Multiprotocol Label Switching (MPLS) – TE usage

Engenharia de Computadores e Telemática
4º ano, 1º semestre, 2020/2021

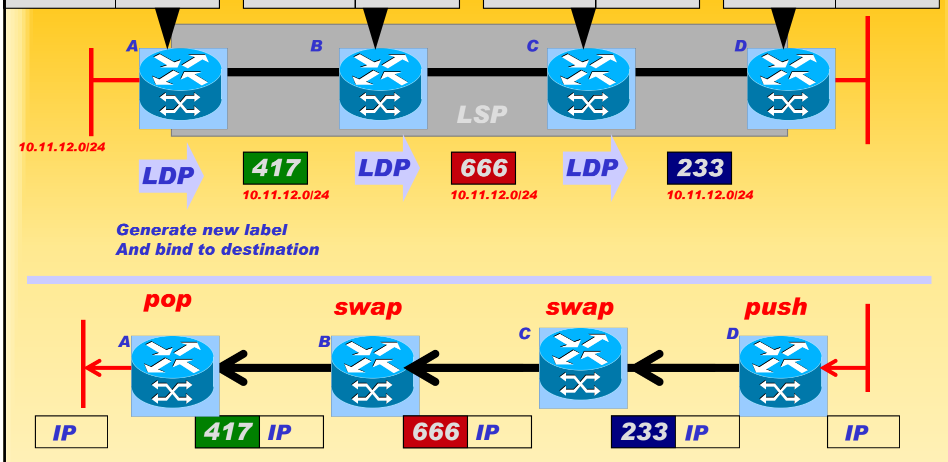
LDP and Hop-by-Hop routing

network	next-hop
10.11.12.0/24	direct

network	next-hop
10.11.12.0/24	A

network	next-hop
10.11.12.0/24	B

network	next-hop
10.11.12.0/24	C



Constrained based Routing

- A cost is associated to each link
- Each link has a further set of attributes that represent performance metrics

The routing objective is to determine the lowest cost path that does not violate the restrictions that were assigned

- Restrictions can be associated to a set of performance characteristics, like for example, bandwidth, delay, priority, etc.
  - For the bandwidth case, the restriction that is imposed to the routing algorithm is that the path must have, on each connection it traverses, a bandwidth higher than a certain threshold.
  - In this case, the connection attribute used is the available bandwidth.

55

Constraint Based Routing?

Basic components

Problem here:
OSPF areas hide information for scalability. So these extensions work best only within an area…

1.Specify path constraints
2.Extend topology database to include resource and constraint information — ***Extend Link State Protocols (IS-IS, OSPF)***
3.Find paths that do not violate constraints and optimize some metric
4.Signal to reserve resources along path
5.Set up LSP along path (with explicit route) — ***Extend RSVP or LDP or both!***
6.Map ingress traffic to the appropriate LSPs

Note: (3) could be offline (e.g. network operational center – NOC - systems) or online (perhaps an extension to OSPF)

Problem here:
what is the "correct" resource model for IP services?

56

Resource Reservation + Label Distribution

Two emerging/competing/dueling approaches:

Add label distribution and explicit routes to a resource reservation protocol

RSVP-TE

RSVP

CR-LDP

LDP

Add explicit routes and resource reservation to a label distribution protocol

Constraint-Based LSP Setup using LDP
draft-ietf-mpls-cr-lpd-05.txt

RSVP-TE: Extensions to RSVP for LSP Tunnels
draft-ietf-mpls-rsvp-lsp-tunnel-08.txt

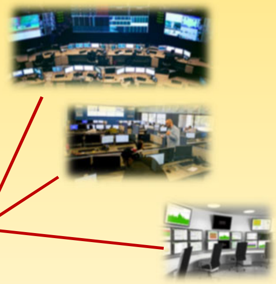
As of February 2003, the IETF MPLS working group deprecated CR-LDP and decided to focus purely on RSVP-TE.

RFC 3468: The Multiprotocol Label Switching (MPLS) Working Group decision on MPLS signalling protocols

57

LSPs establishing protocols

- RSVP-TE (*Resource Reservation Protocol – Traffic Engineering*)
  - Extension of the RSVP protocol
- CR-LDP (*Constrained based Routing – Label Distribution Protocol*)
  - Extension of the LDP protocol, deprecated
- Both protocols enable:
  - The specification of a route to a LSP
  - To chose the labels on each link of the route
  - To make resources reservation for the LSP
- Routes are previously determined:
  - By management (Traffic engineering), in a NOC
  - By a *Constrained based Routing* type protocol



58

RSVP-TE vs. CR-LPD

RSVP-TE

- ***Soft state periodically refreshed***
- ***IntServ QoS model***

CR-LDP

- ***State maintained incrementally***
- ***New QoS model derived from ATM models***

And the QoS model determines the additional information attached to links and nodes and distributed with extended link state protocols…

And what about that other Internet QoS model, diffserv?

59

Resource Reservation Protocol with Traffic Engineering (RSVP-TE)

- Evolution of RSVP
  - RFC 3209: RSVP-TE: Extensions to RSVP for LSP Tunnels. (12/2001)
  - RFC 5151: Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions. (2/2008)
- To map traffic flows onto the physical network topology through label switched paths, resource and constraint network information are required
  - Provided by Extend Link State Protocols (IS-IS or OSPF with TE extensions).
    - RFC 3630: Traffic Engineering (TE) Extensions to OSPF Version 2. (9/2003)
    - RFC 5305: IS-IS Extensions for Traffic Engineering. (10/2008)

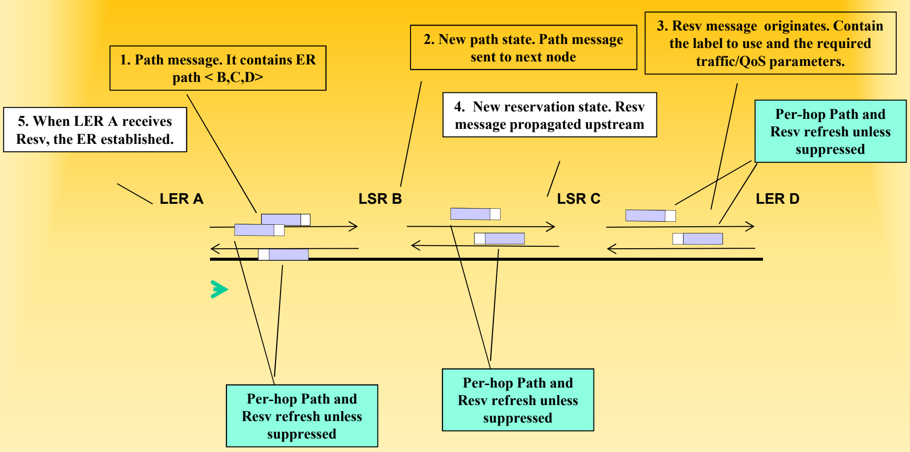
62

RECALL:
ReSerVation Protocol (RSVP)

- ReSerVation Protocol (RSVP) was developed to communicate resource needs between hosts and network devices
  - Associated to the Intserv QoS model
- RSVP allows:
  - The source to describe the characteristics of the IP packets flow.
  - Destinations to describe the reservation they want.
  - Routers to know how to process the packets flow in order to fulfil the requested reservation.
- Encapsulated on IP (protocol type = 46 (0x2E))
- Signalling is based on PATH and RESV messages.
  - PATH announces the traffic characteristics at the sender.
  - RESV achieves reservations that were initiated by the receivers.
  - If the reservation is not possible, a RESV ERR message is sent.
- The routers reservation states have to be periodically refreshed (soft states).

63

Establishment of a LSP with the RSVP-TE protocol



64

REFERENCE:
Extensions to RSVP for LSP Tunnels

- The SENDER_TEMPLATE (or FILTER_SPEC) object together with the SESSION object uniquely identifies an LSP tunnel (flow).
- LSP Tunnel related new objects
  - Explicit Route
    - Carried in PATH and contains a series of variable-length data items called sub-objects.
    - Possible sub-objects: IPv4 prefix, IPv6 prefix, and autonomous system number.
  - Label Request
    - Carried in PATH requesting a label for a specific tunnel/flow.
    - Request cab be without label range, with an ATM label range, or with an Frame Relay label range.
  - Label
    - Carried in RESV messages and contain a single label for a specific tunnel/flow.
  - Record Route
    - Carried in PATH and RESV, used to collect detailed path information and useful for loop detection and diagnostics.
  - Session Attribute
    - Carried in PATH, used to define the type and name of the session/tunnel/flow, also used to define priority values.
- LSP Tunnel related new object types
  - Session object new types
    - LSP_TUNNEL_IPv4 and LSP_TUNNEL_IPv6
  - Sender Template object new types
    - LSP_TUNNEL_IPv4 and LSP_TUNNEL_IPv6
  - Filter Specification object new types
    - LSP_TUNNEL_IPv4 and LSP_TUNNEL_IPv6

65

REFERENCE:
Traffic Engineering Extensions to OSPF

- RFC 3630: Traffic Engineering (TE) Extensions to OSPF Version 2. (9/2003)
- OSPF Traffic Engineering (TE) extensions are used to advertise TE Link State Advertisements (TE-LSAs) containing information about TE-enabled links.
  - Traffic Engineering LSA is a type 10 Opaque LSAs, which have an area flooding scope.
- TE-LSA contains one of two possible top-level Type Length Values (TLVs)
  - **Router Address**: specifies a stable IP address of the advertising router that is always reachable if there is any connectivity to it; this is typically implemented as a "loopback address";
  - **Link**: describes a single link with a a set of sub-TLVs (Link type, Link ID, Local interface IP address, Remote interface IP address, Traffic engineering metric, Maximum bandwidth, Maximum reservable bandwidth, Unreserved bandwidth, and Administrative group.
- The information made available by these extensions can be used to build an extended link state database
  - Can be used to:
    - Monitoring the extended link attributes;
    - Local constraint-based source routing;
    - Global traffic engineering.

66

LSPs priorities

When:

- A new LSP requires resources that are not available on the network, or
- On failure situations (on a link, for example)

The operator can establish different priorities to avoid the "most important" traffic from becoming blocked by the "less important" traffic.

- Each LSP has two priorities assigned: "*setup Priority*" and "*holding Priority*"
- There are 8 different priority levels
- A established LSP can "steal" network resources from the already established LSPs that have a lower "*holding Priority*" than its "*setup Priority*"

67

REVIEW: MPLS - Major Drivers

- Provide IP VPN Services *"the leased line"*
  - Scalable IP VPN service – Build once and sell many
  - Managed Central Services – Building value added services and offering them across VPNs
- Managing traffic on the network using MPLS Traffic Engineering
  - Providing tighter SLA/QoS (Guaranteed BW Services)
  - Protecting bandwidth - Bandwidth Protection Services
- Integrating Layer 2 & Layer 3 Infrastructure
  - Layer 2 services such as ATM (or Frame Relay) over MPLS
  - Mimic layer 2 services over a highly scalable layer 3 infrastructure

68

MPLS QoS

DiffServ over MPLS

- MPLS doesn’t define a new QoS architecture
  - Actually one of the issues between CR-LDP vs RSVP-TE
- Most of the work on MPLS QoS has focused on supporting current IP QoS architectures
- Same traffic conditioning and Per-Hop behaviors as defined by DiffServ

70

Label Header for Packet Media

```
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
```

Label	EXP	S	TTL

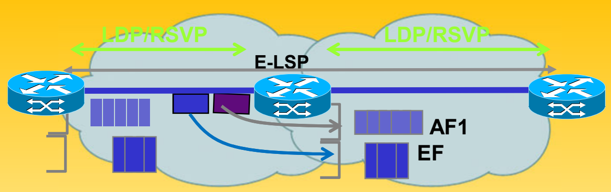
Label	20 bits
EXP	Experimental Field, 3 bits (Class of service information)
S	Bottom of Stack, 1 Bit
TTL	Time to Live, 8 Bits

Shim Header (may or may not exist)

- Can be used over other layer-2 technologies, avoiding its formal existence
- Contains all information needed at forwarding time
- One 32-bit word per label
- EXP field size limitation by standards

71

Diff-Serv Support Over MPLS

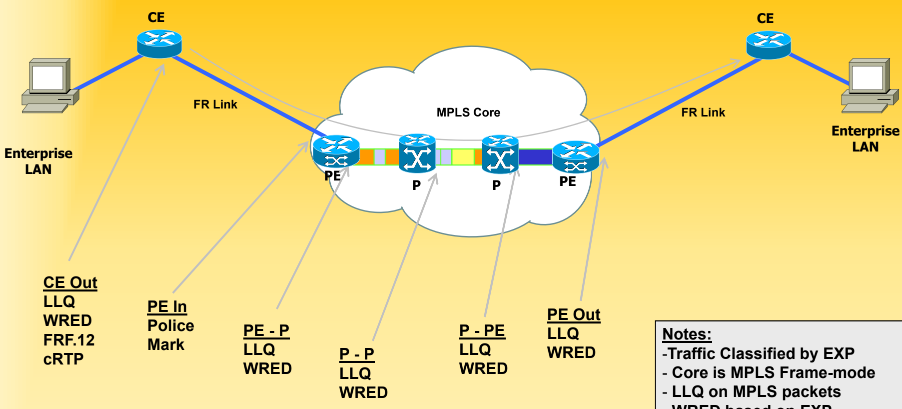


- DiffServ is supported today over MPLS
  - RFC3270
- "plain old" Diff-Serv just over MPLS
- Example above illustrates support of EF and AF1 on single E-LSP
- EF (Expedited Forwarding) and AF1 (Assured Forwarding) packets travel on single LSP (single label)

…. but are queued into different queues (different EXP values)

72

EXAMPLE:
DiffServ MPLS QoS Implementation(s)



LLQ – Low Latency Queing

Notes:
- Traffic Classified by EXP
- Core is MPLS Frame-mode
- LLQ on MPLS packets
- WRED based on EXP
- No need for inbound policy in Core
- LLQ for Min B/W guarantee

Unmanaged Customer Equipment (CE) example shown

73

Relationship between MPLS TE and MPLS Diff-Serv

• Diff-Serv specified independently of Routing/Path Computation
• MPLS Diff-Serv (RFC3270) also specified independently of Routing/Path Computation

• MPLS TE designed as tool to improve backbone efficiency independently of QoS:
– MPLS TE compute routes for aggregates across all Classes
– MPLS TE performs admission control over "global" bandwidth pool for all Classes (i.e., unaware of bandwidth allocated to each queue)
• MPLS TE and MPLS Diff-Serv:
– can run simultaneously
– can provide their own benefit (i.e. TE distributes aggregate load, Diff-Serv provides differentiation)
– are unaware of each other (TE cannot provide its benefit on a per class basis such as CAC and constraint based routing)

74

DiffServ aware Traffic Engineering (DS-TE)

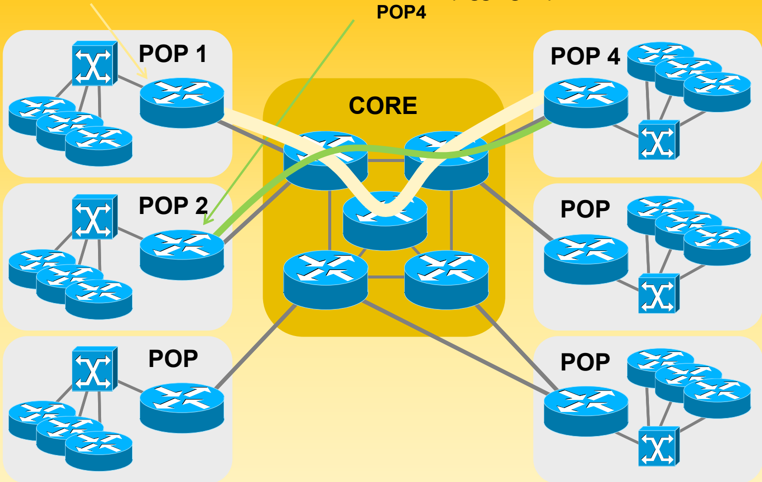
- DS-TE is more than MPLS TE + MPLS DiffServ
- DS-TE makes MPLS TE aware of DiffServ:
  - DS-TE establishes separate tunnels for different classes
  - DS-TE takes into account the "bandwidth" available to each class (e.g. to queue)
  - DS-TE takes into account separate engineering constraints for each class
    - e.g. I want to limit Voice traffic to 70% of link max, but I don't mind having up to 100% of BE traffic.
    - e.g I want overbook ratio of 1 for voice but 3 for BE
- DS-TE ensures specific QoS level of each DiffServ class is achieved

75

MPLS TE with Best Effort Network

Find Route and Set-Up Tunnel for 20 Mb/s (Aggregate) From POP1 to POP4

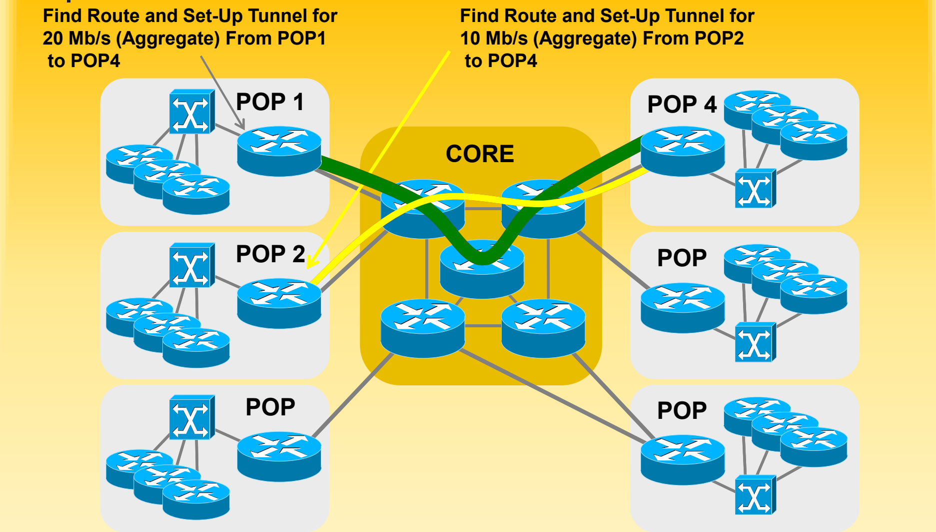
Find Route and Set-Up Tunnel for 10 Mb/s (Aggregate) From POP2 to POP4



No priority in network, selection of path by LSP

76

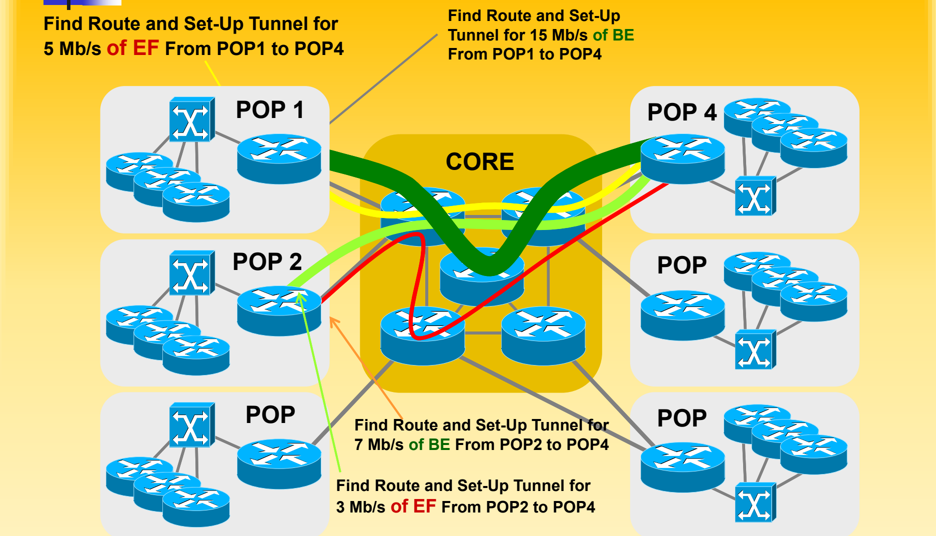
MPLS TE with DiffServ Network



Priority in network, selection of path by LSP, priority in the routers inside each tunnel

77

MPLS DS-TE with DiffServ Network

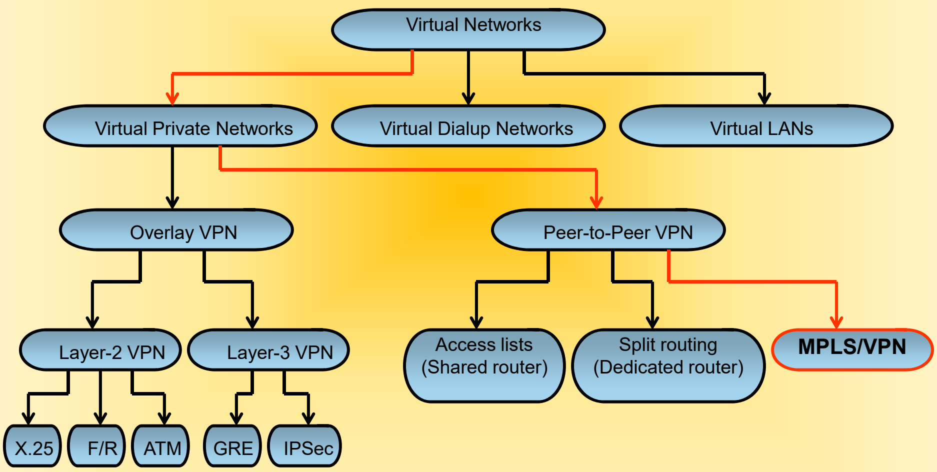


Priority in network, selection of path by LSP **and** priority, priorities in the core may lead to repositioning of the tunnels.

78

MPLS Layer 3 VPNs

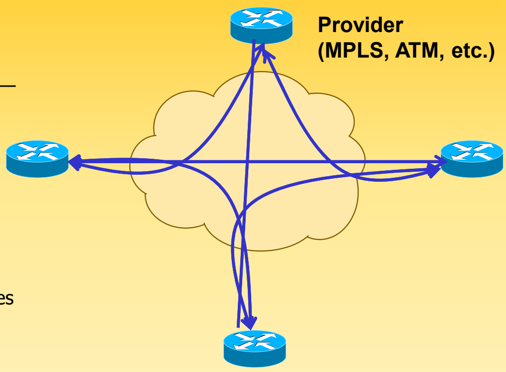
Virtual Network Models

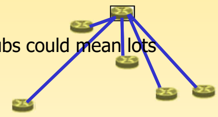


80

Overlay Network

- Provider sells a circuit service
  - customer wants to deploy its own network over those services
- Customers purchases circuits to connect sites, runs IP
  - N sites, (N*(N-1))/2 circuits for full mesh—expensive
- scalability issue because of routing peers in mesh approach
  - N sites, each site has N-1 peers
- Hub and spoke with static routes is simpler,
  - still buying N-1 circuits from hub to spokes
  - suffers from the same N-1 number of routing peers
  - Spokes distant from hubs could mean lots of long-haul circuits

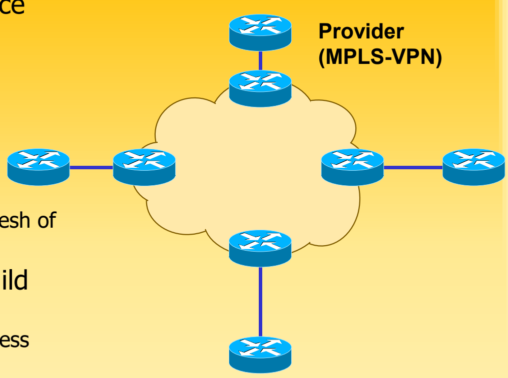




81

Peer Network

- Provider sells an MPLS-VPN service
- Customers purchases circuits to connect sites, runs IP
  - N sites, N circuits into provider
  - Access circuits can be any media at any point (FE, POS, ATM, T1, dial, etc.)
  - Full mesh connectivity without full mesh of L2 circuits
- Hub and spoke is also easy to build
  - Spokes distant from hubs connect to their local provider's POP, lower access charge because of provider's size



The Internet is a large peer network!

82

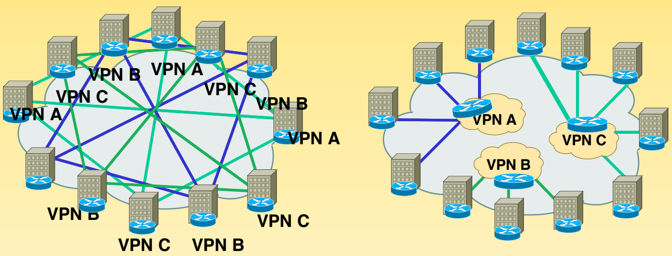
IP/MPLS Applications

MPLS-based VPNs

- MPLS L3VPN – VPN-IP over MPLS
  - VPN is a secure IP-based network between geographically dispersed sites that can communicate securely over a shared backbone;
  - MPLS VPNs provides the capability to deploy and administer scalable Layer 3 VPN backbone services to business customers
- MPLS L2VPN – Any Technology (AT) or Transport over MPLS (e.g.: EoMPLS)
  - AT over MPLS transport Layer2 packets over MPLS network;
  - Allow the use of MPLS network to provide connectivity between customer sites with existing Layer2 networks;
- MPLS-TE
  - Extends existing IP protocols and makes use of MPLS forwarding capabilities to provide TE
  - Brings explicit routing capabilities to MPLS networks

MPLS L3 VPNs using BGP (RFC2547)

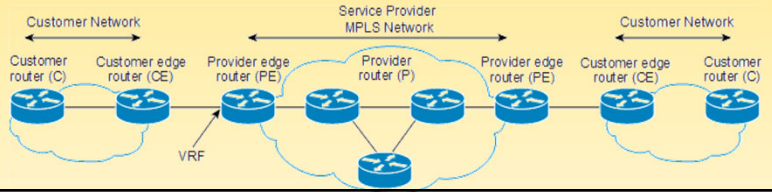
- End user perspective
  - Virtual Private IP service
  - **Simple routing – just point default to provider**
  - Full site-site connectivity without the usual drawbacks (routing complexity, scaling, configuration, cost)
- Major benefit for provider – scalability



84

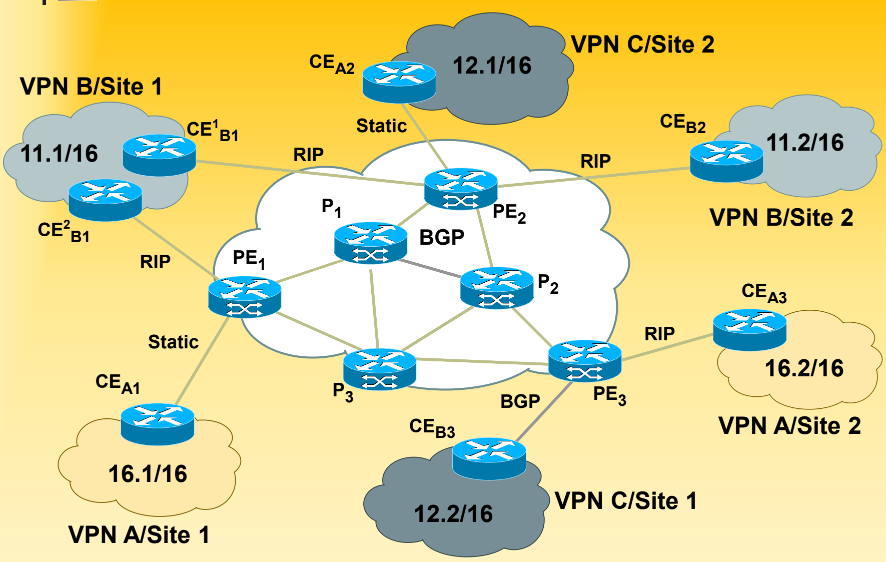
MPLS VPN Terminology

- Customer router (C) is connected only to other customer devices.
- Customer Edge (CE) router peers at Layer 3 to the Provider Edge (PE).
  - The PE-CE Interface runs either a dynamic routing protocol (eBGP, RIPv2, EIGRP, or OSPF) or has static routing (Static, Connected).
- Provider (P) router, resides in the core of the provider network.
  - Participates in the control plane for customer prefixes. The P router is also referred to as a Label Switch Router (LSR), in reference to its primary role in the core of the network, performing label switching/swapping of MPLS traffic.
- Provider Edge (PE) router, sits at the edge of the MPLS SP network.
  - In an MPLS VPN context, separate VRF routing tables are allocated for each user group.
  - Contains a global routing table for routes in the core SP infrastructure.
  - The PE is sometimes referred to as a Label Edge Router (LER) or Edge Label Switch Router (ELSR) in reference to its role at the edge of the MPLS cloud, performing label imposition and disposition.



85

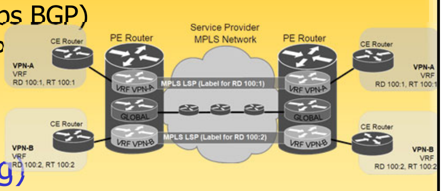
EXAMPLE: MPLS VPN Topology



86

VPN Routing and Forwarding Instance (VRF)

- PE routers maintain separate routing tables.
  - Virtual Routing and Forwarding (VRF) instance is separate from the global routing table that exists on PE routers
- – Global routing table
  - Contains all PE and P routes (perhaps BGP)
  - Populated by the VPN backbone IGP



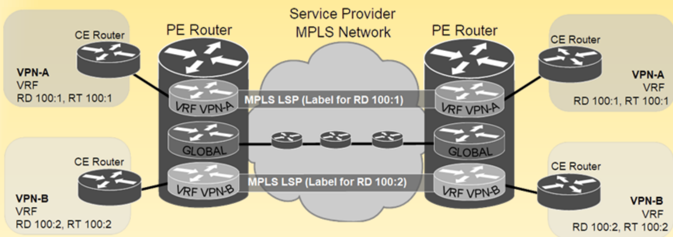
- – VRF (VPN routing and forwarding)
  - Routing and forwarding table associated with one or more directly connected sites (CE routers)
  - VRF is associated with any type of interface, whether logical or physical (e.g. sub/virtual/tunnel)
  - Interfaces may share the same VRF if the connected sites share the same routing information
  - Routes are injected into the VRF from the CE-PE routing protocols for that VRF and any MP-BGP announcements that match the defined VRF.

87

Route Distinguisher

```
!
ip vrf red
rd 1:1
route-target export 1:1
route-target import 1:1
```

- To differentiate 10.0.0.0/8 in VPN-A from 10.0.0.0/8 in VPN-B
  - 64-bit quantity
- Configured as ASN:YY or IPADDR:YY
  - ASN is the most common
- Purely to make a route unique
  - Unique route is now RD:Ipaddr (96 bits) plus a mask on the IPAddr portion
  - So customers don't see each others routes




```
!
ip vrf VPN-A
rd 100:1
route-target export 100:1
route-target import 100:1
```

88

Route Target

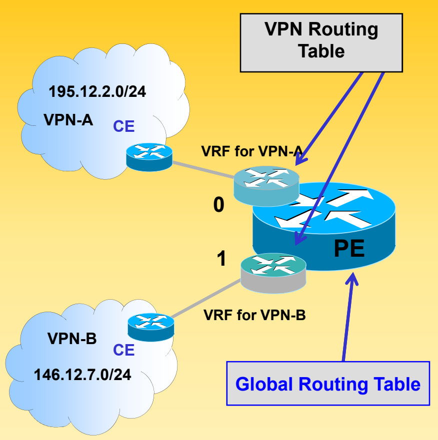
```
!
ip vrf red
rd 1:1
route-target export 1:1
route-target import 1:1
```

- Creates or adds to a list of VPN extended communities used to determine which routes are imported by a VRF
- To control policy about who sees what routes
- 64-bit quantity (2 bytes type, 6 bytes value)
- Carried as an extended community
  - Typically written as ASN:YY
- Each VRF 'imports' and 'exports' one or more RTs
  - Exported RTs are carried in VPNv4 BGP
  - Imported RTs are local to the box
- A PE that imports an RT installs that route in its routing table
- Example: Each VRF in VPN A has the same route target in their import list and export list. Each VPN A VRF accepts only received routes that have this route target attached. Because this route target is attached to each route advertised by VPN A VRFs, every site in VPN A accepts routes only from other sites in VPN A.

89

Virtual Routing and Forwarding Interface definition

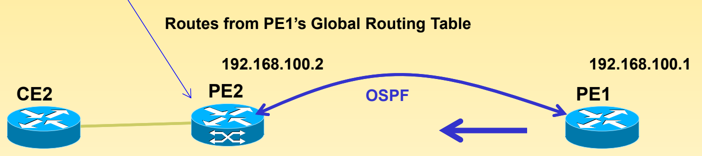
- Define a unique VRF for interface 0
- Define a unique VRF for interface 1
  - Packets will never go between interfaces 0 and 1
- Uses VPNv4 to exchange VRF routing information between PE's



90

PE Router – Global Routing Table Output

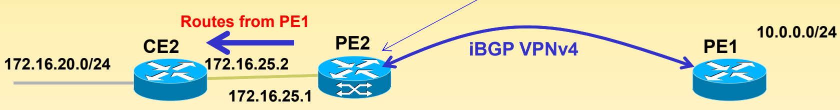
```
PE2#sh ip route

Gateway of last resort is not set

C    192.168.1.0/24 is directly connected, Ethernet0/0
     192.168.100.0/32 is subnetted, 3 subnets
O       192.168.100.1 [110/11] via 192.168.1.1, 00:04:27, Ethernet0/0
C       192.168.100.2 is directly connected, Loopback0
O       192.168.100.3 [110/11] via 192.168.1.3, 00:04:27, Ethernet0/0
```



91

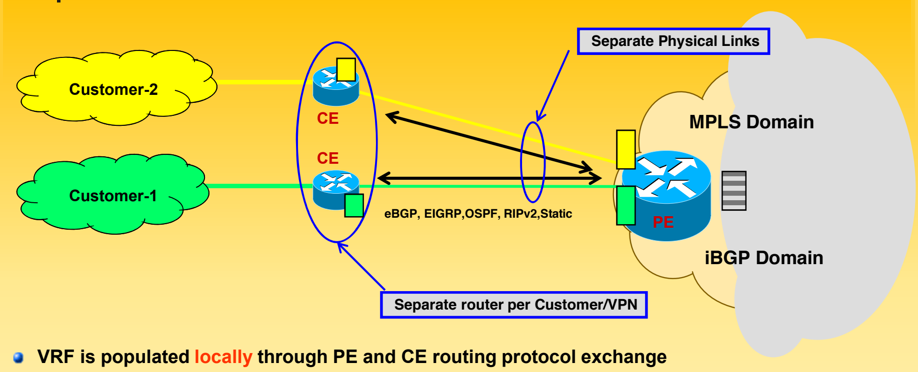
PE Router – VRF Routing Table Output

```
PE2#sh ip route vrf RED
Routing Table: RED

Gateway of last resort is 192.168.100.1 to network 0.0.0.0

     172.16.0.0/16 is variably subnetted, 8 subnets, 3 masks
C       172.16.25.0/30 is directly connected, Serial4/0
C       172.16.25.2/32 is directly connected, Serial4/0
B       172.16.20.0/24 [20/0] via 172.16.25.2, 00:07:04
     10.0.0.0/24 is subnetted, 1 subnets
B       10.0.0.0 [200/307200] via 192.168.100.1, 00:06:28
B*    0.0.0.0/0 [200/0] via 192.168.100.1, 00:07:03
```



92

VRF Route Population



- **VRF is populated locally through PE and CE routing protocol exchange**
  - RIP Version 2, OSPF, BGP-4, EIGRP, & Static routing
  - “connected” is also supported (i.e. Default-gateway is PE)
- **Separate routing context for each VRF**
  - routing protocol context (BGP-4 & RIP V2)
  - separate process (OSPF)

93

Carrying VPN Routes in BGP

- VRFs by themselves aren’t all that useful
  - Need some way to get the VRF routing information off the PE and to other PEs
  - This is done with BGP
- Additions to MP-BGP to Carry MPLS-VPN Info
  - RD: Route Distinguisher
  - RT: Route Target
  - VPNv4 address family
  - MPLS Label

```
Border Gateway Protocol - UPDATE Message
 Marker: ffffffffffffffffffffffffffffffff
 Length: 91
 Type: UPDATE Message (2)
 Withdrawn Routes Length: 0
 Total Path Attribute Length: 68
▿Path attributes
 ▹Path Attribut - ORIGIN: INCOMPLETE
 ▹Path Attribut - AS_PATH: empty
 ▹Path Attribut - MULTI_EXIT_DISC: 0
 ▹Path Attribut - LOCAL_PREF: 100
 ▿Path Attribut - EXTENDED_COMMUNITIES
  ▹Flags: 0xc0: Optional, Transitive, Complete
   Type Code: EXTENDED_COMMUNITIES (16)
   Length: 8
  ▿Carried extended communities: (1 community)
   ▹Community Transitive Two-Octet AS Route Target: 200:1
 ▿Path Attribut - MP_REACH_NLRI
  ▹Flags: 0x80: Optional, Non-transitive, Complete
   Type Code: MP_REACH_NLRI (14)
   Length: 33
   Address family: IPv4 (1)
   Subsequent address family identifier: Labeled VPN Unicast (128)
  ▹Next hop network address (12 bytes)
   Subnetwork points of attachment: 0
  ▿Network layer reachability information (16 bytes)
   ▹Label Stack=24 (bottom) RD=200:1, IPv4=192.1.1.0/25
```

94

MPLS/VPN Packet Forwarding

- Between PE and CE, regular IP packets (currently)
- Within the provider network—label stack
  - Outer label: "get this packet to the egress PE"
  - Inner label: "get this packet to the egress CE"
- Remember: MPLS nodes forward packets based on TOP label
  - any subsequent labels are ignored
  - Penultimate Hop Popping procedures used one hop prior to egress PE router

95

VPNv4

- In BGP for IP, 32-bit address + mask makes a unique announcement
- In BGP for MPLS-VPN, (64-bit RD + 32-bit address) + 32-bit mask makes a unique announcement
- Since the route encoding is different, need a different address family in BGP

  VPNv4 = VPN routes for IPv4
  - As opposed to IPv4 or IPv6 or multicast-RPF, etc…
- VPNv4 announcement carries a label with the route
  - "If you want to reach this unique address, get me packets with this label on them"

96

Terminology

Terminology, 1/2

- RR—Route Reflector
  - A router (usually not involved in packet forwarding) that distributes BGP routes within a provider's network
- PE—Provider Edge router
  - The interface between the customer and the MPLS-VPN network; only PEs (and maybe RRs) know anything about MPLS-VPN routes
- P—Provider router
  - A router in the core of the MPLS-VPN network, speaks LDP/RSVP but not necessarily VPNv4
- CE—Customer Edge router
  - The customer router which connects to the PE; does not know anything about labels, only IP (most of the time)
- LDP—Label Distribution Protocol
  - Distributes labels with a provider's network that mirror the IGP, one way to get from one PE to another
- LSP—Label Switched Path
  - The chain of labels that are swapped at each hop to get from one PE to another

98

Terminology, 2/2

- VPN—Virtual Private Network
  - A network deployed on top of another network, where the two networks are separate and never communicate
- VRF—Virtual Routing and Forwarding instance
  - Mechanism in IOS used to build per-interface RIB and FIB
- VPNv4
  - Address family used in BGP to carry MPLS-VPN routes
- RD
  - Route Distinguisher, used to uniquely identify the same network/mask from different VRFs (i.e., 10.0.0.0/8 from VPN A and 10.0.0.0/8 from VPN B)
- RT
  - Route Target, used to control import and export policies, to build arbitrary VPN topologies for customers

99