Overview

Introduction to MPLS & Applications

- Introduction to MPLS
- MPLS Label Distribution Protocol
- Basic MPLS Configuration (Lab Exercise)
- MPLS Layer 3 VPN Concept & Architecture
- MPLS L3 VPN Configuration (Lab Exercise)
- MPLS Layer 2 VPN Concept & Architecture
- MPLS L2 VPN Configuration (Lab Exercise)
- MPLS Traffic Engineering

APNIC

Knowledge Required

- Intermediate level routing knowledge
- Have good knowledge of configuring and troubleshooting IGPs such as OSPF/ISIS
- Good knowledge of configuring and troubleshooting of BGP
- Terminal software such as Putty, SecureCRT
- For Mac OS X iTerm etc.

APNIC

Questions?

APNIC

Overview

Introduction to MPLS & Applications

- **Introduction to MPLS**
- MPLS Label Distribution Protocol
- Basic MPLS Configuration (Lab Exercise)
- MPLS Layer 3 VPN Concept & Architecture
- MPLS L3 VPN Configuration (Lab Exercise)
- MPLS Layer 2 VPN Concept & Architecture
- MPLS L2 VPN Configuration (Lab Exercise)
- MPLS Traffic Engineering

APNIC

What is MPLS ?

APNIC

Data Communication Network

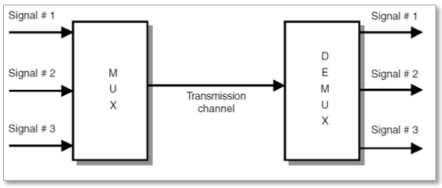
Definitions

- A set of nodes connected by links
- Nodes are basically routers, LAN switches, WAN Switches, Add-drop Multiplexers (ADM)
- Links are communication channel between nodes i.e. it can be 64kbps DS0 circuits to OC 768/40G Ethernet

APNIC

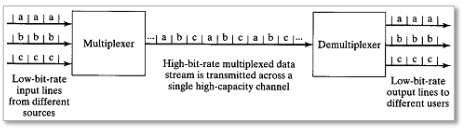
Multiplexing

- Allows multiple connections across a network to share the same transmission infrastructure
- Two main type of multiplexing:
  - Time Division Multiplexing (TDM)
  - Statistical Multiplexing (StatMUX)
- Other types i.e FDM, WDM etc



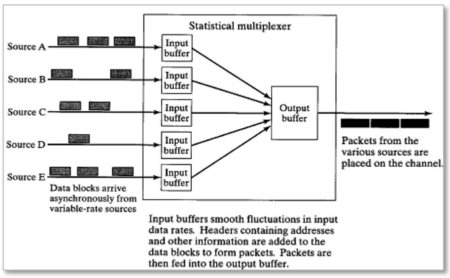
Time Division Multiplexing (TDM)

- Allocate a certain amount of time on a given physical circuit to a number of connection.
- Synchronous technology, data entering to the network control by a master clock source.
- No congestion and ensure dedicated bandwidth.
- Expensive.
- Example: OC-3c / STM-1(155.52Mb) OC-12 / STM-4 (622.08 Mbit/s) etc.



Statistical Multiplexing (StatMUX)

- Divide network traffic into discrete units and deal with each of them units separately.
  - IP call it packet
  - Frame relay call it frame
  - ATM call it cell
- Share transmission bandwidth among all network users
  - Oversubscribe the network
  - Make more money
  - Cheaper access cost them TDM
  - No dedicated Bandwidth
- Considering all user will not transmit at he same time.
  - If they do there is a possibility of congestion.
- Example:
  - IP
  - Frame Relay
  - ATM
  - MPLS (Another type of StatMUX)



Over Subscription Issue in StatMUX

- Resource contention that TDM doesn't have
  - Packets enter the network asynchronously
  - If two packet enter the router at the exact same time from two different interface then one of the packet had to wait for the other packet to be transmitted.
  - On a non-oversubscribed link this is not a big issue.
- Oversubscription introduce following three issues in StatMUX:
  - Buffering
  - Queuing
  - Dropping

Over Subscription Issue in StatMUX

- How frame relay deal with this?
  - CIR
  - FECN & BECN
  - DE Bit
- How IP deal with this?
  - Random Early Discard (RED) which rely on TCP to handle dropping
  - IP ECN bit which are relatively new and limited use so far
- How ATM deal with this?
  - CBR (Constant Bit Rate)
  - rt-VBR (real-time variable bit rate)
  - nrt-VBR (non real-time variable bit rate)
  - ABR (Available Bit Rate)
  - UBR (Unspecified Bit Rate)

APNIC

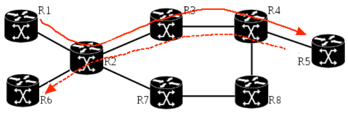
StaMUX Over StatMux

- StatMUX technology generation
  - IP standardize around 1981
  - Frame Relay commercially available around early 1990
  - ATM came around mid of 1990s
- Network operators start introducing StatMUX protocol ie. IP, IPX, DECNet, AppleTalk etc as layer3 protocol
- Replace TDM from layer 2 i.e. Frame Relay, ATM etc
  - It become StatMUX over StatMUX
  - Sub optimal end to end performance
  - Resource contention parameter doesn't translate well between L3 and L2 protocol.
- In times since IP become the single dominant protocol requirement raised to exactly map L3 resource contention parameter in L2.
  - This is how MPLS came into the picture, then started evolving with other features

APNIC

Limitation of Traditional IP Routing

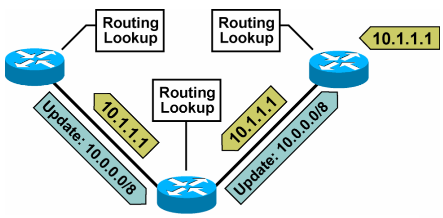
- Routing protocols are used to distribute Layer 3 routing information
- Forwarding is based on the destination address only
- Routing lookups are performed on every hop.



APNIC

Limitation of Traditional IP Routing

- Every router may need full Internet routing information
  - Global Internet routing table size 500,000+ routes
- Destination-based routing lookup is needed on every hop.

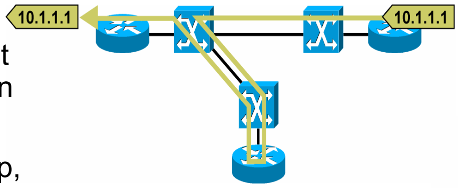


APNIC

Limitation of Traditional IP Routing

IP over ATM

- Layer 2 devices have no knowledge of Layer 3 routing information—virtual circuits must be manually established.
- Layer 2 topology may be different from Layer 3 topology, resulting in suboptimal paths and link use.
- Even if the two topologies overlap, the hub-and-spoke topology is usually used because of easier management.

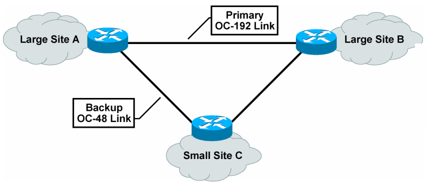


APNIC

Limitation of Traditional IP Routing

Traffic Engineering

- Most traffic goes between large sites A and B, and uses only the primary link.
- Destination-based routing does not provide any mechanism for load balancing across unequal paths.
- Policy-based routing can be used to forward packets based on other parameters, but this is not a scalable solution.



APNIC

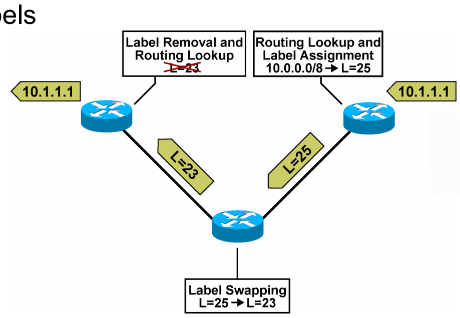
How MPLS Enhanced Traditional IP Routing Limitation?

- MPLS is a new forwarding mechanism in which packets are forwarded based on labels.
- Labels usually correspond to IP destination networks (equal to traditional IP forwarding).
- Labels can also correspond to other parameters, such as QoS or source address.
- MPLS was designed to support forwarding of other protocols as well.

APNIC

Basic MPLS Concepts

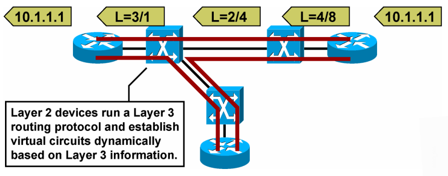
- Only the edge routers will perform a routing lookup
- Core routers switch packets based on simple label lookups and swap labels



APNIC

Differences Between MPLS and IP over ATM

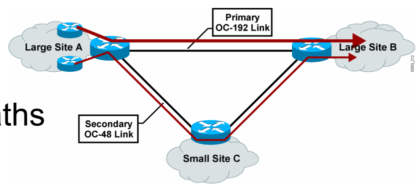
- Layer 2 devices are IP-aware and run a routing protocol.
- There is no need to manually establish virtual circuits.
- MPLS provides a virtual full mesh topology.



APNIC

Traffic Engineering with MPLS

- Traffic can be forwarded based on other parameters
  – (QoS, source, and so on).
- Load sharing across unequal paths can be achieved.
- TE determines the path at the source based on additional parameters, such as available resources and constraints in the network



APNIC

So What is MPLS?

Brief Summary

- It's all about labels …
- Use the best of both worlds
  - Layer-2 (ATM/FR): efficient forwarding and traffic engineering
  - Layer-3 (IP): flexible and scalable
- MPLS forwarding plane
  - Use of labels for forwarding Layer-2/3 data traffic
  - Labeled packets are being switched instead of routed
    - Leverage layer-2 forwarding efficiency
- MPLS control/signaling plane
  - Use of existing IP control protocols extensions + new protocols to exchange label information
    - Leverage layer-3 control protocol flexibility and scalability

APNIC

Technology Comparison

Key Characteristics of IP, Native Ethernet, and MPLS

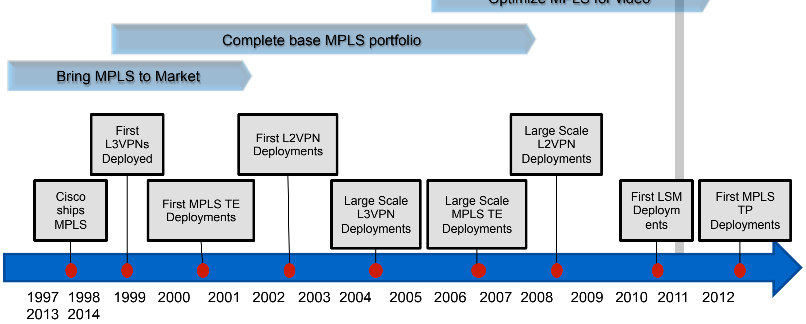
	IP	Native Ethernet	MPLS
Forwarding	Destination address based Forwarding table learned from control plane TTL support	Destination address based Forwarding table learned from data plane No TTL support	Label based Forwarding table learned from control plane TTL support
Control Plane	Routing Protocols	Ethernet Loop avoidance and signaling protocols	Routing Protocols MPLS protocols
Packet Encapsulation	IP Header	802.3 Header	MPLS shim header
QoS	8 bit TOS field in IP header	3-bit 802.1p field in VLAN tag	3 bit TC field in label
OAM	IP ping, traceroute	E-OAM	MPLS OAM

APNIC

Evolution of MPLS

Technology Evolution and Main Growth Areas

- Evolved from tag switching in 1996 to full IETF standard, covering over 130 RFCs
- Key application initially were Layer-3 VPNs, followed by Traffic Engineering (TE), and Layer-2 VPNs



Market Segments

MPLS Business Drivers and Typical Deployment Characteristics

	Business Drivers	Business Goals	MPLS Capabilities
Service Provider	• Networking service reliability • Cost effective service bandwidth • Flexible enablement of existing and new services	• Leverage single network for scalable delivery of multiple services • Optimize network capacity to meet current and future growth of service bandwidth • Deliver premium services with guaranteed SLAs	Layer-3 VPN Layer-2 VPN MPLS TE MPLS OAM, QoS
Enterprise	• Mergers and acquisitions • Network consolidation • Shared services • Compliance	• Network Segmentation • Network integration	Layer-3 VPN
Data Center	• Multi-tenant hosting • Data Center Interconnect	• Leverage single data center infrastructure for multiple users and services • Deliver geographic independent services from any data center	Layer-2 VPN Layer-3 VPN

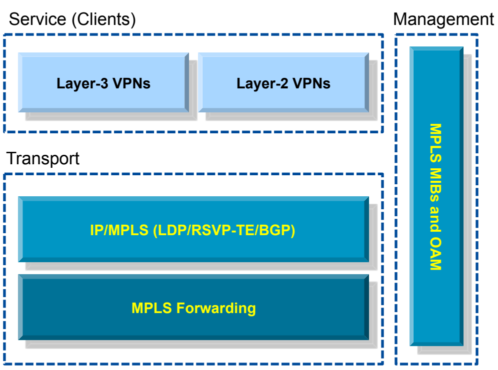
MPLS Technology Basics

Technology Building Blocks of MPLS

APNIC

Technology Building Blocks of MPLS

- MPLS reference architecture
- MPLS Labels
- MPLS signaling and forwarding operations
- MPLS Traffic Engineering
- MPLS OAM and MIBs

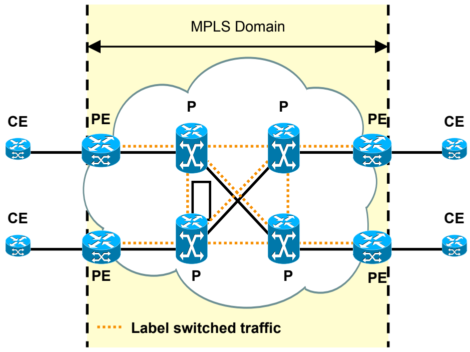


APNIC

MPLS Reference Architecture

Different Type of Nodes in a MPLS Network

- P (Provider) router
  - Label switching router (LSR)
  - Switches MPLS-labeled packets
- PE (Provider Edge) router
  - Edge router (LER)
  - Imposes and removes MPLS labels
- CE (Customer Edge) router
  - Connects customer network to MPLS network

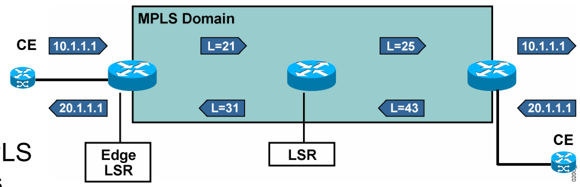


APNIC

LER & LSR Router Functions

Different Type of Nodes in a MPLS Network

- LSR (P) primarily forwards labeled packets (label swapping).
- Edge LSR (PE) primarily labels IP packets and forwards them into the MPLS domain, or removes labels and forwards IP packets out of the MPLS domain.
- CE routers are not aware about MPLS. They work on traditional IP routing protocol



APNIC

MPLS Labels

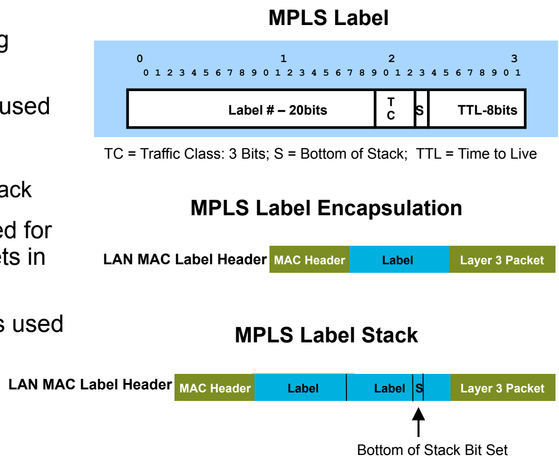
- MPLS technology is intended to be used anywhere regardless of Layer 1 media and Layer 2 protocol.
- MPLS uses a 32-bit label field that is inserted between Layer 2 and Layer 3 headers (frame-mode MPLS).
- MPLS over ATM uses the ATM header as the label (cell-mode MPLS)

APNIC

MPLS Shim Labels

Label Definition and Encapsulation

- Labels used for making forwarding decision
- Multiple labels can be used for MPLS packet encapsulation
  – Creation of a label stack
- Outer label always used for switching MPLS packets in network
- Remaining inner labels used to specific services (e.g., VPNs)



APNIC

MPLS TTL Operation

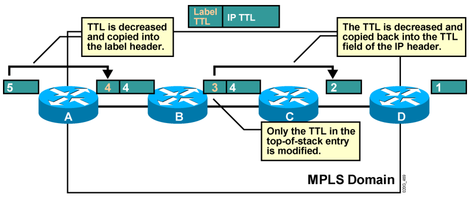
TTL Marking in MPLS Labels

On ingress LER: TTL is copied from IP header to label header.

On egress LER: TTL is copied from label header to IP header.

Cisco routers have TTL propagation enabled by default.

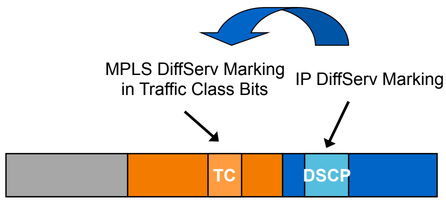
Disabling TTL propagation causes routers to set the value 255 into the TTL field of the label when an IP packet is labeled



MPLS QoS

QoS Marking in MPLS Labels

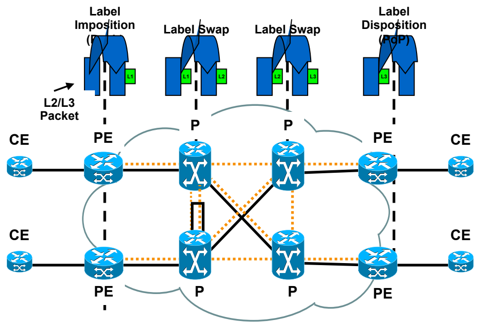
- MPLS label contains 3 TC bits
- Used for packet classification and prioritization
  - Similar to Type of Service (ToS) field in IP packet (DSCP values)
- DSCP values of IP packet mapped into TC bits of MPLS label
  - At ingress PE router
- Most providers have defined 3–5 service classes (TC values)
- Different DSCP <-> TC mapping schemes possible



Basic MPLS Forwarding Operations

How Labels Are Being Used to Establish End-to-end Connectivity

- Label imposition (PUSH)
  - By ingress PE router; classify and label packets
  - Based on Forwarding Equivalence Class (FEC)
- Label swapping or switching
  - By P router; forward packets using labels; indicates service class & destination
- Label disposition (POP)
  - By egress PE router; remove label and forward original packet to destination CE



APNIC

MPLS Path (LSP) Setup and Traffic Forwarding

MPLS Traffic Forwarding and MPLS Path (LSP) Setup

- LSP signaling
  - Either LDP* or RSVP
  - Leverages IP routing
  - Routing table (RIB)
- Exchange of labels
  - Label bindings
  - Downstream MPLS node advertises what label to use to send traffic to node
- MPLS forwarding
  - MPLS Forwarding table (FIB)

	IP	MPLS
Forwarding	Destination address based Forwarding table learned from control plane TTL support	Label based Forwarding table learned from control plane TTL support
Control Plane	OSPF, IS-IS, BGP	OSPF, IS-IS, BGP LDP, RSVP
Packet Encapsulation	IP Header	One or more labels
QoS	8 bit TOS field in IP header	3 bit TC field in label
OAM	IP ping, traceroute	MPLS OAM

* LDP signaling assumed for next the examples

APNIC

MPLS Path (LSP) Setup

Signaling Options

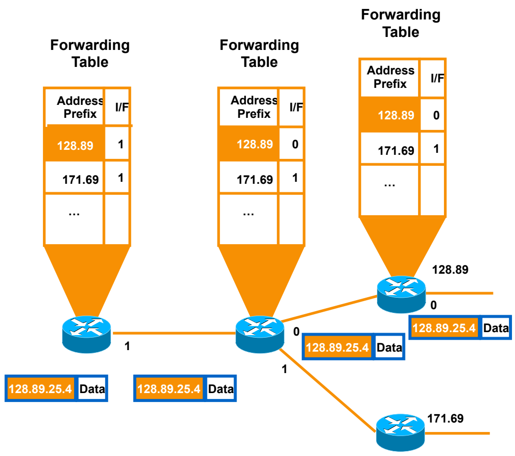
- LDP signaling
  - Leverages existing routing
- RSVP signaling
  - Aka MPLS RSVP/TE
  - Enables enhanced capabilities, such as Fast ReRoute (FRR)

	LDP	RSVP
Forwarding path	LSP	LSP or TE Tunnel Primary and, optionally, backup
Forwarding Calculation	Based on IP routing database Shortest-Path based	Based on TE topology database Shortest-path and/or other constraints (CSPF calculation)
Packet Encapsulation	Single label	One or two labels
Signaling	By each node independently Uses existing routing protocols/information	Initiated by head-end node towards tail-end node Uses routing protocol extensions/information Supports bandwidth reservation Supports link/node protection

IP Packet Forwarding Example

Basic IP Packet Forwarding

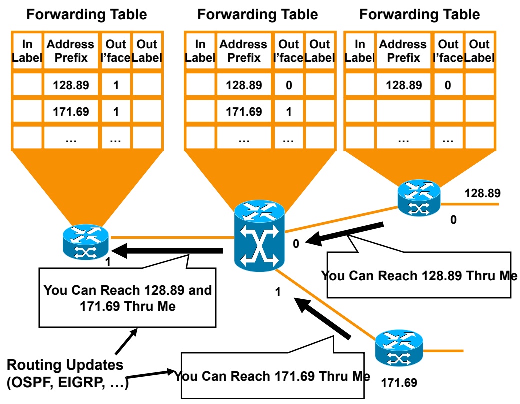
- IP routing information exchanged between nodes
  - Via IGP (e.g., OSFP, IS-IS)
- Packets being forwarded based on destination IP address
  - Lookup in routing table (RIB)



MPLS Path (LSP) Setup

Step 1: IP Routing (IGP) Convergence

- Exchange of IP routes
  - OSPF, IS-IS, EIGRP, etc.
- Establish IP reachability

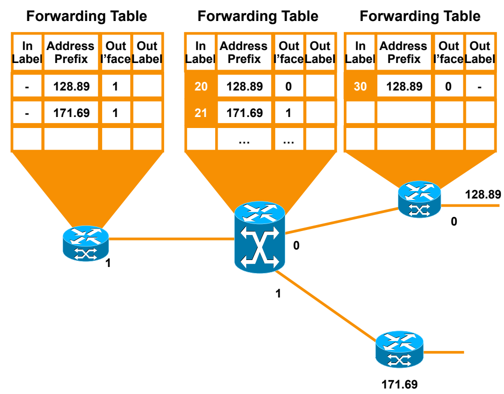


APNIC 37

MPLS Path (LSP) Setup

Step 2A: Assignment of Local Labels

- Each MPLS node assigns a local label to each route in local routing table
  - In label

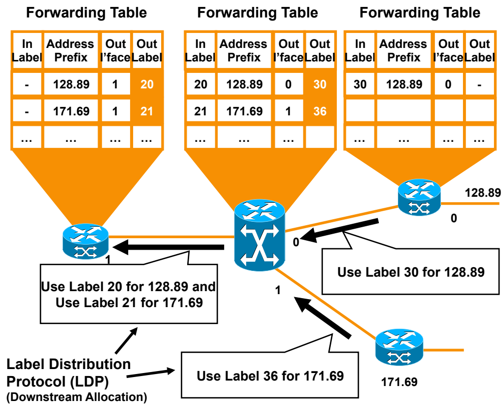


APNIC 38

MPLS Path (LSP) Setup

Step 2B: Assignment of Remote Labels

- Local label mapping are sent to connected nodes
- Receiving nodes update forwarding table
  - Out label

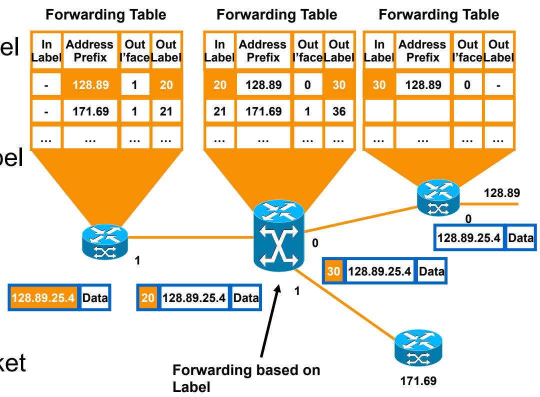


APNIC

MPLS Traffic Forwarding

Hop-by-hop Traffic Forwarding Using Labels

- Ingress PE node adds label to packet (push)
  - Via forwarding table
- Downstream node use label for forwarding decision (swap)
  - Outgoing interface
  - Out label
- Egress PE removes label and forwards original packet (pop)



APNIC

Summary

- MPLS Layer-3 VPNs provide IP connectivity among CE sites
  - MPLS VPNs enable full-mesh, hub-and-spoke, and hybrid IP connectivity
- CE sites connect to the MPLS network via IP peering across PE-CE links
- MPLS Layer-3 VPNs are implemented via VRFs on PE edge nodes
  - VRFs providing customer routing and forwarding segmentation
- BGP used for signaling customer VPN (VPNv4) routes between PE nodes
- To ensure traffic separation, customer traffic is encapsulated in an additional VPN label when forwarded in MPLS network
- Key applications are layer-3 business VPN services, enterprise network segmentation, and segmented layer-3 Data Center access

APNIC 41

Questions?

APNIC

Overview

Introduction to MPLS & Applications

- Introduction to MPLS
- **MPLS Label Distribution Protocol**
- Basic MPLS Configuration (Lab Exercise)
- MPLS Layer 3 VPN Concept & Architecture
- MPLS L3 VPN Configuration (Lab Exercise)
- MPLS Layer 2 VPN Concept & Architecture
- MPLS L2 VPN Configuration (Lab Exercise)
- MPLS Traffic Engineering

APNIC

Things to Discuss

- LDP Overview
- LDP Protocol Details
- LDP Configuration and Monitoring

APNIC

LDP Overview

Label Binding Protocol of MPLS

APNIC

Label Distribution Protocol

Overview

- MPLS nodes need to exchange label information with each other
  - Ingress PE node (Push operation)
    - Needs to know what label to use for a given FEC to send packet to neighbor
  - Core P node (Swap operation)
    - Needs to know what label to use for swap operation for incoming labeled packets
  - Egress PE node (Pop operation)
    - Needs to tell upstream neighbor what label to use for specific FEC type LDP used for exchange of label (mapping) information
- Label Distribution Protocol (LDP)
  - Defined in RFC 3035 and RFC3036; updated by RFC5036
  - LDP is a superset of the Cisco-specific Tag Distribution Protocol
- Note that, in addition LDP, also other protocols are being used for label information exchange
  - Will be discussed later

APNIC

Label Distribution Protocol

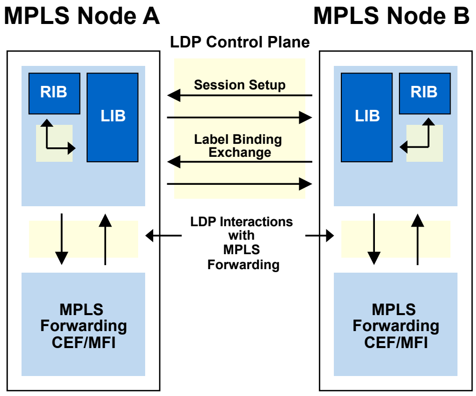
Some More Details

- Assigns, distributes, and installs (in forwarding) labels for prefixes advertised by unicast routing protocols
  - OSPF, IS-IS, EIGRP, etc.
- Also used for Pseudowire/PW (VC) signaling
  - Used for L2VPN control plane signaling
- Uses UDP (port 646) for session discovery and TCP (port 646) for exchange of LDP messages
- LDP operations
  - LDP Peer Discovery
  - LDP Session Establishment
  - MPLS Label Allocation, Distribution, and Updating MPLS forwarding
- Information repositories used by LDP
  - LIB: Label Information Database (read/write)
  - RIB: Routing Information Database/routing table (read-only)

Label Distribution Protocol

Operations Details

- LDP startup
  - Local labels assigned to RIB prefixes and stored in LIB
  - Peer discovery and session setup
  - Exchange of MPLS label bindings
- Programming of MPLS forwarding
  - Based on LIB info
  - CEF/MFI updates



LDP Protocol Details

Label Binding Protocol of MPLS

APNIC

Agenda LDP Protocol Details

- LDP Concepts
- LDP Identifier
- LDP PDU
- LDP Messages
- LDP Session Establishment
- LDP Sessions between ATM LSRs
- Targeted LDP sessions
- Summary

APNIC

LDP Concepts

- Label Distribution Protocol
  - LDP works between adjacent/non-adjacent peers
  - LDP sessions are established between peers
  - LDP messages sent in the form of TLVs
    - <Type, Length, Value>
- Standardized via RFC 3036 (Updated by RFC 5036)

APNIC

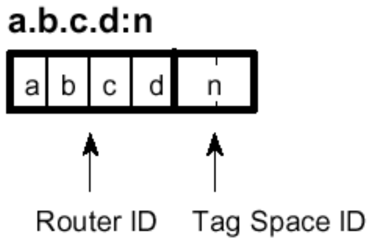
TDP/LDP Transport

- Uses TCP for reliable transport
- Well-known TCP port
  - *LDP (port 646)*
  - ~~*TDP (port 711)*~~
- LSR with higher LDP router-id opens a connection to port 646 of other LSR
- Design Choice:
  - *One TDP/LDP session per TCP connection*

APNIC

LDP Identifier

- Identifies tag space
- 6 bytes (4 bytes =>IP address, 2 bytes =>Label space ID)



APNIC

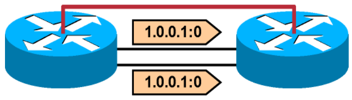
LDP Identifier: Label Space

- LSRs establish one LDP session per label space.
  - Per-platform label space requires only one LDP session, even if there are multiple parallel links between a pair of LSRs.
- Per-platform label space is announced by setting the label space ID to 0, for example:
  - LDP ID = 1.0.0.1:0
- A combination of frame-mode and cell-mode MPLS, or multiple cell-mode links, results in multiple LDP sessions.

APNIC

Label Space and number of LDP sessions

Frame Mode



Frame Mode

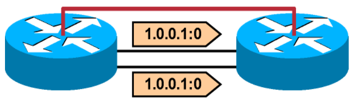
Cell Mode

Mixed Mode

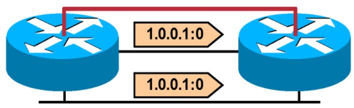
APNIC

Label Space and number of LDP sessions (Cont.)

Frame Mode



Frame Mode

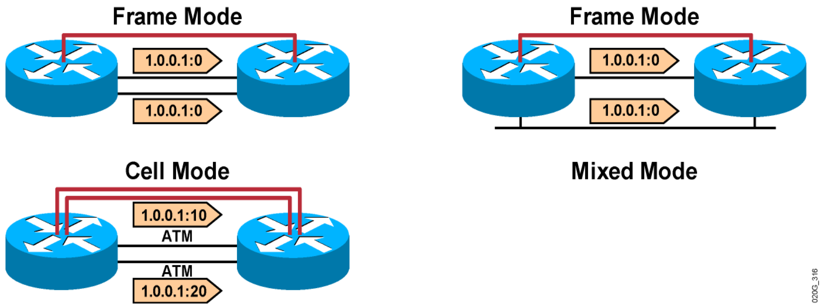


Cell Mode

Mixed Mode

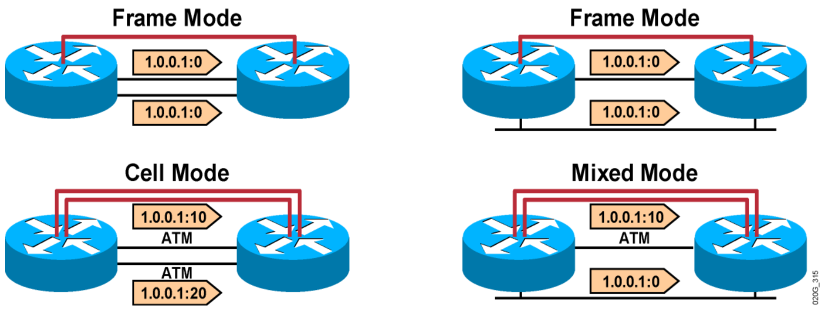
APNIC

Label Space and number of LDP sessions (Cont.)



APNIC

Label Space and number of LDP sessions (Cont.)



–One LDP session is established for each announced LDP identifier (router ID + label space).

–The number of LDP sessions is determined by the number of different label spaces.

–The bottom right example is not common, because ATM LSRs do not use Ethernet for packet forwarding, and frame-mode MPLS across ATM uses per-platform label space.

APNIC

LDP Messages Types

DISCOVERY messages

ADJACENCY messages deal with initialization, keepalive & shutdown of sessions

LABEL ADVERTISEMENT messages deal with label binding, requests, withdrawal & release

NOTIFICATION messages provide advisory information & signal errors

Discovery Message

- Used to discover and maintain the presence of new peers using HELLO messages
- Hello packets (UDP) sent to all-routers multicast address (224.0.0.2)
- Direct unicast hello is sent to non-adjacent neighbors
- Once session is established, HELLO messages serve as link integrity messages
- Session is bi-directional

Adjacency Messages

INITIALIZATION

Two LSRs negotiate on various parameters & options

These include keepalive timer values, Label ranges, Unsolicited vs. On-demand label advertisement, Ordered vs. Independent mode, Liberal vs. Conservative Label retention

KEEPALIVE

LDP message that indicates that neighbor is alive

APNIC

Label Advertisement related messages

- LABEL RELEASE

  An LSR releases a Label Binding that it previously got from it's LDP peer. Used in Conservative Label Retention mode

- LABEL REQUEST

  Used by an upstream LSR to request a Label binding for a prefix from the downstream LDP peer. Used in downstream on-demand mode

- LABEL ABORT REQUEST

  Send to abort the LABEL REQUEST message

- LABEL MAPPING

  Are the TLV object containing <Label, prefix> information

- LABEL WITHDRAWAL

  Used to revoke a previously advertised label binding

APNIC

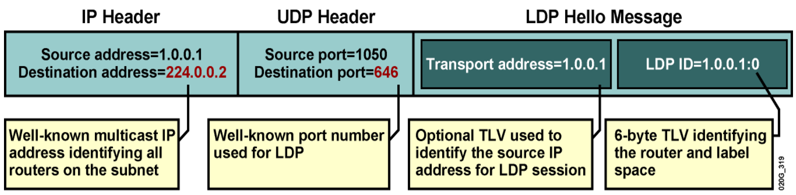
Notification message

- NOTIFICATION
  Used for Error Notification and Advisory

LDP Session Establishment

- LDP establishes a session by performing the following:
  - Hello messages are periodically sent on all interfaces that are enabled for MPLS.
  - If there is another router connected to that interface, that it also has MPLS enabled, it will respond by trying to establish a session with the source of the hello messages.
- UDP is used for hello messages. It is targeted at “all routers on this subnet” multicast address (224.0.0.2).
- TCP is used to establish the session.
- Both TCP and UDP use well-known LDP port number 646 (711 for TDP).

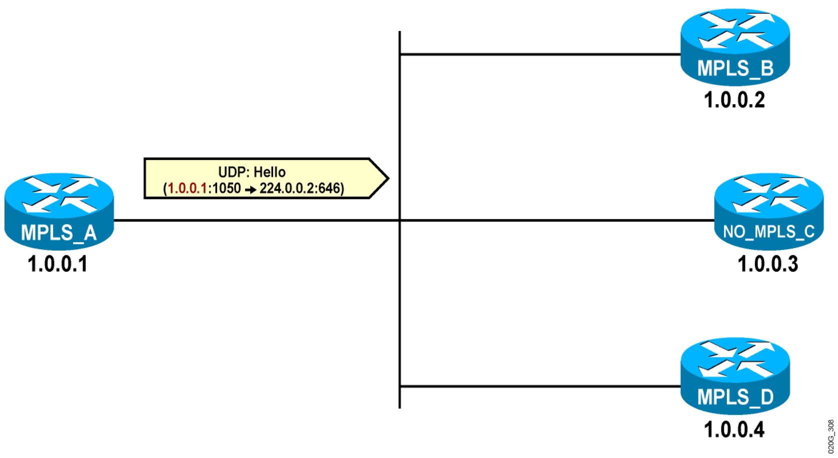
LDP Hello Message



- –Hello messages are targeted at all routers reachable through an interface.
- –LDP uses well-known (UDP and TCP) port number 646.
- –The source address used for an LDP session can be set by adding the transport address TLV to the hello message.
- –A 6-byte LDP identifier (TLV) identifies the router (first four bytes) and label space (last two bytes).

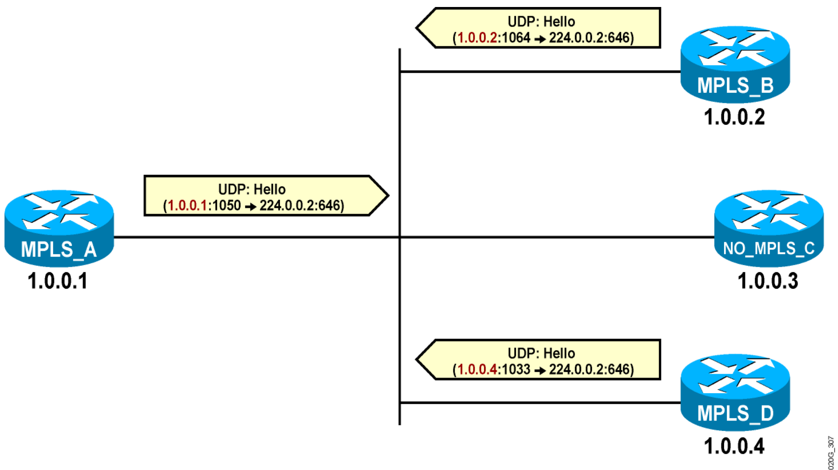
APNIC

LDP Neighbor Discovery



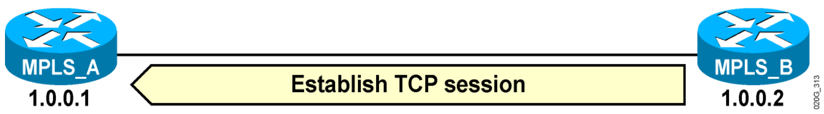
APNIC

LDP Neighbor Discovery



APNIC

LDP Session: Transport Connection



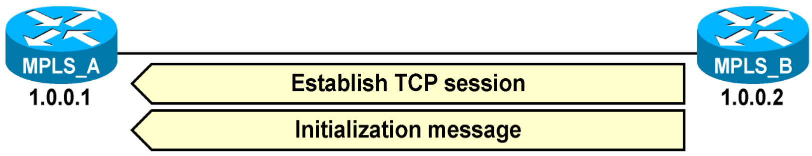
–Once LDP peers receive hellos, they establish a TCP connection

–Peer with higher LDP router-id is active LSR and the peer with lower LDP router-id is the passive LSR

–Active LSR tries to open a TCP connection to the well-known LDP port number 646 of the passive LSR, while the passive LSR waits for the active LSR to initiate the connection

APNIC

LDP Session:
Session Initialization

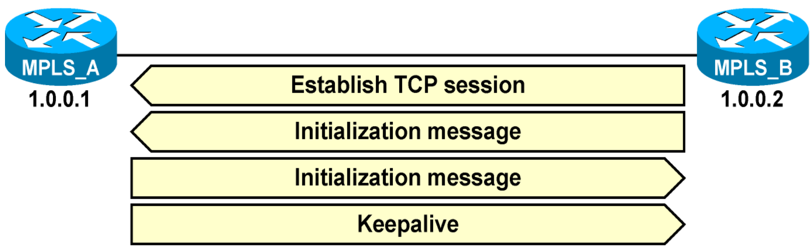


–Active LDP peer (1.0.0.2) sends Initialization message to passive LDP peer

–Initialization message contains important parameters:

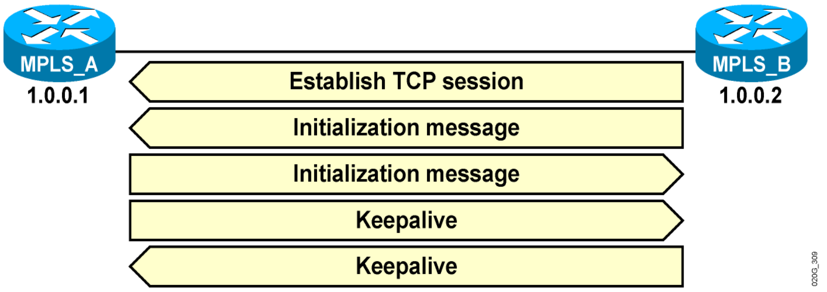
- Session keepalive time (default=180 sec)
- Label distribution method: Downstream unsolicited
- Max PDU length
- Receiver’s LDP Identifier
- Whether Loop Detection is enabled
- Some optional parameters

LDP Session:
Session Initialization (cont.)



- Passive LDP peer sends Initialization message and/or keepalive message to active LDP peer if Initialization message parameters are acceptable
- Passive LDP peer could also send Error Notification & close the LDP connection if something was unacceptable

LDP Session: Session Initialization (cont.)

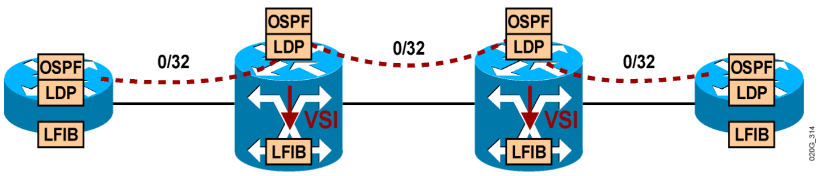


–Active LDP peer sends keepalive to passive LDP peer & the LDP session is up

–The session is ready to exchange label mappings after receiving the first keepalive.

APNIC

LDP Sessions Between ATM LSRs



–An IP adjacency between ATM LSRs is established through the control virtual circuit (0/32).

–The control virtual circuit is used for LDP as well as for IP routing protocols.

–VSI protocol is used to populate the ATM switching matrix (LFIB) in the data plane of some ATM switches (Cisco implementation).

APNIC

Targeted LDP Sessions

–LDP neighbor discovery of nonadjacent neighbors differs from normal discovery only in the addressing of hello packets:

- **Hello packets use unicast IP addresses instead of multicast addresses.**

–When a neighbor is discovered, the mechanism to establish a session is the same.

APNIC

Summary

- TCP is used to establish LDP sessions between neighbors.
- LDP uses PDUs to carry messages
- LDP hello messages contain an identifier field that uniquely identifies the neighbor and the label space.
- Per-platform label space requires only one LDP session.
- Routers that have the higher IP address must initiate the TCP session.
- LDP session negotiation is a three-step process.
- LDP sessions between ATM LSRs use the control VPI/VCI, which by default is 0/32.
- Nonadjacent neighbor discovery is accomplished by using unicast IP addresses instead of multicast.

APNIC

Questions?

Overview

Introduction to MPLS & Applications

- Introduction to MPLS
- MPLS Label Distribution Protocol
- **Basic MPLS Configuration (Lab Exercise)**
- MPLS Layer 3 VPN Concept & Architecture
- MPLS L3 VPN Configuration (Lab Exercise)
- MPLS Layer 2 VPN Concept & Architecture
- MPLS L2 VPN Configuration (Lab Exercise)
- MPLS Traffic Engineering

Configuration and Monitoring

Label Binding Protocol of MPLS

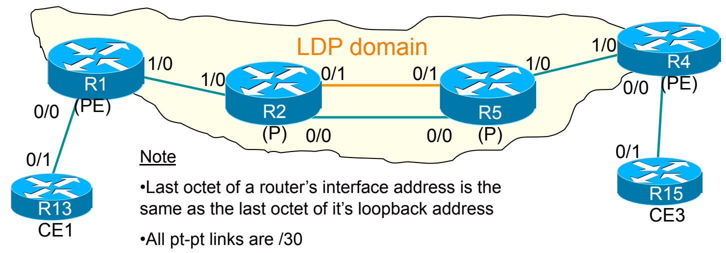
APNIC

Configuring MPLS

- Mandatory
  - Enable CEF switching if this is not default (Only Cisco Router)
  - Configure TDP or LDP on every label-enabled interface
- Optional
  - Configure MTU size for labeled packets
  - Configure IP TTL propagation
  - Configure conditional label advertising

APNIC

Network Topology



	R1	R13	R2	R5	R4	R15
R1	172.16.15.1(lo0)	172.16.11.0	172.16.10.0			
R13	172.16.11.0	172.16.16.254(lo0)				
R2	172.16.10.0		172.16.15.2 (lo0)	172.16.12.0 (0/0) 172.16.13.0 (0/1)		
R5			172.16.12.0 (0/0) 172.16.13.0 (0/1)	172.16.15.5 (lo0)	172.16.10.24	
R4				172.16.10.24	172.16.15.4 (lo0)	172.16.11.64
R15					172.16.11.64	172.16.20.254(lo0)

APNIC

Configuring LDP (IOS)

Global

`ip cef <distributed>`

`mpls label protocol <ldp | tdp | both>`

`tag-switching tdp router-id Loopback0`

`mpls ldp explicit-null` (optional)

`no mpls ip propagate-ttl` (optional)

Interface

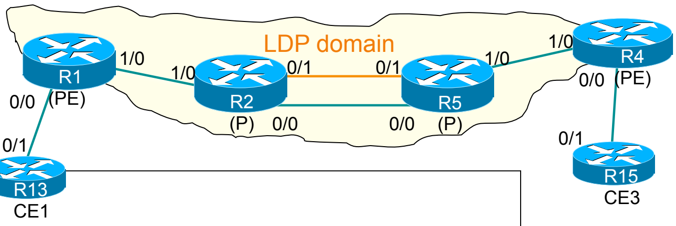
`mpls ip` or `tag-switching ip` (enables this interface for MPLS forwarding)

`mpls label protocol ldp`

(optional, if you want to run LDP on this interface only, while other interfaces don't run LDP or run another label protocol such as TDP)

APNIC

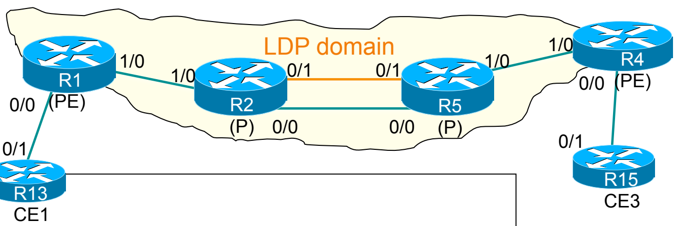
Configuring LDP [R1]




```
R1
config t
ip cef
mpls ldp router-id loopback 0 force
int e1/0
mpls ip
mpls label protocol ldp
mpls mtu override 1512
exit
exit
wr
```

APNIC

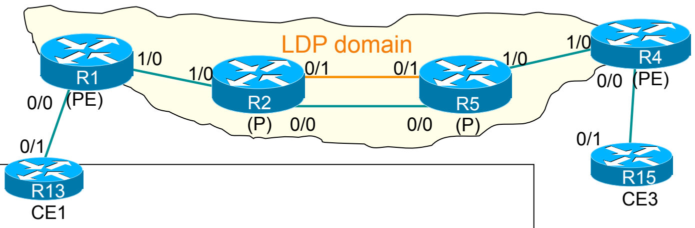
Configuring LDP [R2]




```
R2
config t
ip cef
mpls ldp router-id loopback 0 force
int e1/0
mpls ip
mpls label protocol ldp
mpls mtu override 1512
int fa0/0
mpls ip
mpls label protocol ldp
mpls mtu override 1512
int fa0/1
mpls ip
mpls label protocol ldp
mpls mtu override 1512
exit
exit
Wr
```

APNIC

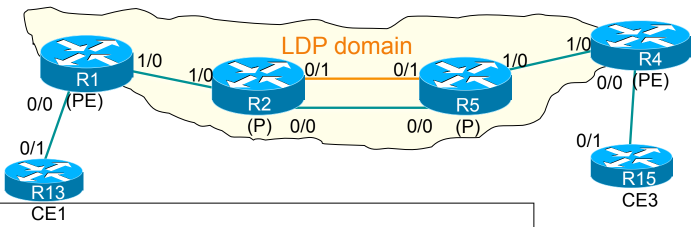
Configuring LDP [R5]



R5

```
config t
ip cef
mpls ldp router-id loopback 0 force
int e1/0
mpls ip
mpls label protocol ldp
mpls mtu override 1512
int fa0/0
mpls ip
mpls label protocol ldp
mpls mtu override 1512
int fa0/1
mpls ip
mpls label protocol ldp
mpls mtu override 1512
exit
exit
wr
```

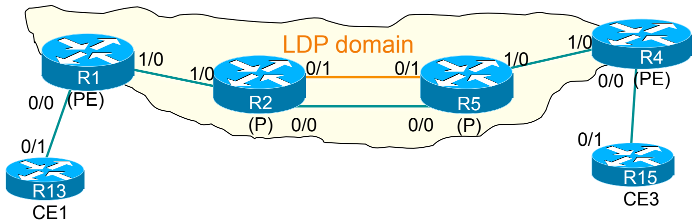
Configuring LDP [R4]



R4

```
config t
ip cef
mpls ldp router-id loopback 0 force
int e1/0
mpls ip
mpls label protocol ldp
mpls mtu override 1512
exit
exit
wr
```

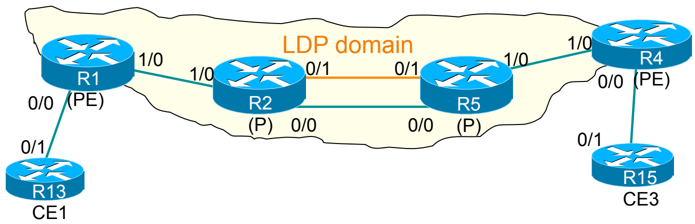
MPLS TTL Operation



- On ingress LER: TTL is copied from IP header to label header.
- On egress LER: TTL is copied from label header to IP header.
- Cisco routers have TTL propagation enabled by default.
  - Traceroute can reveal all PE and P routers along the path
- Disable TTL propagation
  - Traceroute will hide all PE and P routers along the path

APNIC

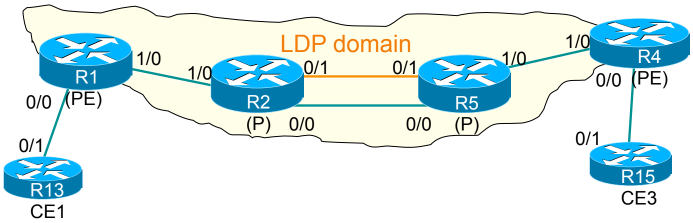
Configuring LDP [R1, R2, R5, R4]




```
R1, R2, R5, R4
config t
no mpls ip propagate-ttl forwarded
exit
wr
```

APNIC

Configuring LDP [R1, R2, R5, R4]



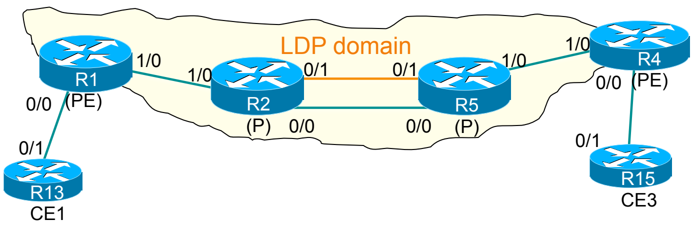

```
Traceroute before disable TTL propagation

R13#traceroute 172.16.20.254 [R15 lo0]

1 172.16.11.1 [R1 0/0]
2 172.16.10.1 [R2 1/0] [MPLS: Label 28 Exp 0]
3 172.16.12.2 [R5 0/1] [MPLS: Label 29 Exp 0]
4 172.16.10.26 [R4 1/0]
5 172.16.11.66 [R15 0/1]
```

APNIC

Configuring LDP [R1, R2, R5, R4]




```
Traceroute after disable TTL propagation

R13#traceroute 172.16.20.254 [R15 lo0]

1 172.16.11.1 [R1 0/0]
2 172.16.10.26 [R4 1/0]
3 172.16.11.66 [R15 0/1]
```

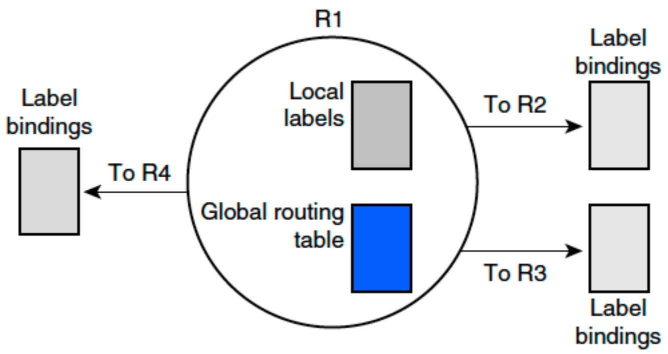
APNIC

Conditional Label Distribution (Local Label Allocation Filtering)

- Enable filtering for selective local label binding by LDP
- It improve LDP scalability and convergence
- LDP local label allocation filtering works on either prefix list or host route on the global routing table of an LSR
  – Prefix list use tree based matching technique which is more efficient
- Remote bindings are retained on LDP binding table
  – So need to apply local label allocation filter on all neighbouring LSR
- Controlling local label allocation could off-load LDP processing for non-VPN LSP on the SP network

APNIC

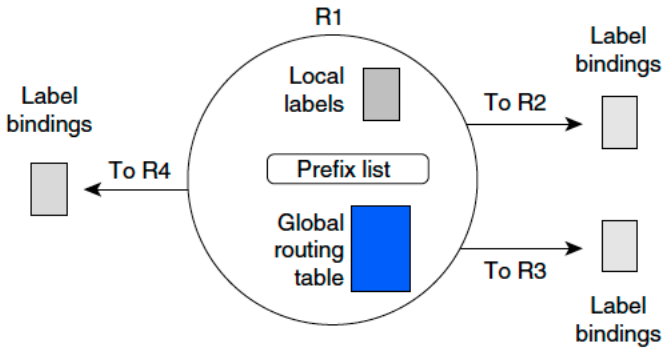
Conditional Label Distribution (Local Label Allocation Filtering)



Default LDP Local Label Allocation Behavior

APNIC

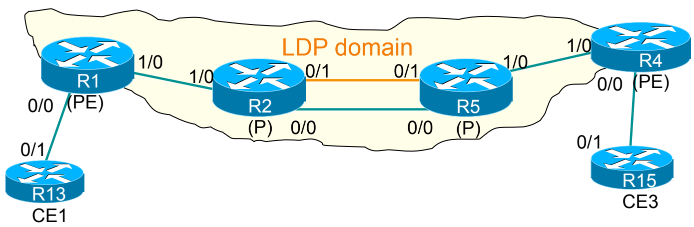
Conditional Label Distribution (Local Label Allocation Filtering)



Controlled LDP Local Label Allocation Behavior

APNIC

Conditional Label Distribution (Local Label Allocation Filtering)




```
R1, R2, R5, R4

config t
ip prefix-list ALL-LOOPBACK seq 15 permit 172.16.15.0/24 le 32
mpls ldp label
allocate global prefix-list ALL-LOOPBACK
exit
exit
wr
```

APNIC

Verify Your Configuration

Label Binding Protocol of MPLS

APNIC

Monitoring LDP

- `show mpls interface <x> detail`
- `show mpls ldp discovery`
- `show mpls ldp neighbor`
- `show mpls ip/ldp binding <prefix> <prefix-length>`
- `show mpls forwarding-table <prefix> <prefix-length>`
- `sh ip cef <prefix>`
- `show mpls ldp parameters`

APNIC

Show mpls interface

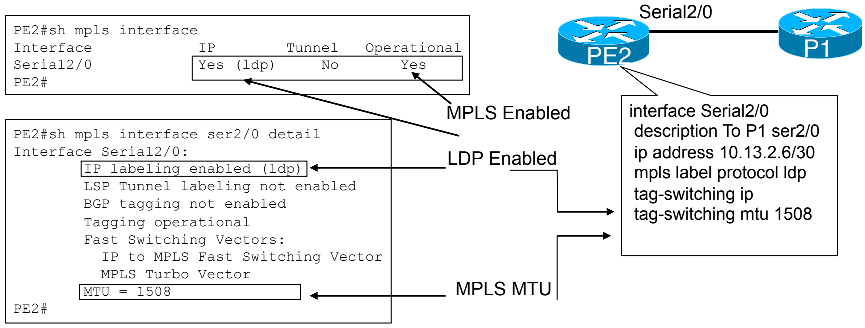
```
mpls-7200a#sh mpls interface
Interface          IP          Tunnel   Operational
Ethernet3/0          Yes (ldp)     No       Yes

mpls-7200a#sh mpls interface ethernet3/0 detail
Interface Ethernet3/0:
      IP labeling enabled (ldp)
      ........<snip>.......
        Fast Switching Vectors:
        IP to MPLS Fast Switching Vector
        MPLS Turbo Vector
        MTU = 1500
```

APNIC

Show mpls interface (contd..)

- “sh mpls interface [detail]”
  - –Lists whether MPLS is enabled and the application that enabled MPLS on the interface



APNIC

Show mpls interface (contd..)

- This slide is to show that BGPipv4+label (or MP-eBGP) is another application that can enable MPLS; *what's different here -*

```
RSP-PE-SOUTH-6#sh mpls int
Interface              IP           Tunnel    Operational
Fddi1/0/0              Yes (ldp)    No        Yes
ATM1/1/0.108           No           No        Yes
RSP-PE-SOUTH-6#
```

MPLS is Operational.

LDP not enabled

```
RSP-PE-SOUTH-6#sh mpls int ATM1/1/0.108 detail
Interface ATM1/1/0.108:
        IP labeling not enabled
        LSP Tunnel labeling not enabled
        BGP tagging enabled
        Tagging operational
        Optimum Switching Vectors:
          IP to MPLS Feature Vector
          MPLS Feature Vector
        Fast Switching Vectors:
          IP to MPLS Fast Feature Switching Vector
          MPLS Feature Vector
        MTU = 4470
RSP-PE-SOUTH-6#
```

LDP not enabled

BGP+Label Enabled

MPLS MTU

LDP Discovery/Adjacency: Commands and Debugs

- `show mpls ldp discovery`
- `debug mpls ldp transport`
- `debug mpls ldp session io`

LDP Discovery

Interface eth3/0 configured with LDP

```
mpls-7200a#sh mpls ldp discovery
 Local LDP Identifier:
    4.4.4.4:0
    Discovery Sources:
    Interfaces:
        Ethernet3/0 (ldp): xmit/recv
            LDP Id: 5.5.5.5:0
```

My LDP id

we are transmitting & receiving LDP messages

Neighbor's LDP id

"debug mpls ldp transport events"

- Should give information regarding whether the HELLOS are advertised/ received

APNIC

LDP Neighbor

```
mpls-7200a#sh mpls ldp neighbor
    Peer LDP Ident: 5.5.5.5:0; Local LDP Ident 4.4.4.4:0
       TCP connection: 5.5.5.5.11000 - 4.4.4.4.646
       State: Oper; Msgs sent/rcvd: 268/264; Downstream Up time:
03:41:45
       LDP discovery sources:
          Ethernet3/0, Src IP addr: 10.0.3.5
       Addresses bound to peer LDP Ident:
          10.0.3.5        10.0.4.5        10.0.5.5        5.5.5.5
```

APNIC

LDP Neighbor (contd..)

- LDP session is a TCP session (port = 646)
- Multiple links between two routers still mean single LDP session.

```
PE1#sh mpls ldp neighbor
    Peer LDP Ident: 10.13.1.101:0; Local LDP Ident 10.13.1.61:0
        TCP connection: 10.13.1.101.11031 - 10.13.1.61.646
        State: Oper; Msgs sent/rcvd: 58/60; Downstream
        Up time: 00:39:27
        LDP discovery sources:
          Ethernet0/0, Src IP addr: 10.13.1.5
          Ethernet1/0, Src IP addr: 10.13.1.9
        Addresses bound to peer LDP Ident:
          10.13.1.9       10.13.1.5       10.13.2.5
10.13.1.101
PE1#
```

LDP_ID

*Unsolicited Label Distribution**

Interfaces on which peer is discovered

Peer's Connected interface

```
PE1#sh tcp brief| i 646
43ABB020  10.13.1.101.11031        10.13.1.61.646         ESTAB
PE1#
```

APNIC

LDP Binding Commands

- "sh mpls ip binding detail"
  - –Lists all prefixes with labels & LDP neighbors
- "sh mpls ip binding <*prefix*> <mask> detail"
  - –Lists ACLs (if any), *prefix* bindings, and LDP neighbors. <u>Notice "Advertised to:" field</u>.
- "sh mpls ip binding advertisement-acls"
  - –Lists LDP filter, if there is any, on the first line. Prefixes followed by "Advert acl(s):" are advertised via LDP, others are not.

APNIC

LIB Information

```
mpls-7200a#sh mpls ip binding 12.12.12.12 32
  12.12.12.12/32
        in label:     21
        out label:    19         lsr: 5.5.5.5:0    in use


mpls-7200a#sh mpls ldp binding 12.12.12.12 32
  tib entry: 12.12.12.12/32, rev 48
        local binding:  tag: 21
        remote binding: tsr: 5.5.5.5:0, tag: 19
```

APNIC

LFIB Information

```
show mpls forwarding-table <prefix>
  <prefix-length>
sh ip cef <prefix> internal
```

APNIC

Looking at LFIB

Looking at LFIB on 12008a

Ethertype=
8847
Label Value in MPLS shim=
13 Hex=19 dec

```
mpls-12008a#sh mpls forwarding 12.12.12.12 32 detail
Local  Outgoing   Prefix            Bytes tag  Outgoing   Next Hop
tag    tag or VC  or Tunnel Id      switched   interface
19     19         12.12.12.12/32    498        Et2/0      10.0.4.11
        MAC/Encaps=14/18, MTU=1500, Tag Stack{19}
        AABBCC000502AABBCC0004028847 00013000
        No output feature configured
 Per-destination load-sharing, slots: 0 2 4 6 8 10 12 14
       19         12.12.12.12/32    498        Et3/0      10.0.5.11
        MAC/Encaps=14/18, MTU=1500, Tag Stack{19}
        AABBCC000503AABBCC0004038847 00013000
        No output feature configured
 Per-destination load-sharing, slots: 1 3 5 7 9 11 13 15
```

Destination MAC=
AABBCC000502
Source MAC=
AABBCC000402

APNIC

CEF Command

```
mpls-12008a#sh ip cef 12.12.12.12 internal
12.12.12.12/32, version 24, epoch 0, per-destination sharing
0 packets, 0 bytes
  tag information set, local tag: 19
  via 10.0.4.11, Ethernet2/0, 0 dependencies
    traffic share 1
    next hop 10.0.4.11, Ethernet2/0
    valid adjacency
    tag rewrite with Et2/0, 10.0.4.11, tags imposed: {19}
  via 10.0.5.11, Ethernet3/0, 0 dependencies
    traffic share 1
    next hop 10.0.5.11, Ethernet3/0
    valid adjacency
    tag rewrite with Et3/0, 10.0.5.11, tags imposed: {19}
0 packets, 0 bytes switched through the prefix
......................(contd..)
```

tmstats: external 0 packets, 0 bytes

internal 0 packets, 0 bytes

Load distribution: 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 (refcount 1)

Hash	OK	Interface	Address	Packets	Tags imposed
1	Y	Ethernet2/0	10.0.4.11	0	{19}
2	Y	Ethernet3/0	10.0.5.11	0	{19}
3	Y	Ethernet2/0	10.0.4.11	0	{19}
4	Y	Ethernet3/0	10.0.5.11	0	{19}
5	Y	Ethernet2/0	10.0.4.11	0	{19}
6	Y	Ethernet3/0	10.0.5.11	0	{19}
7	Y	Ethernet2/0	10.0.4.11	0	{19}
8	Y	Ethernet3/0	10.0.5.11	0	{19}
9	Y	Ethernet2/0	10.0.4.11	0	{19}
10	Y	Ethernet3/0	10.0.5.11	0	{19}
11	Y	Ethernet2/0	10.0.4.11	0	{19}
12	Y	Ethernet3/0	10.0.5.11	0	{19}
13	Y	Ethernet2/0	10.0.4.11	0	{19}
14	Y	Ethernet3/0	10.0.5.11	0	{19}
15	Y	Ethernet2/0	10.0.4.11	0	{19}
16	Y	Ethernet3/0	10.0.5.11	0	{19}

APNIC

Monitoring LDP: LDP Parameters

```
mpls-7200a#sh mpls ldp parameters
Protocol version: 1
Downstream label generic region: min label: 16; max label: 100000
Session hold time: 180 sec; keep alive interval: 60 sec
Discovery hello: holdtime: 15 sec; interval: 5 sec
Discovery targeted hello: holdtime: 180 sec; interval: 5 sec
Downstream on Demand max hop count: 255
TDP for targeted sessions
LDP initial/maximum backoff: 15/120 sec
LDP loop detection: off
```

APNIC

Forwarding Traffic Down The LSP

```
mpls-7200a#sh mpls forwarding-table 12.12.12.12
Local  Outgoing    Prefix            Bytes tag  Outgoing   Next Hop
tag    tag or VC   or Tunnel Id      switched   interface
21     19          12.12.12.12/32    0          Et3/0      10.0.3.5
Note: Bytes tag switched this will increment if packets are being tag switched
using this entry
```

```
mpls-12008a#sh mpls forwarding-table label 19
Local  Outgoing    Prefix            Bytes tag  Outgoing   Next Hop
tag    tag or VC   or Tunnel Id      switched   interface
19     19          12.12.12.12/32    498        Et2/0      10.0.4.11
       19          12.12.12.12/32    1176       Et3/0      10.0.5.11
```

```
mpls-12008b#sh mpls forwarding-table labels 19
Local  Outgoing    Prefix            Bytes tag  Outgoing   Next Hop
tag    tag or VC   or Tunnel Id      switched   interface
19     Pop tag     12.12.12.12/32    4176       Et1/0      10.0.17.12
```

APNIC

Questions?

APNIC

Overview

Introduction to MPLS & Applications

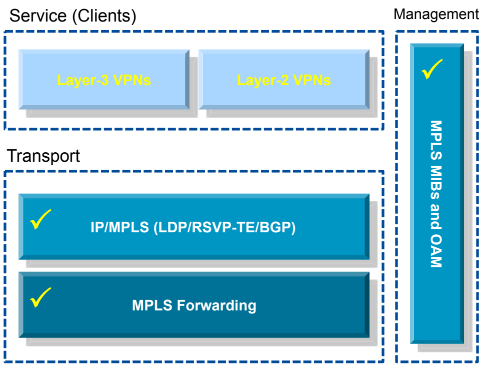
- Introduction to MPLS
- MPLS Label Distribution Protocol
- Basic MPLS Configuration (Lab Exercise)
- **MPLS Layer 3 VPN Concept & Architecture**
- MPLS L3 VPN Configuration (Lab Exercise)
- MPLS Layer 2 VPN Concept & Architecture
- MPLS L2 VPN Configuration (Lab Exercise)
- MPLS Traffic Engineering

APNIC

MPLS Virtual Private Networks

Topics

- Definition of MPLS VPN service
- Basic MPLS VPN deployment scenario
- Technology options



What Is a Virtual Private Network?

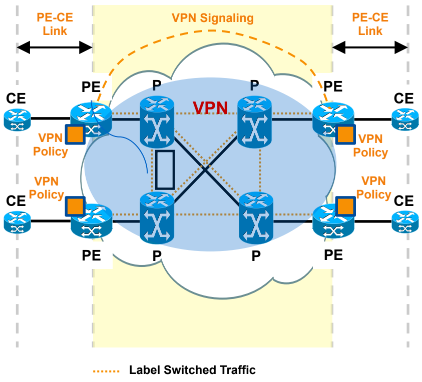
Definition

- Set of sites which communicate with each other in a secure way
  - Typically over a shared public or private network infrastructure
- Defined by a set of administrative policies
  - Policies established by VPN customers themselves (DIY)
  - Policies implemented by VPN service provider (managed/unmanaged)
- Different inter-site connectivity schemes possible
  - Full mesh, partial mesh, hub-and-spoke, etc.
- VPN sites may be either within the same or in different organizations
  - VPN can be either intranet (same org) or extranet (multiple orgs)
- VPNs may overlap; site may be in more than one VPN

MPLS VPN Example

Basic Building Blocks

- VPN policies
  - Configured on PE routers (manual operation)
- VPN signaling
  - Between PEs
  - Exchange of VPN policies
- VPN traffic forwarding
  - Additional VPN-related MPLS label encapsulation
- PE-CE link
  - Connects customer network to MPLS network; either layer-2 or layer-3

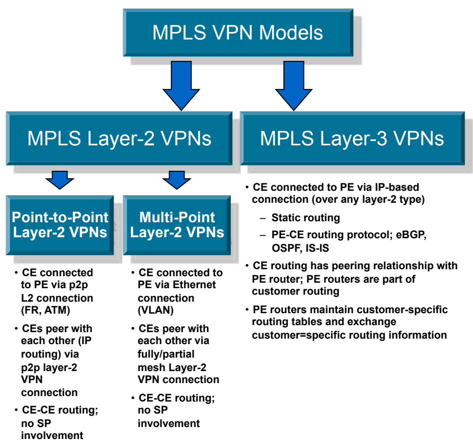


APNIC

MPLS VPN Models

Technology Options

- MPLS Layer-3 VPNs
  - Peering relationship between CE and PE
- MPLS Layer-2 VPNs
  - Interconnect of layer-2 Attachment Circuits (ACs)



APNIC

MPLS Layer-3 Virtual Private Networks

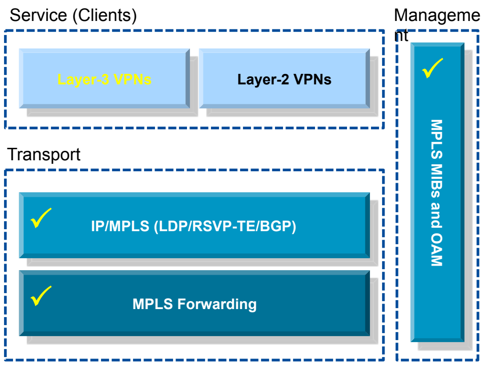
End-to-end Layer-3 Services Over MPLS Networks

APNIC

MPLS Layer-3 Virtual Private Networks

Topics

- Technology components
- VPN control plane mechanisms
- VPN forwarding plane
- Deployment use cases
  - Business VPN services
  - Network segmentation
  - Data Center access



APNIC

MPLS Layer-3 VPN Overview

Technology Components

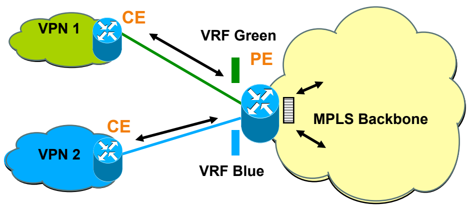
- VPN policies
  - Separation of customer routing via virtual VPN routing table (VRF)
  - In PE router, customer interfaces are connected to VRFs
- VPN signaling
  - Between PE routers: customer routes exchanged via BGP (MP-iBGP)
- VPN traffic forwarding
  - Separation of customer VPN traffic via additional VPN label
  - VPN label used by receiving PE to identify VPN routing table
- PE-CE link
  - Can be any type of layer-2 connection (e.g., FR, Ethernet)
  - CE configured to route IP traffic to/from adjacent PE router
  - Variety of routing options; static routes, eBGP, OSPF, IS-IS

APNIC

Virtual Routing and Forwarding Instance

Virtual Routing Table and Forwarding to Separate Customer Traffic

- Virtual routing and forwarding table
  - On PE router
  - Separate instance of routing (RIB) and forwarding table
- Typically, VRF created for each customer VPN
  - Separates customer traffic
- VRF associated with one or more customer interfaces
- VRF has its own routing instance for PE-CE configured routing protocols
  - E.g., eBGP

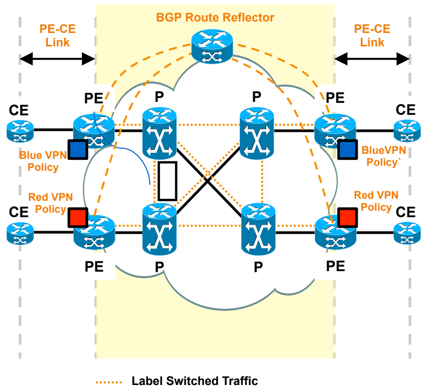


APNIC

VPN Route Distribution

Exchange of VPN Policies Among PE Routers

- Full mesh of BGP sessions among all PE routers
  - BGP Route Reflector
- Multi-Protocol BGP extensions (MP-iBGP) to carry VPN policies
- PE-CE routing options
  - Static routes
  - eBGP
  - OSPF
  - IS-IS



APNIC

VPN Control Plane Processing

VRF Parameters

Make customer routes unique:

- **Route Distinguisher (RD):** 8-byte field, VRF parameters; unique value to make VPN IP routes unique
- **VPNv4 address**: RD + VPN IP prefix

Selective distribute VPN routes:

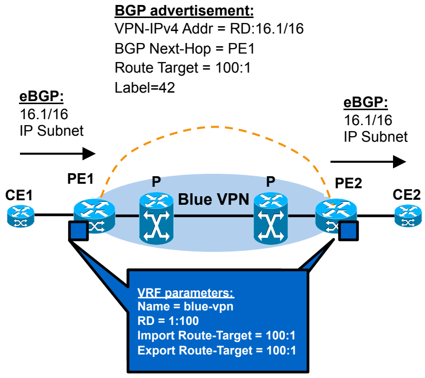
- **Route Target (RT):** 8-byte field, VRF parameter, unique value to define the import/ export rules for VPNv4 routes
- MP-iBGP: advertises VPNv4 prefixes + labels

APNIC

VPN Control Plane Processing

Interactions Between VRF and BGP VPN Signaling

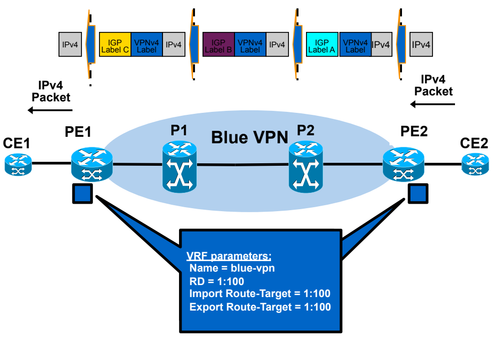
1. CE1 redistribute IPv4 route to PE1 via eBGP
2. PE1 allocates VPN label for prefix learnt from CE1 to create unique VPNv4 route
3. PE1 redistributes VPNv4 route into MP-iBGP, it sets itself as a next hop and relays VPN site routes to PE2
4. PE2 receives VPNv4 route and, via processing in local VRF (green), it redistributes original IPv4 route to CE2



VPN Forwarding Plane Processing

Forwarding of Layer-3 MPLS VPN Packets

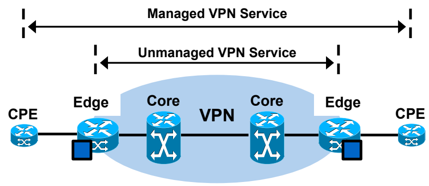
1. CE2 forwards IPv4 packet to PE2
2. PE2 imposes pre-allocated VPN label to IPv4 packet received from CE2
   - Learned via MP-IBGP
3. PE2 imposes outer IGP label A (learned via LDP) and forwards labeled packet to next-hop P-router P2
4. P-routers P1 and P2 swap outer IGP label and forward label packet to PE1
   - A->B (P2) and B->C (P1)
5. Router PE1 strips VPN label and IGP labels and forwards IPv4 packet to CE1



Service Provider Deployment Scenario

MPLS Layer-3 VPNs for Offering Layer-3 Business VPN Services

- Deployment Use Case
  - Delivery of IP VPN services to business customers
- Benefits
  - Leverage same network for multiple services and customers (CAPEX)
    - Highly scalable
  - Service enablement only requires edge node configuration (OPEX)
  - Different IP connectivity can be easily configured; e.g., full/ partial mesh

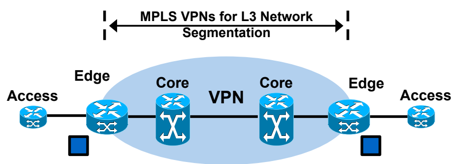


Network Segment	CPE	Edge	Core
MPLS Node	CE	PE	P
Typical Platforms	ASR1K ISR/G2	ASR9K 7600 ASR1K ASR903 ME3800X	CRS-1 GSR ASR9K

Enterprise Deployment Scenario

MPLS Layer-3 VPNs for Implementing Network Segmentation

- Deployment Use Case
  - Segmentation of enterprise network to provide selective connectivity for specific user groups and organizations
- Benefits
  - Network segmentation only requires edge node configuration
  - Flexible routing; different IP connectivity can be easily configured; e.g., full/partial mesh

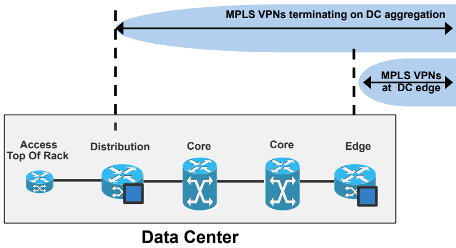


Network Segment	Access	Edge	Core
MPLS Node	CE	PE	P
Typical Platforms	ASR1K ISR/G2	7600 ASR1K	CRS-1 GSR ASR9K 7600 6500

Data Center Deployment Scenario

MPLS Layer-3 VPNs for Segmented L3 Data Center Access and Interconnect

- **Deployment Use Case**
  - Segmented WAN Layer-3 at Data Center edge
  - Layer-3 segmentation in Data Center
- **Benefits**
  - Only single Data Center edge node needed for segmented layer-3 access
  - Enables VLAN/Layer-2 scale (> 4K)



Network Segment	Distribution	Core	Edge
MPLS Node	CE or PE	P or CE	PE
Typical Platforms	N7K 6500	N7K 6500	ASR9K 7600

APNIC

Questions?

APNIC