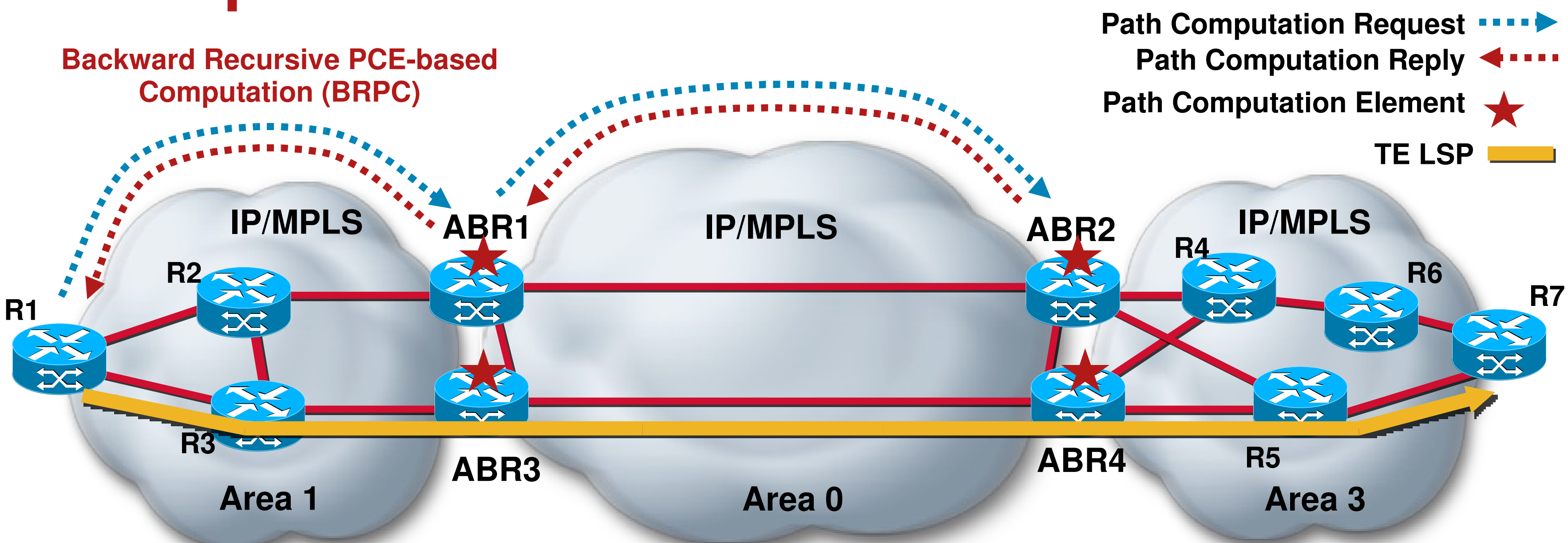
Authentication key

Add ASBR link to TE topology database

Enable RSVP authentication

Process signaling from AS 65016 if FRR not requested and 10M or less

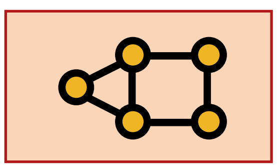
Distributed Path Computation Using Path Computation Element



R1

Path (cost 500):
R3, ABR3, ABR4, R5, R7

R1
Topology
database

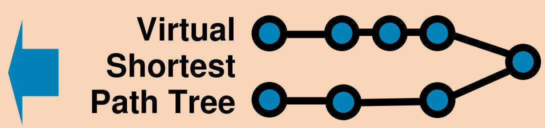


ABR1

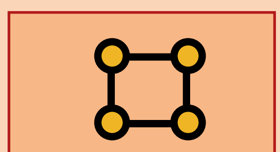
Path1 (cost 400): ABR1, ABR2, R4, R6 R7

Path2 (cost 300): ABR3, ABR4, R5, R7

Virtual
Shortest
Path Tree



ABR1
Topology
database
(area 0)



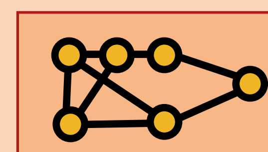
ABR2

Path1 (cost 300): ABR2, R4, R6 R7

Path2 (cost 200): ABR4, R5, R7

Virtual
Shortest
Path Tree

ABR2
Topology
database
(area 3)



Distributed Path Computation with Backward Recursive PCE-Based Computation (BRPC)

- Head-end sends request to a path computation element (PCE)
- PCE recursively computes virtual shortest path tree (SPT) to destination
- Head-end receives reply with virtual SPT if a path exists
- Head-end uses topology database and virtual SPT to compute end-to-end path
- Head-end can discover PCEs dynamically or have them configured statically

Configuring PCE

Headend

```
interface tunnel-te1
 description FROM-ROUTER-TO-DST2
 ipv4 unnumbered Loopback0
 destination 172.16.255.1
 path-option 10 dynamic pce
!
router static
 address-family ipv4 unicast
  172.16.255.1/32 tunnel-te1
 !
!
```

- **path-option 10 dynamic pce**: Use discovered PCEs for path computation
- **address-family ipv4 unicast 172.16.255.1/32 tunnel-te1**: Static route mapping IP traffic to `tunnel-te1`

PCE

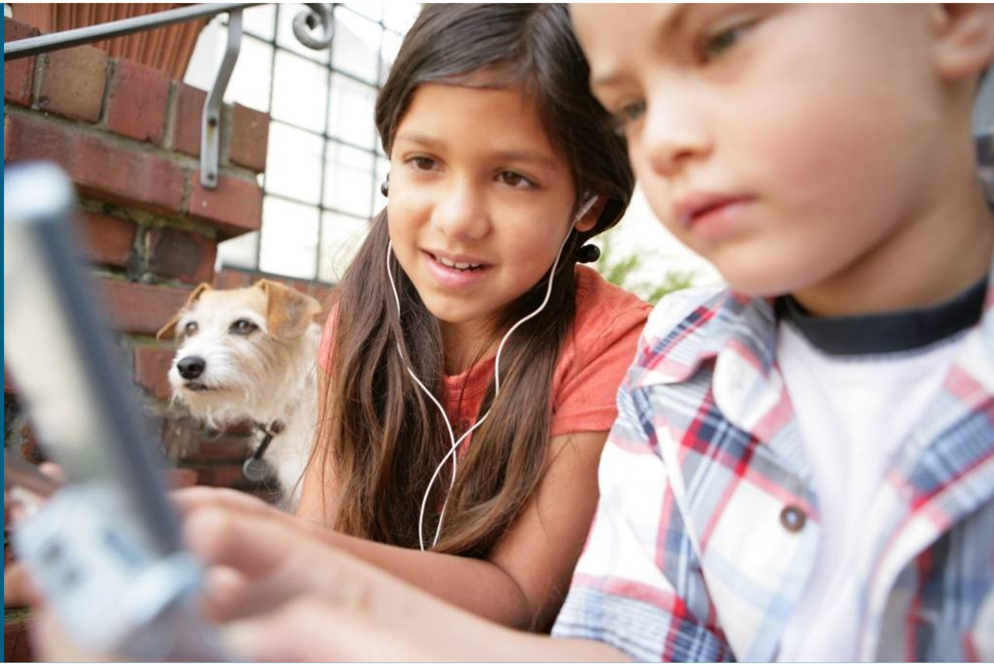
```
mpls traffic-eng
 pce deadtimer 30
 pce address ipv4 172.16.255.129
 pce keepalive 10
!
```

- **pce deadtimer 30**: Declare peer down if no keepalive in 30s
- **pce address ipv4 172.16.255.129**: Advertise PCE capability with address 172.16.255.129
- **pce keepalive 10**: Send per keepalive every 10s

Inter-Domain TE
Take into Account before Implementing

- Semantics of link attributes across domain boundaries
- Semantics of TE-Classes across domain boundaries for DS-TE
- Auto-route not possible for traffic selection

Deployment Best Practices

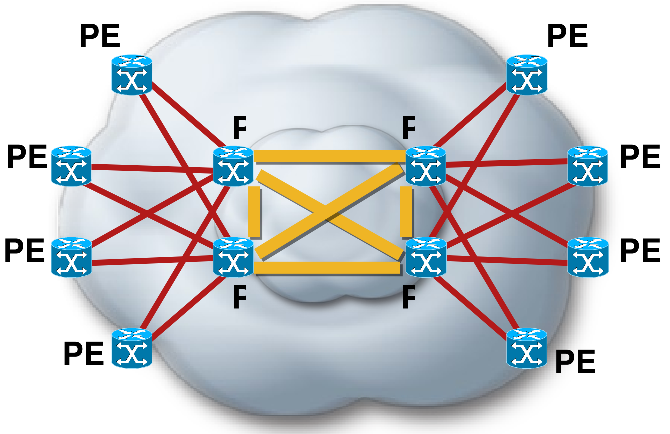


Should RSVP-TE and LDP Be Used Simultaneously?

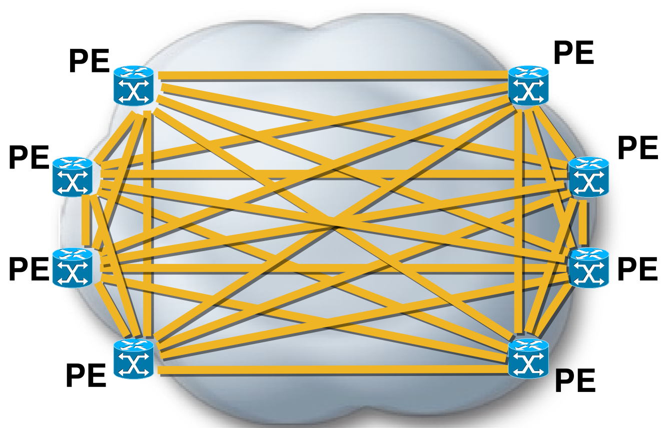
- Guarantees forwarding of VPN traffic if a TE LSP fails
- May be required if full mesh of TE LSPs not in use
- Increased complexity

How Far Should Tunnels Span?

12 TE LSP



56 TE LSP



- PE-to-PE Tunnels

  More granular control on traffic forwarding

  Larger number of TE LSPs

- P-to-P Tunnels

  Requires IP tunnels or LDP over TE tunnels to carry VPN traffic

  Fewer TE LSPs

MPLS TE on Link Bundles

- Different platforms support different link bundles
  - Ethernet
  - POS
  - Multilink PPP
- Bundles appear as single link in topology database
- Same rules for link state flooding
- Hard TE LSP preemption if bundle bandwidth becomes insufficient
- Configurable minimum number of links to maintain bundle active
- Bundle failure can act as trigger for FRR

MPLS TE on Ethernet Bundle (Cisco IOS)

```
interface Port-channel1
 ip address 172.16.0.0 255.255.255.254
 mpls traffic-eng tunnels
 mpls traffic-eng attribute-flags 0xF
 mpls traffic-eng administrative-weight 20
 ip rsvp bandwidth percent 100
!
interface GigabitEthernet2/0/0
 no ip address
 channel-protocol lacp
 channel-group 1 mode active
!
interface GigabitEthernet2/0/1
 no ip address
 channel-protocol lacp
 channel-group 1 mode active
!
```

- Enable MPLS TE on this interface
- Attribute flags
- TE metric
- Maximum reservable bandwidth (100% of total bundle bandwidth)
- LACP as channel protocol
- Associate with `Port-channel1` and enable LACP (non-passive)
- LACP as channel protocol
- Associate with `Port-channel1` and enable LACP (non-passive)

Scaling Signaling (Refresh Reduction)

	MSG_Id	Path State
LSP1	22	…
LSP2	62	…
:	:	…
LSPn	94	…

	MSG_Id	Resv State
LSP1	43	…
LSP2	37	…
:	:	…
LSPn	29	…

- Message Identifier associated with Path/Resv state
- Summary Refresh (SRefresh) message with message_id list to refresh soft state
- SRefresh only replaces refresh Path/Resv messages

Summary

- Technology Overview
  - Explicit and constrained-based routing
  - TE protocol extensions (OSPF, ISIS and RSVP)
  - P2P and P2MP TE LSP
- Bandwidth optimization
  - Strategic (full mesh, auto-tunnel)
  - Tactical
- Traffic Protection
  - Link/node protection (auto-tunnel)
  - Bandwidth protection
- TE for QoS
  - DS-TE (MAM, RDM)
  - CBTS
- Inter-Domain Traffic Engineering
  - Inter-Area
  - Inter-AS (Authentication, policy control)
- General Deployment Considerations
  - MPLS TE and LDP
  - PE-to-PE vs. P-to-P tunnels
  - TE over Bundles
  - Scaling signaling

MPLS Layer 2 VPN

Agenda

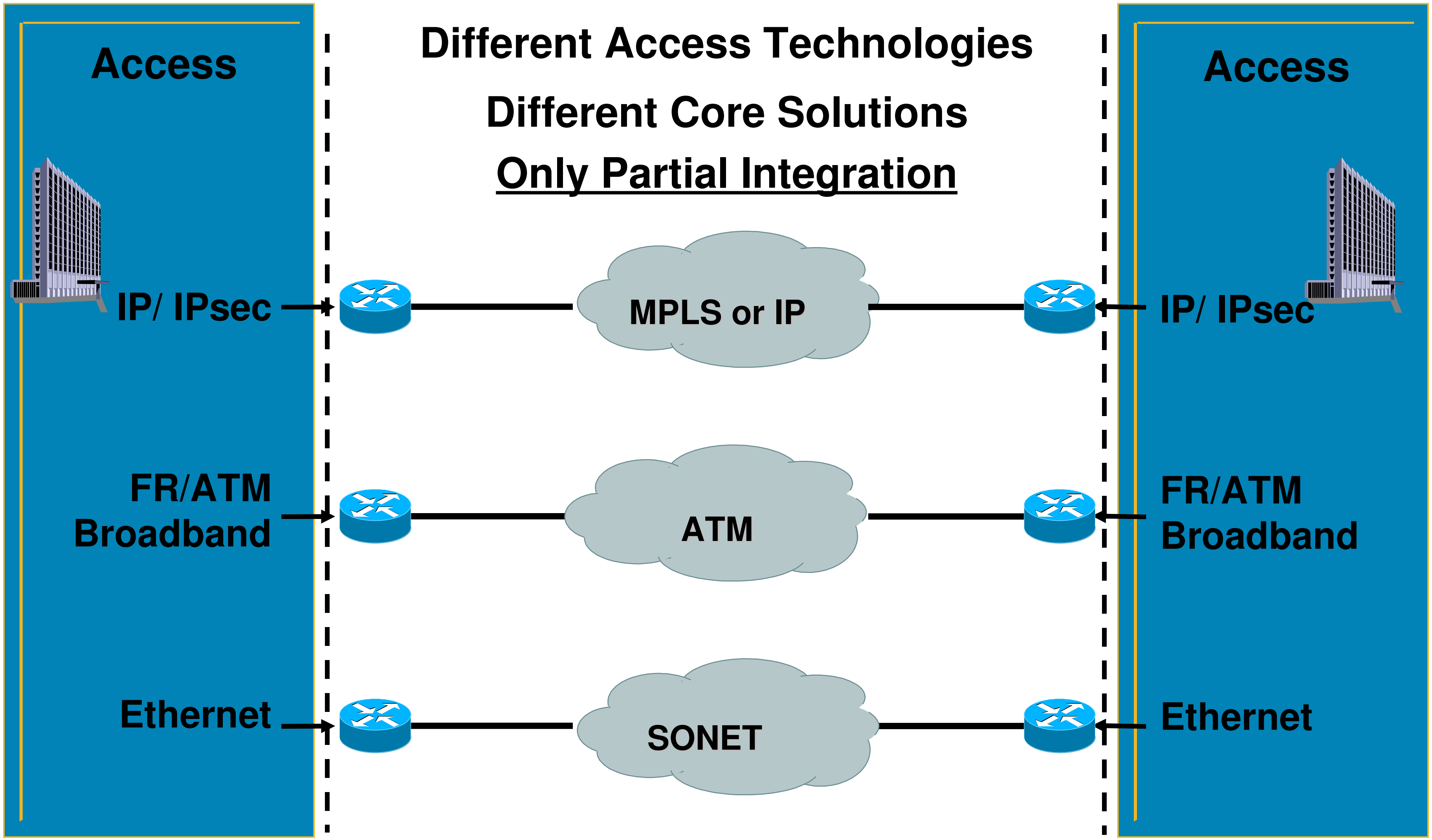
- L2VPN Technology Overview
    - L2VPN Fundamentals
    - PWE3 Signaling Concepts
    - L2VPN Transports
    - EVC Infrastructure
- VPLS
    - VPLS Fundamentals
    - H-VPLS Deployment Models
- EoMPLS/VPLS Network Resiliency

L2VPN Technology Overview

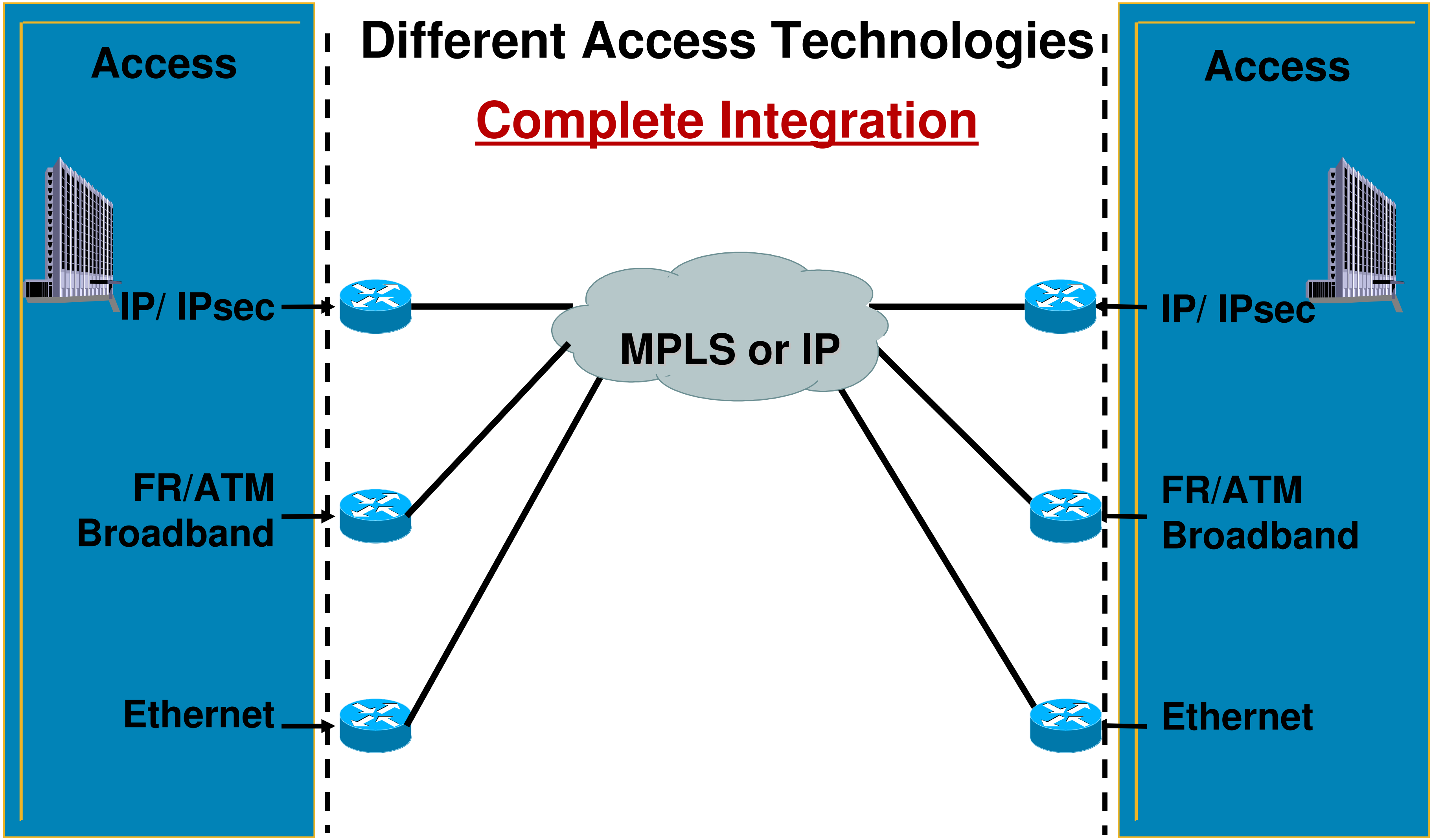
VPN – Types, Layers and Implementations

VPN Type	Layer	Implementation
Leased Line	1	TDM/SDH/SONET
Frame Relay switching	2	DLCI
ATM switching	2	VC/VP
Ethernet/ATM/FR	2	VPWS/VPLS
GRE/UTI/L2TPv3	3	IP Tunnel
IP	3	MP-BGP/RFC2547
IP	3	IPSec

VPN Deployments Today
Technology & VPN Diversity



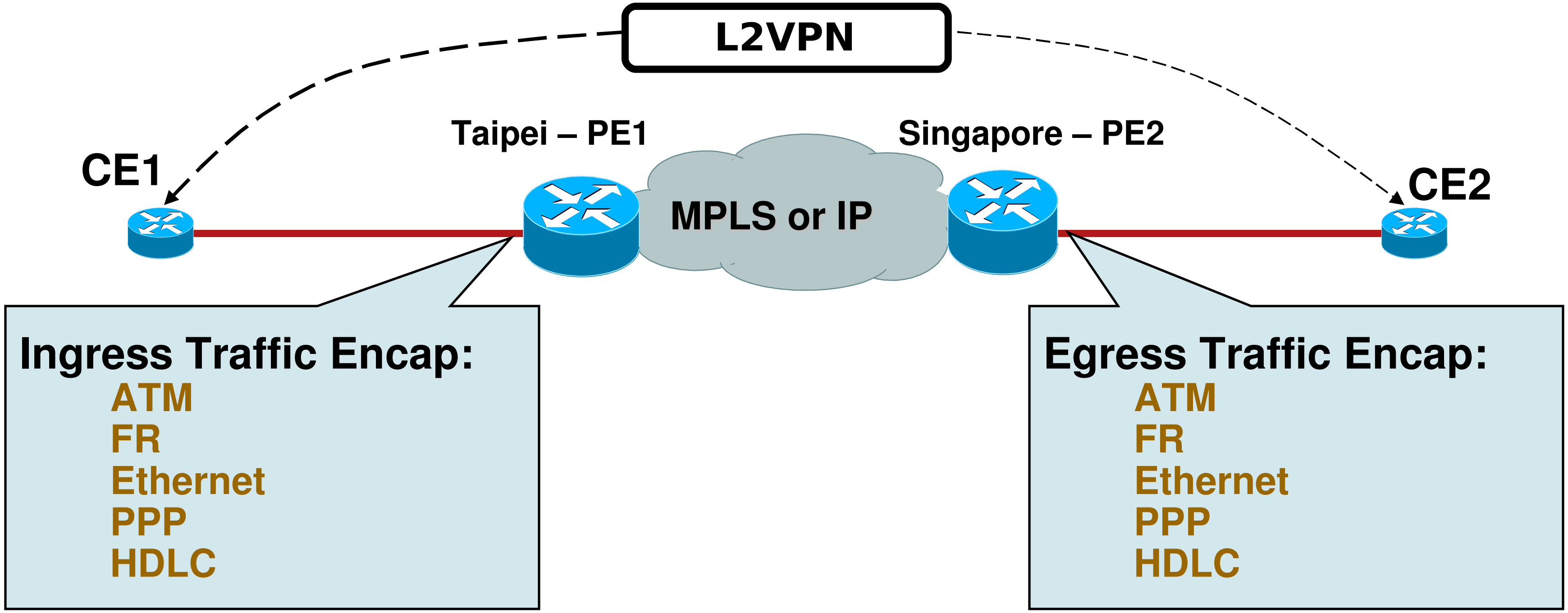
Consolidated Core supports …



Why is L2VPN needed?

- Allows SP to have a single infrastructure for both IP and legacy services
  - Migration
  - Provisioning is incremental
  - Network Consolidation
  - Capital and Operational savings
- Customer can have their own routing, qos policies, security mechanisms, etc
  - Layer 3 (IPv4, IPX, OSPF, BGP, etc …) on CE routers is transparent to MPLS core
  - CE1 router sees CE2 router as next-hop
  - No routing involved with MPLS core
- Open architecture and vendor interoperability

L2VPN - Simple definition



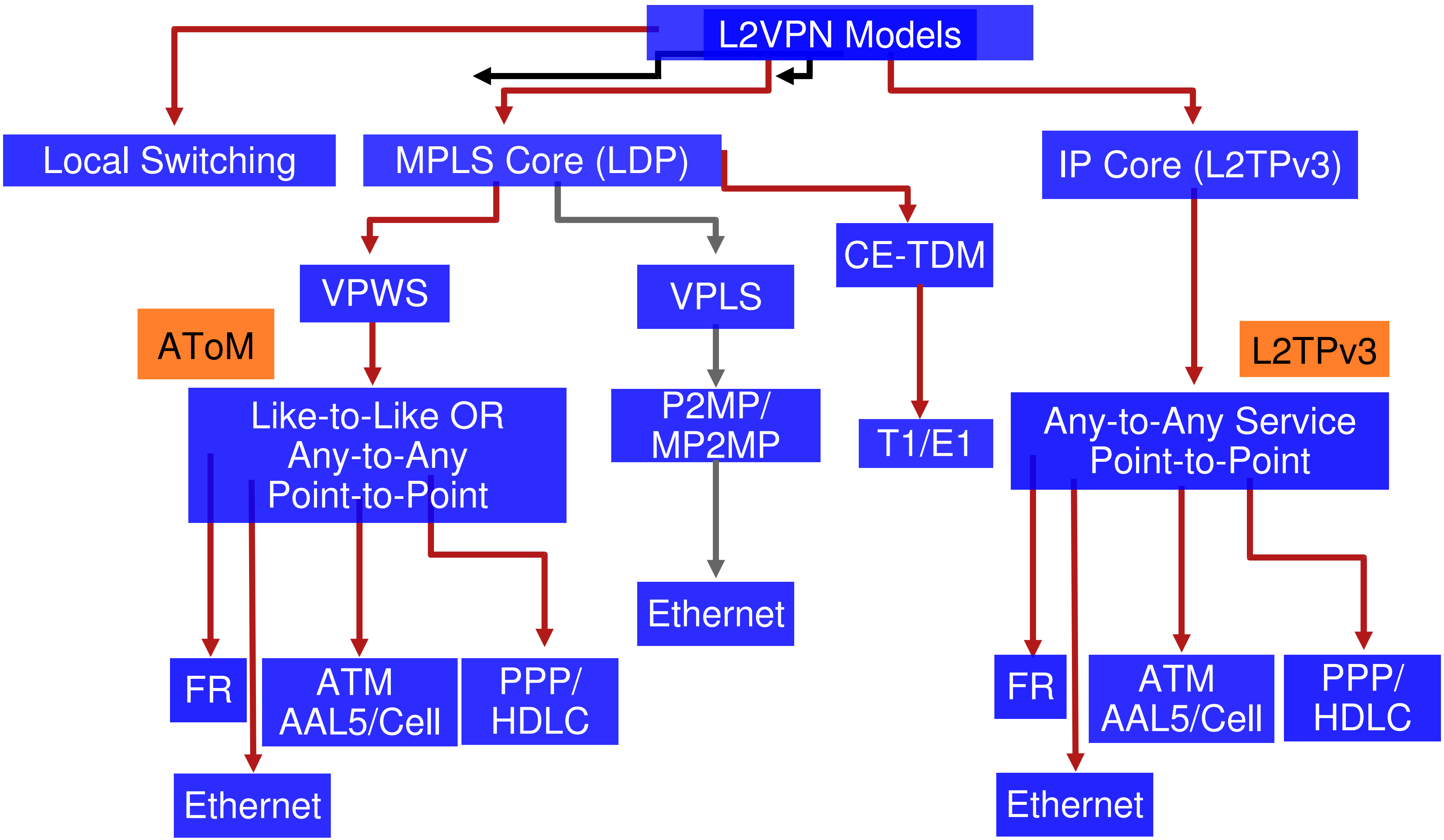
L2VPN provides an end-to-end layer 2 connection to an enterprise office in Taipei and Singapore over a SP's MPLS or IP core

Layer 3 and Layer 2 VPN Characteristics

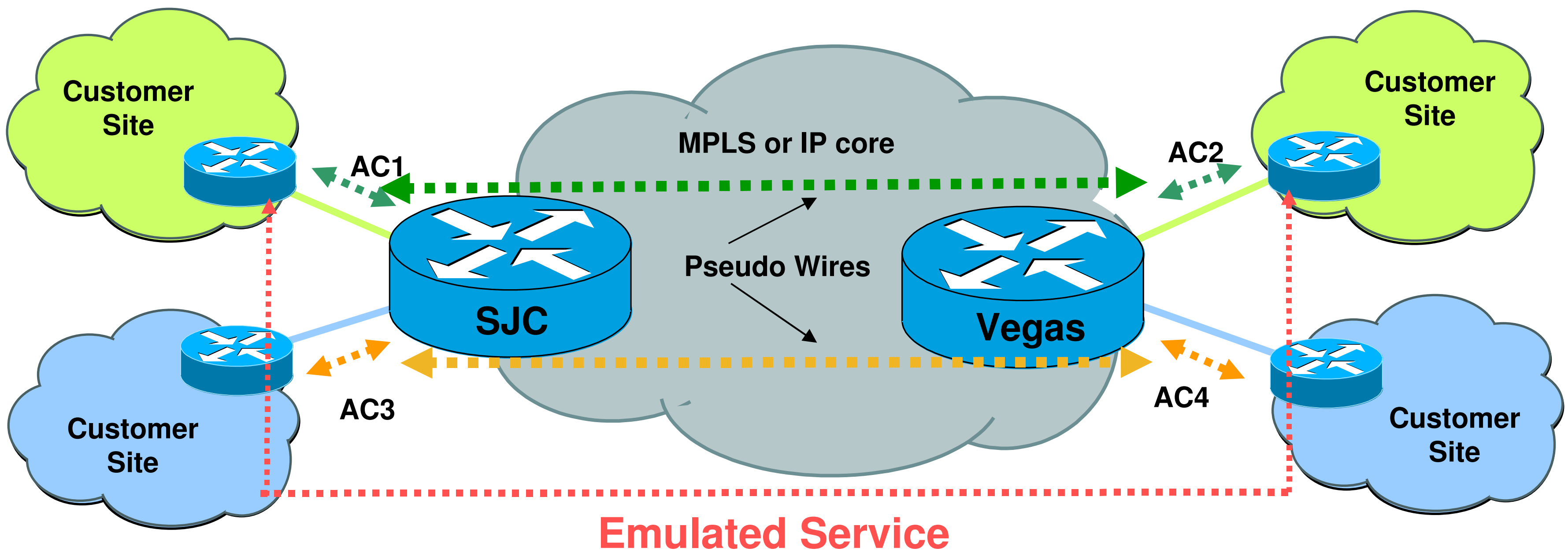
LAYER 3 VPNs	LAYER 2 VPNs
2. Packet based forwarding e.g. IP	2. Frame Based forwarding e.g. DLCI,VLAN, VPI/VCI
3. SP is involved	3. No SP involvement
4. IP specific	4. Multiprotocol support
5. Example: RFC 2547bis VPNs (L3 MPLS-VPN)	5. Example: FR—ATM—Ethernet

The Choice of L2VPN over L3VPN Will Depend on How Much Control the Enterprise Wants to Retain.
L2 VPN Services Are Complementary to L3 VPN Services

L2VPN Models

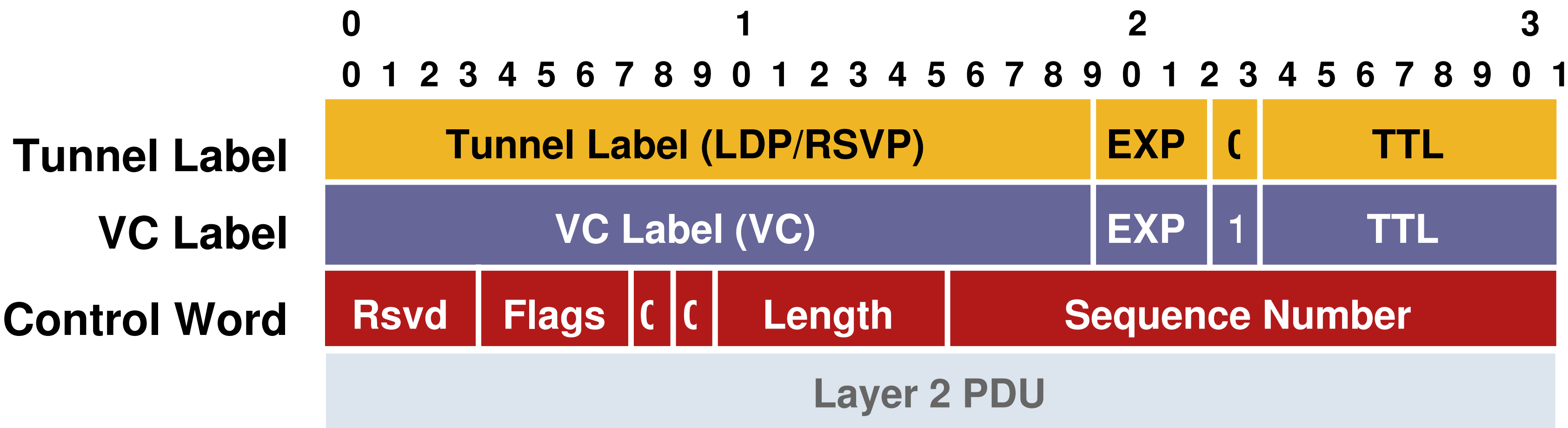


Pseudo Wire Reference Model



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

L2VPN – Label Stacking



- Three Layers of Encapsulation
  - Tunnel Label – Determines path through network
  - VC Label – Identifies VC at endpoint
  - Control Word – Contains attributes of L2 payload (optional)

Control Word	
Encap.	**Required**
CR	No
AAL5	Yes
Eth	No
FR	Yes
HDLC	No
PPP	No

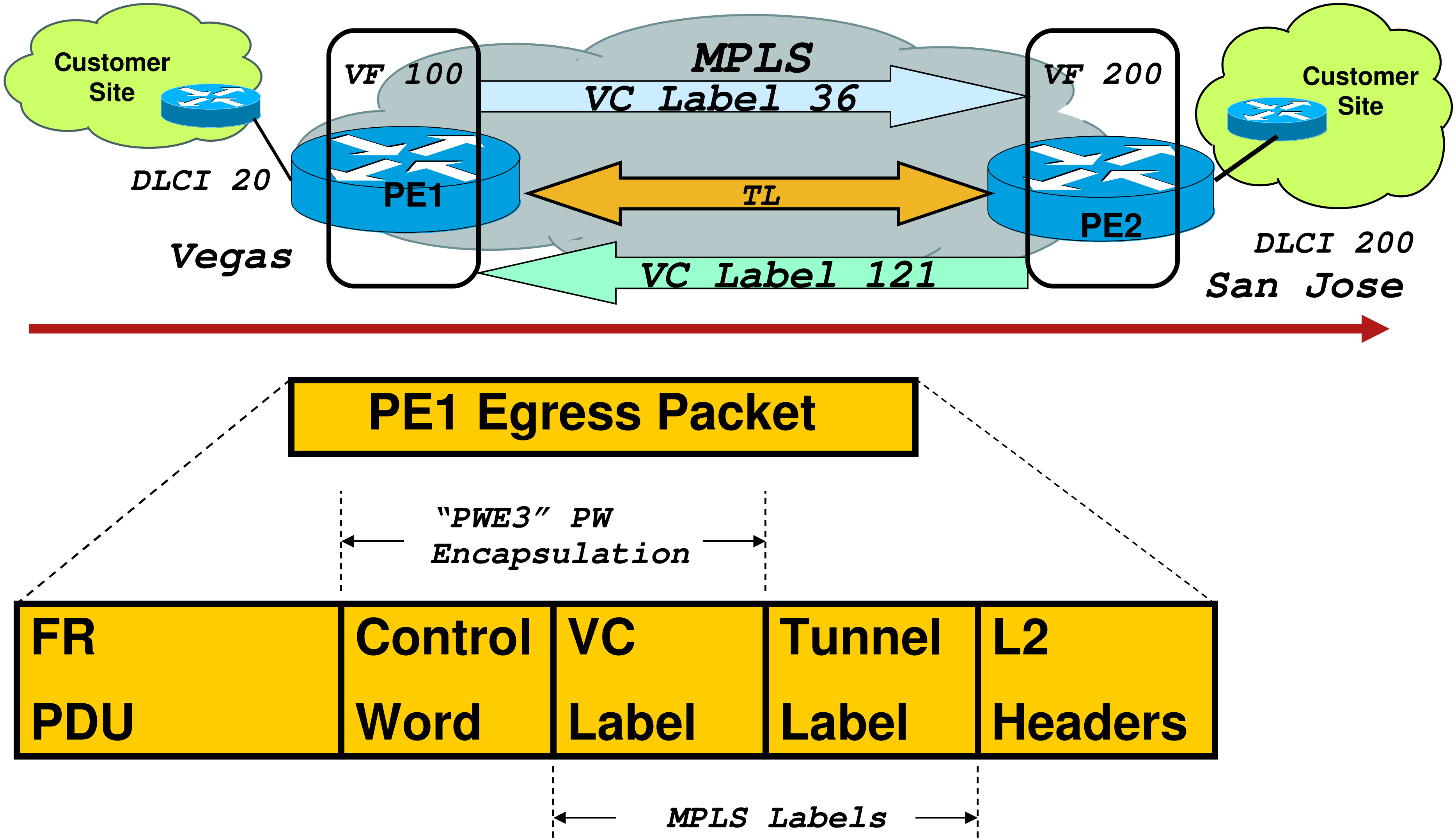
Control Word

bits	4	4	8	16
	Rsvd	**Flags**	**Length**	**Sequence Number**

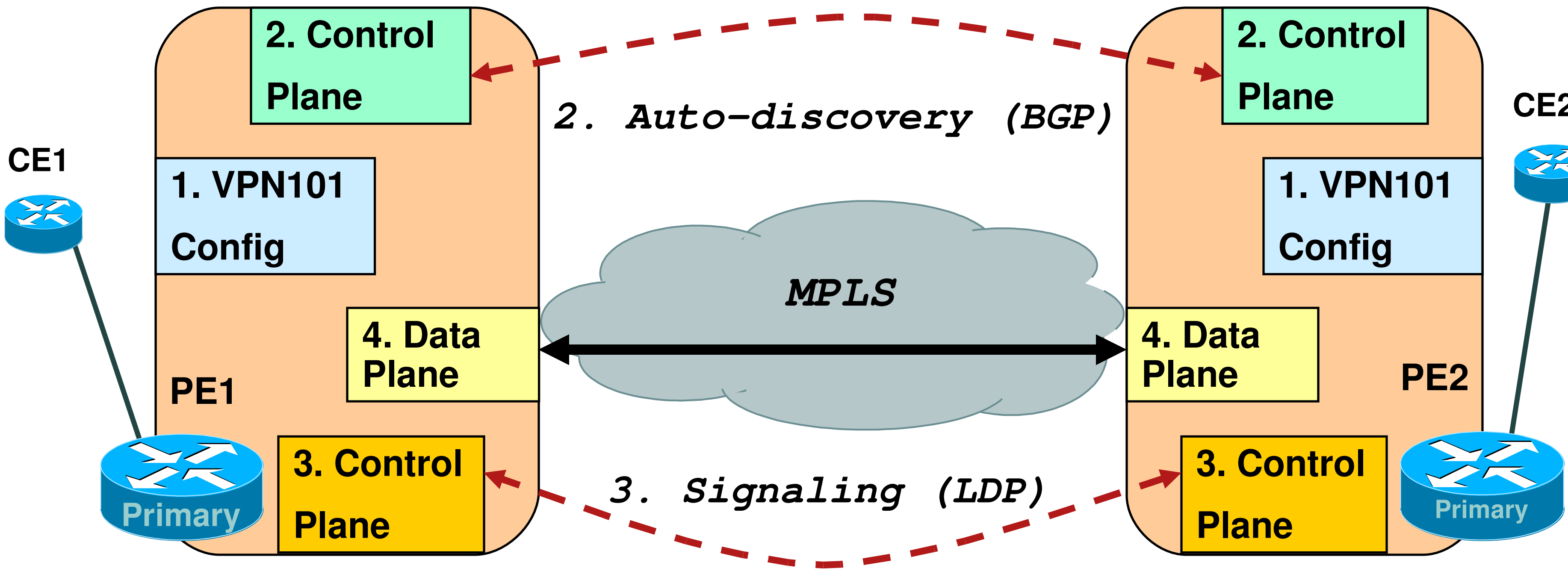
- Use of control word is optional
- Flags - Carries "flag" bits depending on encapsulation

  (FR; FECN, BECN, C/R, DE, ATM; CLP, EFCI, C/R, etc)
- Length - Required for padding small frames when < interface MTU
- Sequence number – Used to detect out of order delivery of frames

Data Plan Components – FR Example



Building Blocks for L2VPNs – Control Plane



1. Provision – Config VPN
2. Auto-discovery – Advertise loopback & vpn members
3. Signaling – Setup pseudowire
4. Data Plane – Packet forwarding

LDP Signaling Overview

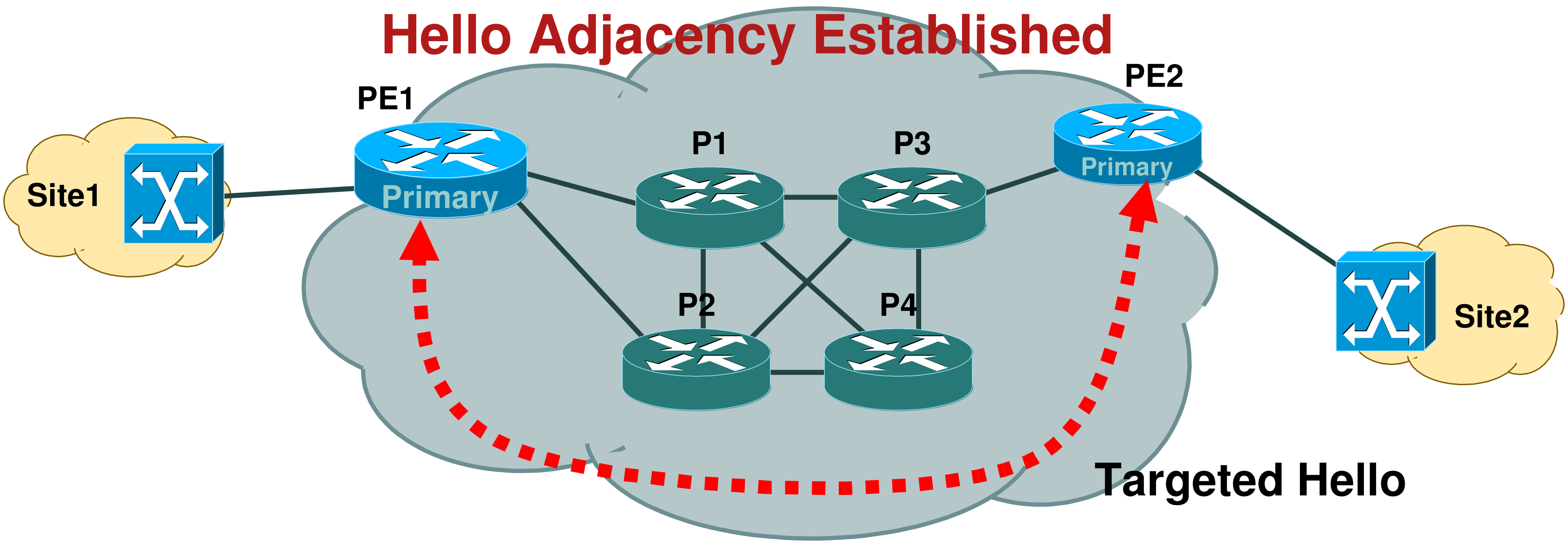
Four Classes of LDP messages:

- Peer discovery
  - LDP link hello message
  - Targeted hello message
- LDP session
  - LDP initialization and keepalive
  - Setup, maintain and disconnect LDP session
- Label advertisement
  - Create, update and delete label mappings
- LDP Notification
  - Signal error or status info

UDP (Peer discovery)

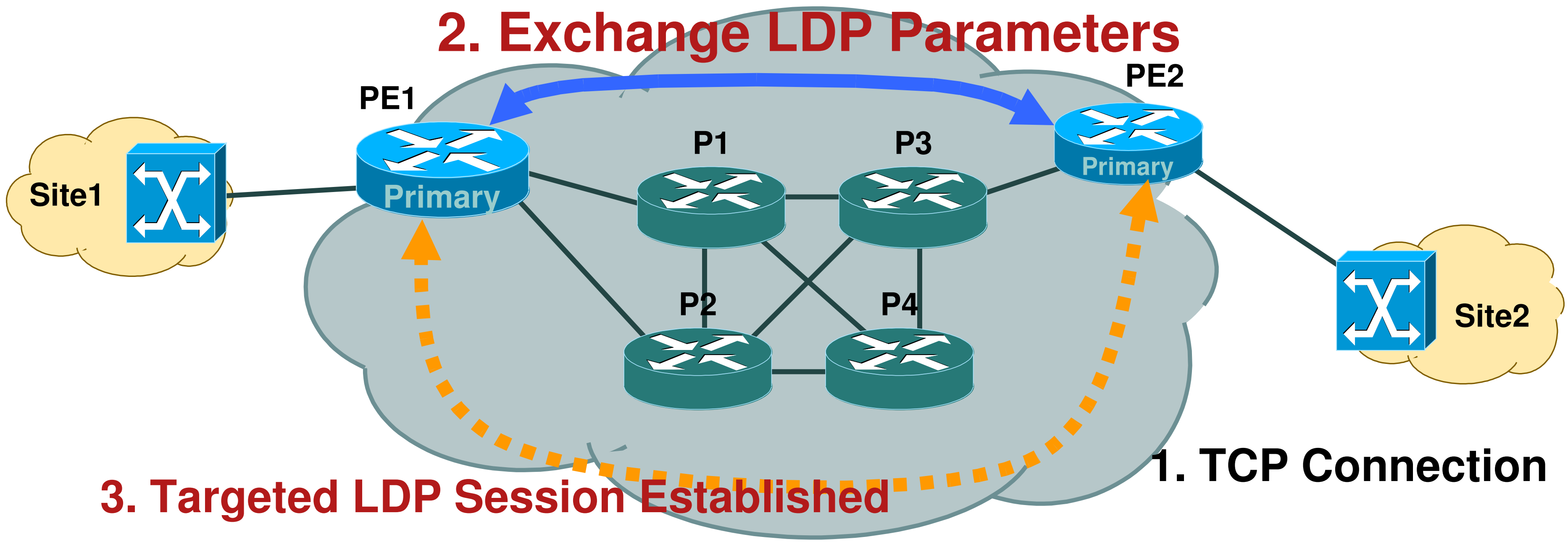
TCP (LDP session, Label advertisement, LDP Notification)

L2VPN LDP Extended Discovery



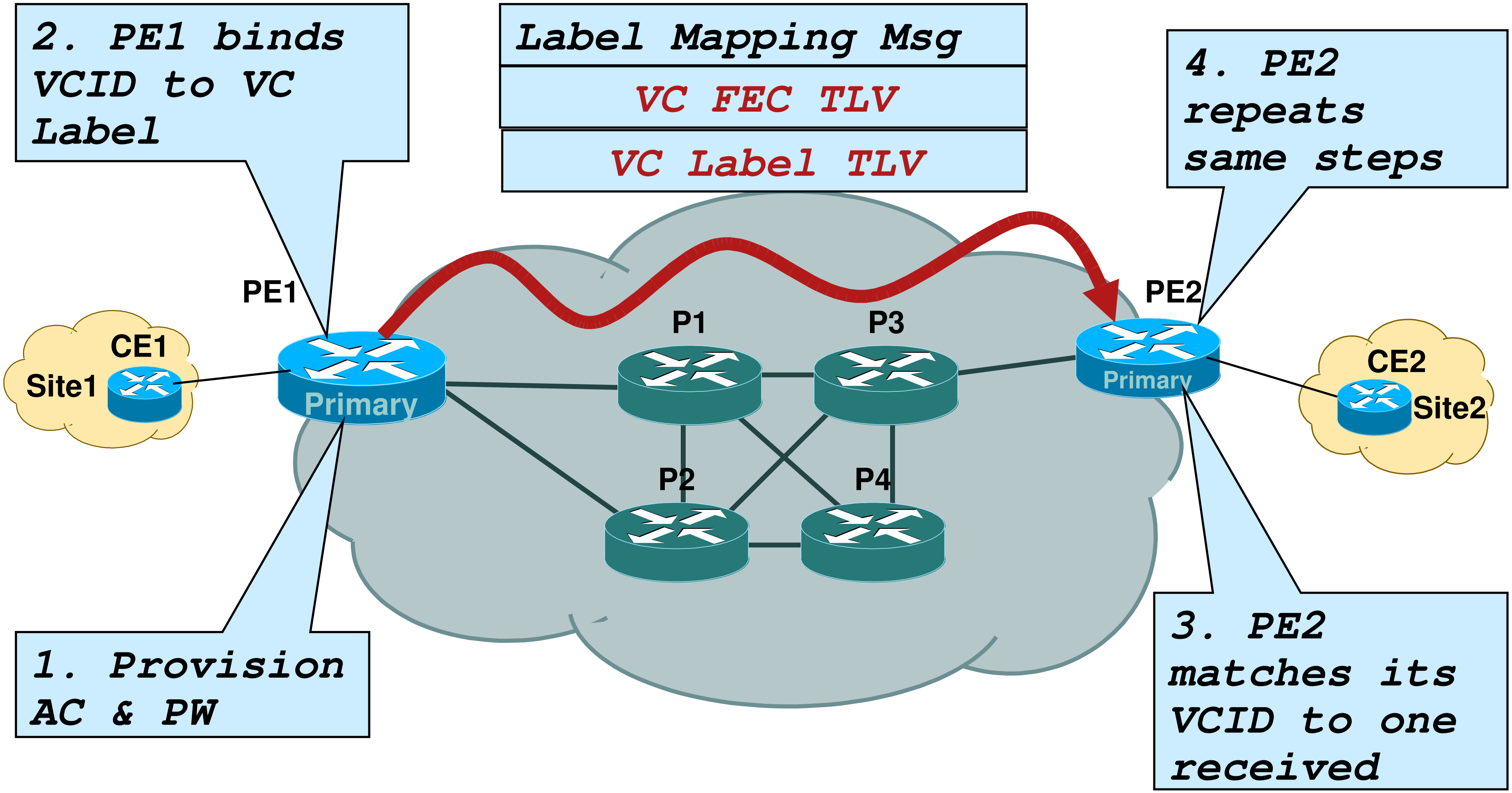
- Targeted hello messages are exchanged as UDP packets on port 646 consisting of router-id and label space

L2VPN LDP Session Establishment



1. Active role PE - establishes TCP connection using port 646
2. LDP peers exchange and negotiate session parameters such as the protocol version, label distribution methods, timer values, label ranges, and so on
3. LDP session is operational

L2VPN – Pseudo-Wire Label Binding



Uni-directional PW LSP Established

Virtual Circuit FEC Element

VC TLV	C	VC Type	VC Info Length
Group ID			
VC ID			
Interface Paramaters			

- Virtual Circuit FEC Element

  C – Control word present

  VC Type – ATM, FR, Ethernet, HDLC, PPP, etc …

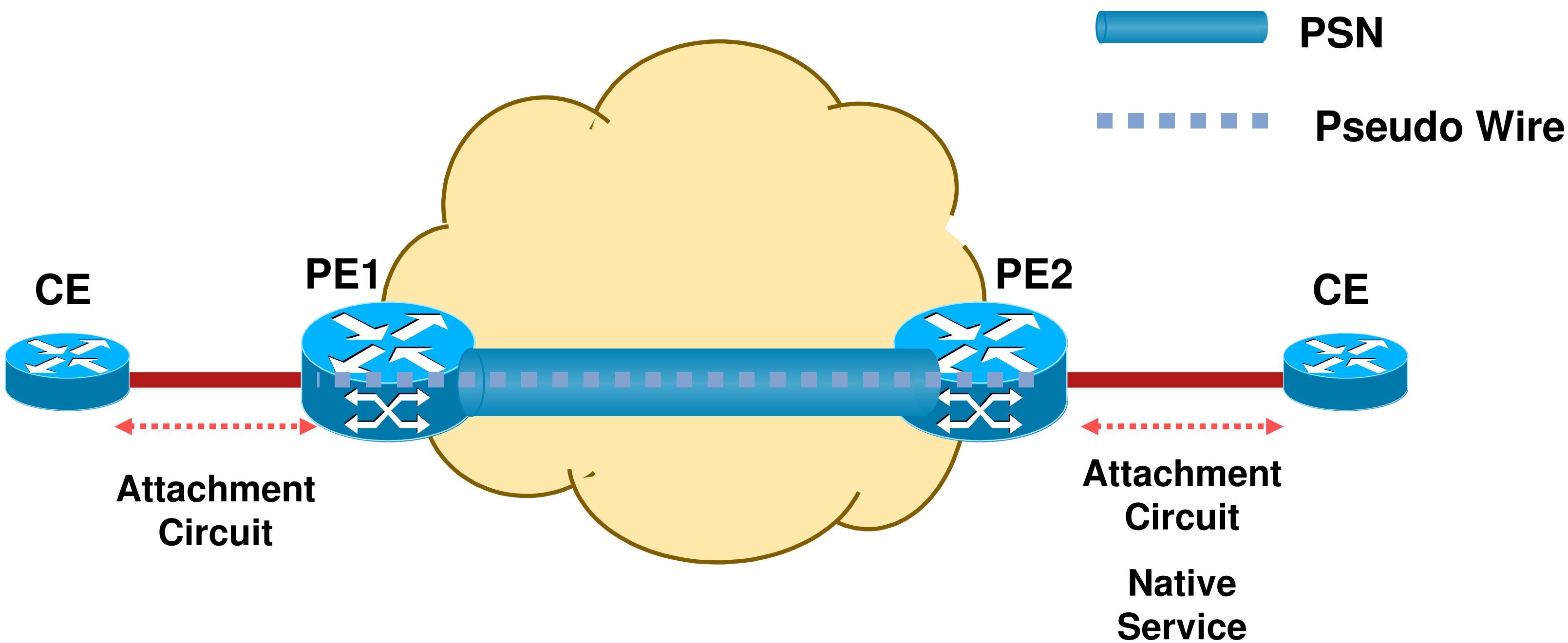
  VC Info Length – Length of VCID

  Group ID – Group of VCs referenced by index (user configured)

  VC ID – Identify PW

  Interface Parameters – MTU, etc ….

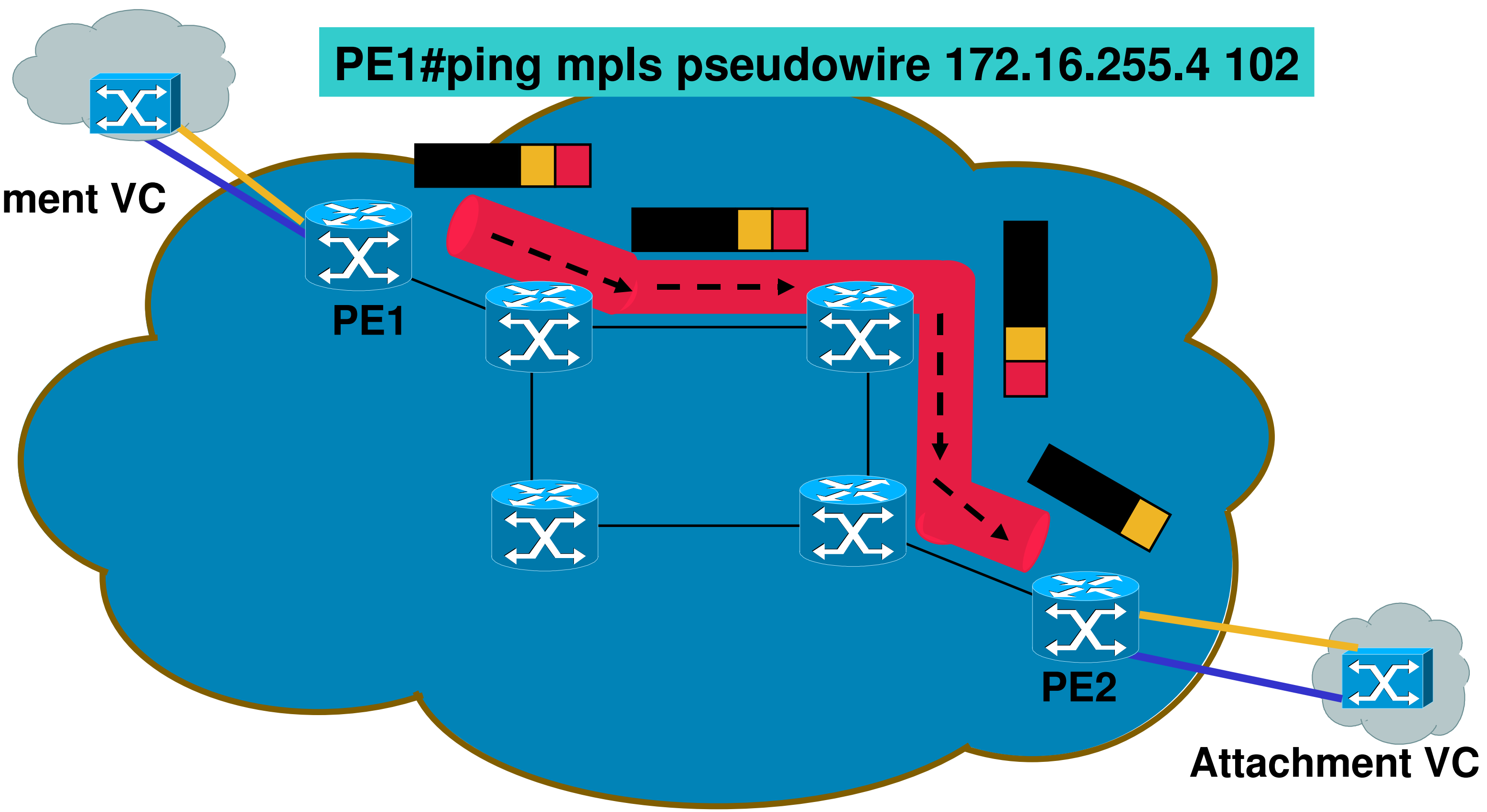
MPLS OAM –
Virtual Circuit Connection Verification (VCCV)



- Motivation

  One tunnel can serve many pseudo-wires.
  MPLS LSP ping is sufficient to monitor the PSN tunnel (PE-PE connectivity), but not VCs inside of tunnel.

MPLS Embedded Management –
Connectivity Trace Using VCCV



L2VPN Transports – Encapsulations

- Ethernet / 802.1Q VLAN (EoMPLS)

  draft-ietf-pwe3-ethernet-encap-xx.txt

- Frame Relay (FRoMPLS)

  draft-ietf-pwe3-frame-relay-encap-xx.txt

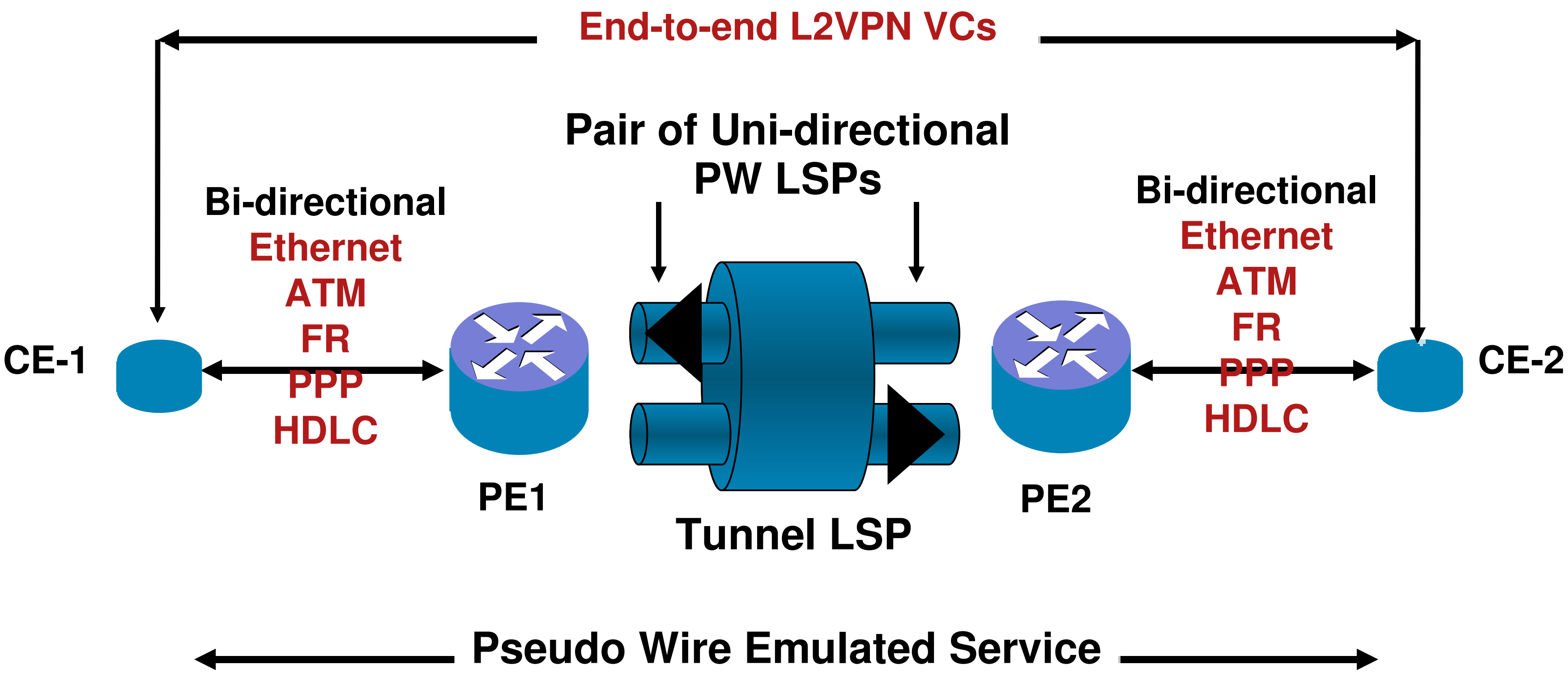
- ATM AAL5 and ATM Cell (ATMoMPLS)

  draft-ietf-pwe3-atm-encap-xx.txt

- PPP / HDLC (PPPoMPLS / HDLCoMPLS)

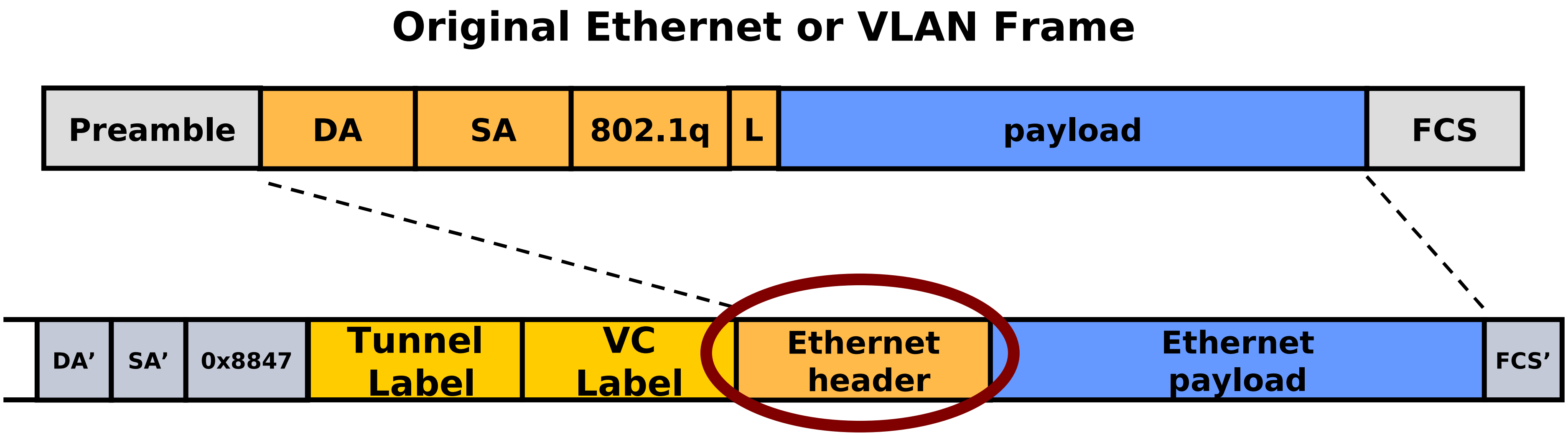
  draft-ietf-pwe3-hdlc-ppp-encap-mpls-xx.txt

L2VPN Transports Service: Reference Model



- Pseudowire transport (across PEs) applications
- Local switching (within a PE) applications

L2VPN EoMPLS –
draft-ietf-pwe3-ethernet-encap-xx.txt



- VC type-0x0004 is used for VLAN over MPLS application
- VC type-0x0005 is used for Ethernet port tunneling application (port transparency)

L2VPN EoMPLS –
draft-ietf-pwe3-ethernet-encap-xx.txt

- **The control word is optional**
- **If the control word is used then the flags must be set to zero**
  - The VLAN tag is transmitted unchanged but may be overwritten by the egress PE router (VLAN Rewrite)

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

Rsvd	0	0	0	0	0	0	Length	Sequence number
Ethernet PDU								

Optional

EVC – New Ethernet Infrastructure

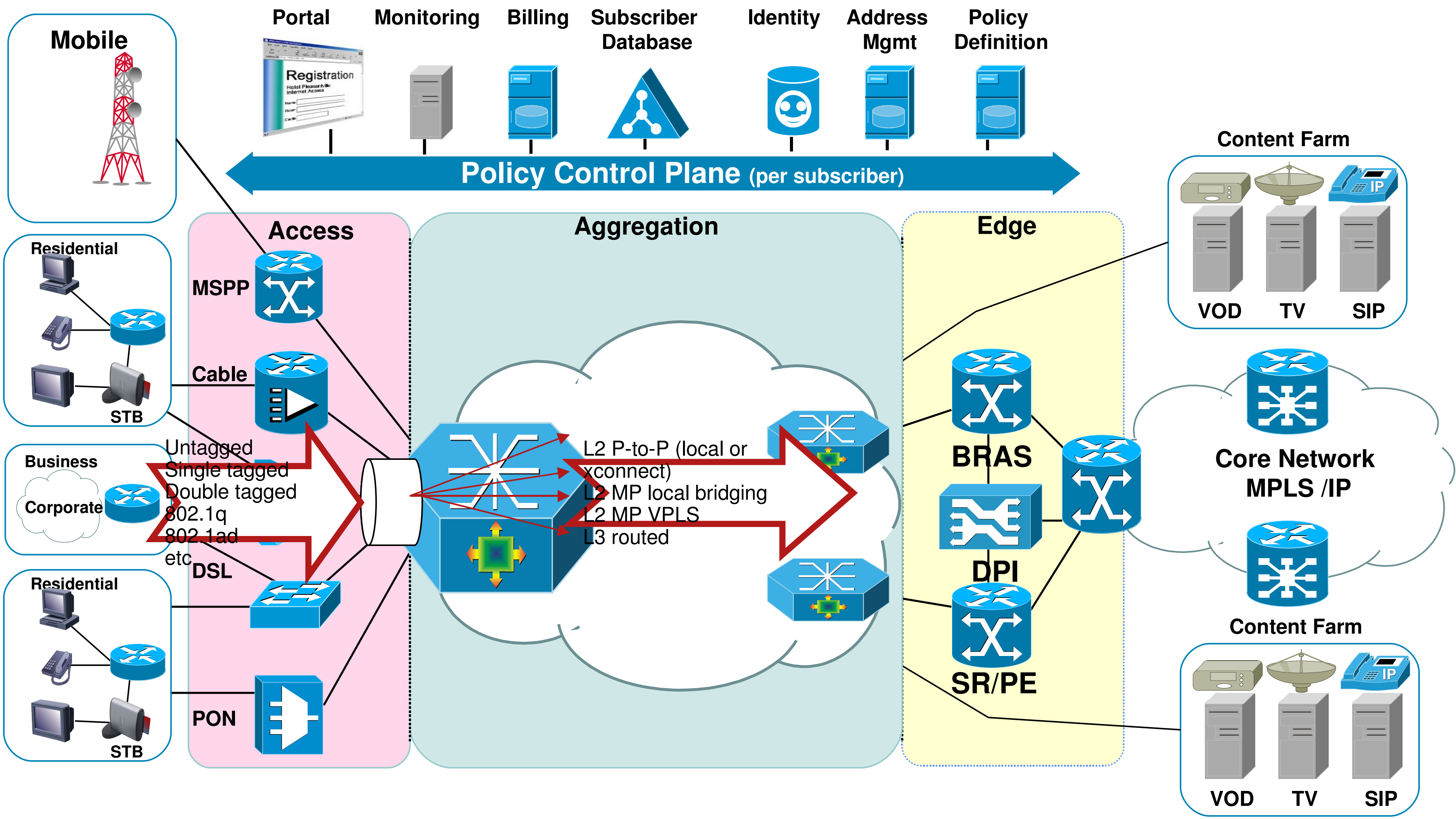
EVC = Ethernet Virtual Circuit (.1q/QinQ/.1ad/.1ah, EoMPLS, VPLS, local connect, etc)

- Provide uniform platform independent framework to make Ethernet carrier-class
- Alignment with MEF, IEEE, IETF standards
- Scalability, High Availability, Manageable and Distribution
- Structured CLI for next-gen Ethernet services (Ethernet VCs)
- Defined Ethernet Flow Points for identifying traffic on UNI

EVC Infrastructure provides

- Classification (VLAN matching)
- VLAN Translation
- Services Mapping
- L2 Split-horizon privacy
- H-QOS support

Flexible Ethernet Edge →
New EVC Ethernet Infrastructure



EVC – End User CLI

interface <type><slot/port>

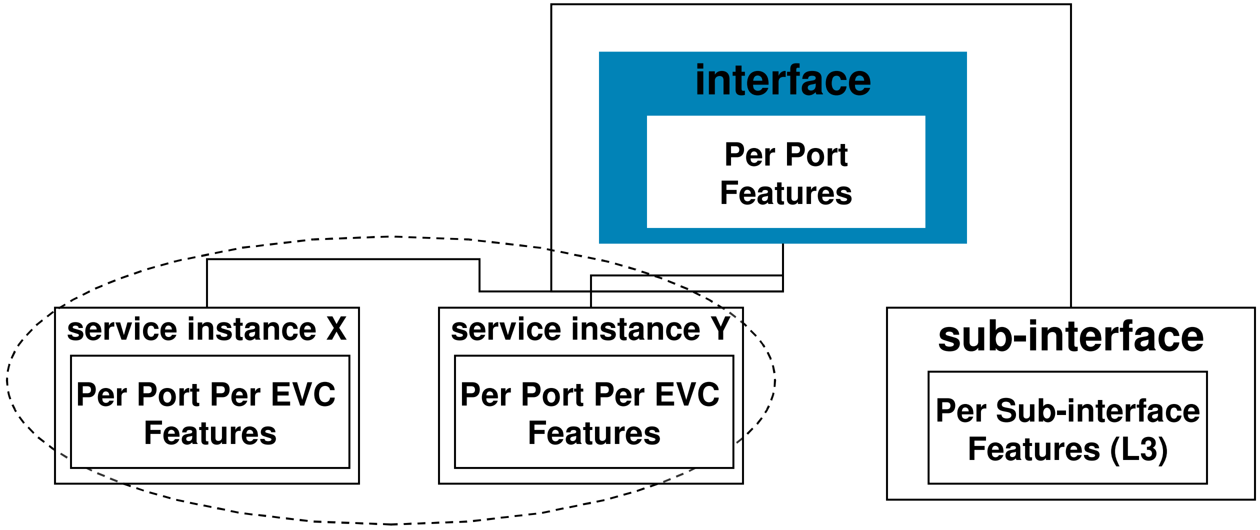
service instance <id> ethernet <evc-name> ← ID is per interface scope

<match criteria commands> ← VLAN tags, MAC, CoS, Ethertype

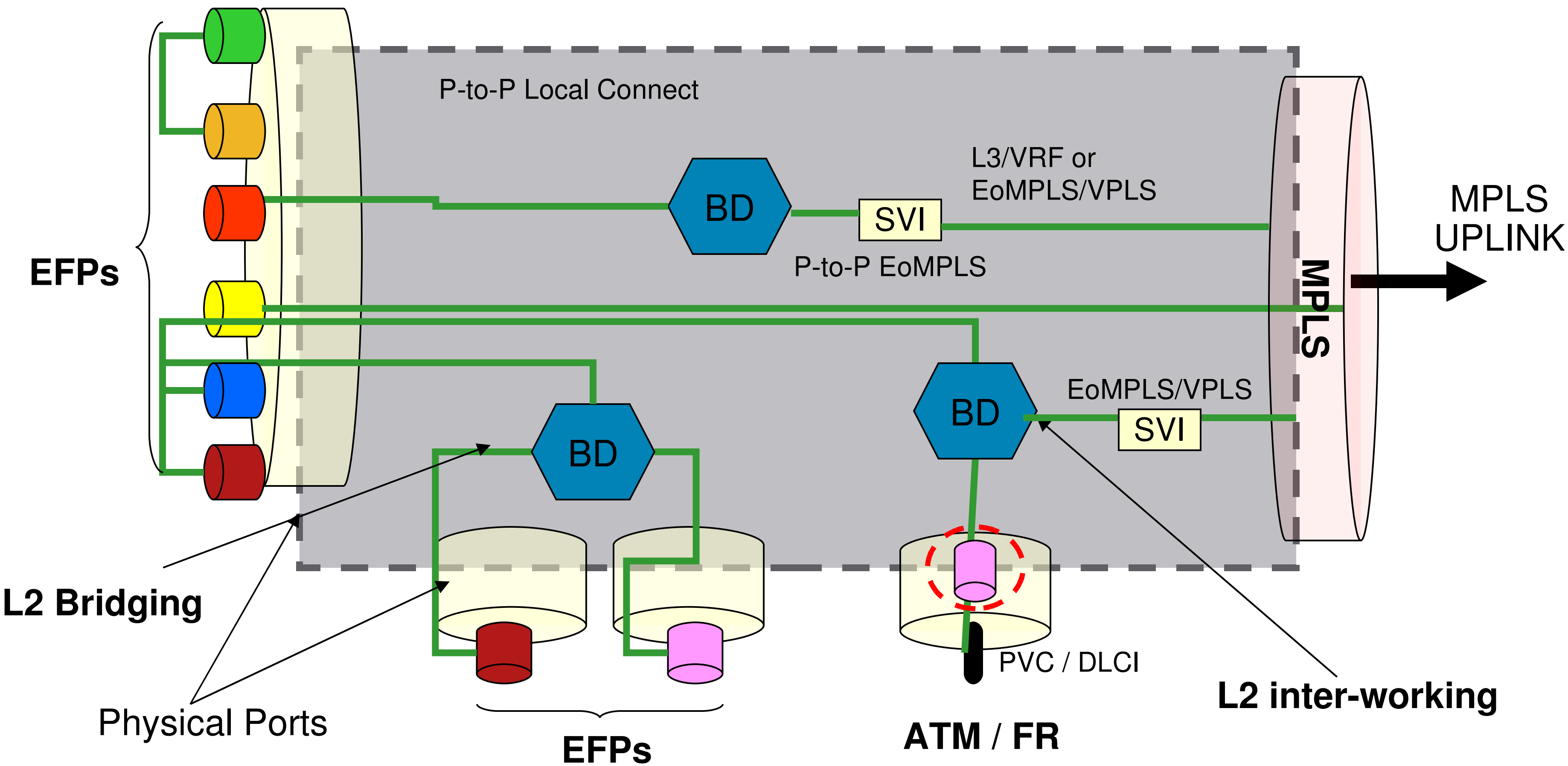
<rewrite commands> ← VLAN tags pop/push/translation

<forwarding commands> ← bridge-domain, xconnect or local connect

<feature commands> ← QoS, BPDUs, ACL, etc

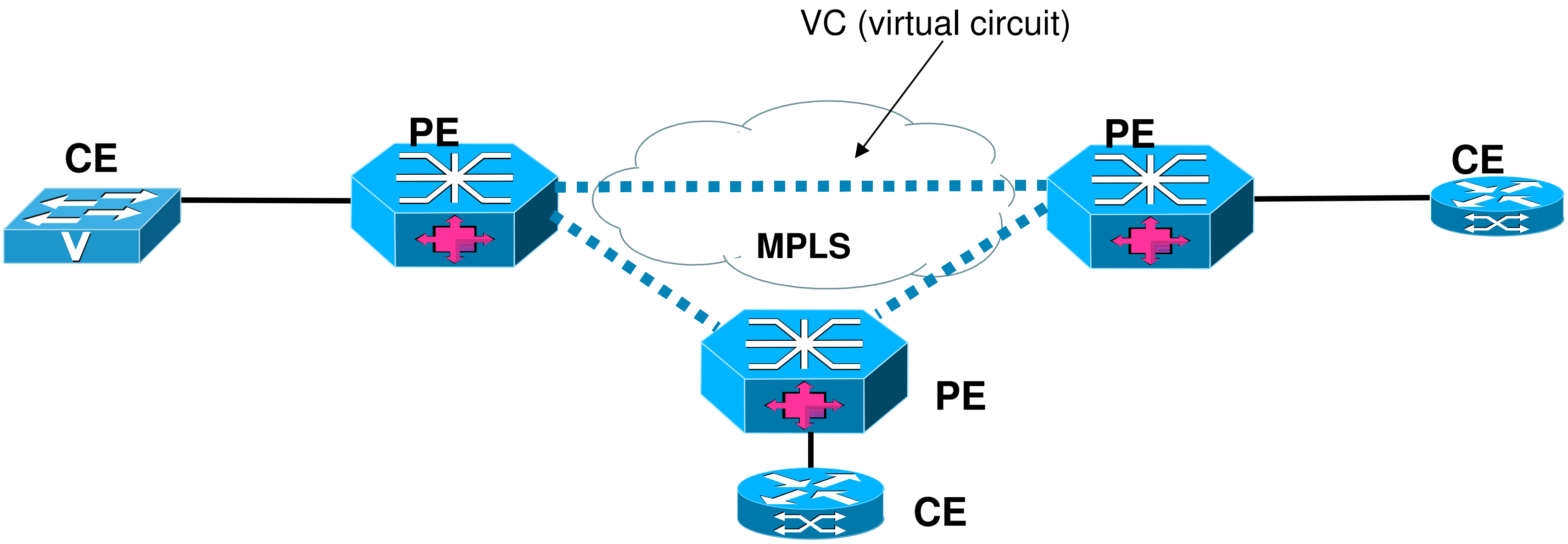


EVC – Flexible Forwarding Model



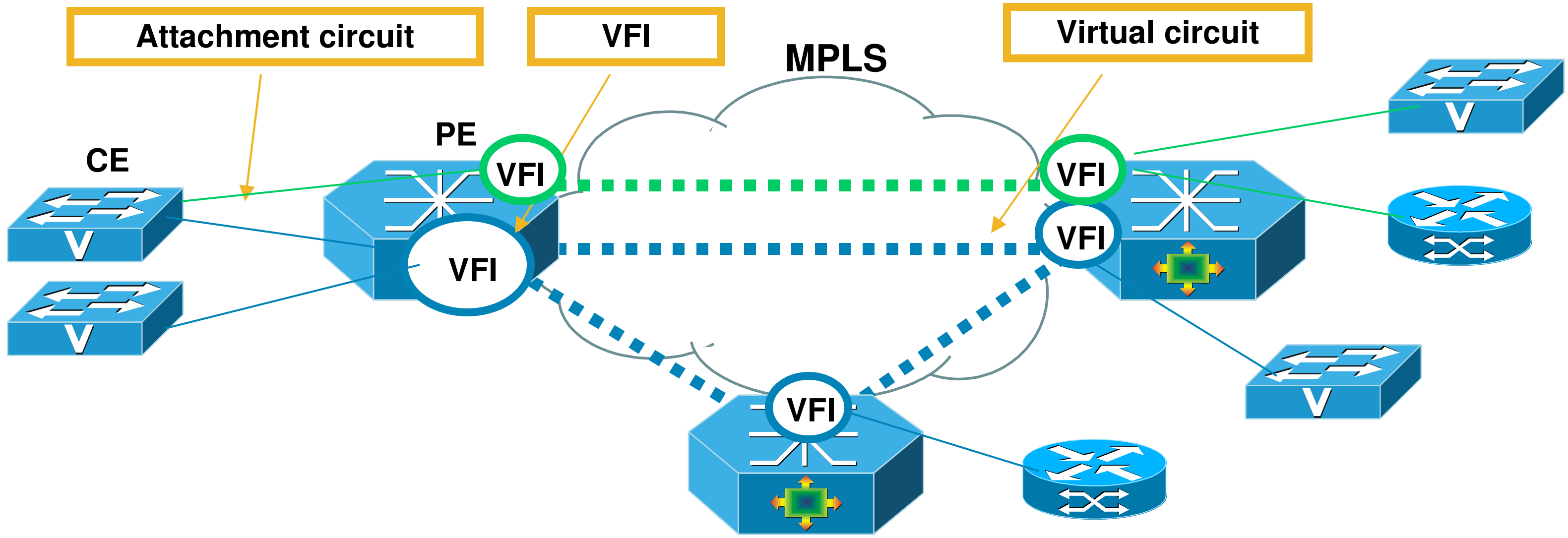
VPLS

What's VPLS (Virtual Private LAN Services) ?



- End-to-end architecture that allows IP/MPLS networks to provide multipoint Ethernet services
- Virtual – multiple instances of this services share the same SP physical infrastructure
- Private – each instance of the service is independent and isolated from one another
- LAN service – provides a multipoint connectivity among the participant endpoints across a MAN/WAN that looks like a LAN

VPLS Components



- AC (Attachment Circuit)

  Connect to CE device, it could be Ethernet physical or logical port, ATM bridging (RFC1483), FR bridging (RFC1490), even AToM pseudo wire. One or multiple ACs can belong to same VFI

- VC (Virtual Circuit)

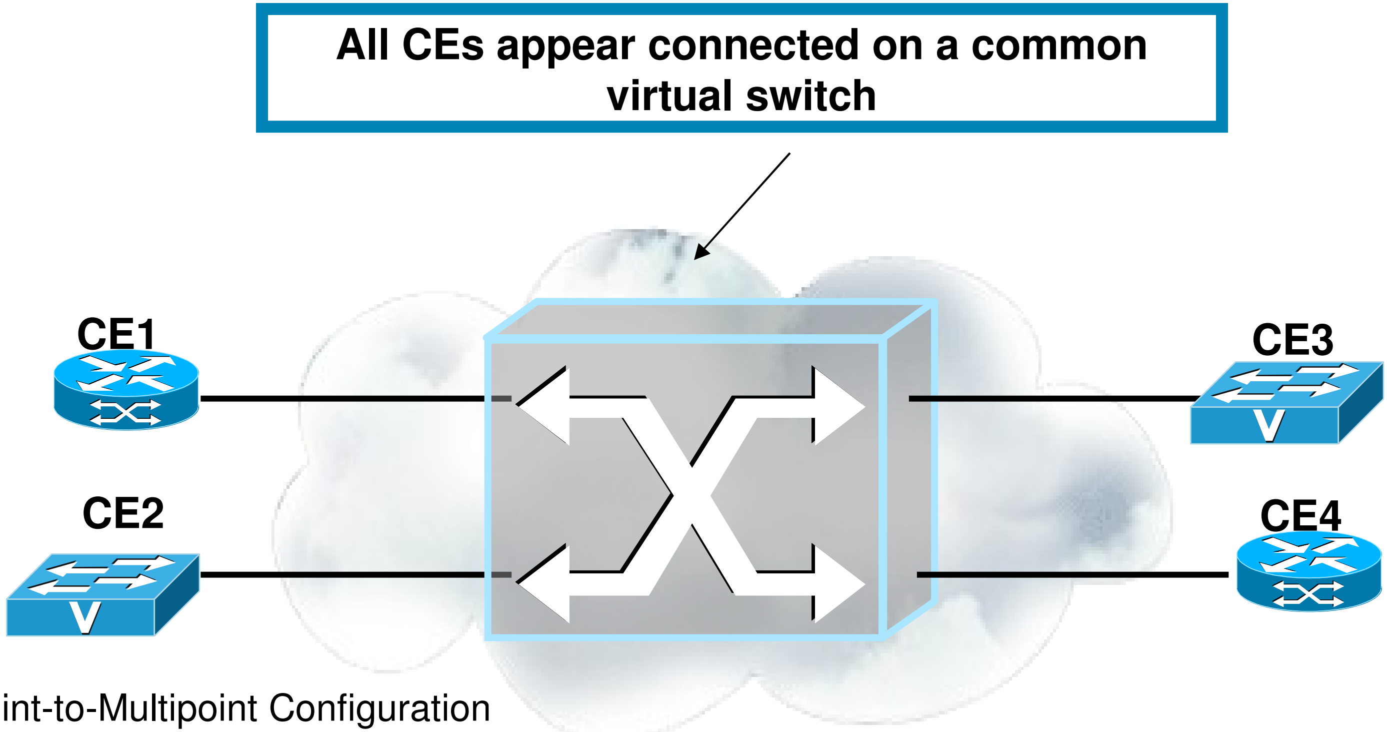
  EoMPLS data encapsulation, tunnel label is used to reach remote PE, VC label is used to identify VFI. One or multiple VCs can belong to same VFI

- VFI (Virtual Forwarding Instance)

  Also called VSI (Virtual Switching Instance). VFI create L2 multipoint bridging among all ACs and VCs. It's L2 broadcast domain like VLAN

  Multiple VFI can exist on the same PE box to separate user traffic like VLAN

VPLS Customer Perspective



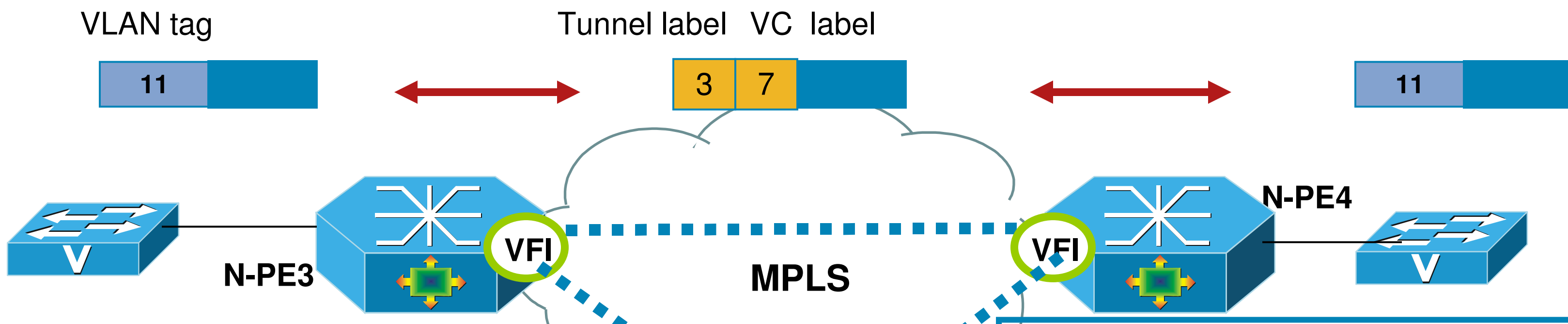
- Multipoint-to-Multipoint Configuration
- Forwarding of Frames based on Learned MAC addresses
- Uses a Virtual Forwarding Instances (VFI, like VLAN) for customer separation

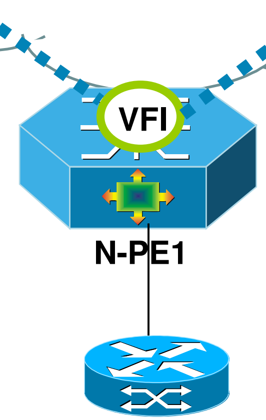
Multipoint Bridging Requirements

VPLS simulate a virtual LAN service, it MUST operate like a traditional L2 LAN switch as well

- Flooding/Forwarding
  - Forwarding based on [VLAN, Destination MAC Address]
  - Unknwon Ucast/Mcast/Broadcast – Flood to all ports (IGMP snooping can be used to constrict multicast flooding)
- MAC Learning/Aging/Withdrawal
  - Dynamic learning based on Source MAC and VLAN
  - Refresh aging timers with incoming packet
  - MAC withdrawal upon topology changes
- Loop Prevention
  - Split Horizon to avoid loop
  - Spanning Tree (possible but not desirable)

A Simple VPLS Configuration Example

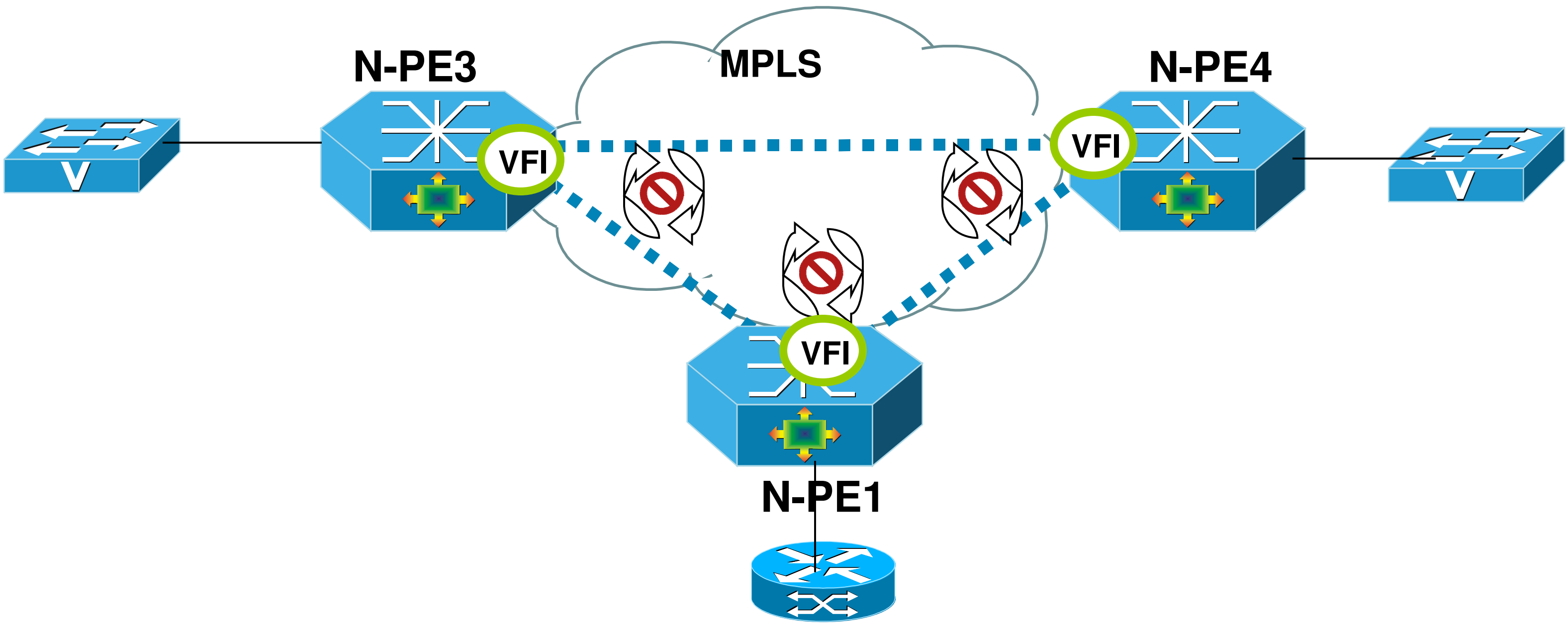





```
N-PE3
interface Loopback0
 ip address 10.0.0.3 255.255.255.255

! Define VPLS VFI
l2 vfi vpls11 manual
 vpn id 11 ← global significant
 neighbor 10.0.0.1 encapsulation mpls
 neighbor 10.0.0.4 encapsulation mpls

! Attach VFI to VLAN interface
! VLAN ID is local PE significant
interface Vlan11
 xconnect vfi vpls11

! Attachment circuit config
interface GigabitEthernet5/1
 switchport
 switchport trunk encapsulation dot1q
 switchport mode trunk
```

```
N-PE4
interface Loopback0
 ip address 10.0.0.4 255.255.255.255

l2 vfi vpls11 manual
 vpn id 11
 neighbor 10.0.0.1 encapsulation mpls
 neighbor 10.0.0.3 encapsulation mpls

interface Vlan11
 xconnect vfi vpls11

! Attachment circuit
interface GigabitEthernet5/1
 switchport
 switchport trunk encapsulation dot1q
 switchport mode trunk
```

Loop Prevention – Split-horizon



How to avoid loop in VPLS (multipoint bridging) network?

- Spanning tree is possible but not desirable
- VPLS use split-horizon to avoid loop
    - Packet received on VPLS VC can only be forwarded to ACs, not the other VPLS VCs (H-VPLS is exception)
    - Require full mesh VCs among all PEs

VPLS Data Plane and Control Plane

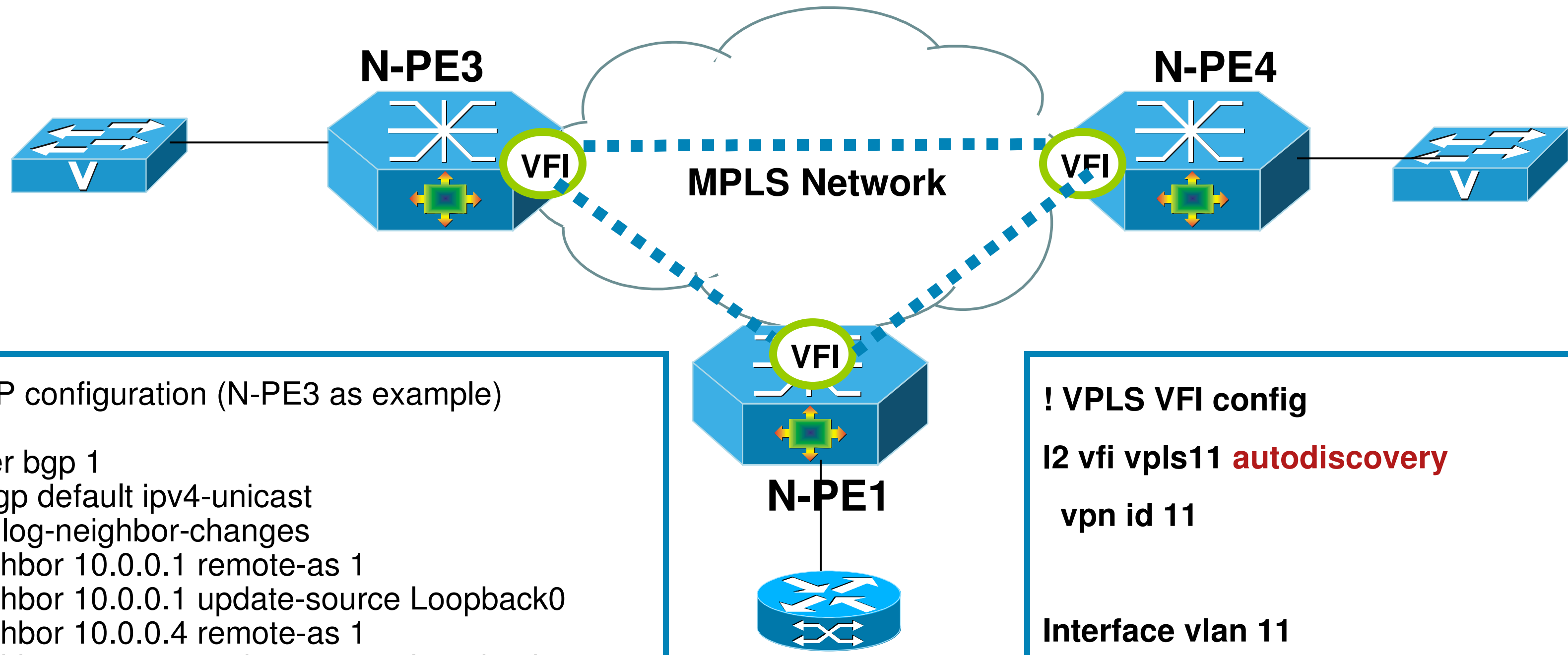
- Data Plane

  Although VPLS simulate multipoint virtual LAN service, the individual VC is still point-to-point EoMPLS. It uses the same data encapsulation as point-to-point EoMPLS

- Control plane Signalling

  Same as EoMPLS, using directed LDP session to exchange VC information

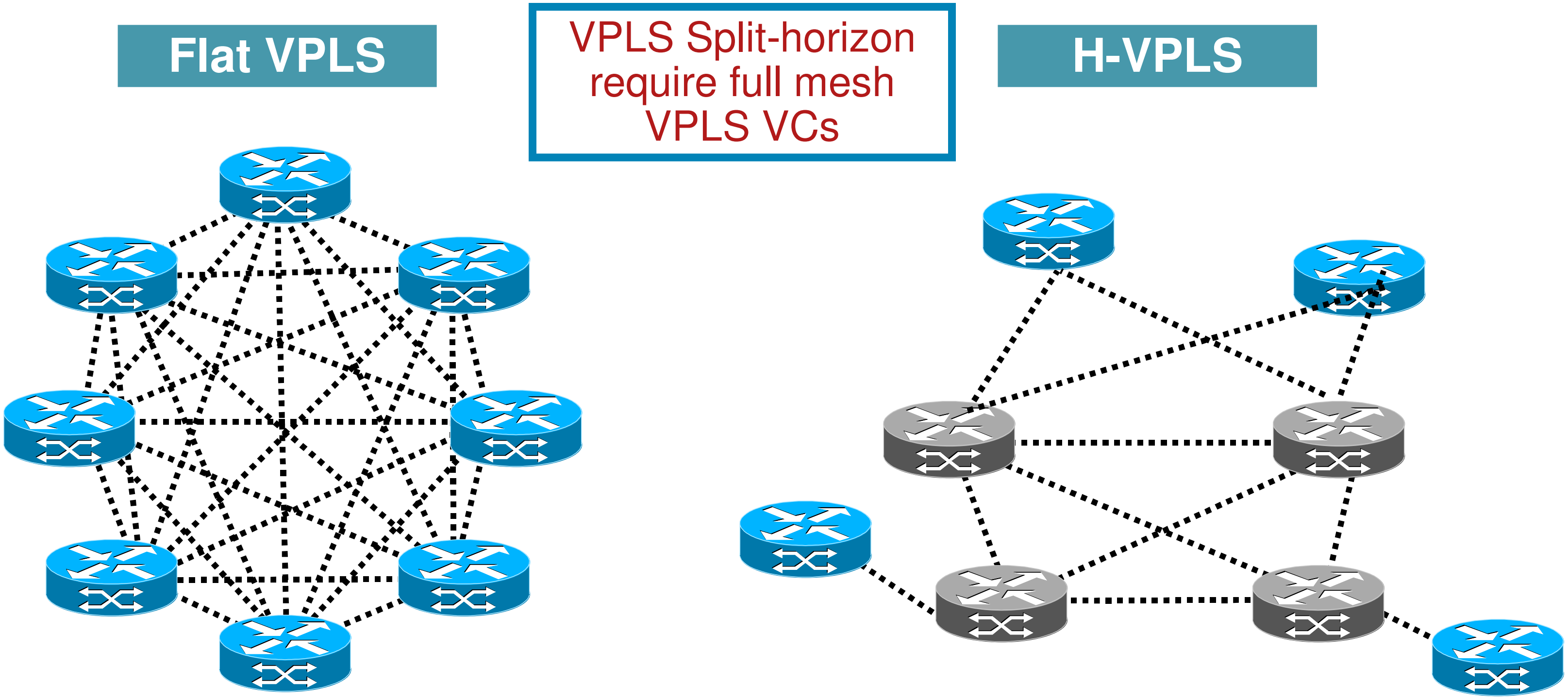
BGP-based VPLS Auto Discovery –
Configuration Example




```
! BGP configuration (N-PE3 as example)

router bgp 1
no bgp default ipv4-unicast
 bgp log-neighbor-changes
 neighbor 10.0.0.1 remote-as 1
 neighbor 10.0.0.1 update-source Loopback0
 neighbor 10.0.0.4 remote-as 1
 neighbor 10.0.0.4 update-source Loopback0
 !

 !
 address-family l2vpn vpls
 neighbor 10.0.0.1 activate
 neighbor 10.0.0.1 send-community extended
 neighbor 10.0.0.4 activate
 neighbor 10.0.0.4 send-community extended
 exit-address-family
!
```

```
! VPLS VFI config
l2 vfi vpls11 autodiscovery
  vpn id 11

Interface vlan 11
  xconnect vfi vpls11

! AC config, the same as before
```

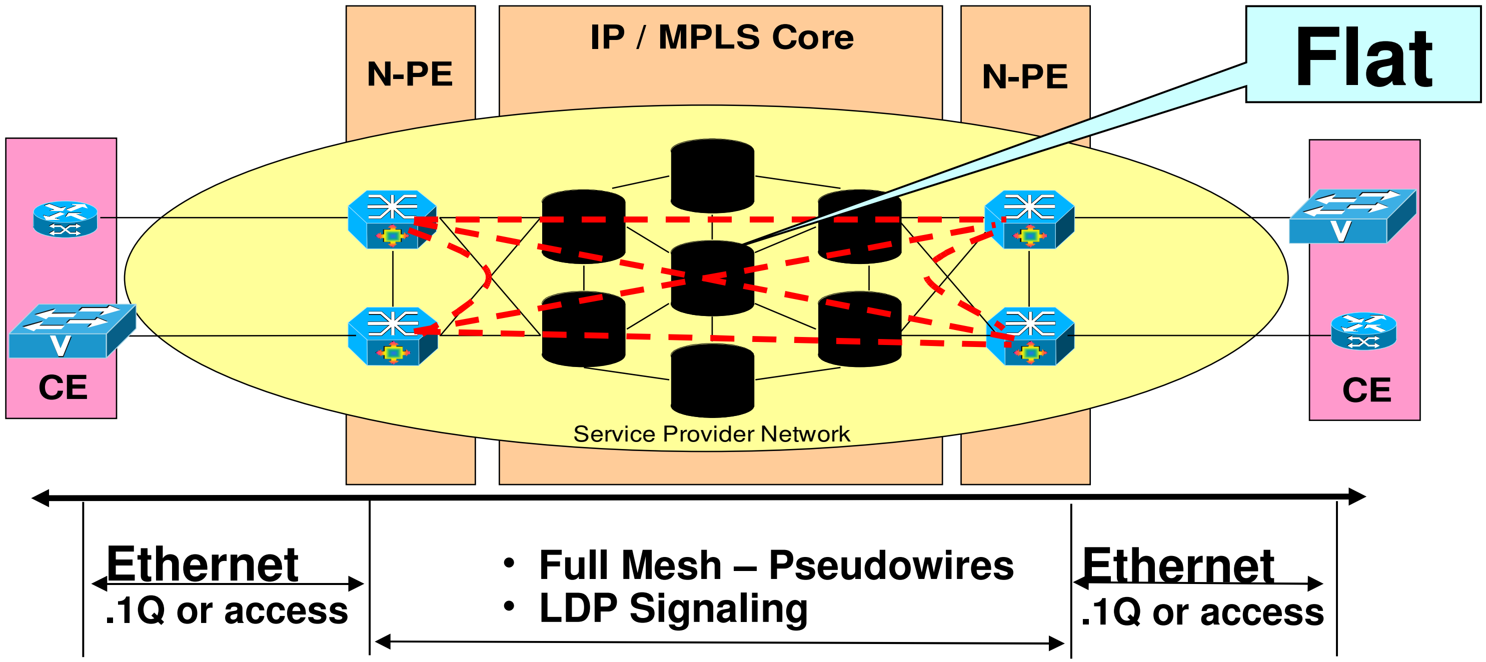
Why H-VPLS?



- Potential signaling overhead
- Full PW mesh from the edge
- Packet replication done at the edge
- Node discovery and provisioning extends end-to-end

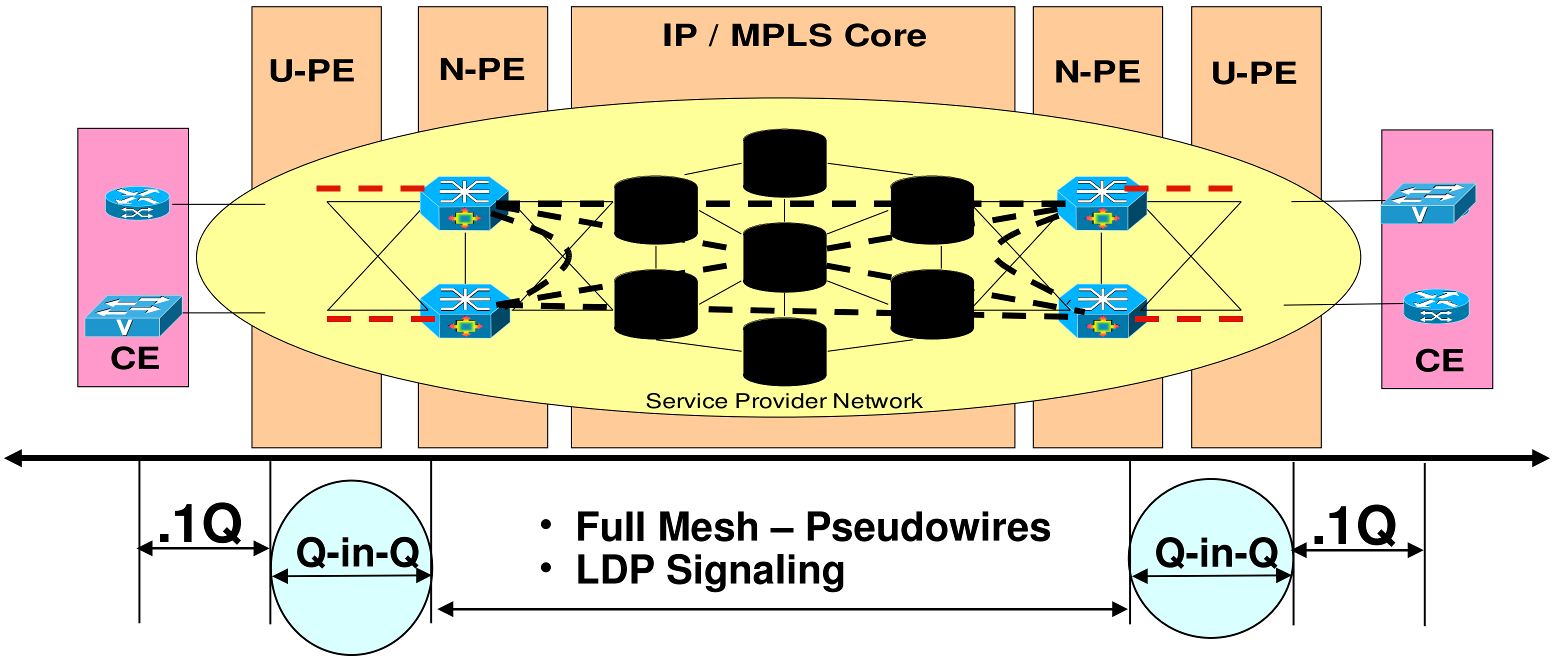
- Minimizes signaling overhead
- Full PW mesh among core devices only
- Packet replication done the core only

Flat VPLS – Ethernet access without QinQ



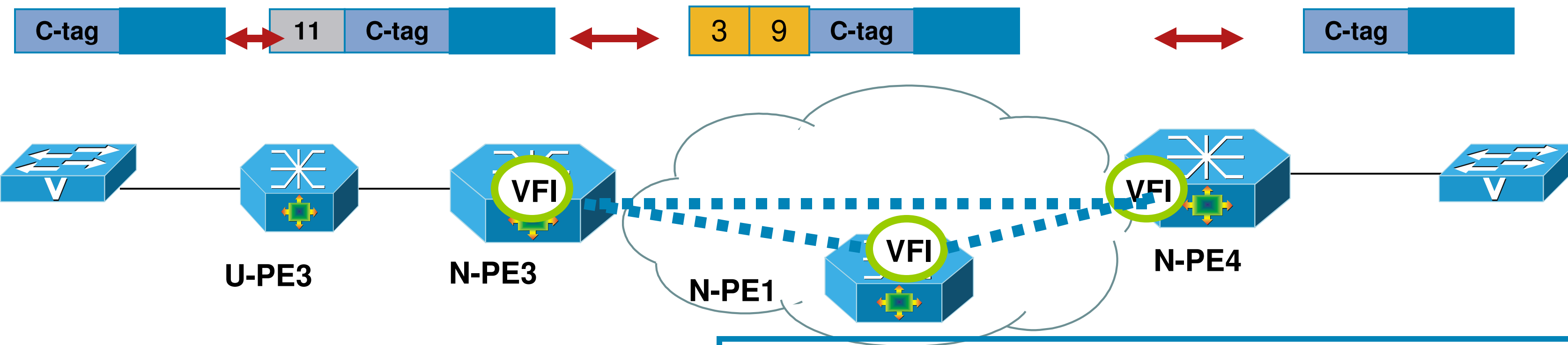
- Full mesh of directed LDP sessions required between participating PEs
- N*(N-1)/2 ; N = number of PE nodes
- Limited scalability
- Potential signaling and packet replication overhead
- Suitable for smaller networks, simple provisioning
- Customer VLAN tag is used as VPLS VFI service delimiter

H-VPLS with Ethernet Access QinQ



- Best for larger scale deployment
- Reduction in packet replication and signaling overhead
- Full mesh for Core tier (Hub) only
- Expansion affects new nodes only (no re-configuring existing PEs)
- QinQ frame in Ethernet access network. S-tag is used as VPLS VFI service delimiter. Customer tag is invisible. Each Ethernet access network can have 4K customers, 4K*4K customer vlans

H-VPLS with QinQ Access Example



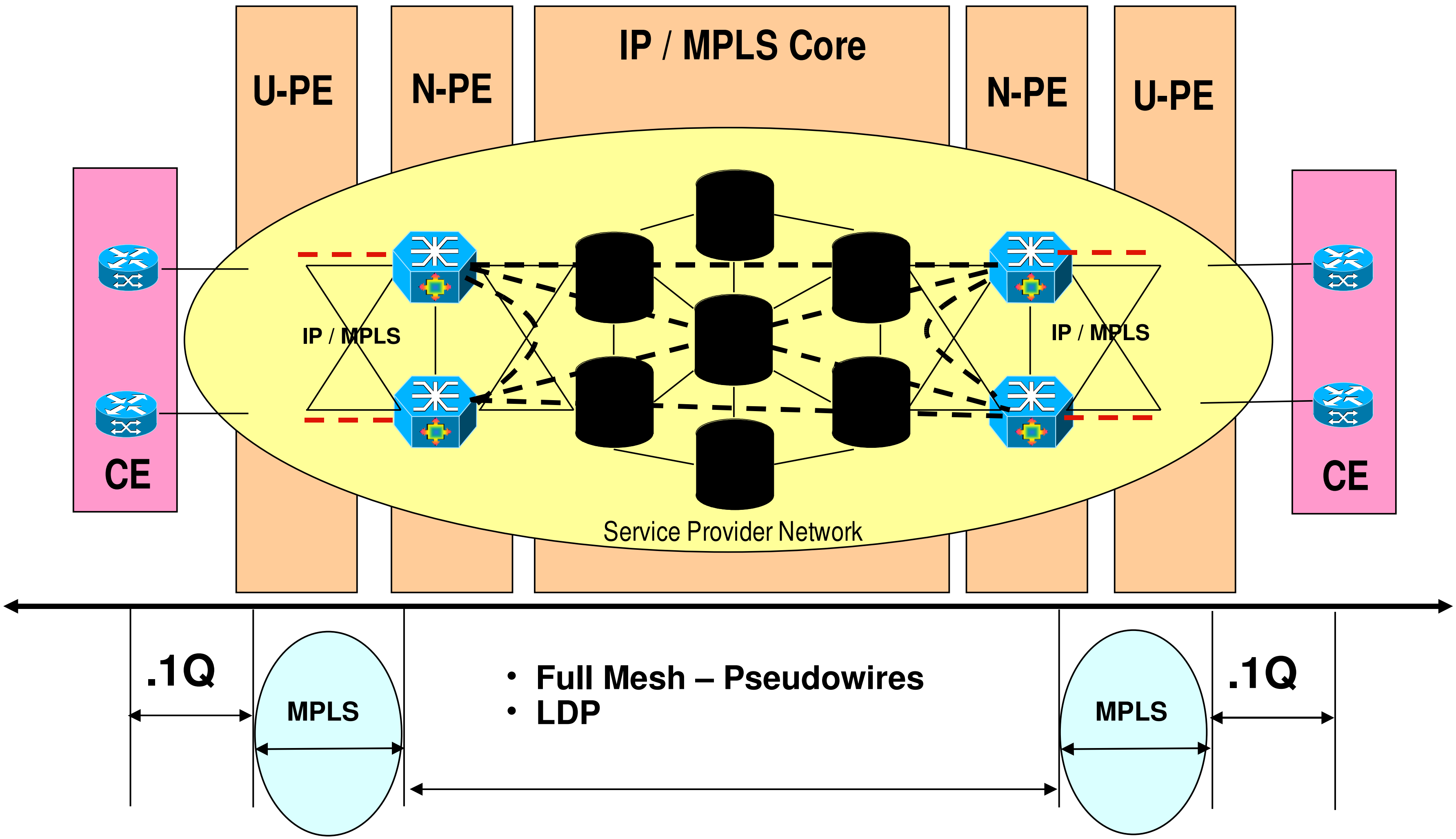
U-PE Configuration

```
! Interface connected to CE
! It's dot1q-tunnel port
interface GigabitEthernet2/13
 switchport
 switchport access vlan 11
 switchport mode dot1q-tunnel
 spanning-tree bpdufilter enable

! Interface connected to N-PE
! It's regular dot1q trunk port
interface GigabitEthernet2/47
 switchport
 switchport trunk encapsulation dot1q
 switchport mode trunk
```

N-PE (3&4) Configuration

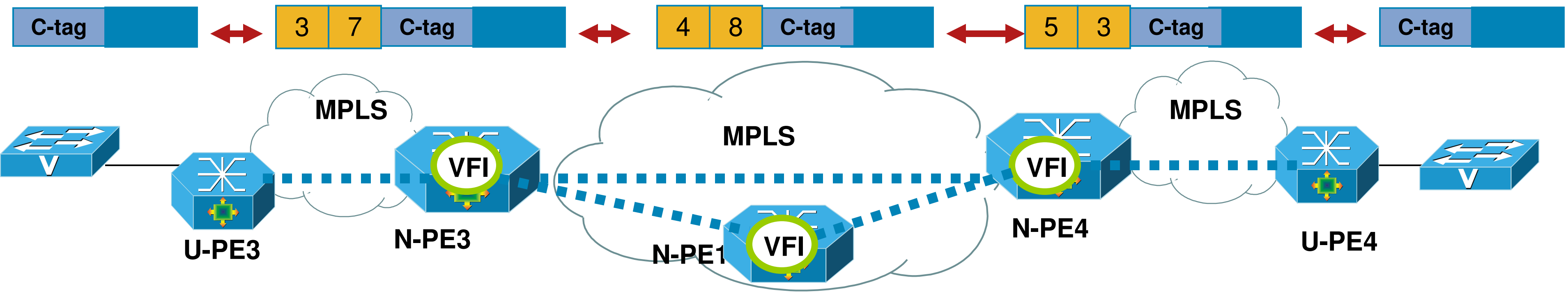
```
! Same VPLS VFI config as flat VPLS

! Attachment circuit has two config options

! Option 1 – dot.1q trunk if it connected to U-PE like N-PE3

interface GigabitEthernet5/1
 switchport
 switchport trunk encapsulation dot1q
 switchport mode trunk

! Option 2 – dot1q tunnel if it connected to CE directly, like N-PE4
interface GigabitEthernet5/1
 switchport
 switchport access vlan 11
 switchport mode dot1q-tunnel
Spanning-tree bpdufilter enable
```

H-VPLS with MPLS Access



- Full Mesh – Pseudowires
- LDP

H-VPLS with MPLS Access Example



U-PE3 Configuration

```
! Regular EoMPLS configuration on U-PE
! Use port-mode in this example

interface GigabitEthernet2/13
 xconnect 10.0.0.3 11 encap mpls


! Uplink is MPLS/IP  to support EoMPLS

interface GigabitEthernet2/47
 ip address 10.0.57.2 255.255.255.252
 mpls ip
```

N-PE3 Configuration

```
! Define VPLS VFI
l2 vfi vpls11 manual
 vpn id 11
 neighbor 10.0.0.1 encapsulation mpls
 neighbor 10.0.0.4 encapsulation mpls
 neighbor 10.0.0.7 encapsulation mpls no-split-horizon

! Attach VFI to VLAN interface
interface Vlan11
 xconnect vfi vpls11

! Attachment circuit is spoke PW for H-VPLS MPLS access
! Downlink is MPLS/IP configuration to support H-VPLS
interface GigabitEthernet4/0/1
 ip address 10.0.57.1 255.255.255.252
 mpls ip
```

H-VPLS with MPLS Access Example

show CLI

```
NPE3#sh mpls l2 vc 11

Local intf     Local circuit              Dest address    VC ID      Status
-------------  -------------------------- --------------- ---------- ----------
VFI vpls11     VFI                        10.0.0.1        11         UP
VFI vpls11     VFI                        10.0.0.4        11         UP
VFI vpls11     VFI                        10.0.0.7        11         UP


NPE3#sh vfi vpls11

Legend: RT=Route-target, S=Split-horizon, Y=Yes, N=No

VFI name: vpls11, state: up, type: multipoint
  VPN ID: 11
  Local attachment circuits:
    Vlan11
  Neighbors connected via pseudowires:
  Peer Address     VC ID        S
  10.0.0.1         11           Y
  10.0.0.4         11           Y
  10.0.0.7         11           N
```

H-VPLS with MPLS Access Example

show CLI

```
NPE3#sh mac-add vlan 11
Legend: * - primary entry
        age - seconds since last seen
        n/a - not available

  vlan   mac address     type    learn     age              ports
------+----------------+--------+-----+----------+--------------------------
   11  2222.2211.1111   dynamic  Yes          0   10.0.0.1, 11
   11  2222.2233.3333   dynamic  Yes          0   10.0.0.7, 11 ← spoke PW
   11  2222.2244.4444   dynamic  Yes          0   10.0.0.4, 11
```

```
UPE3#sh mpl l2 vc 11

Local intf     Local circuit              Dest address    VC ID      Status
------------- ------------------------- --------------- ---------- ----------
Gi2/13         Ethernet                   10.0.0.5        11         UP
```

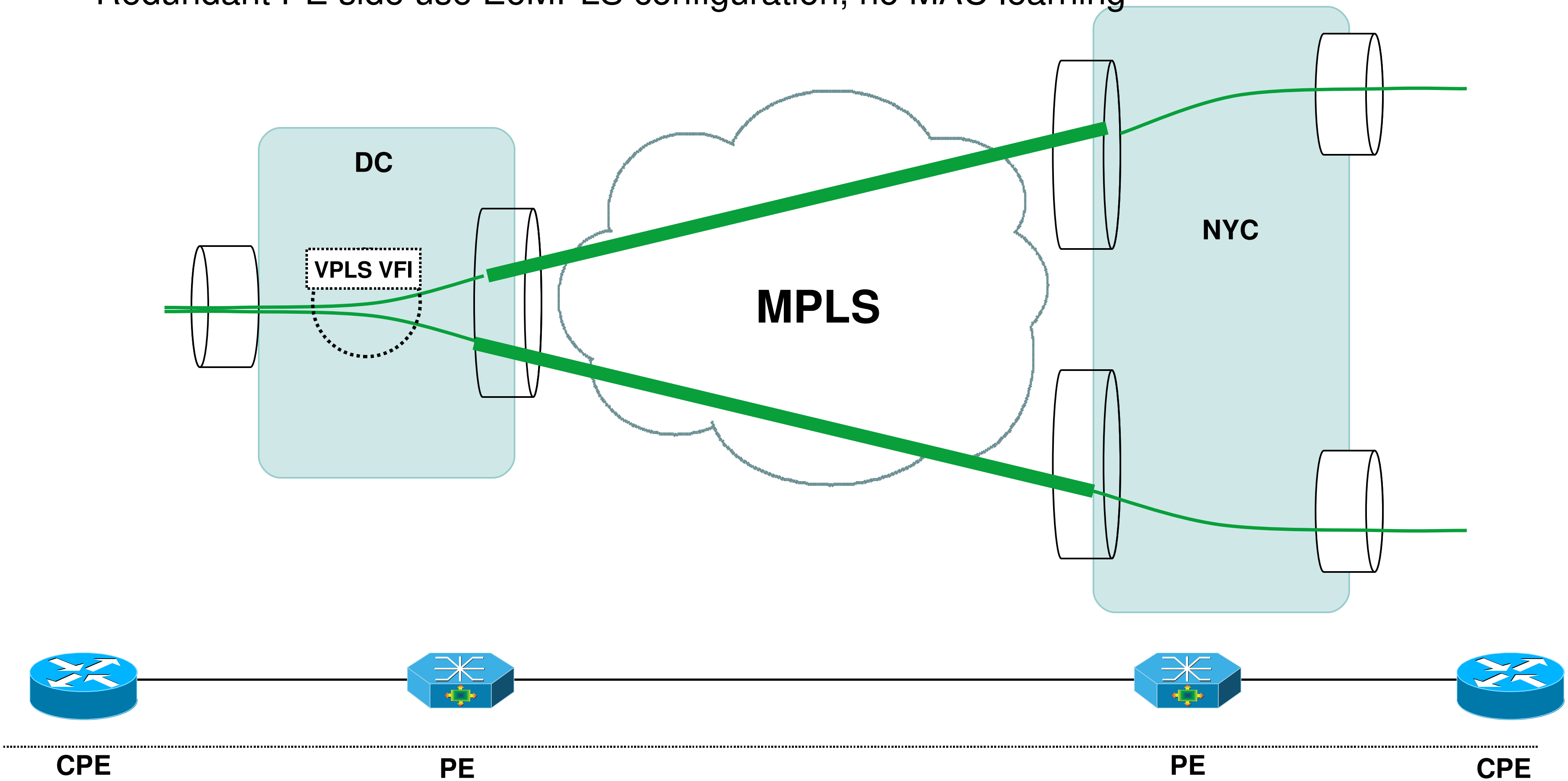
H-VPLS/VPLS Topology Comparison

	Flat VPLS – Ethernet access without QinQ	H-VPLS – Ethernet access with QinQ	H-VPLS - MPLS access
Pros	•Ethernet network benefit – simple, high bandwidth, cheap, efficient local switching and broadcast/multicast distribution	•Same Ethernet network benefit as flat VPLS •Hierarchical support via QinQ at access •Scalable customer VLANs (4kx4k) •4k customer limit per Ethernet island	•Fast L3 IGP convergence •MPLS TE and FRR (50msec convergence time) •Advanced MPLS QoS •Hierarchical support via spoke PW at access • Spoke PE can have QinQ attachment circuit for additional level of hierarchy
Cons	•Not hierarchical, not scalable •Customer VLAN can't over lap (with exception of VLAN translation). 4K customer VLAN limit in Ethernet access domain •High STP re-convergence time	•High STP re-convergence time (potentially improved by different L2 protocols)	•More complicated provisioning •Requires MPLS to u-PE, potentially more expensive u-PE device

Flexible Design with H-VPLS (1)

Node Redundancy

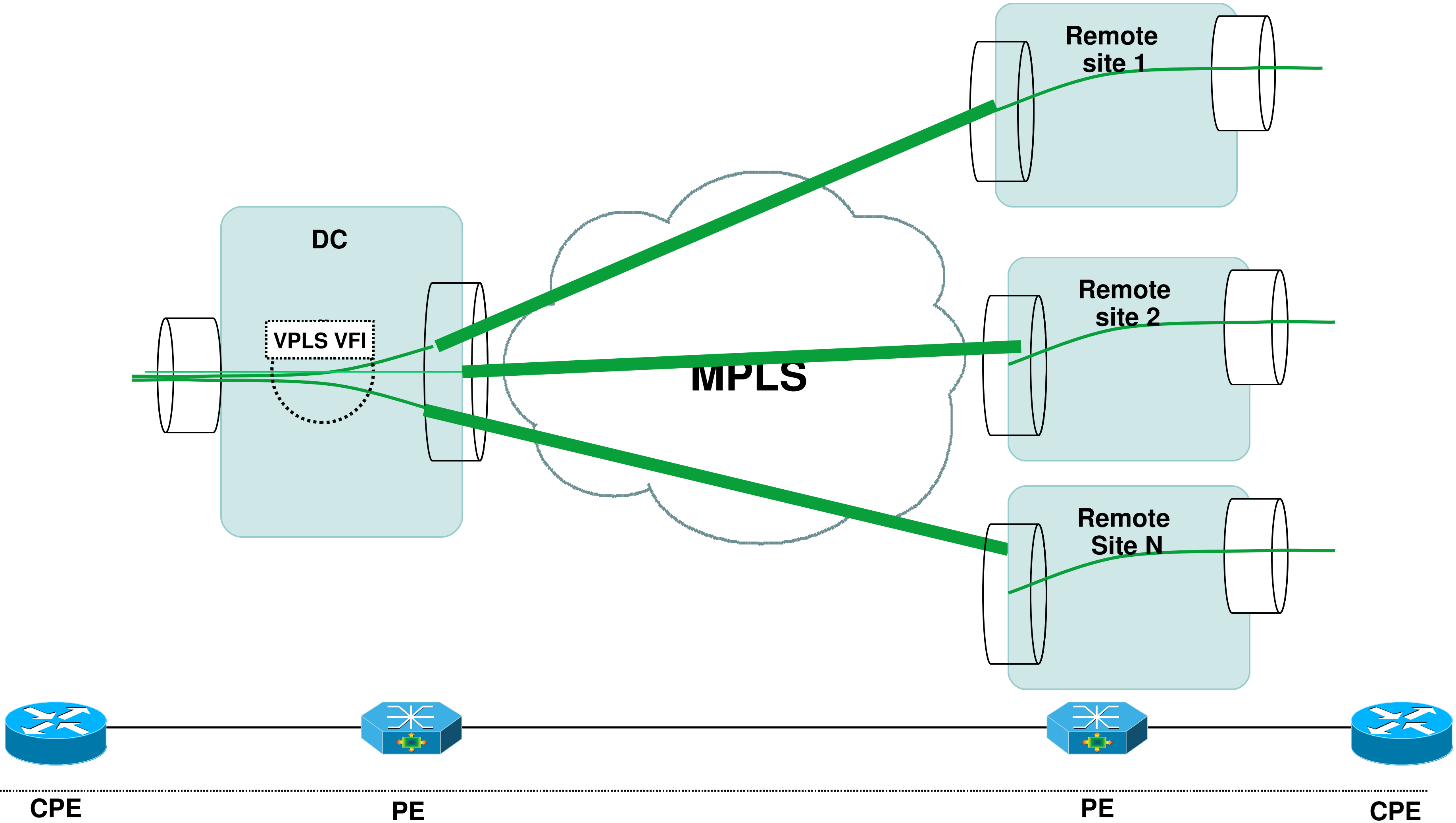
- Site-to-site L2 circuit. One side have redundant PEs, the other side has single PE
- Single PE side use H-VPLS configuration to have two active PWs going to redundant PEs. MAC learning and forwarding are involved
- Redundant PE side use EoMPLS configuration, no MAC learning



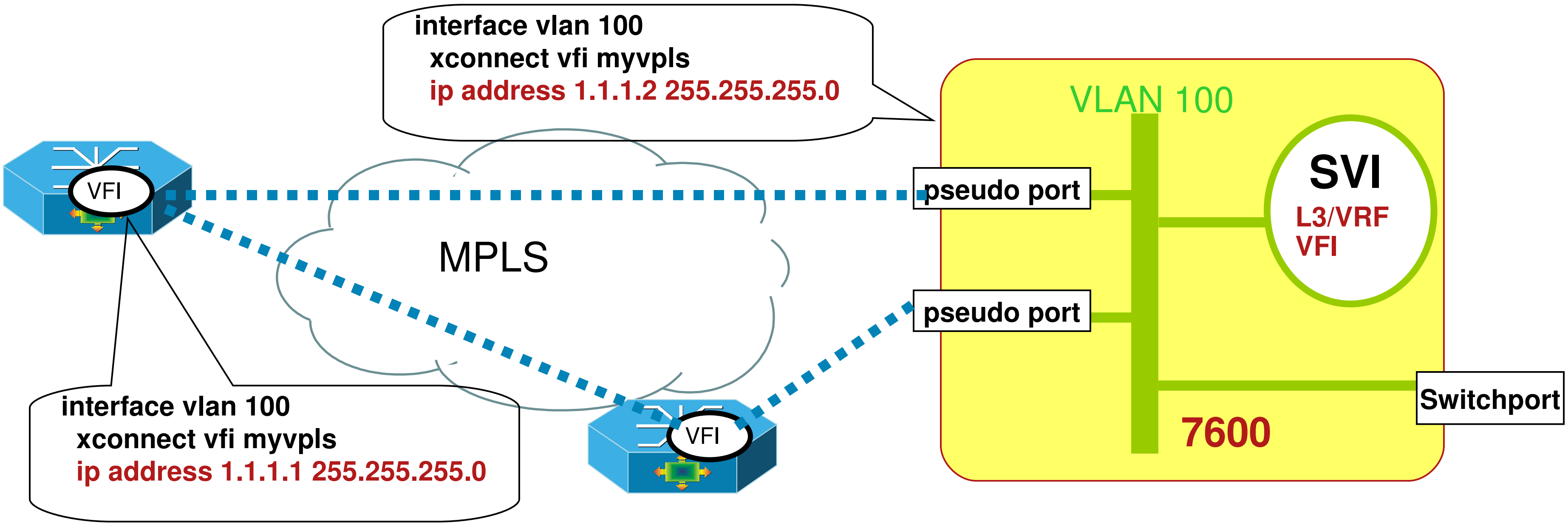
Flexible Design with H-VPLS (2)

VPLS-on-a-stick Design

- Use H-VPLS for spoke-and-hub topology, point-to-multipoint design



Routed PW (VPLS) – What's it?



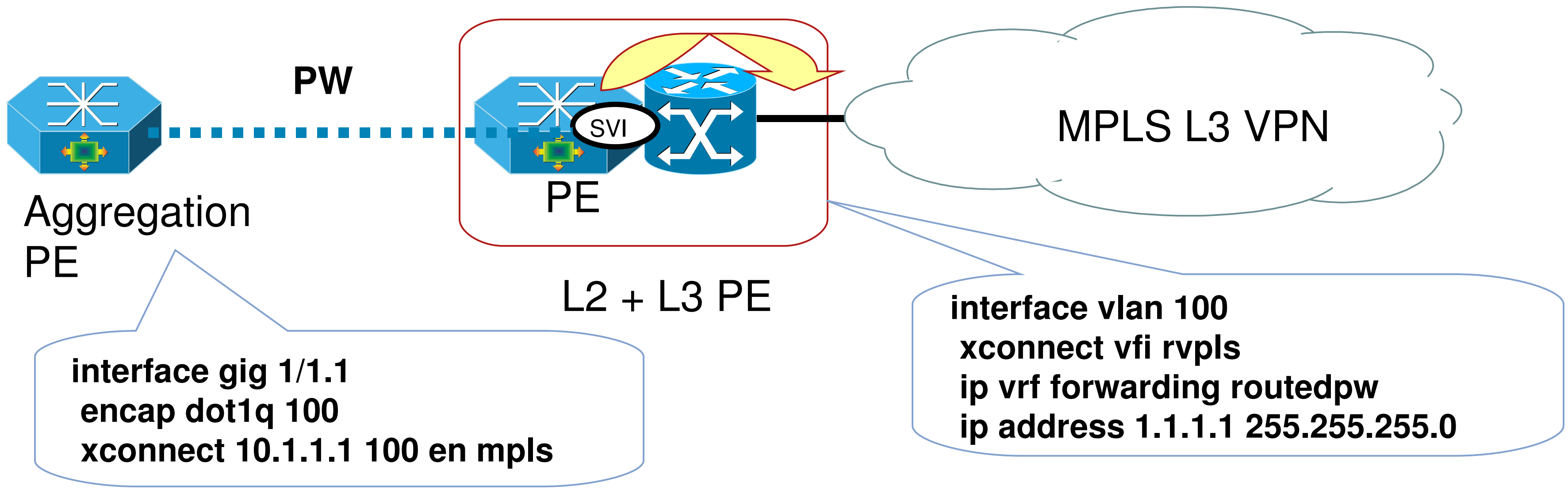
- L2 switching among L2 switchport and L2 pseuo port (PW) with MAC learning/forwarding
- L3 routing via SVI for both L2 switchport and L2 pseuo port (PW)
- Same L3 attributes (IP, Routing) on SVIs in addition to a xconnect configuration
- "Routed PW" is the ability to L3 route in addition to L2 bridge frames to and from PW

Routed PW (VPLS) – Features Supported

In general, Routed PW can now offer same functionality as other L3 tunnels like GRE tunnel. Virtually all the listed L3 features should work.

- **IP address and IP VRF**
- **ACLs**
- **PBR**
- **Routing protocols, OSPF, RIP, EIGRP,ISIS, BGP**
- **Netflow**
- **QoS Policing for SVI**
- **IP unnumbered**
- **Mcast routing, IGMP, PIM**
- **HSRP/VRRP/GLBP**

Routed PW (VPLS) Application Scenario –
PW Terminated into L3/VRF



PE receive EoMPLS frame from PW

After EoMPLS decap'd, it become normal IP packet

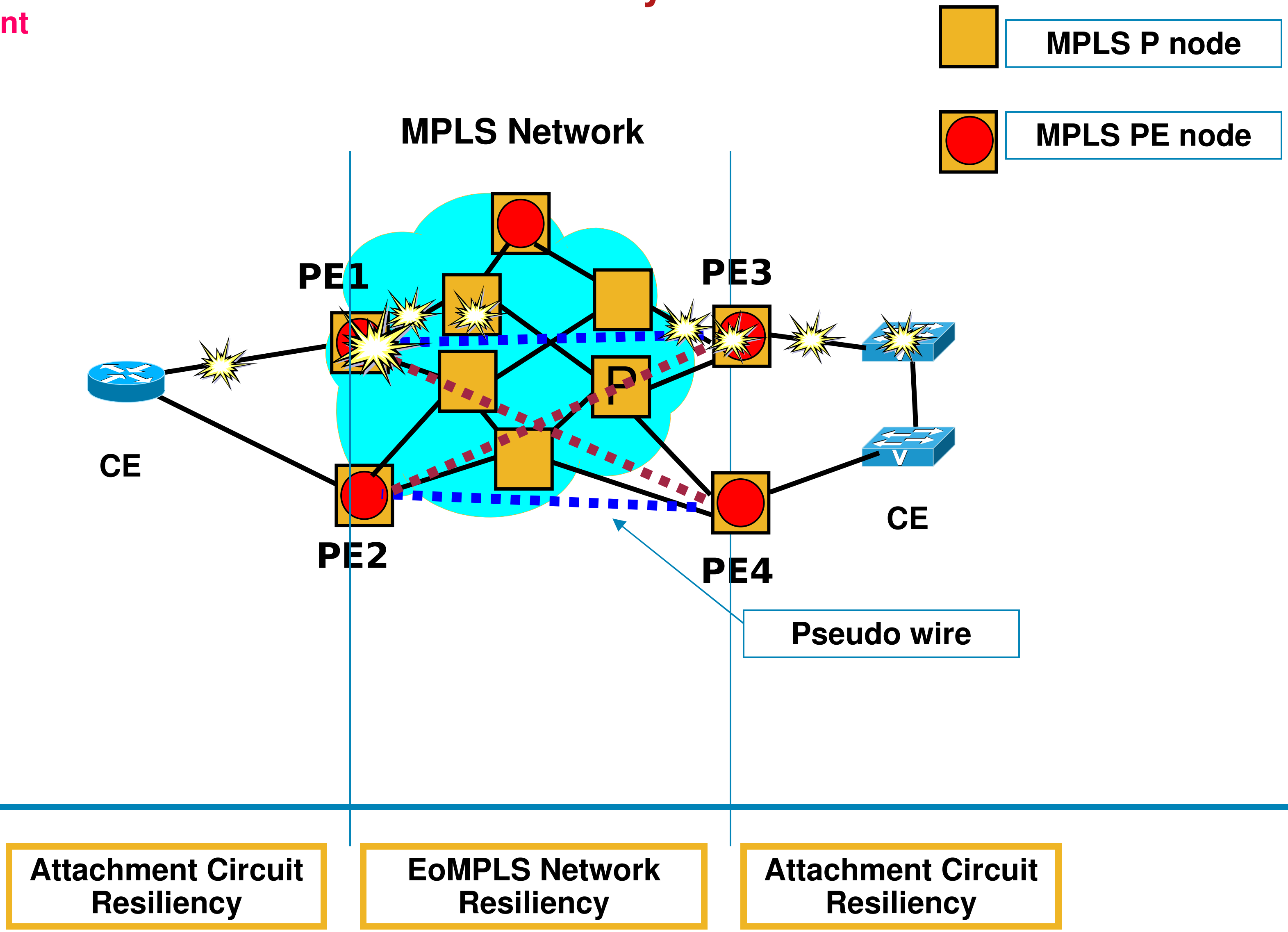
IP packet is L3 routed into L3 VPN cloud via SVI

Single box solution!

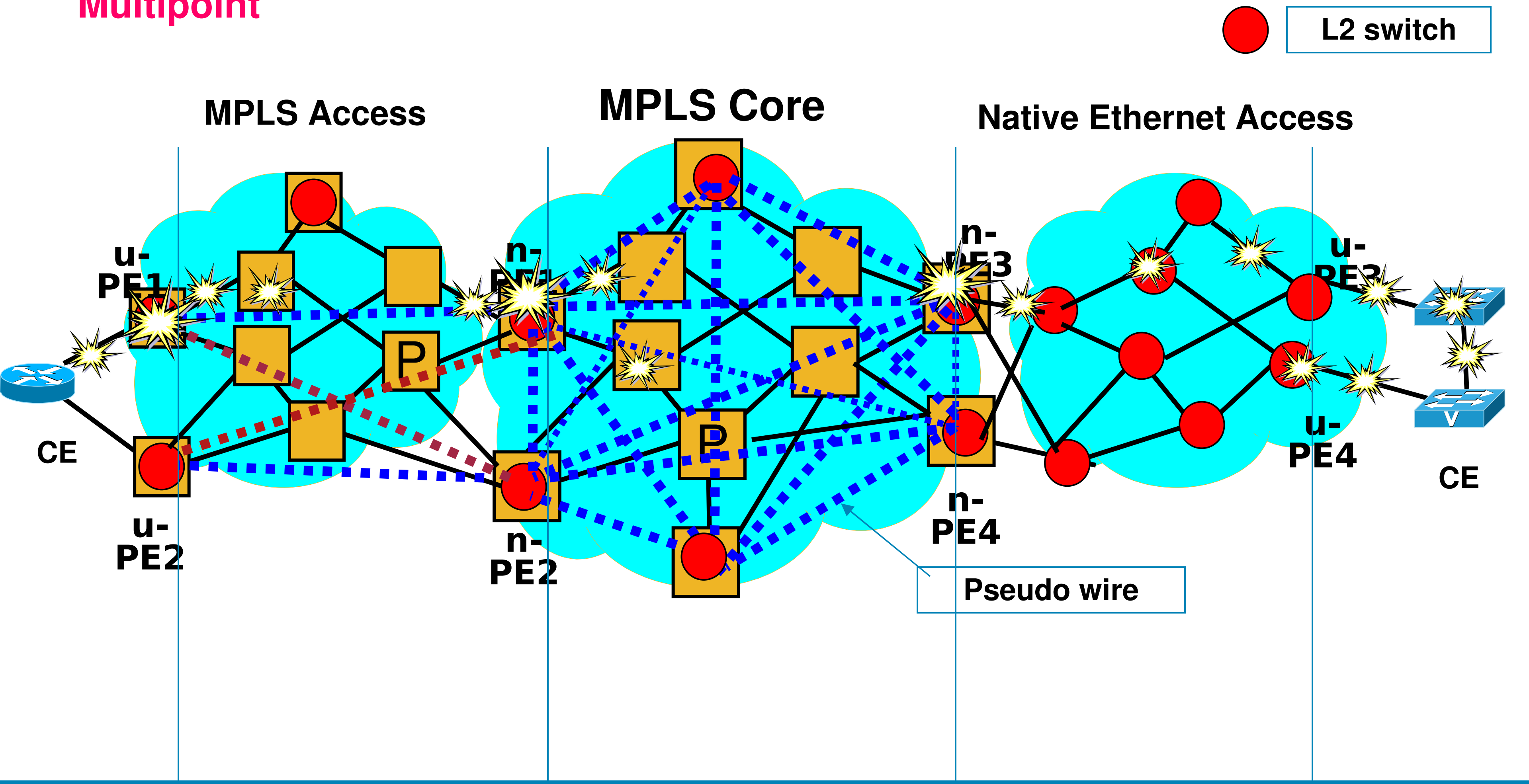
EoMPLS/VPLS Network Resiliency

End-to-End EoMPLS Network Resiliency

Point-to-Point



End-to-End VPLS/H-VPLS Network Resiliency
Multipoint



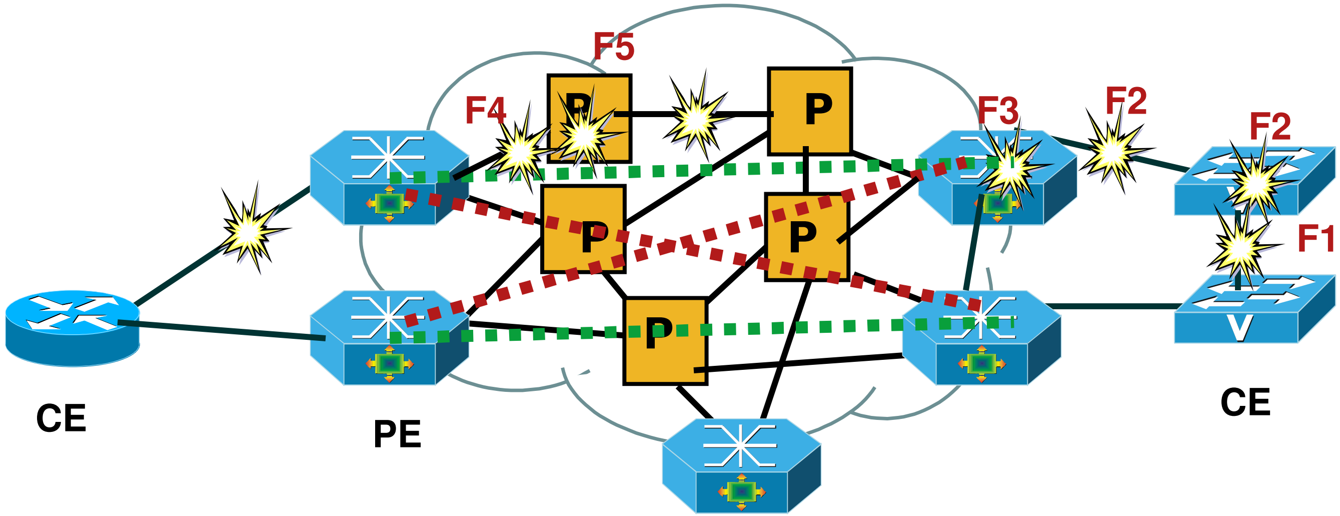
Attachment Circuit Resiliency	H-VPLS with MPLS Access Network Resiliency	VPLS Core Network Resiliency	H-VPLS with Native Ethernet Access Network Resiliency	Attachment Circuit Resiliency

EoMPLS Network Resiliency

Feature Highlights

- MPLS TE FRR (fast re-route)
- EoMPLS PW Redundancy

EoMPLS PW Failure Scenarios

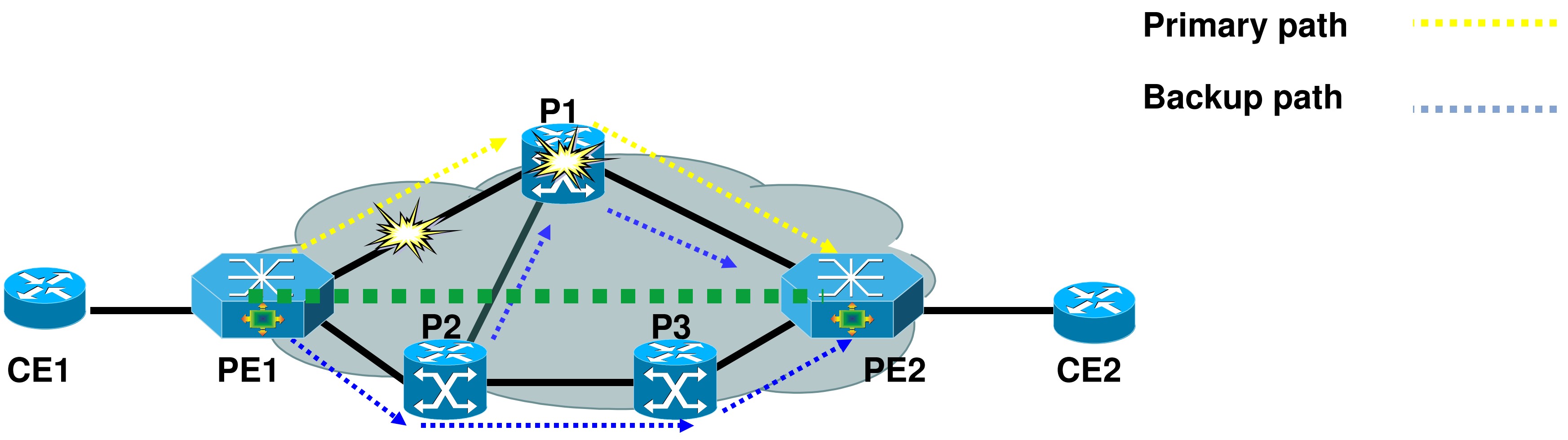


Failure Scenarios

- Failure 1 – CE to CE link failure (out of the scope of this presentation)
- Failure 2 – CE node failure or CE to PE link (or attachment) failure → PW redundancy
- Failure 3 – PE node failure → PW redundancy
- Failure 4 – MPLS link failure (P-P, PE-P, PE-PE) → MPLS TE/FRR
- Failure 5 – P node failure → MPLS TE/FRR

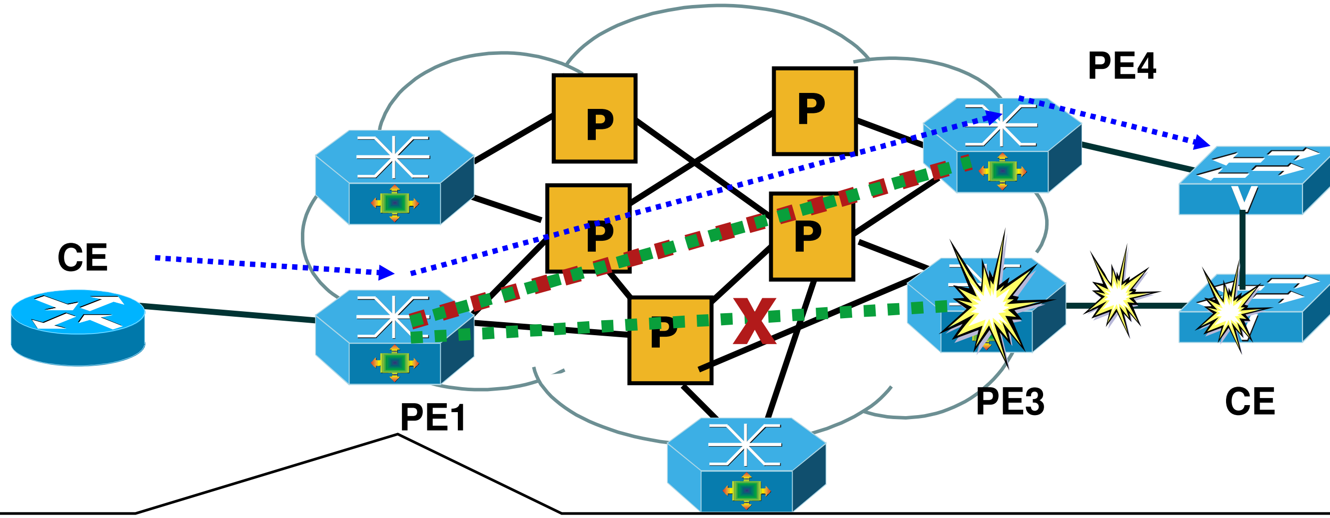
MPLS-TE/FRR (fast re-route) for PW Protection

Address F4 and F5



- FRR builds an alternate path to be used in case of a network failure (Link or P Node) / local repair negates convergence delays
- No special configuration for AToM PWs. FRR protected tunnel will support all the traffic traversing the link no matter if it's AToM PW or not
- When tied to POS or certain GE/10GE link ~50ms restore times are achievable
- No PW control or forwarding plane changes during TE FRR

PW Redundancy
Address F2 and F3



pe1(config)#int gig 1/1.1
pe1(config-subif)#encapsulation dot1q 10
pe1(config-subif)# xconnect <PE3 router ID> <VCID> encapsulation mpls
pe1(config-subif-xconn)#backup peer <PE4 router ID> <VCID>

- PW between PE1 and PE3
- If PE3 fail or PE3 attachment circuit fail, PW will go down. TE/FRR won't help this failure scenario
- Solution – create backup PW between PE1 and PE4. When primary PW goes down, backup PE will come up. Traffic will continue between CEs
- Primary and backup PW can be between same pair of PEs, with different Attachment Circuit, or between different pair of PE like this example

PW redundancy- Config Examples

- Example 1 – The debounce timer is set to 3 seconds so that we don't allow a switchover until the connection has been deemed down for 3 seconds.

```
interface gig1/1
 xconnect 10.0.0.1 100 encapsulation mpls
  backup peer 10.0.0.2 200
  backup delay 3 10
```

- Example 2 – xconnect with 1 redundant peer. In this example, once a switchover occurs, we will not fallback to the primary until the secondary xconnect fails.

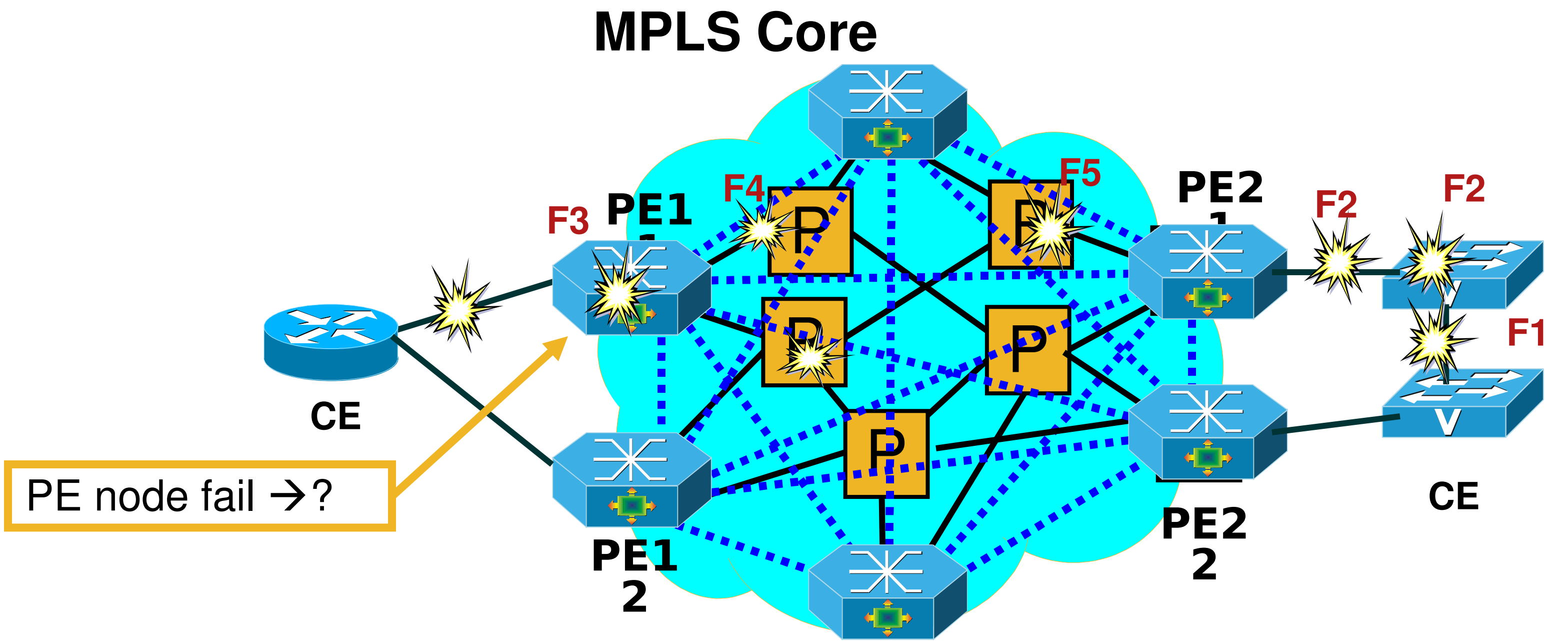
```
Interface gig 1/1
 xconnect 20.0.0.1 50 encapsulation mpls
 backup peer 20.0.0.2 50
 backup delay 0 never
```

VPLS core Network Resiliency

Highlights

- VPLS core has full mesh PWs among all PEs. This provide PE node redundancy natively

VPLS Core Failure Scenario



Failure Scenarios

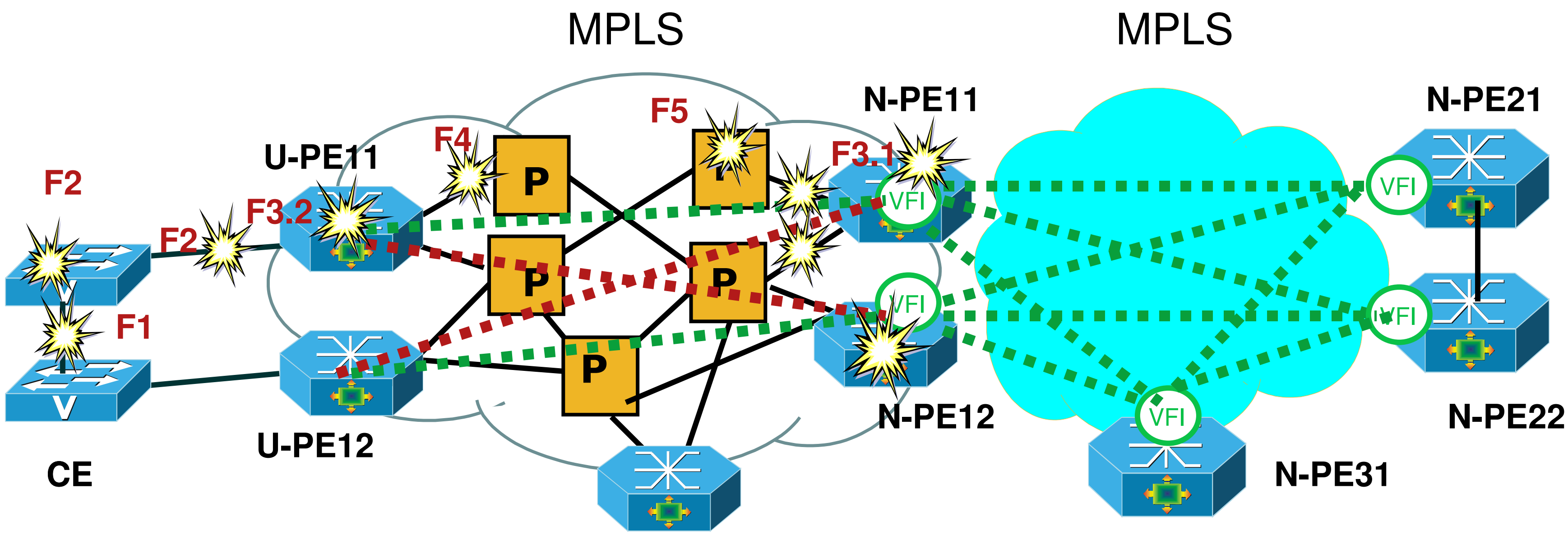
- Failure 1 – CE to CE link failure (out of the scope of this presentation)
- Failure 2 – CE node or CE to PE link (or attachment) failure → attachment circuit resiliency
- Failure 3 – PE node failure → CE re-direct traffic to the redundant PE which still has active PW to the remote PEs. No special configuration needed
- Failure 4 – MPLS link failure (P-P, PE-P, PE-PE) → TE/FRR
- Failure 5 – P node failure → TE/FRR

H-VPLS with MPLS Access Network Resiliency

Highlights

- u-PE use PW redundancy to create primary/backup PW dual home to two n-PEs
- Upon PW switchover to different n-PE, it need MAC withdrawal on the peer n-PEs

H-VPLS with MPLS Access Network Resiliency

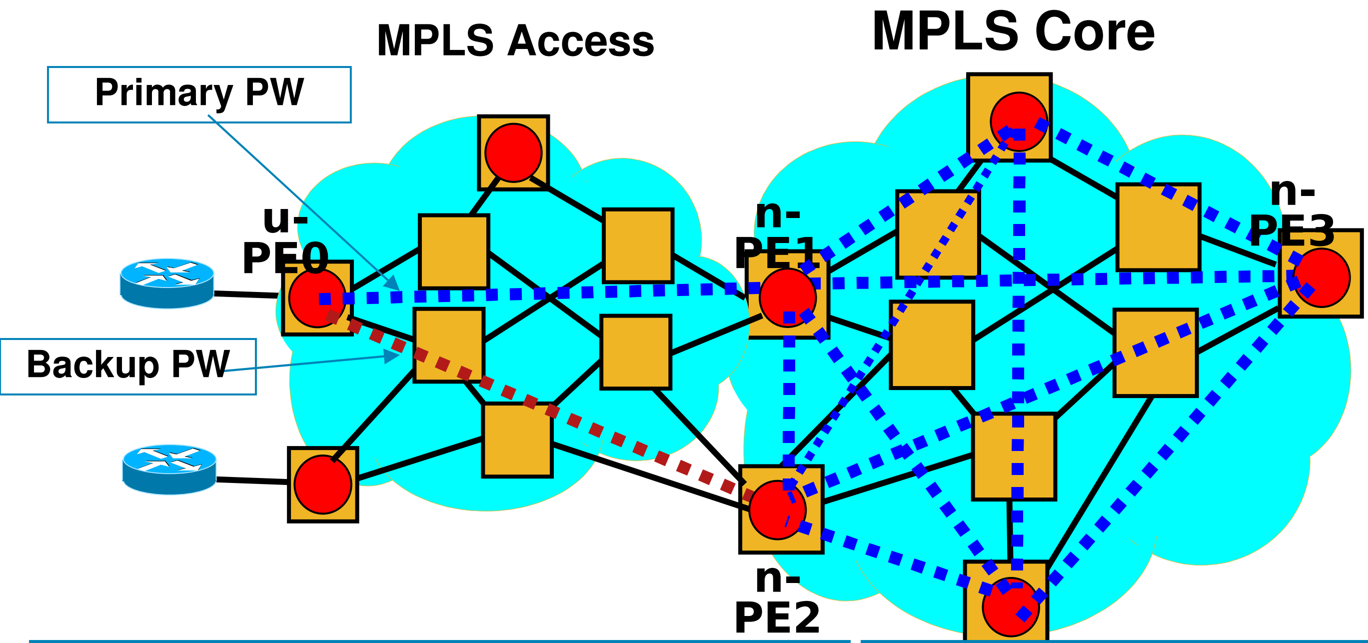


Failure Scenarios

- Failure 1 – CE to CE link failure (out of the scope of this presentation)
- Failure 2 – CE node or CE to PE link (or attachment) failure → attachment circuit resiliency
- Failure 3.1 – N-PE node failure → PW redundancy
- Failure 3.2 – U-PE node failure → attachment circuit resiliency
- Failure 4 – MPLS link failure (P-P, PE-P, PE-PE) → MPLS TE/FRR
- Failure 5 – P node failure → MPLS TE/FRR

H-VPLS with MPLS Access Network Resiliency

F3.1 n-PE node redundancy



U-PE has primary/backup PW to two n-PEs

If primary PW fail (for example, primary N-PE fail), u-PE will switchover to the backup PW

U-PE Configuration

U-PE use PW redundancy configuration to create primary/backup PWs to two N-PEs

```
interface gig1/1
 no ip address
 xconnect 10.0.2.1 998 encapsulation mpls
  backup peer 10.0.2.2 998
```

N-PE Configuration (N-PE1)

N-PE use regular H-VPLS configuration, no PW redundancy configuration is involved

```
l2 vfi red-vpls manual
 vpn id 998
 neighbor 10.0.2.10 encapsulation mpls no-split-horizon
 neighbor 10.0.2.2 encapsulation mpls

interface Vlan998
 no ip address
 xconnect vfi red-vpls
```

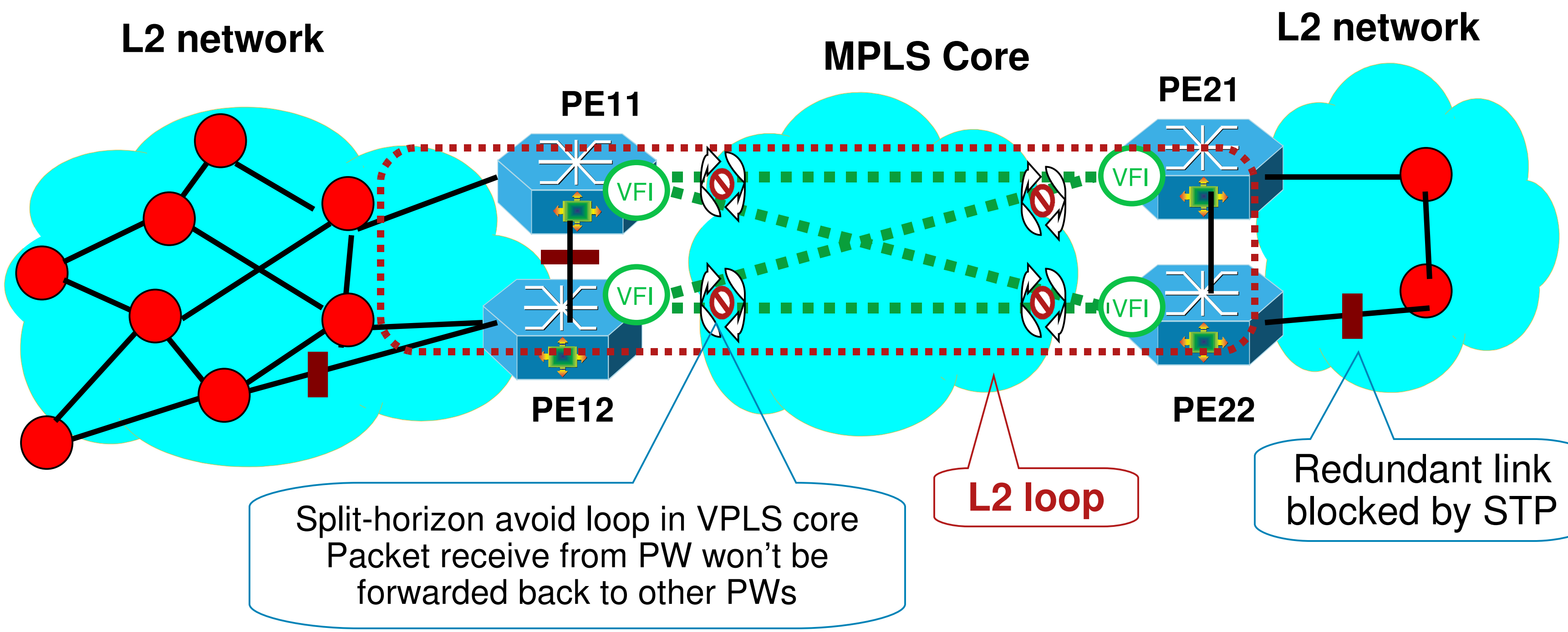
H-VPLS with Native Ethernet Access Network Resiliency Overview

Highlights

2 possible approaches

H-VPLS with Ethernet Access PE Redundancy

Key Requirement – How to avoid the L2 loop?

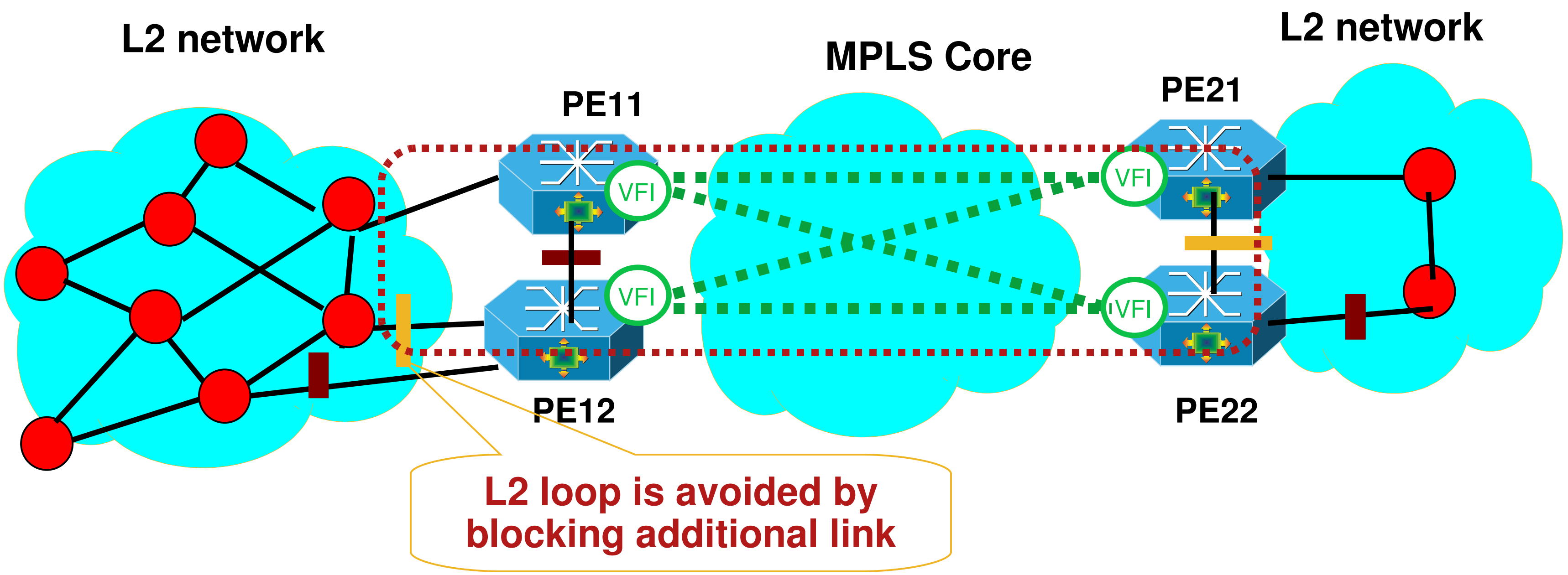


- VPLS core – full mesh PWs, use split-horizon to avoid loop, not run STP
- L2 network – run STP or other L2 protocols to avoid loop
- Fundamental requirement – STP or L2 protocols are not across VPLS core to make L2 domain locally
- With redundant N-PEs, VPLS full mesh PWs + local L2 network → L2 loop

- **Key requirement – how to avoid the L2 loop?**

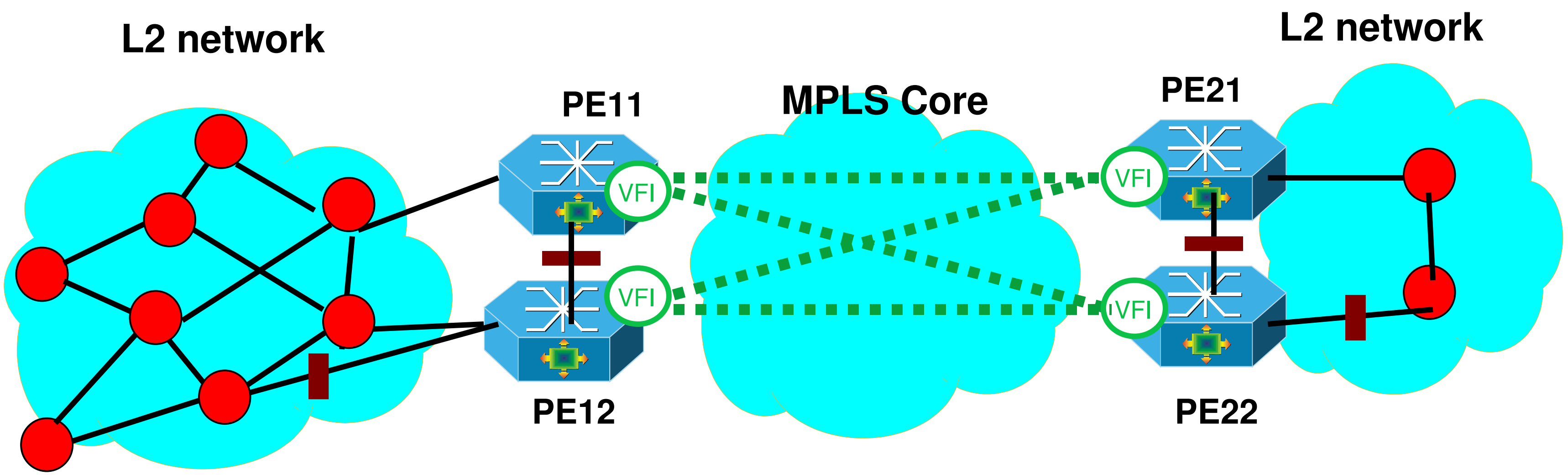
L2 Loop Prevention Approach 1 –

Single L2 Path between L2 network and N-PE group



- ALWAYS ONLY ONE available L2 data path between L2 network and N-PE group
- Redundant links are blocked by L2 protocols
- Packet sent from N-PE group to L2 network wont be loop'd back, thus NO loop
- Traditional L2 protocols can't achieve this goal, need special BPDU relay mechanism

L2 Loop Prevention Approach 2 –
End-to-End STP



- VPLS PW tunnel BPDU across sites
- End-to-End STP will break the loop accordingly
- More complex, not scale to more than two sites
- Topology changes in one site can affect other sites
- Exist today already

- Not recommended

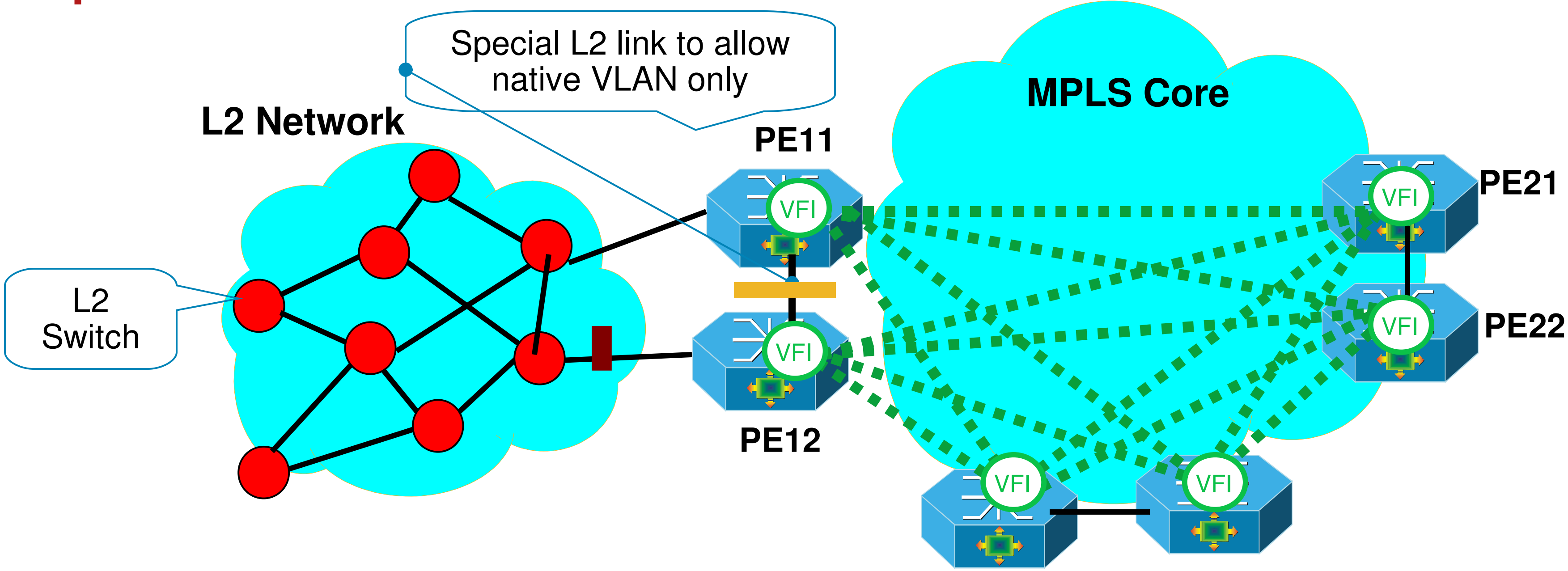
H-VPLS with Native Ethernet Access Network Resiliency (first approach)

Highlights

Two options exist today, enhancement in the future release

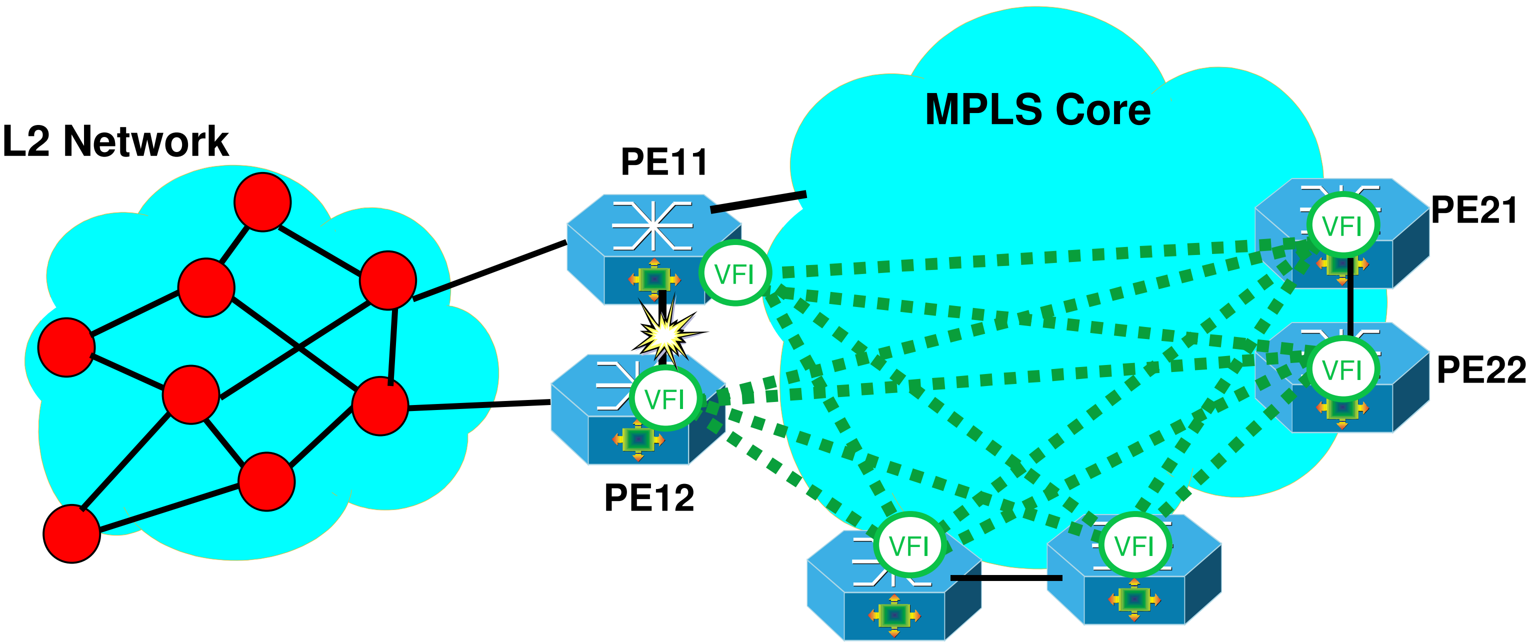
- N-PE participate STP, require dedicated L2 link between two N-PEs
- N-PE doesn’t run STP, but relay BPDU through dedicated PW

Option 1 - Dedicated L2 link between two n-PEs



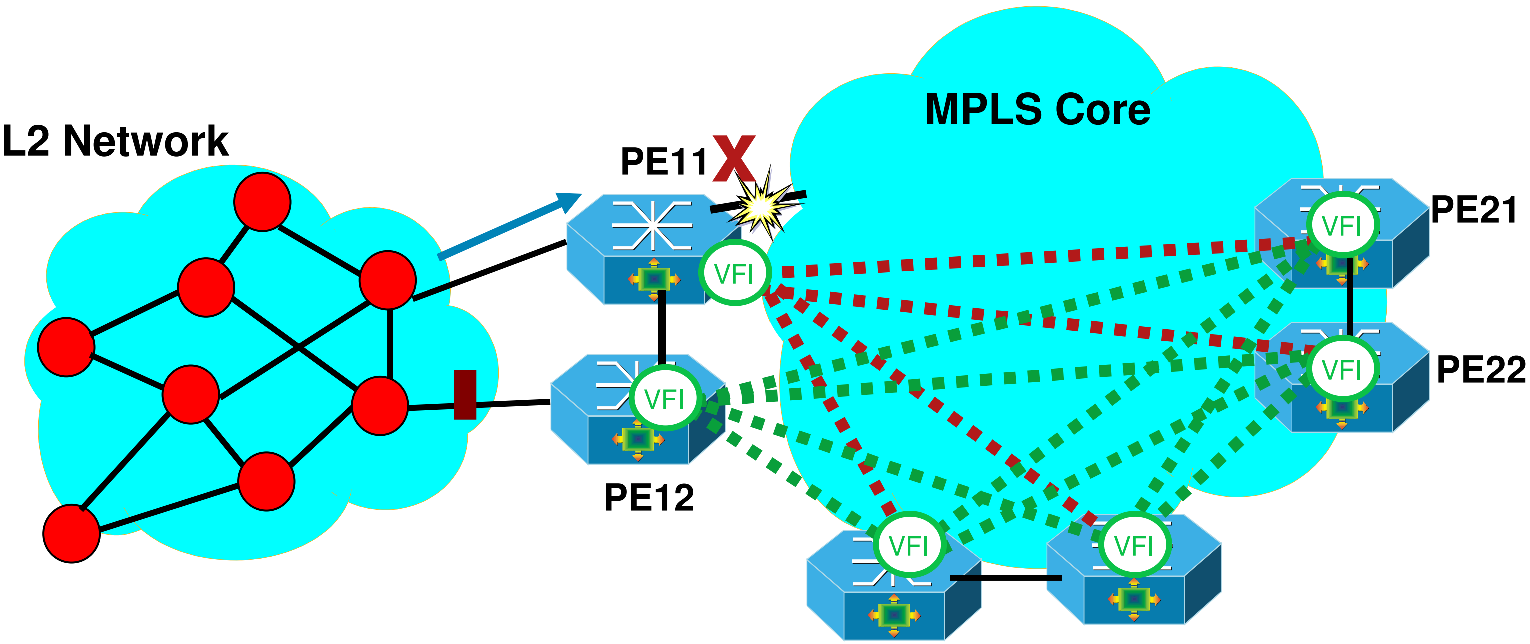
- Dedicated L2 link between two PEs. This link only allow native VLAN. Thus MST BPDU packet will pass through this link. No user data is allowed in native VLAN, thus user data can't pass through this link
- PE must be STP root or by configuring STP port cost to make sure this special link is not "blocked" by STP.
- The trick is to let STP put this special link into forwarding state, but actually no user data packet pass through
- Convergence time is determined by STP. With rapid STP, it can get 1-2 seconds convergence time

Option 1 - issue #1



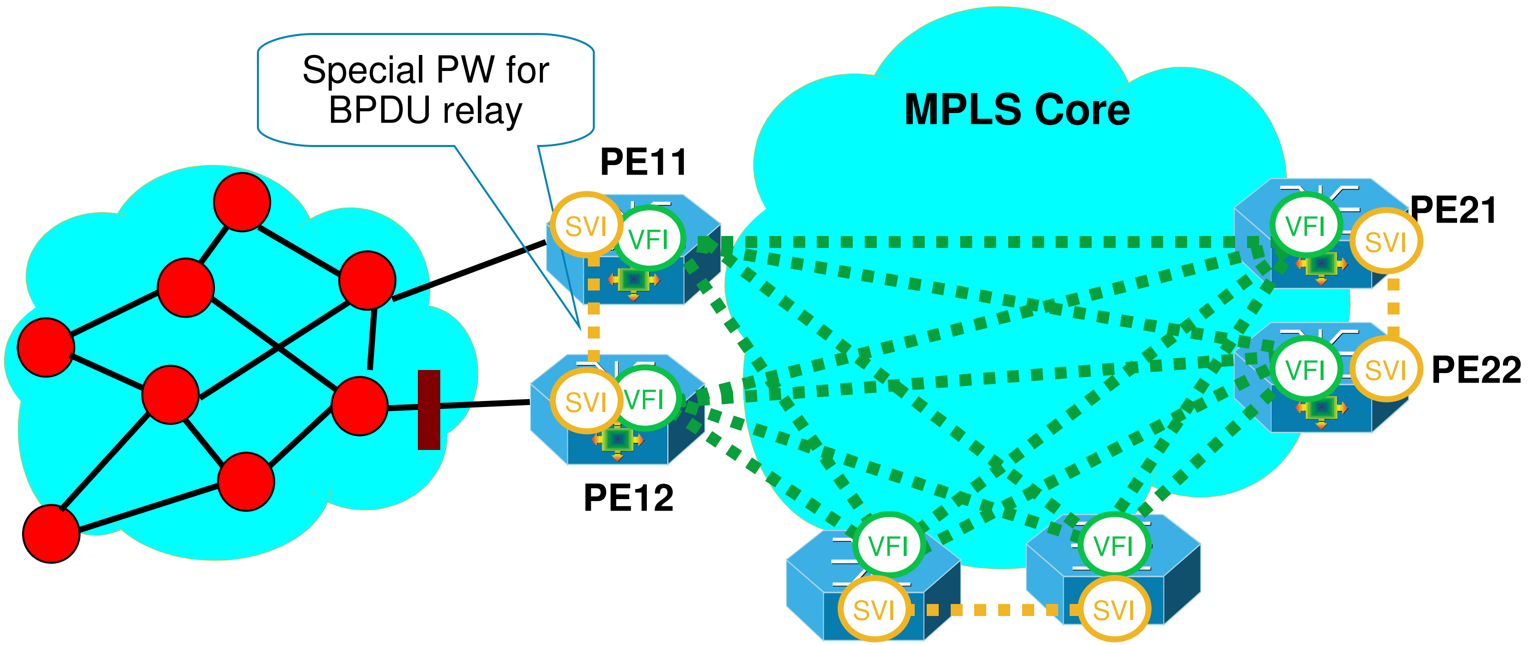
- Issue – if dedicated L2 link fail, then redundant link will be unblocked. This will create duplicated packets and possible L2 loop.
- Solution – Use port channel between two PEs, make sure the link between two PEs are always up. The drawback is wasting the port and link

Option 1 – issue#2



- Issue – if primary PE's MPLS uplinks fail, L2 protocol is not aware. Packet is still forwarded to original primary PE. Since there is no active PWs on the primary PE, packet get dropped
- Solution – have redundant L3 MPLS link between two PEs. If one MPLS link is down, it can have backup link going through the other PE

Option 2 - Dedicated PW between two n-PEs



- No L2 link between two PEs. Instead, a dedicated PW is created on native VLAN
- Native VLAN is not used to pass data traffic. For MST mode, BPDU is sent through native VLAN. As result, BPDU is relayed through this dedicated PW
- Redundant links from L2 switches to PE will be blocked by STP
- To tunnel BPDU through PW, STP must be disable in current release
- Require u-PE run MST mode

Option 2 – Sample configuration

U-PE Sample Configuration

```
spanning-tree mode mst
spanning-tree extend system-id
!
spanning-tree mst configuration
 name cisco
 instance 1 vlan 11, 13, 15, 17
 instance 2 vlan 12, 14, 16, 18

! Tune the STP timer
spanning-tree mst hello-time 1
spanning-tree mst forward-time 4
spanning-tree mst max-age 6

interface GigabitEthernet2/47
 switchport
 switchport trunk encapsulation dot1q
 switchport mode trunk
 spanning-tree cost 50000 ← configure high cost
on the link to N-PE, make sure the blocked link
is between U-PE and N-PE, instead of U-PEs
internal links

! Configure STP root for load balancing
spanning-tree mst 1 priority 4096
spanning-tree mst 2 priority 8192
```

N-PE Sample Configuration

```
spanning-tree mode pvst
spanning-tree extend system-id
no spanning-tree vlan 1-4094 ← disable
spanning-tree on N-PE

l2 vfi bpdu-pw manual
 vpn id 1
 neighbor 10.0.0.6 encapsulation mpls

interface Vlan1 ← special PW peering with
the other N-PE to relay BPDU
 xconnect vfi bpdu-pw
```

Attachment Circuit Resiliency

Attachment Circuit Redundancy Scenarios

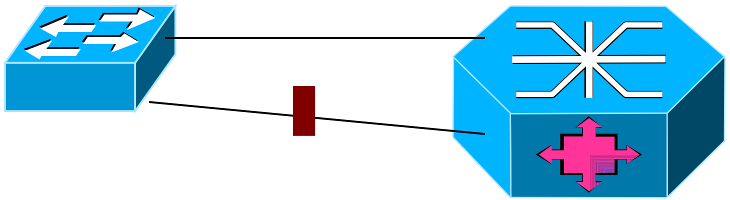
- Single CE dual home to single PE
- Single CE dual home to two PEs
- L2 ring/Network connect to single PE
- L2 ring/Network connect to two PEs

Possible Redundancy Solution

- Etherchannel
- Flexible Link
- STP
- BPDU Relay

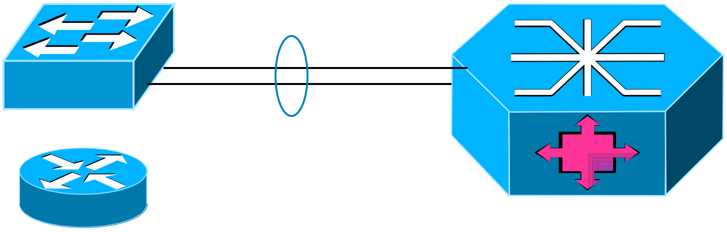
Attachment Circuit Resiliency – 1

Single CE dual home to single PE



FlexLink on CE side

Simple, well known, active-backup model

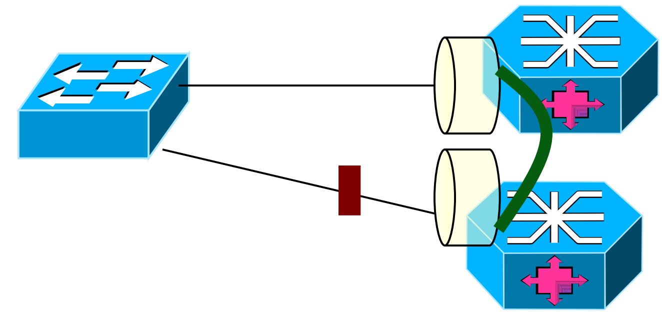


Port Channel

Simple, well known, active-active model

Attachment Circuit Resiliency – 2

Single CE dual home to two PEs

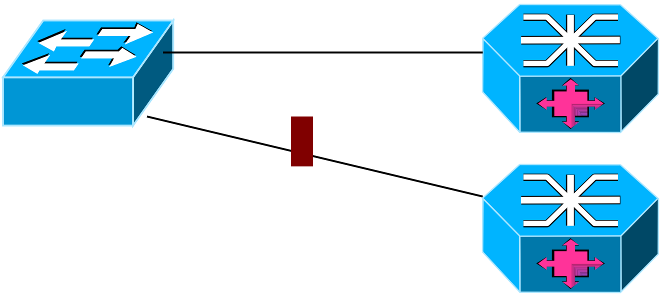


BPDU relay

CE run STP MST mode

PE doesn't run STP

CE BPDU is relay via special PW



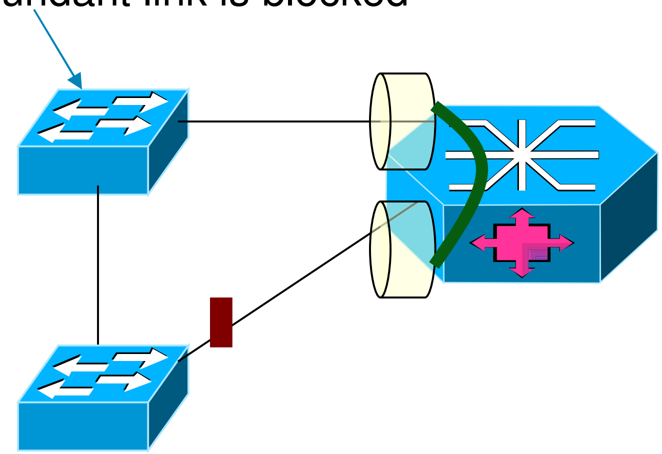
FlexLink on CE side

Simple, well known

Attachment Circuit Resiliency – 3

L2 Networks connect to single PE

Run MST. BPDU is relay by PE, redundant link is blocked



BPDU relay

CEs run STP MST mode

PE doesn't run STP

CEs BPDU is relay by PE

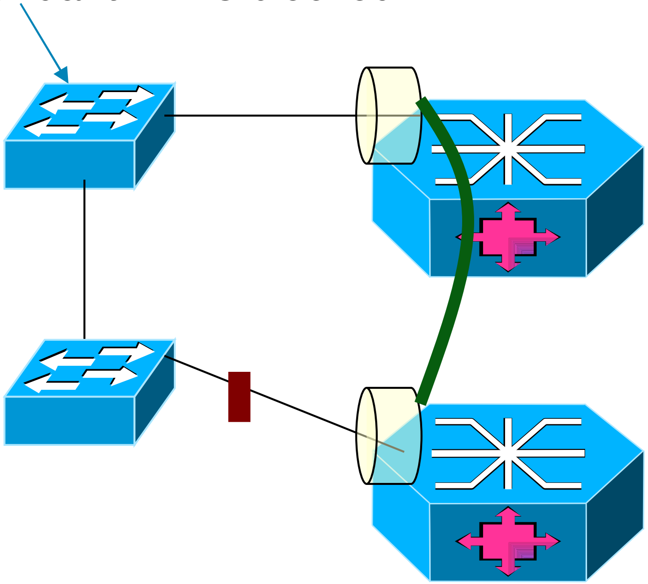
STP

CEs and PE run STP

Attachment Circuit Resiliency – 4

L2 ring connect to two PEs

Run MST. BPDU is relay by PE, redundant link is blocked



BPDU relay

CEs run STP MST mode

PE doesn't run STP

CEs BPDU is relay via special PW

With MST over PW feature, PEs and CEs can participate the same STP domain


MPLS Fundamentals
A Comprehensive Introduction to MPLS Theory and Practice
Luc De Ghein, CCIE No. 1897
Troubleshooting Virtual Private Networks
Master advanced troubleshooting techniques for IPSec, MPLS Layer-3, MPLS Layer-2 (AToM), L2TPv3, L2TPv2, PPTP, and L2F VPNs
Mark Lewis, CCIE No. 6280
Next-Generation Network Services
MPLS Configuration on Cisco IOS Software
DEPLOYING IP AND MPLS QoS FOR MULTISERVICE NETWORKS
THEORY AND PRACTICE
JOHN EVANS • CLARENCE FILSFILS
Lancy Lobo, CCIE No. 4690
Umesh Lakshman
Developing IP-Based Services
SOLUTIONS FOR SERVICE PROVIDERS AND VENDORS
Monique Morrow
Kateel Vijayananda
JEAN PHILIPPE VASSEUR • MARIO PICKAVET • PIET DEMEESTER
NETWORK RECOVERY
PROTECTION AND RESTORATION
OF OPTICAL, SONET-SDH, IP, AND MPLS
Selecting MPLS VPN
Chris Lewis
Steve Pickavance
MPLS VPN Security
A practical guide to hardening MPLS networks
Michael H. Behringer
Monique J. Morrow
Definitive MPLS Network Designs
Field-proven MPLS designs covering MPLS VPNs, pseudowire, QoS, traffic engineering, IPv6, network recovery, and multicast
Jim Guichard, CCIE No. 2069
François Le Faucheur
Jean-Philippe Vasseur
Next-Generation Network Services
A guide to building service-oriented networks to differentiate and grow your business
Robert Wood
MPLS and VPN Architectures
CCIP Edition
Prepare for CCIP certification as you learn to design and deploy MPLS-based VPNs
Ivan Pepelnjak, CCIE No. 1354
Jim Guichard, CCIE No. 2069
Fault-Tolerant IP and MPLS Networks
Iftekhar Hussain
MPLS and Next-Generation Networks
Foundations for NGN and Enterprise Virtualization
Monique J. Morrow, CCIE No. 1711
Azhar Sayeed
MPLS Technology and Applications
QoS for IP/MPLS Networks
A comprehensive guide to implementing QoS in IP/MPLS networks using Cisco IOS and Cisco IOS XR Software
Santiago Alvarez, CCIE No. 3621
Traffic Engineering with MPLS
Design, configure, and manage MPLS TE to optimize network performance
Eric Osborne, CCIE #4122
Ajay Simha, CCIE #2970
Cisco WAN Switching Professional Reference
Edited by Tracy Thorpe
Advanced MPLS Design and Implementation
Vivek Alwayn, CCIE No. 2995
MPLS Configuration on Cisco IOS Software
Lancy Lobo, CCIE No. 4690
Umesh Lakshman

Questions?

Thanks for your time & attention!
Enjoy the rest of the Program!

Acknowledgement

- Santiago Alvarez, Javed Asghar, Rajiv Asati