NETWORKERS 2004

MPLS ADVANCED CONCEPTS AND DEVELOPMENTS IN MPLS

SESSION RST-2T09

References

- **RST-1601 Introduction to MPLS**
- **RST-1607 QoS in MPLS Networks**
- **RST-2602 Deploying MPLS VPNs**
- **RST-2603 Deploying MPLS Traffic Engineering**
- **RST-2606 Understanding Convergence in MPLS VPN Networks**
- **RST-3605 Troubleshooting MPLS Networks**
- **RST-3606 Troubleshooting MPLS VPNs**
- **RST-4607 Advanced Topics and Future Directions in MPLS**
- **NMS-4012 MPLS Embedded Management Tools**

Today’s Agenda

- **Recap**
  - **Multi-Protocol Label Switching (MPLS) Review**
- **MPLS Aware Services**
  - **Virtual Routing Forwarding (VRF) Aware IP Services**
- **Network Integration with MPLS**
  - **Advanced L3 VPN Services Models and Mechanisms**
  - **Advanced Traffic Engineering (TE) Applications**
  - **MPLS High Availability**
- **MPLS and Layer 1**
  - **Generalized MPLS and Optical User Network Interface (OUNI)**
- **MPLS Deployment Experience**

MPLS REVIEW

Agenda

- MPLS Basics
- MPLS L3 VPNs
- MPLS TE
- AToM

MPLS Basics

- MPLS fundamentals
- MPLS components
- MPLS operation

MPLS Fundamentals

- Based on the label-swapping and forwarding paradigm
- As a packet enters an MPLS network, it is assigned a label based on its Forwarding Equivalence Class (FEC) as determined at the edge of the MPLS network
- FECs are groups of packets forwarded over the same Label Switched Path (LSP)
- Need a mechanism that will create and distribute labels to establish LSP paths
- Separated into two planes:

    Control Plane—responsible for maintaining correct label tables among Label Switching Routers

    Forwarding Plane—uses label carried by packet and label table maintained by LSR to forward the packet

MPLS Components

- Label Distribution Protocol (LDP)
- Label Switching Paths (LSP)
- Label Switching Routers (LSR)
- Edge label switching routers (ELSR)
- MPLS labels and label stacking

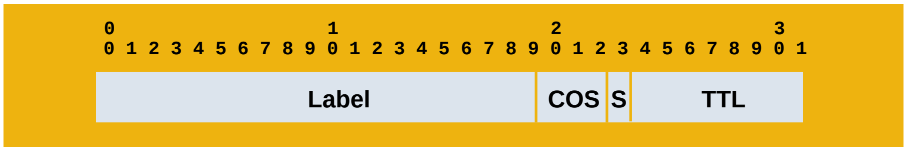
Label Distribution Protocol

Cisco.com

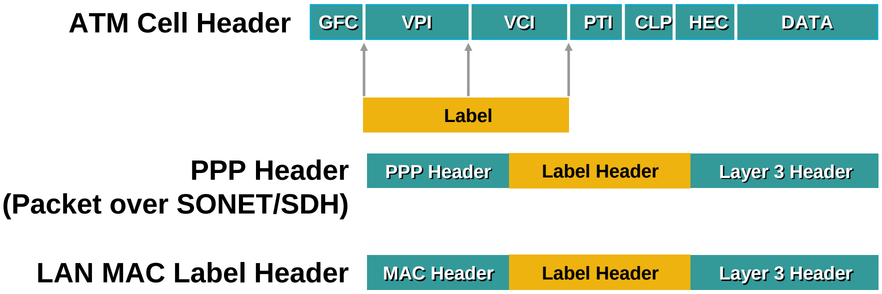
- Defined in RFC 3036 and 3037
- Used to distribute labels in an MPLS network
- Forwarding Equivalence Class (FEC)
  - How packets are mapped to LSPs (Label Switched Paths)
- Advertise labels per FEC
  - Reach destination a.b.c.d with label x
- Neighbor discovery
  - Basic and extended discovery

Label Header for Packet Media

Cisco.com


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
COS
S
TTL

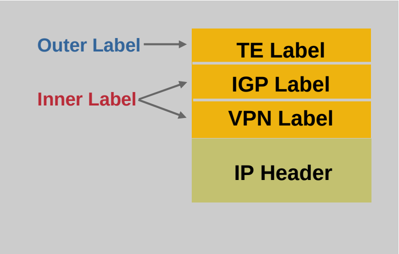
Label = 20 Bits   COS = Class of Service, 3 Bits S = Bottom of Stack, 1 Bit
TTL = Time to Live, 8 Bits


ATM Cell Header
GFC
VPI
VCI
PTI
CLP
HEC
DATA
Label
PPP Header
(Packet over SONET/SDH)
PPP Header
Label Header
Layer 3 Header
LAN MAC Label Header
MAC Header
Label Header
Layer 3 Header

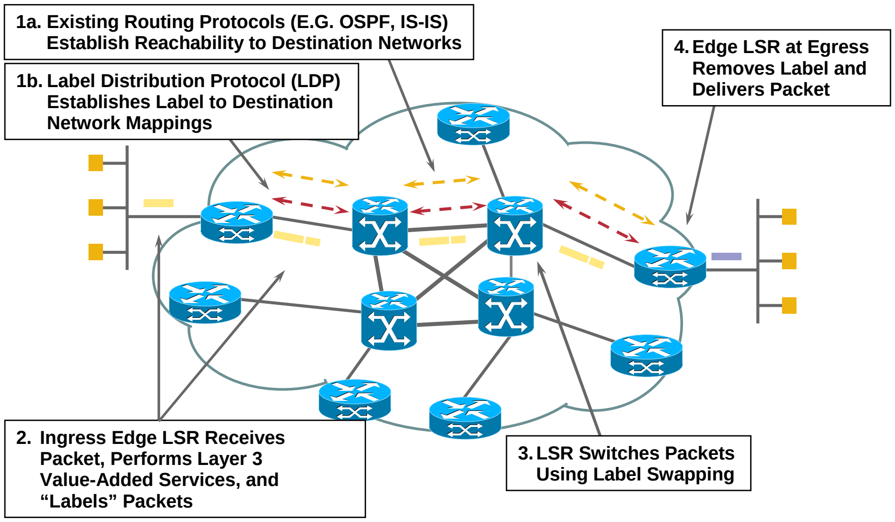
Label Stacking

Labels are arranged in a stack to support multiple services:

- MPLS VPNs: basic and InterProvider services
    - Any transport over MPLS
    - TE and fast reroute
    - VPNs over traffic engineered core
- Inner labels are used to designate services, FECs, etc.
- Outer label is used to switch the packets in MPLS core



MPLS Operation



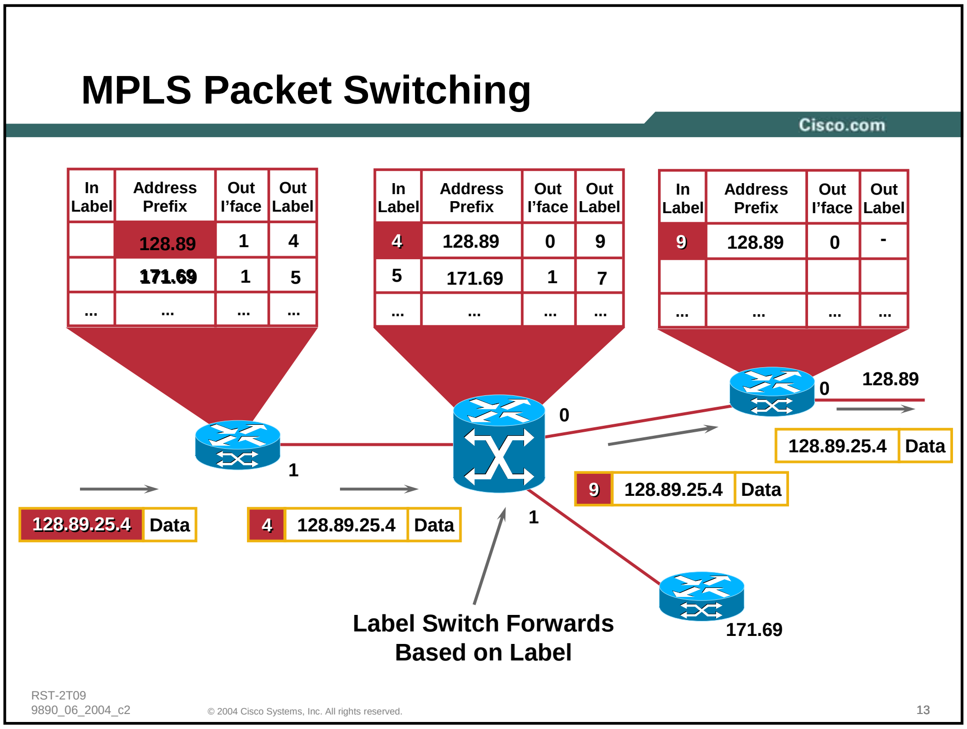
1a. Existing Routing Protocols (E.G. OSPF, IS-IS) Establish Reachability to Destination Networks

1b. Label Distribution Protocol (LDP) Establishes Label to Destination Network Mappings

2. Ingress Edge LSR Receives Packet, Performs Layer 3 Value-Added Services, and "Labels" Packets

3. LSR Switches Packets Using Label Swapping

4. Edge LSR at Egress Removes Label and Delivers Packet



MPLS Packet Switching

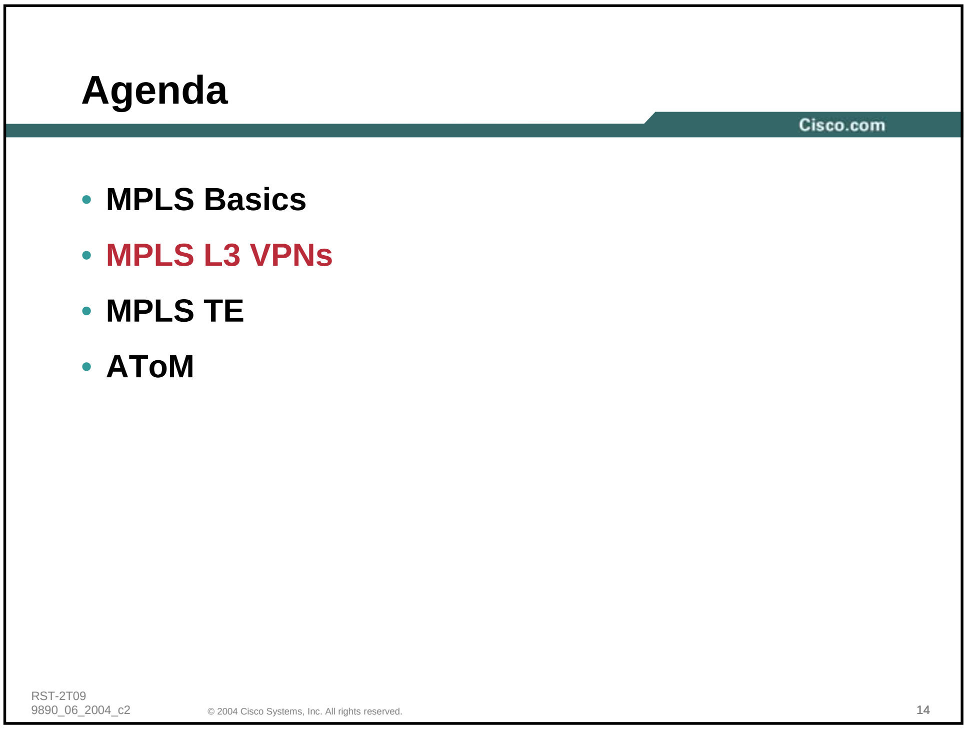
In Label	Address Prefix	Out I'face	Out Label
	128.89	1	4
	171.69	1	5
...	...	...	...

In Label	Address Prefix	Out I'face	Out Label
4	128.89	0	9
5	171.69	1	7
...	...	...	...

In Label	Address Prefix	Out I'face	Out Label
9	128.89	0	-
...	...	...	...

128.89.25.4 | Data
1
4 | 128.89.25.4 | Data
0
1
9 | 128.89.25.4 | Data
0
128.89
128.89.25.4 | Data
171.69

Label Switch Forwards Based on Label



Agenda

- MPLS Basics
- MPLS L3 VPNs
- MPLS TE
- AToM

MPLS L3 VPN Components

- Control plane components
    - Virtual Routing Forwarding (VRF) name
    - Route Target (RT)
    - Route Distinguisher (RD)
    - VRF table
    - MP-BGP
- Forwarding plane components
    - VPN label
    - Routing Information Base (RIB)
    - Forwarding Information Base (FIB)
    - Label Forwarding Information Base (LFIB)

Route Distinguisher

- Purely to make a route unique so customers don’t see each other’s routes
    - Ex: differentiate 10.0.0.0/8 in VPN-A from 10.0.0.0/8 in VPN-B
    - Makes IPv4 route a VPNv4 routes: VPNv4=RD:IPv4
    - Unique route is now RD:IPaddr (96 bits) plus a mask on the IPAddr portion
    - So route reflectors make a bestpath decision on something other than 32-bit network + 32-bit mask
- 64-bit quantity
- Configured as ASN:YY or IPADDR:YY
    - Almost everyone uses Autonomous System Number (ASN)

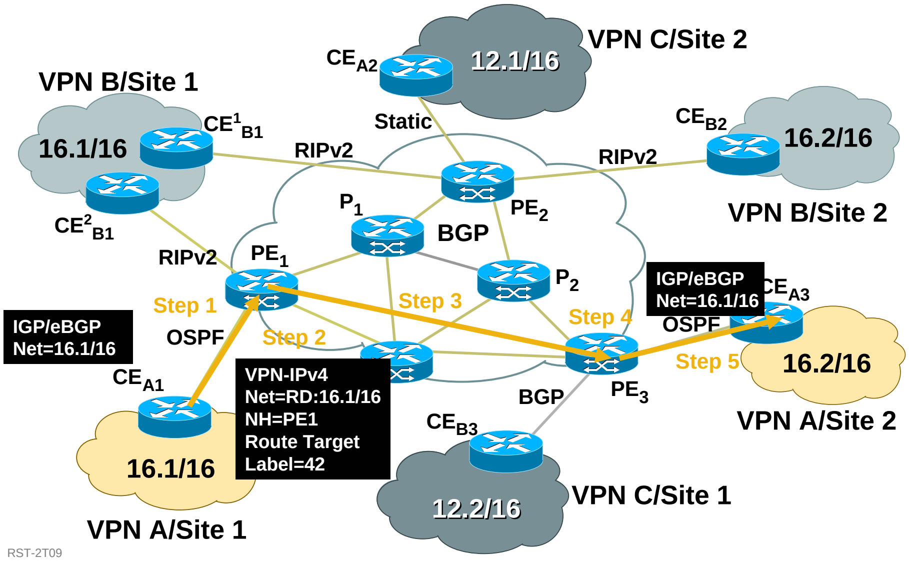
Route Target

- To control policy about who sees what routes
- 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 associated VRF table
- 64-bit quantity (2 bytes type, 6 bytes value) carried as an extended community
- Typically written as ASN:YY
- For deployment model:
    - Full mesh:
        - All sites import X:Y and export X:Y
    - Hub-and-spoke:
        - Hub exports X:H and imports X:S
        - Spokes export X:S and import X:H

MP-BGP and VPNv4

- MP-BGP session facilitates the advertisement of VPNv4* prefixes + labels between MP-BGP peers
- At the advertising PE, BGP allocates labels for VPN prefixes and installs them in the LFIB (MPLS forwarding table)
- At the receiving PE, IF BGP accepts VPN prefixes with labels, THEN BGP installs VPN prefixes in the VRF FIB (CEF table)
- VPNv4 announcement carries a label with the route
    - "If you want to reach this unique address, get me packets with this label on them"

MPLS L3 VPN Control Plane



VPN C/Site 2

12.1/16

CE$_{A2}$

Static

VPN B/Site 1

16.1/16

CE^1$_{B1}$

CE^2$_{B1}$

RIPv2

RIPv2

PE$_2$

CE$_{B2}$

16.2/16

VPN B/Site 2

P$_1$

BGP

RIPv2

PE$_1$

P$_2$

IGP/eBGP
Net=16.1/16

CE$_{A3}$

OSPF

Step 1

Step 3

Step 4

IGP/eBGP
Net=16.1/16

OSPF

Step 2

Step 5

16.2/16

CE$_{A1}$

VPN-IPv4
Net=RD:16.1/16
NH=PE1
Route Target
Label=42

BGP

PE$_3$

CE$_{B3}$

VPN A/Site 2

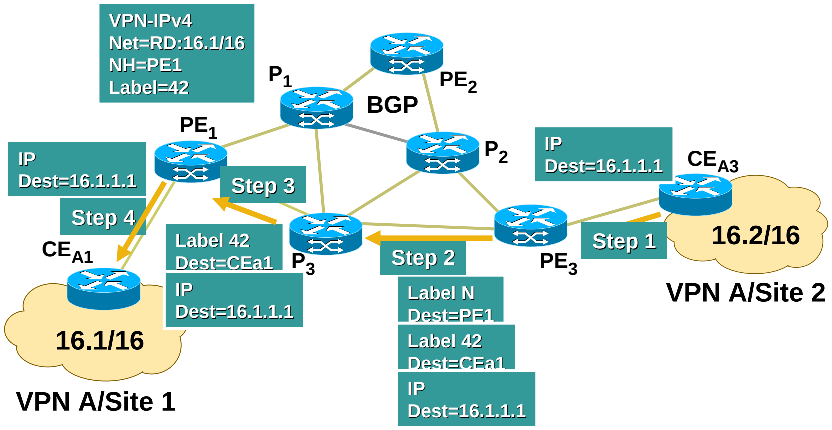
16.1/16

12.2/16

VPN C/Site 1

VPN A/Site 1

MPLS L3 VPN Forwarding Plane



VPN-IPv4
Net=RD:16.1/16
NH=PE1
Label=42

P$_1$

PE$_2$

BGP

PE$_1$

IP
Dest=16.1.1.1

P$_2$

IP
Dest=16.1.1.1

CE$_{A3}$

Step 3

Step 4

CE$_{A1}$

Label 42
Dest=CEa1

P$_3$

Step 2

PE$_3$

Step 1

16.2/16

VPN A/Site 2

IP
Dest=16.1.1.1

Label N
Dest=PE1

Label 42
Dest=CEa1

IP
Dest=16.1.1.1

16.1/16

VPN A/Site 1

MPLS L3 VPN Service Models

Cisco.com

Various VPN Service Models Are Available Based on the Following Topologies

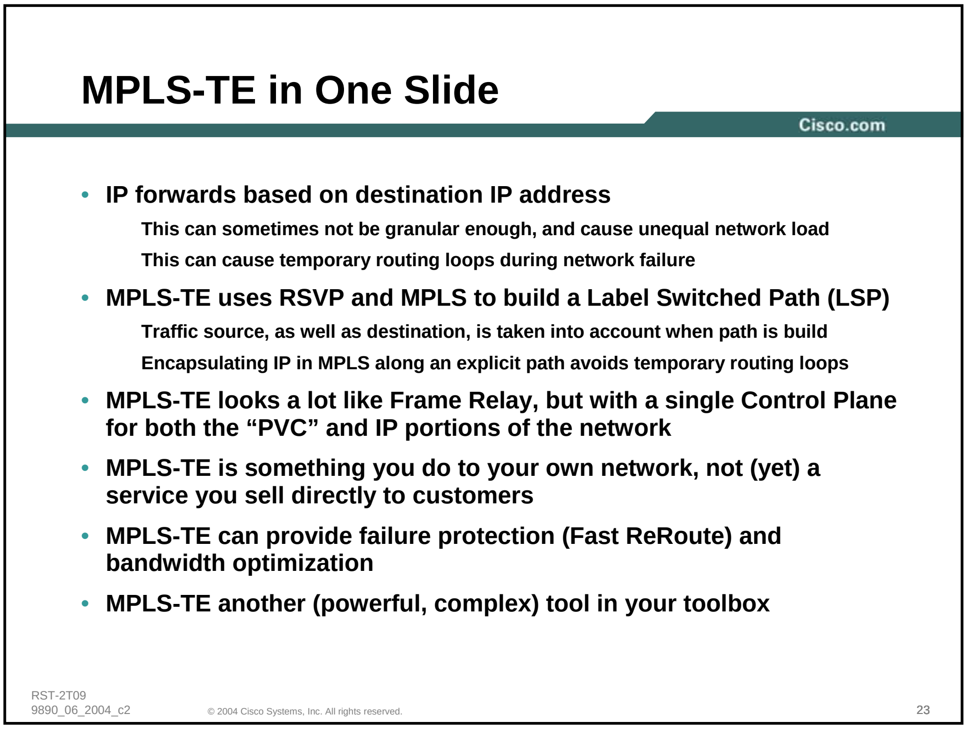
- Full mesh simple intranet
- Hub/spoke central site
- Advanced extranet
- Multi-VPN (per VLAN… etc.)
- Multi-AS VPNs

Deployment Examples Are Covered in RST2061

Agenda

Cisco.com

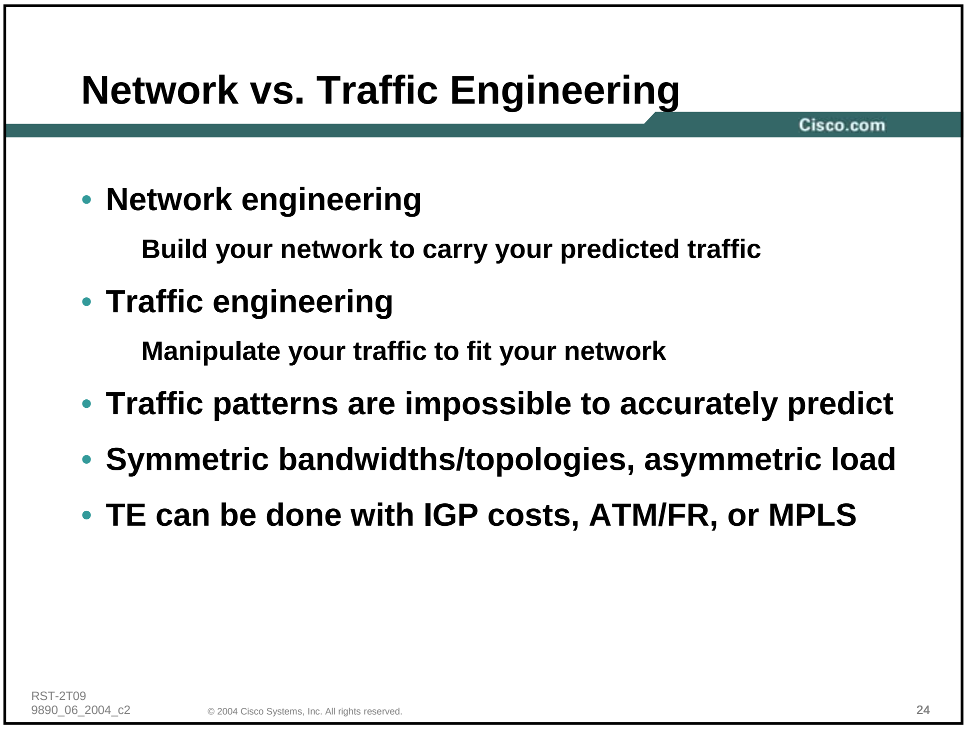
- MPLS Basics
- MPLS L3 VPNs
- MPLS TE
- AToM



MPLS-TE in One Slide

Cisco.com

- IP forwards based on destination IP address
  - This can sometimes not be granular enough, and cause unequal network load
  - This can cause temporary routing loops during network failure
- MPLS-TE uses RSVP and MPLS to build a Label Switched Path (LSP)
  - Traffic source, as well as destination, is taken into account when path is build
  - Encapsulating IP in MPLS along an explicit path avoids temporary routing loops
- MPLS-TE looks a lot like Frame Relay, but with a single Control Plane for both the “PVC” and IP portions of the network
- MPLS-TE is something you do to your own network, not (yet) a service you sell directly to customers
- MPLS-TE can provide failure protection (Fast ReRoute) and bandwidth optimization
- MPLS-TE another (powerful, complex) tool in your toolbox



Network vs. Traffic Engineering

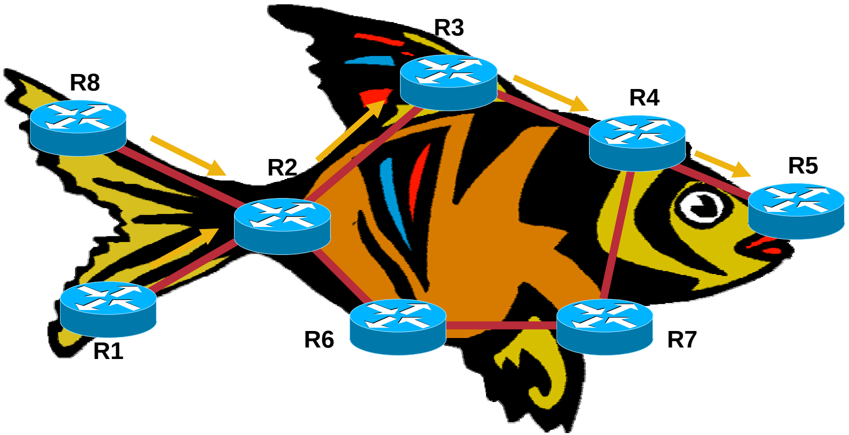
Cisco.com

- Network engineering
  - Build your network to carry your predicted traffic
- Traffic engineering
  - Manipulate your traffic to fit your network
- Traffic patterns are impossible to accurately predict
- Symmetric bandwidths/topologies, asymmetric load
- TE can be done with IGP costs, ATM/FR, or MPLS

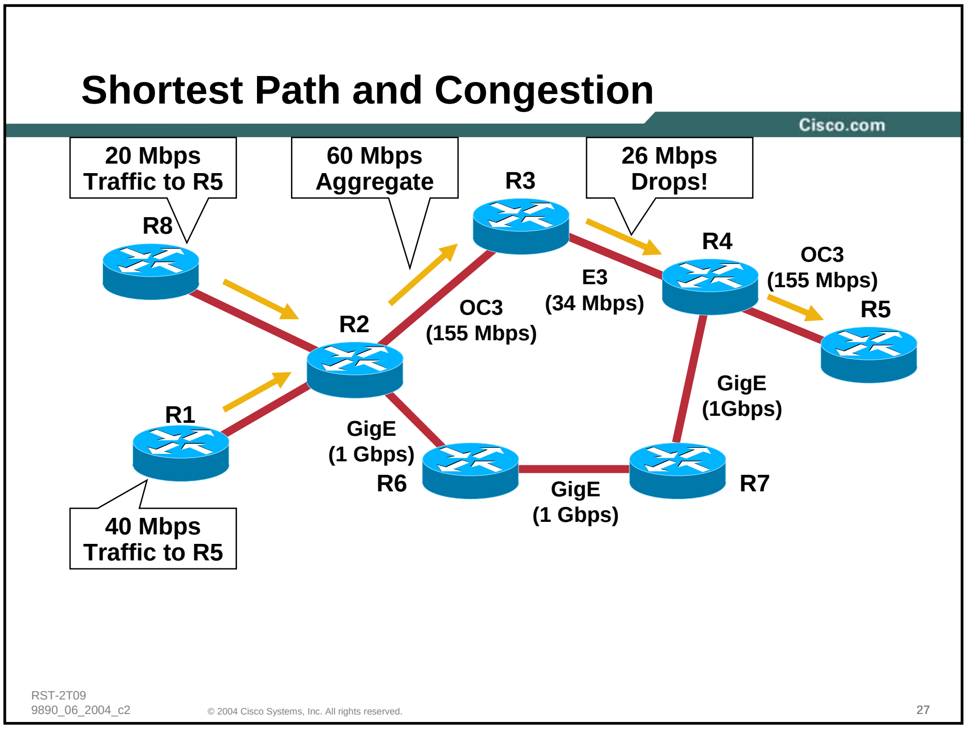
Motivation for Traffic Engineering

- Increase efficiency of bandwidth resources
  - Prevent over-utilized (congested) links while other links are underutilized
- Ensure the most desirable/appropriate path for some/all traffic
  - Override the shortest path selected by the IGP
- Replace ATM/FR Cores
  - PVC-like traffic placement without IGP full mesh and associated O(N^2) flooding
- The ultimate goal is COST SAVING
  - Service development also progressing

The “Fish” Problem (Shortest Path)

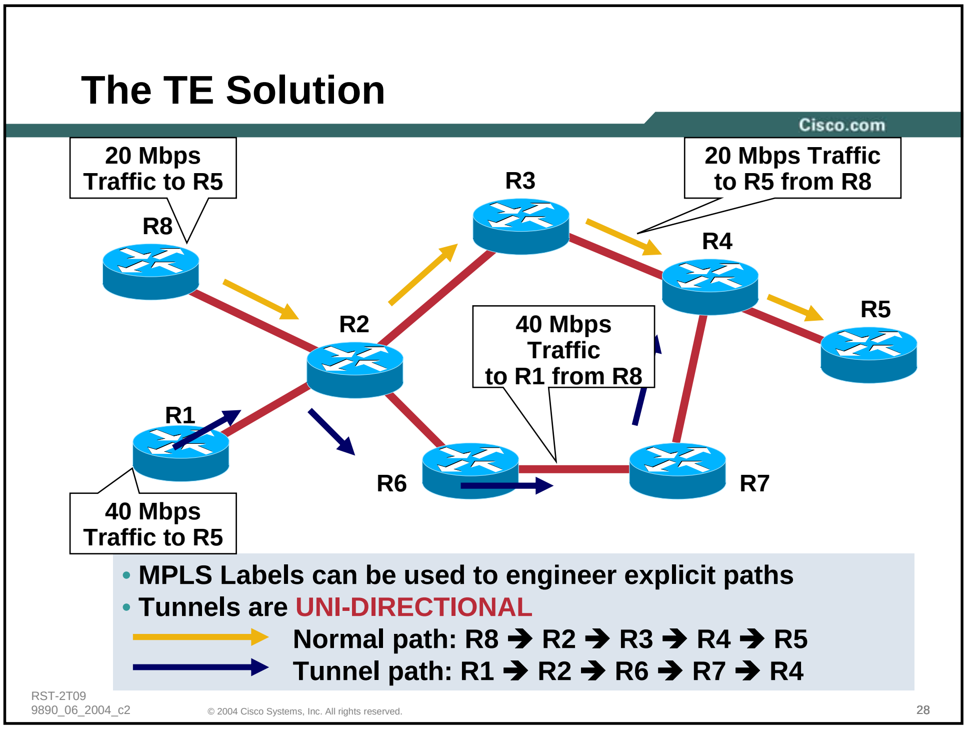


- IP uses shortest path destination-based routing
- Shortest path may not be the only path
- Alternate paths may be underutilized
- While the shortest path is over-utilized



Shortest Path and Congestion

Cisco.com
20 Mbps
Traffic to R5
60 Mbps
Aggregate
R3
26 Mbps
Drops!
R8
R4
OC3
(155 Mbps)
E3
(34 Mbps)
R5
OC3
(155 Mbps)
R2
GigE
(1Gbps)
R1
GigE
(1 Gbps)
R6
R7
GigE
(1 Gbps)
40 Mbps
Traffic to R5



The TE Solution

Cisco.com
20 Mbps
Traffic to R5
R3
20 Mbps Traffic
to R5 from R8
R8
R4
R2
40 Mbps
Traffic
to R1 from R8
R5
R1
R6
R7
40 Mbps
Traffic to R5

- MPLS Labels can be used to engineer explicit paths
- Tunnels are UNI-DIRECTIONAL

Normal path: R8 ➔ R2 ➔ R3 ➔ R4 ➔ R5

Tunnel path: R1 ➔ R2 ➔ R6 ➔ R7 ➔ R4

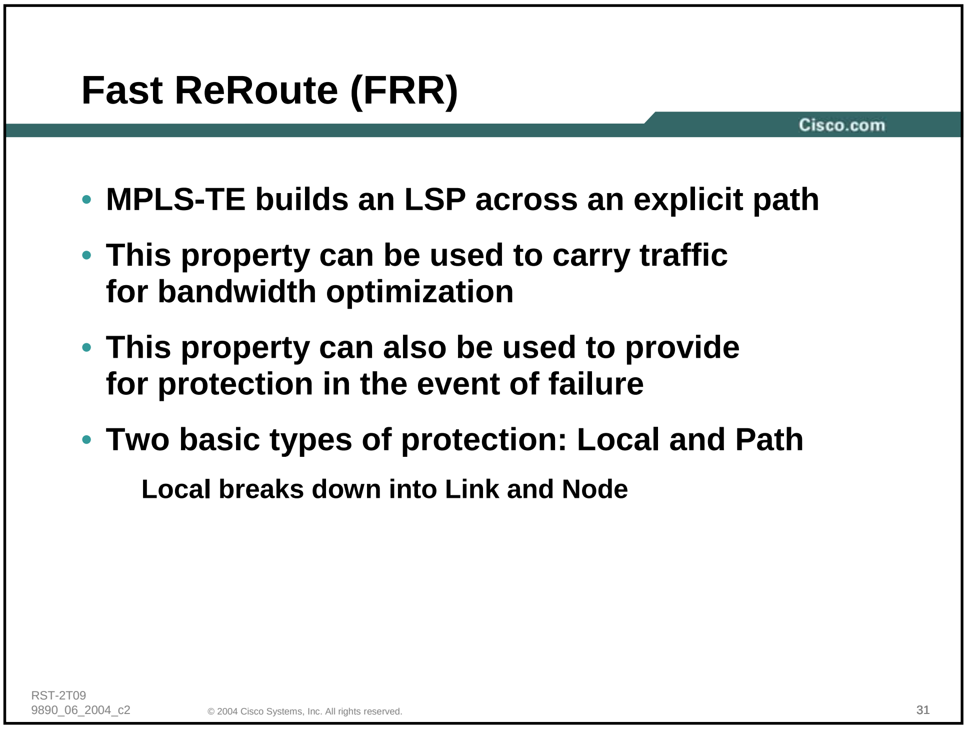
Design Approach and Scalability

- **Two methods to deploy MPLS-TE**
- **Tactical**
  - As needed to clear up congestion
  - You only have tunnels when there is a problem (and you must remember to remove them)
- **Strategic**
  - Mesh of TE tunnels between a level of routers
  - Typically P-to-P but can be PE-to-PE in smaller networks
  - N(N-1) LSPs (one in each direction)

Scalability

Code	Number of Headend Tunnels	Number of Mid-Points	Number of Tails
12.0ST	600	10,000	5,000

- In late 2000/early 2001, we said, “how many tunnels come up in 5 minutes?”
- Answer as above
- With latest code, above converges in 2–3 minutes
- Haven’t needed to formalize larger number testing
  - Largest customer we know of has a network about ¼ this size
- Bottom line: MPLS-TE scalability is not the gating factor in scaling your network

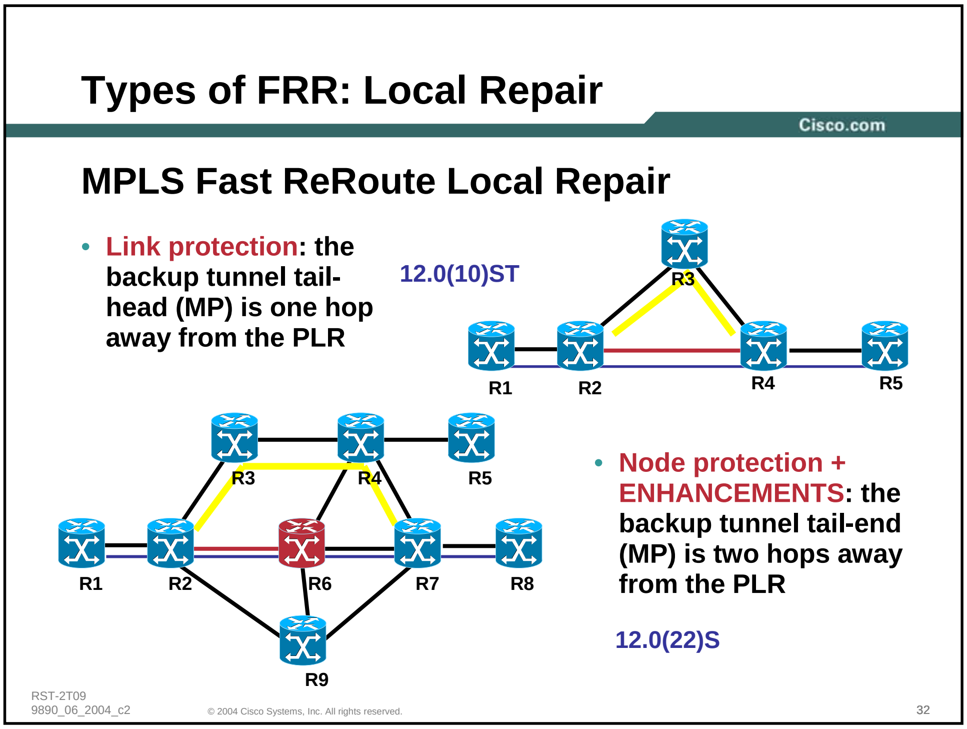


Fast ReRoute (FRR)

Cisco.com

- MPLS-TE builds an LSP across an explicit path
- This property can be used to carry traffic for bandwidth optimization
- This property can also be used to provide for protection in the event of failure
- Two basic types of protection: Local and Path
  - Local breaks down into Link and Node

RST-2T09
9890_06_2004_c2



Types of FRR: Local Repair

Cisco.com

MPLS Fast ReRoute Local Repair

- **Link protection**: the backup tunnel tail-head (MP) is one hop away from the PLR

12.0(10)ST

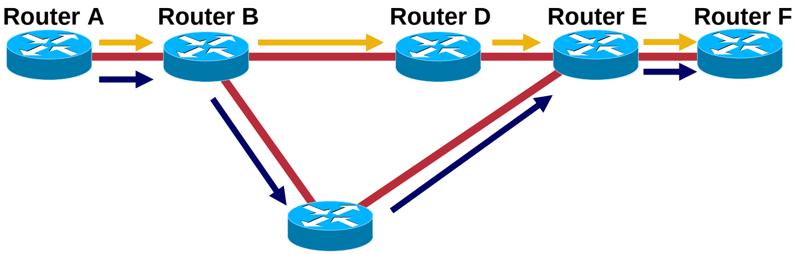
- **Node protection + ENHANCEMENTS**: the backup tunnel tail-end (MP) is two hops away from the PLR

12.0(22)S

RST-2T09
9890_06_2004_c2

Types of FRR: Path Protection

- Link, Node protection: pre-establish 1 TE tunnel per {NHop/NNHop}
- Path protection: pre-establish 2 LSPs for 1 TE tunnel, taking two diverse paths



Summary: MPLS TE

- Helps optimize network utilization (strategic)
- Assists in handling unexpected congestion (tactical)
- Provides fast reroute for link and node failures
- TE is only part of a method of guaranteeing bandwidth
  - It is a control plane mechanism only
  - Must be used with traditional QoS mechanisms

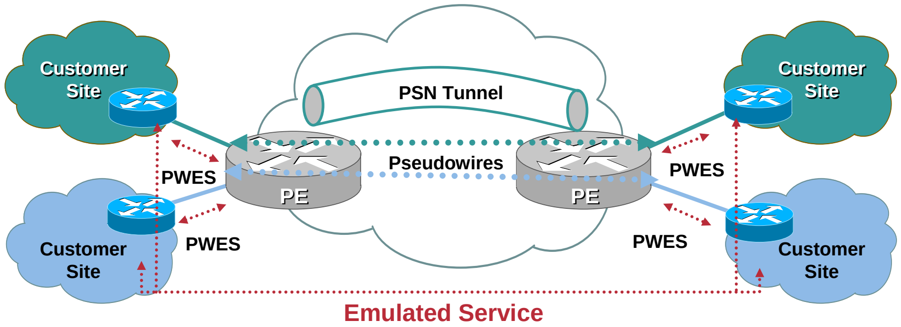
Agenda

- MPLS Basics
- MPLS L3 VPNs
- MPLS TE
- AToM

Motivation for AToM

- Similar to existing circuit switched environment
- Leverage the existing installed gear
- Provide circuit-based services in addition to packet/IP-based services
- Transparent trunking of customer IGP
- Provide any-to-any connectivity
- Trunking Layer 2 over an MPLS network:
    - Ethernet
    - Frame Relay
    - ATM—AAL5, cell mode
    - PPP
    - Cisco HDLC
    - SONET

Pseudowire Reference Model


Customer Site
Customer Site
PSN Tunnel
Pseudowires
PE
PE
PWES
PWES
PWES
PWES
Customer Site
Customer Site
Emulated Service

A Pseudowire (PW) Is a Connection Between Two Provider Edge (PE) Devices which Connects Two Pseudowire End-Services (PWESs) of the Same Type

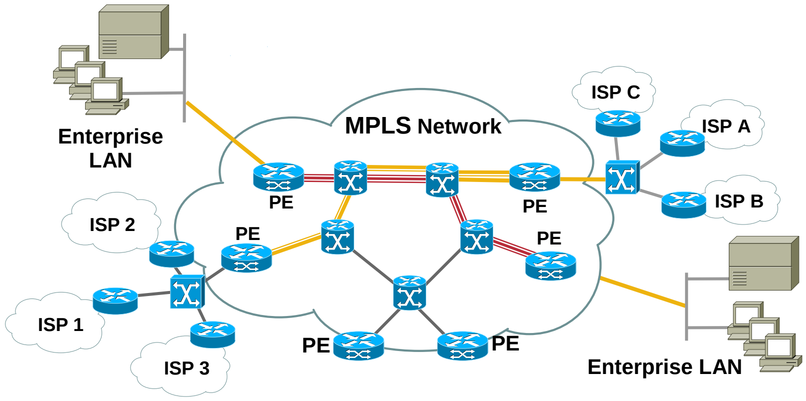
Service Types:

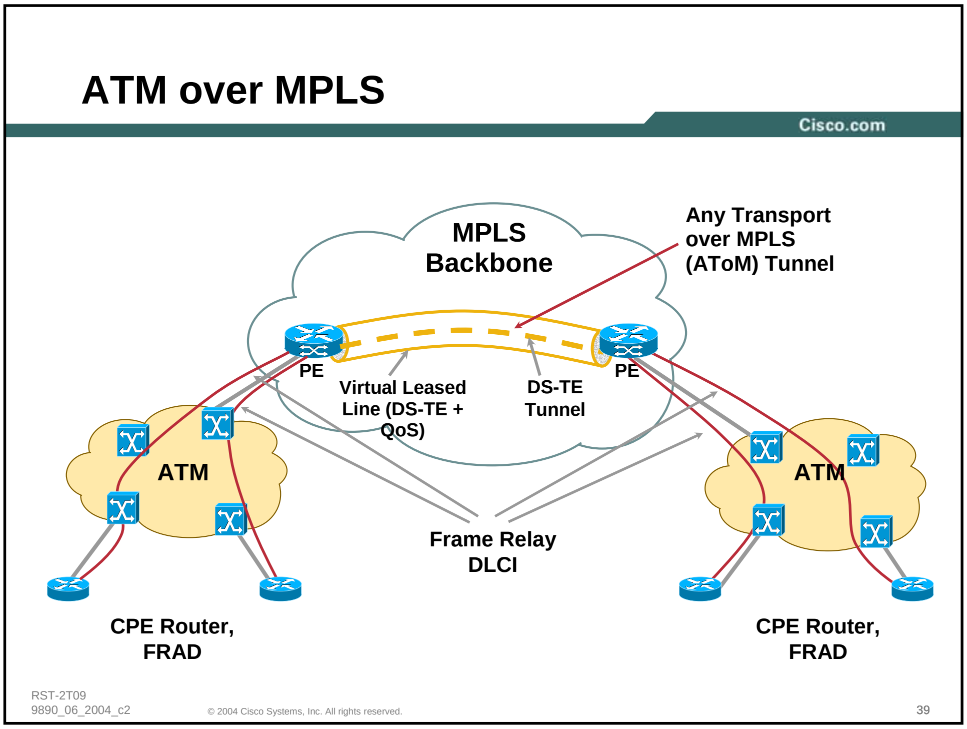
- Ethernet
- 802.1Q (VLAN)
- ATM VC or VP

PWES

- HDLC
- PPP
- Frame Relay VC

Ethernet over MPLS


Enterprise LAN
MPLS Network
ISP C
ISP A
ISP B
ISP 2
ISP 1
ISP 3
PE
PE
PE
PE
PE
PE
Enterprise LAN



ATM over MPLS

Cisco.com

MPLS Backbone

Any Transport over MPLS (AToM) Tunnel

PE

PE

Virtual Leased Line (DS-TE + QoS)

DS-TE Tunnel

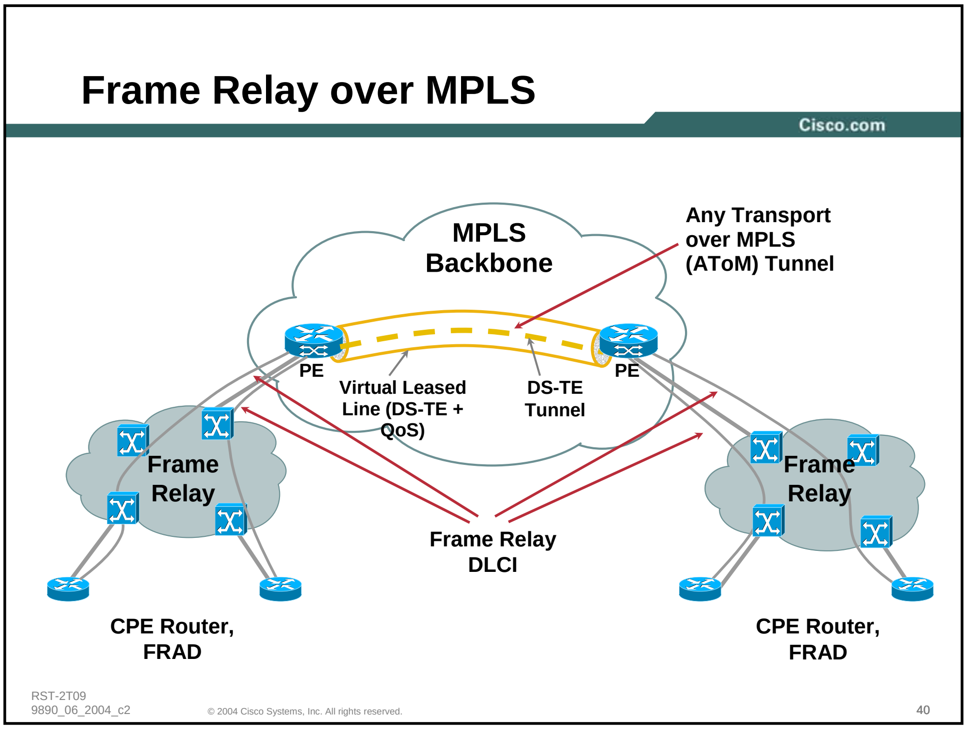
ATM

ATM

Frame Relay DLCI

CPE Router, FRAD

CPE Router, FRAD



Frame Relay over MPLS

Cisco.com

MPLS Backbone

Any Transport over MPLS (AToM) Tunnel

PE

PE

Virtual Leased Line (DS-TE + QoS)

DS-TE Tunnel

Frame Relay

Frame Relay

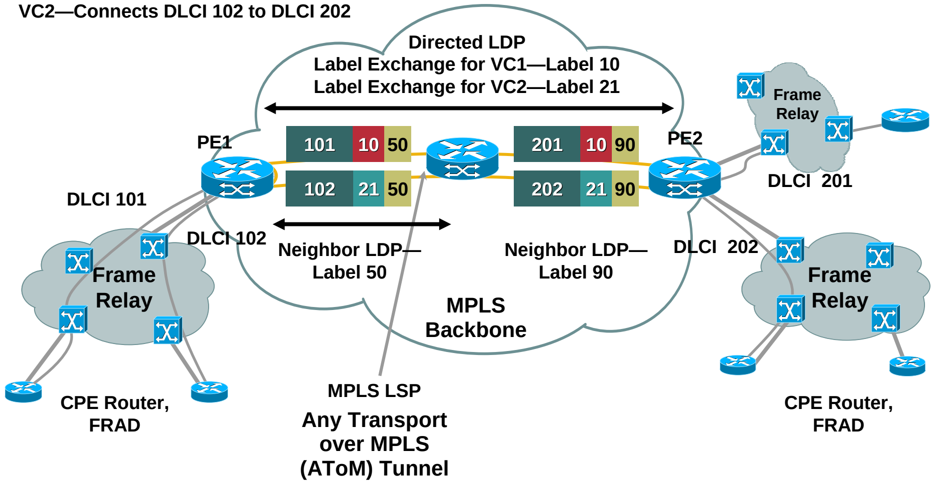
Frame Relay DLCI

CPE Router, FRAD

CPE Router, FRAD

Frame Relay over MPLS: Example

VC1—Connects DLCI 101 to DLCI 201
VC2—Connects DLCI 102 to DLCI 202



Directed LDP
Label Exchange for VC1—Label 10
Label Exchange for VC2—Label 21

PE1

PE2

101	10	50
102	21	50

201	10	90
202	21	90

DLCI 101

DLCI 102

DLCI 201

DLCI 202

Frame Relay

Frame Relay

Frame Relay

Neighbor LDP—
Label 50

Neighbor LDP—
Label 90

MPLS
Backbone

MPLS LSP

Any Transport
over MPLS
(AToM) Tunnel

CPE Router,
FRAD

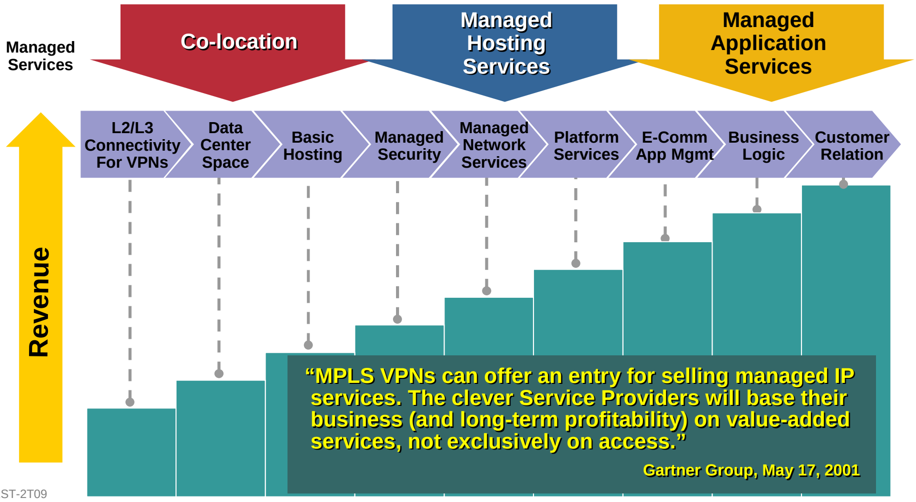
CPE Router,
FRAD

CISCO SYSTEMS

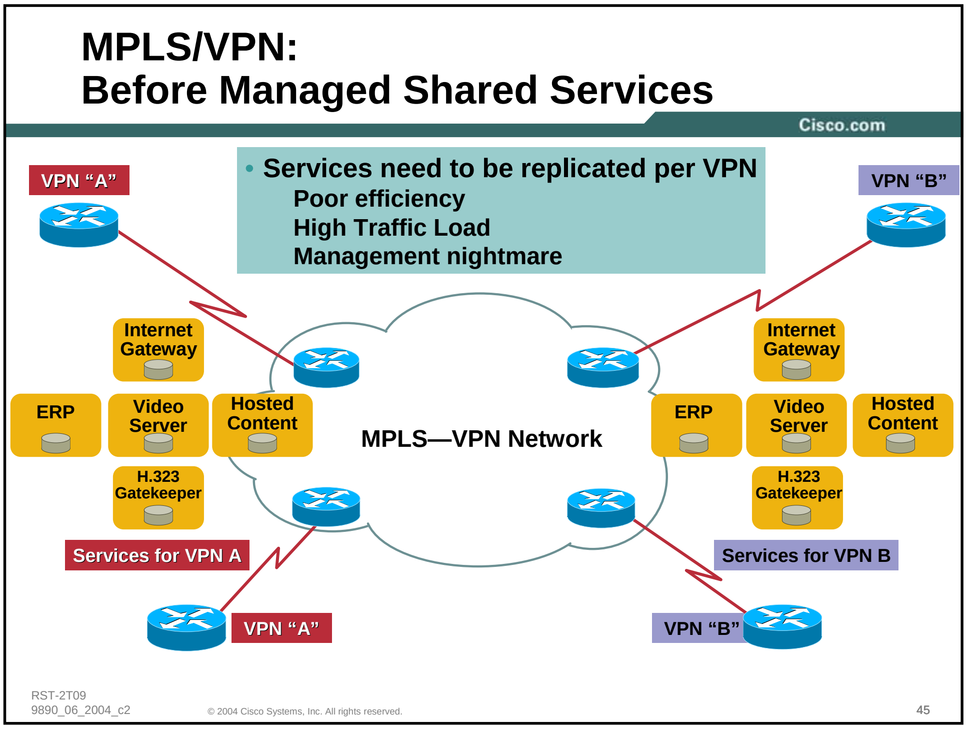
MPLS AWARE IP SERVICES

SP Managed Service Offerings

The Key Is Moving Up the Value Chain by Providing New Services


Managed Services
Co-location
Managed Hosting Services
Managed Application Services
L2/L3 Connectivity For VPNs
Data Center Space
Basic Hosting
Managed Security
Managed Network Services
Platform Services
E-Comm App Mgmt
Business Logic
Customer Relation
Revenue

> "MPLS VPNs can offer an entry for selling managed IP services. The clever Service Providers will base their business (and long-term profitability) on value-added services, not exclusively on access."
>
> Gartner Group, May 17, 2001



MPLS/VPN: Before Managed Shared Services

Cisco.com

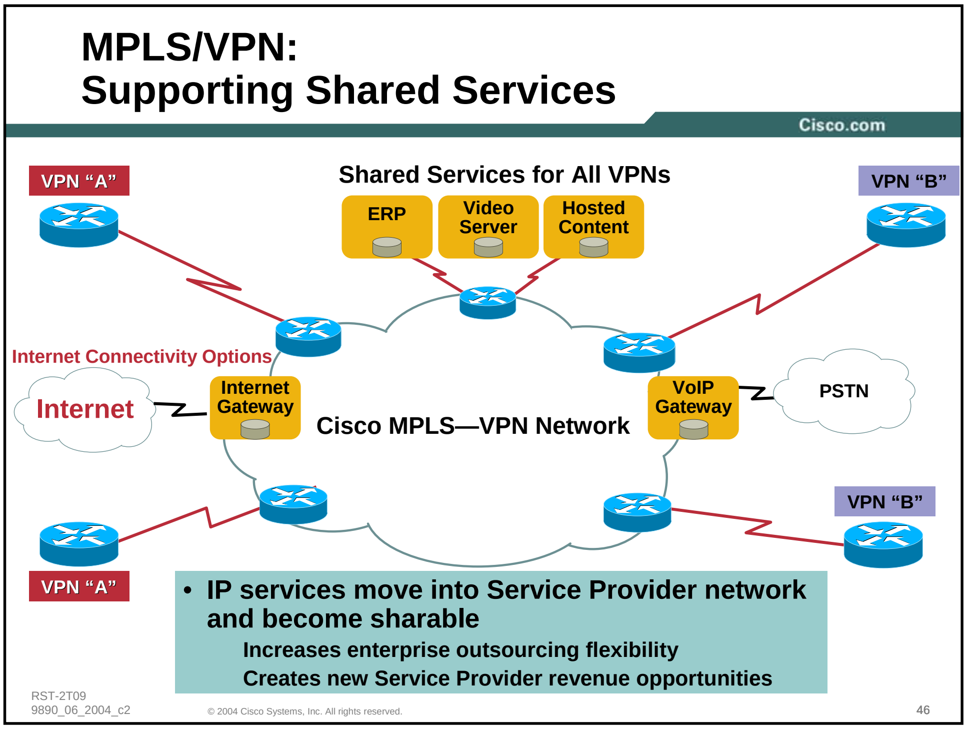
- Services need to be replicated per VPN
  - Poor efficiency
  - High Traffic Load
  - Management nightmare

VPN "A"
VPN "B"
Internet Gateway
ERP
Video Server
Hosted Content
H.323 Gatekeeper
MPLS—VPN Network
Internet Gateway
ERP
Video Server
Hosted Content
H.323 Gatekeeper
Services for VPN A
Services for VPN B
VPN "A"
VPN "B"

RST-2T09
9890_06_2004_c2

© 2004 Cisco Systems, Inc. All rights reserved.

45



MPLS/VPN: Supporting Shared Services

Cisco.com

VPN "A"
Shared Services for All VPNs
VPN "B"
ERP
Video Server
Hosted Content
Internet Connectivity Options
Internet
Internet Gateway
Cisco MPLS—VPN Network
VoIP Gateway
PSTN
VPN "B"
VPN "A"

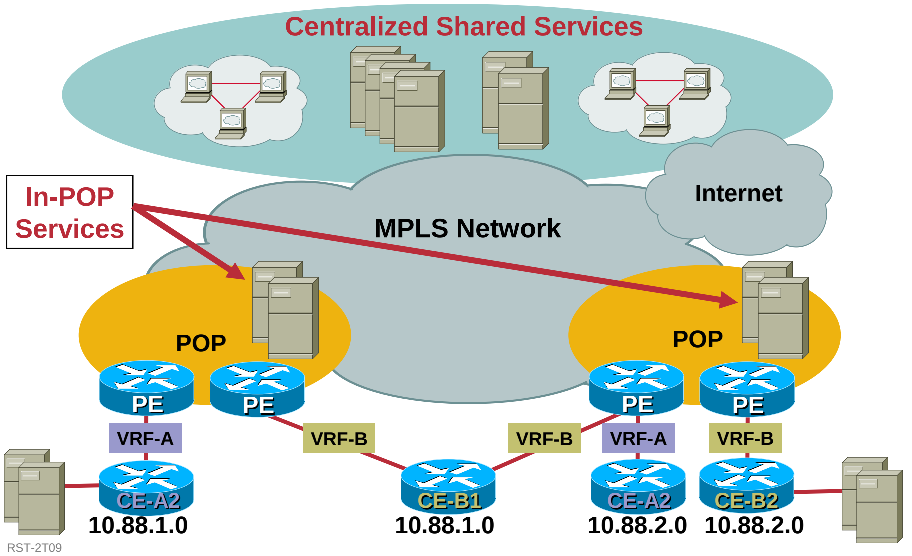
- IP services move into Service Provider network and become sharable
  - Increases enterprise outsourcing flexibility
  - Creates new Service Provider revenue opportunities

RST-2T09
9890_06_2004_c2

© 2004 Cisco Systems, Inc. All rights reserved.

46

Flexible MPLS-VPN Shared Services Model


Centralized Shared Services
In-POP Services
Internet
MPLS Network
POP
POP
PE
PE
PE
PE
VRF-A
VRF-B
VRF-B
VRF-A
VRF-B
CE-A2
CE-B1
CE-A2
CE-B2
10.88.1.0
10.88.1.0
10.88.2.0
10.88.2.0

Agenda

- Managed IP Services
- Managed Security Services
- SP Edge Redundancy
- Shared Management

IP Addressing

- Dynamic Host Configuration Protocol (DHCP) relay for MPLS VPNs
- On Demand Address Pools (ODAP) for MPLS VPNs
- Network Address Translation (NAT)

VRF-AWARE DHCP

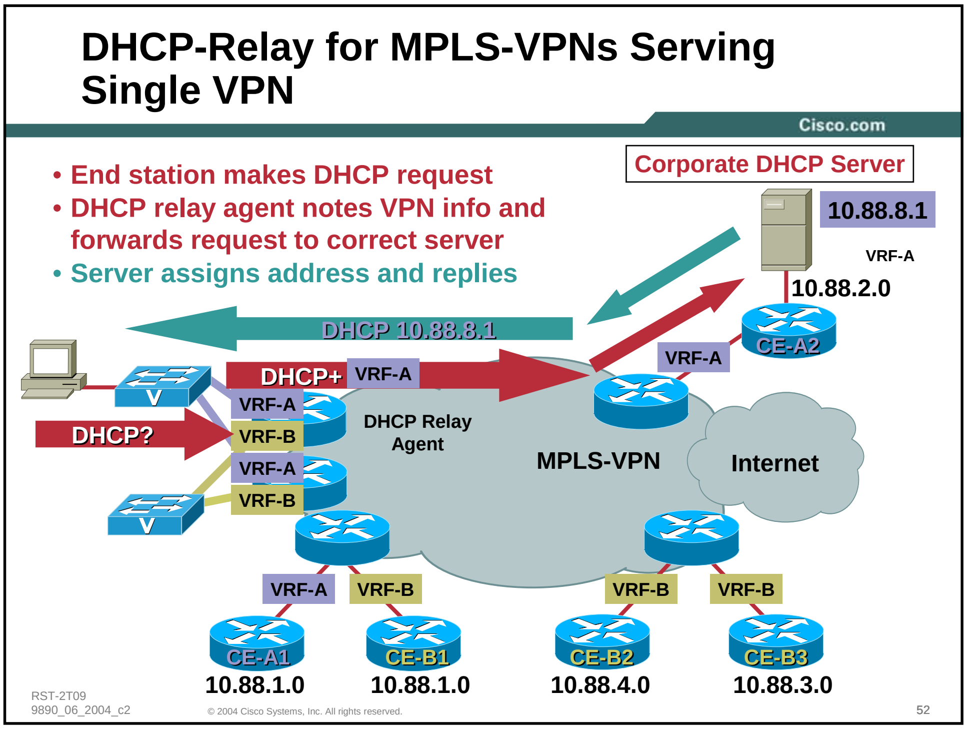
Why DHCP-Relay for MPLS VPNs?

Cisco.com

- DHCP Server can't distinguish requests from hosts belonging to different VRFs
- Assign IP addresses from a shared DHCP server
- Need to be able to assign subnets per VRF
- It's required for a DCHP server to include VPN information DHCP requests/reply
- DHCP-relay uses the VPN identifier sub-option
- The VPN identifier (sub option) also allows any DHCP reply to be properly forwarded back to the relay agent
- VRF/VPNID support in V5.5 CNR
- Cisco IOS® supports DHCP-relay as well as a full router-based DHCP server

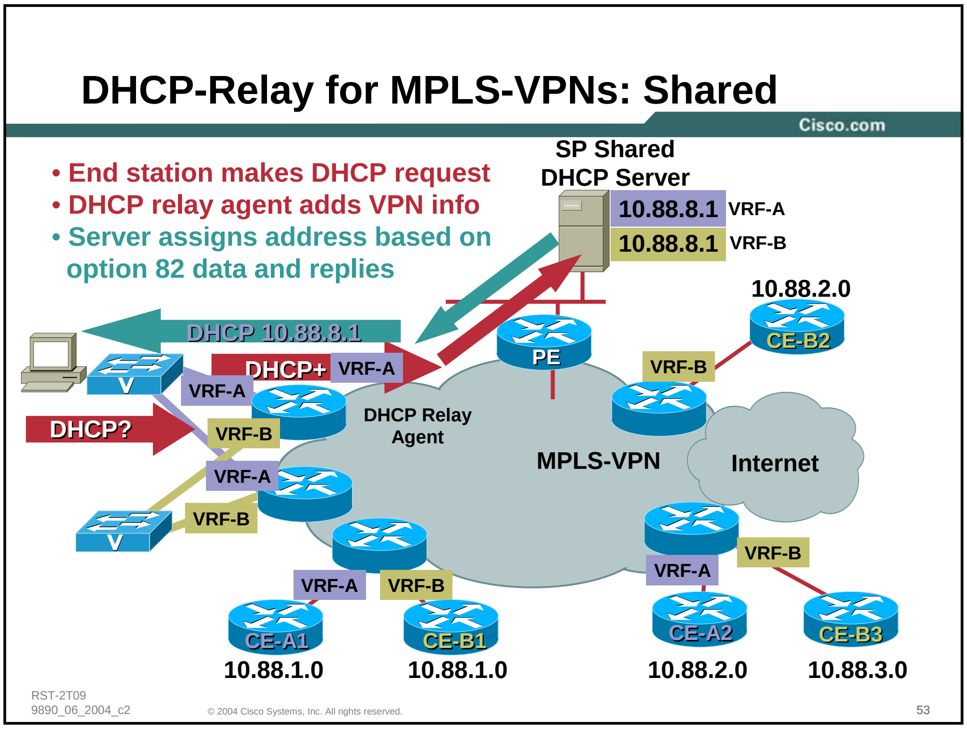
DHCP-Relay for MPLS-VPNs Serving Single VPN

Cisco.com



- End station makes DHCP request
- DHCP relay agent notes VPN info and forwards request to correct server
- Server assigns address and replies

Corporate DHCP Server
10.88.8.1
VRF-A
10.88.2.0
DHCP 10.88.8.1
CE-A2
VRF-A
DHCP+
VRF-A
VRF-A
VRF-B
DHCP?
VRF-A
VRF-B
V
V
DHCP Relay Agent
MPLS-VPN
Internet
VRF-A
VRF-B
VRF-B
VRF-B
CE-A1
CE-B1
CE-B2
CE-B3
10.88.1.0
10.88.1.0
10.88.4.0
10.88.3.0



DHCP-Relay for MPLS-VPNs: Shared

- End station makes DHCP request
- DHCP relay agent adds VPN info
- Server assigns address based on option 82 data and replies

SP Shared
DHCP Server
10.88.8.1 VRF-A
10.88.8.1 VRF-B
10.88.2.0
CE-B2
DHCP 10.88.8.1
DHCP+ VRF-A
PE
VRF-B
V
VRF-A
DHCP?
VRF-B
DHCP Relay
Agent
MPLS-VPN
Internet
VRF-A
VRF-B
V
VRF-B
VRF-A
VRF-A
VRF-B
CE-A1
CE-B1
CE-A2
CE-B3
10.88.1.0
10.88.1.0
10.88.2.0
10.88.3.0

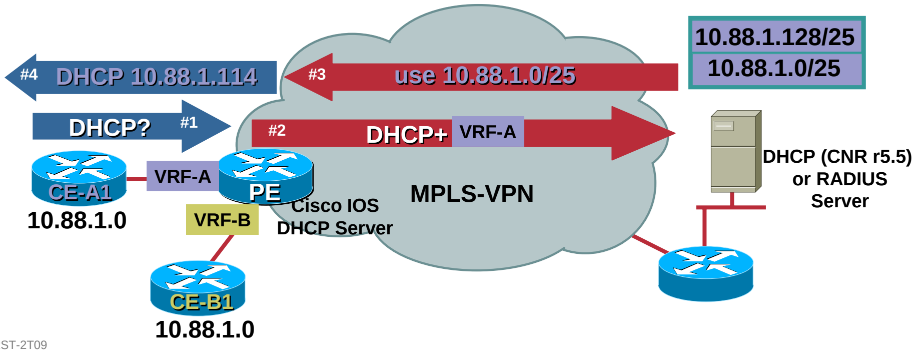
VRF-AWARE ODAP

On-Demand Address Pools (ODAP)

- Dynamically creates and associates IP address pool to an interface
- Addresses are assigned automatically with DHCP (Option 82 sub options)
- Support for DHCP clients and PPP sessions on per interface basis
- Each ODAP is configured and associated with a particular MPLS VPN
- The VPN identifier allows replies to be properly forwarded back to the relay agent
- Works with Cisco Network Registrar (CNR) 5.5 (DHCP) and/or Access Registrar 1.7 (RADIUS)

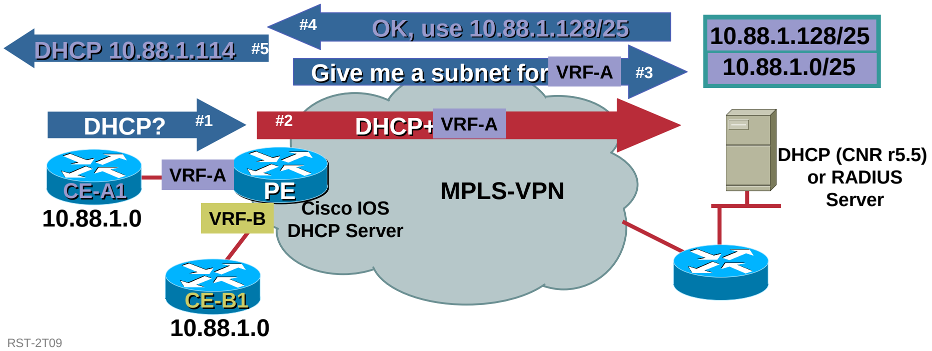
ODAP for MPLS-VPNs: Provisioning and Startup

- PE router is configured for ODAP
- ODAP requests initial pool for VRF-A from server
- CE router is installed and PPP link established to PE router
- CE router uses DHCP proxy to obtain addresses for downstream devices
- PE adds subnet routing information to VRF


#4 DHCP 10.88.1.114
#3 use 10.88.1.0/25
10.88.1.128/25
10.88.1.0/25
DHCP? #1
#2 DHCP+ VRF-A
CE-A1
VRF-A
PE
10.88.1.0
VRF-B
Cisco IOS DHCP Server
MPLS-VPN
DHCP (CNR r5.5) or RADIUS Server
CE-B1
10.88.1.0

ODAP for MPLS VPNs: Address Pool Management

- ODAP requests initial pool for VRF-A from server
- End station makes DHCP request
- DHCP server fulfills request from pool—reaches 90%
- ODAP pool manager requests expansion
- Server allocates another subnet and replies
- PE adds subnet routing information to VRF


#4
OK, use 10.88.1.128/25
10.88.1.128/25
10.88.1.0/25
DHCP 10.88.1.114
#5
Give me a subnet for VRF-A
#3
DHCP?
#1
#2
DHCP+ VRF-A
CE-A1
10.88.1.0
VRF-A
PE
VRF-B
Cisco IOS DHCP Server
MPLS-VPN
DHCP (CNR r5.5) or RADIUS Server
CE-B1
10.88.1.0

MPLS VPN ODAP Configuration

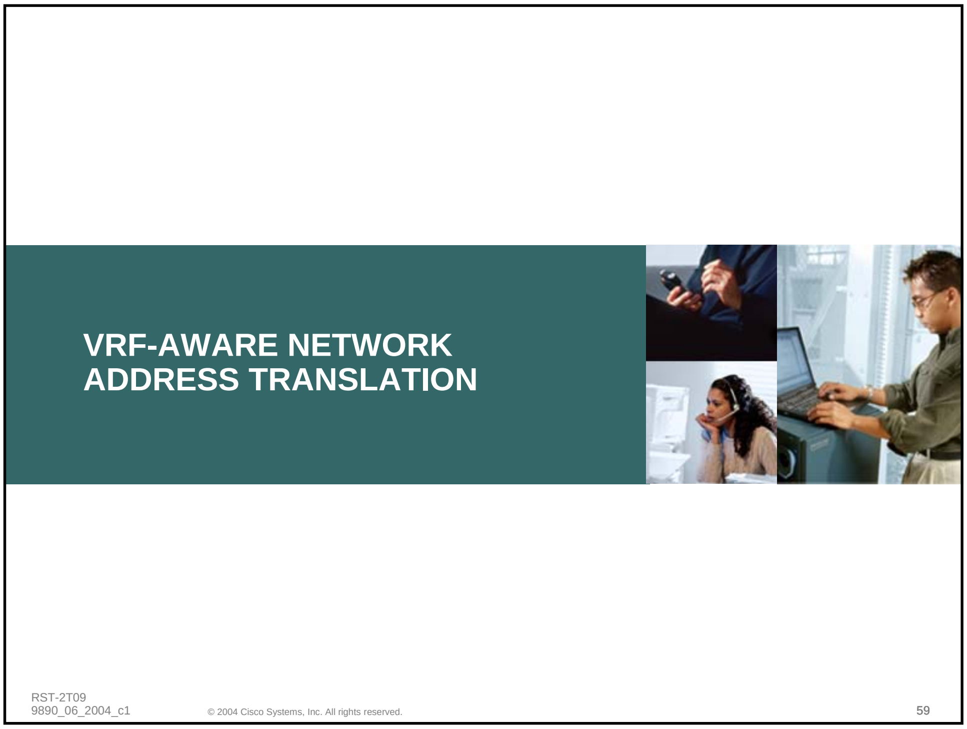
- Configuration

  Set initial pool size

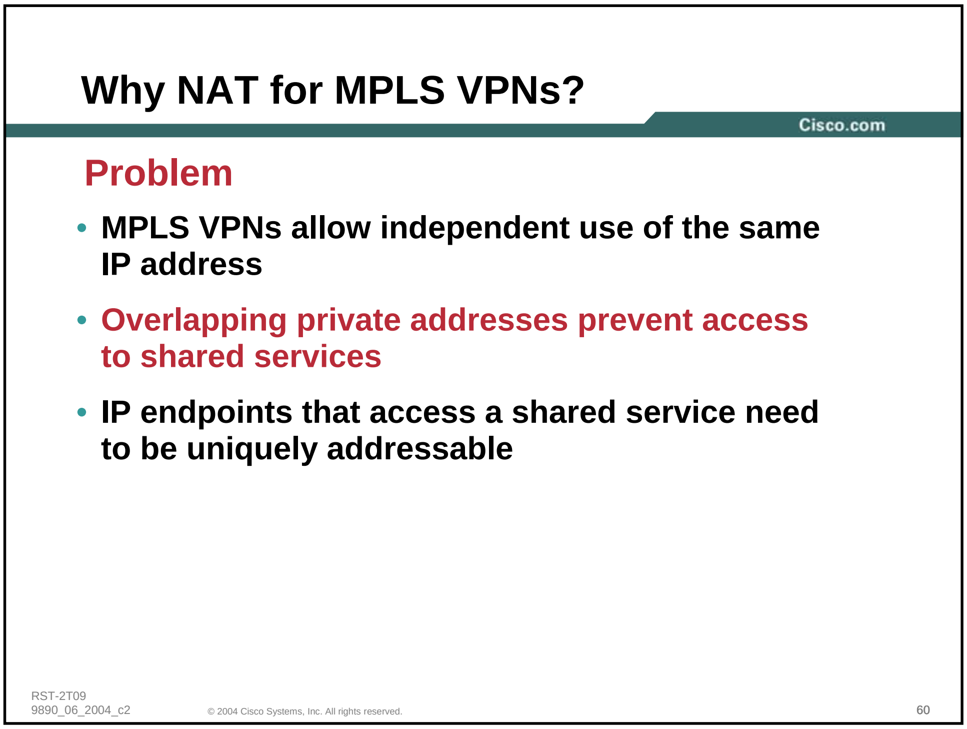
  Expansion/Contraction increment

  High/Low utilization mark (% of pool)

```
ip dhcp pool green_pool
 vrf Green
 utilization mark high 60
 utilization mark low 40
 origin dhcp subnet size initial /24
autogrow /24
```



VRF-AWARE NETWORK ADDRESS TRANSLATION



Why NAT for MPLS VPNs?

Problem

- MPLS VPNs allow independent use of the same IP address
- Overlapping private addresses prevent access to shared services
- IP endpoints that access a shared service need to be uniquely addressable

Why NAT for MPLS VPNs? (Cont.)

Solution

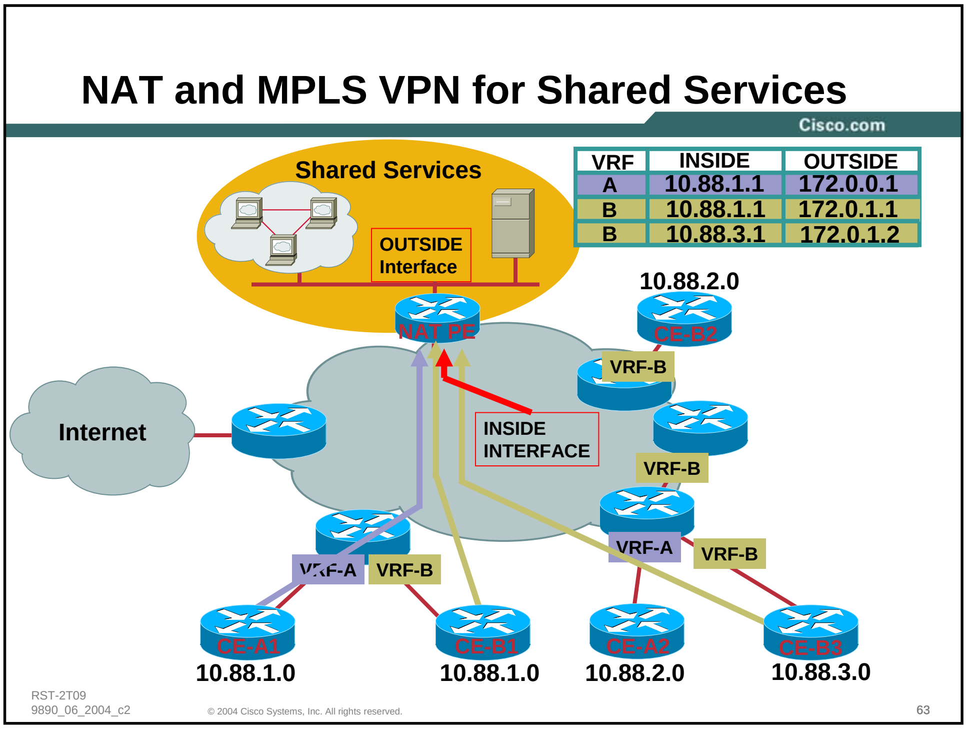
- Overlapping addresses are distinguished by their associated VPN ID
- VRF-Aware NAT creates different translations per VPN by including VRF info in the address translation entries
- VRF-Aware NAT allows access to shared services from multiple MPLS VPN customers, even though IP addresses overlap

Outcome

- ISPs can offer centralized address translation services

NAT and MPLS VPN Integration

- Have concept of outside/inside interfaces in NAT
- Outside/inside interfaces could be any type of interface (VRF, IP, MPLS)
- NAT will inspect all traffic routed VRF-to-VRF or VRF-to-Global
- NAT can be configured on 1 or more PE's for redundancy
    - The 'shared service' does not need to be physically connected to the PE device performing NAT
    - NAT Pools must be unique
- All native NAT applications are supported
- NAT command changes

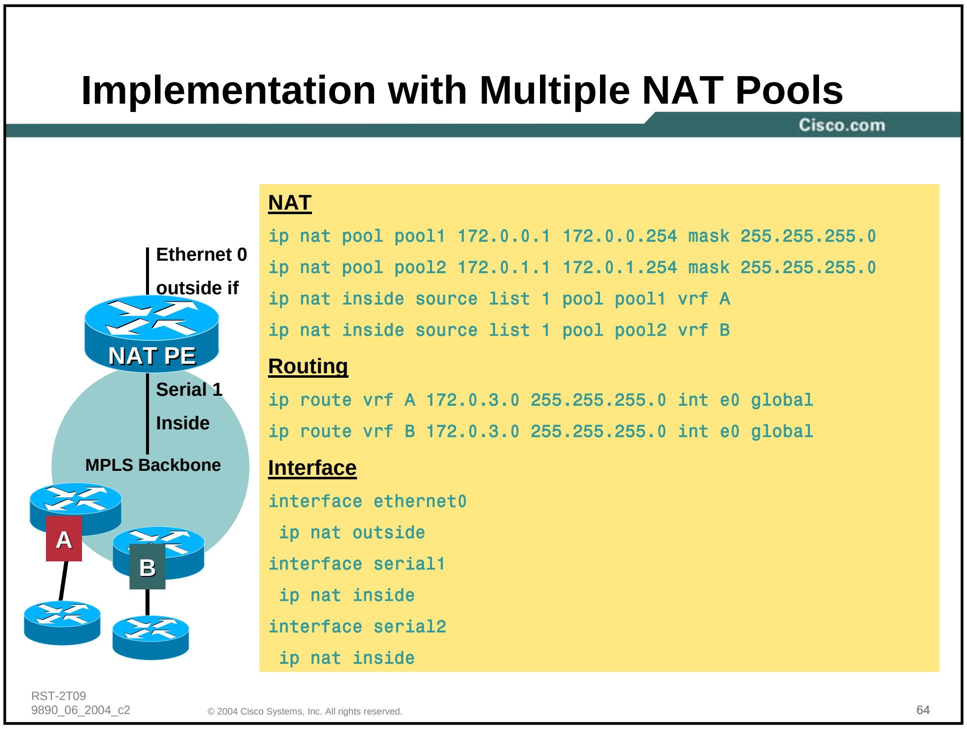


NAT and MPLS VPN for Shared Services

Cisco.com

VRF	INSIDE	OUTSIDE
A	10.88.1.1	172.0.0.1
B	10.88.1.1	172.0.1.1
B	10.88.3.1	172.0.1.2

Shared Services
OUTSIDE Interface
NAT PE
10.88.2.0
CE-B2
VRF-B
Internet
INSIDE INTERFACE
VRF-B
VRF-A
VRF-B
VRF-A
VRF-B
CE-A1
CE-B1
CE-A2
CE-B3
10.88.1.0
10.88.1.0
10.88.2.0
10.88.3.0



Implementation with Multiple NAT Pools

Cisco.com

Ethernet 0
outside if
NAT PE
Serial 1
Inside
MPLS Backbone
A
B

NAT

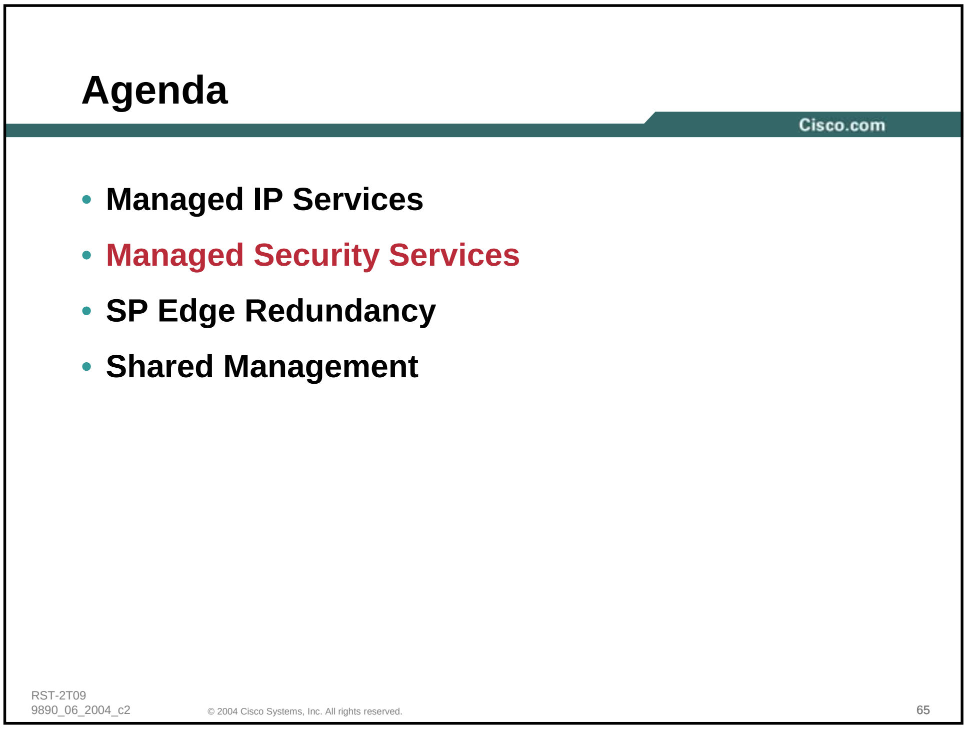
```
ip nat pool pool1 172.0.0.1 172.0.0.254 mask 255.255.255.0
ip nat pool pool2 172.0.1.1 172.0.1.254 mask 255.255.255.0
ip nat inside source list 1 pool pool1 vrf A
ip nat inside source list 1 pool pool2 vrf B
```

Routing

```
ip route vrf A 172.0.3.0 255.255.255.0 int e0 global
ip route vrf B 172.0.3.0 255.255.255.0 int e0 global
```

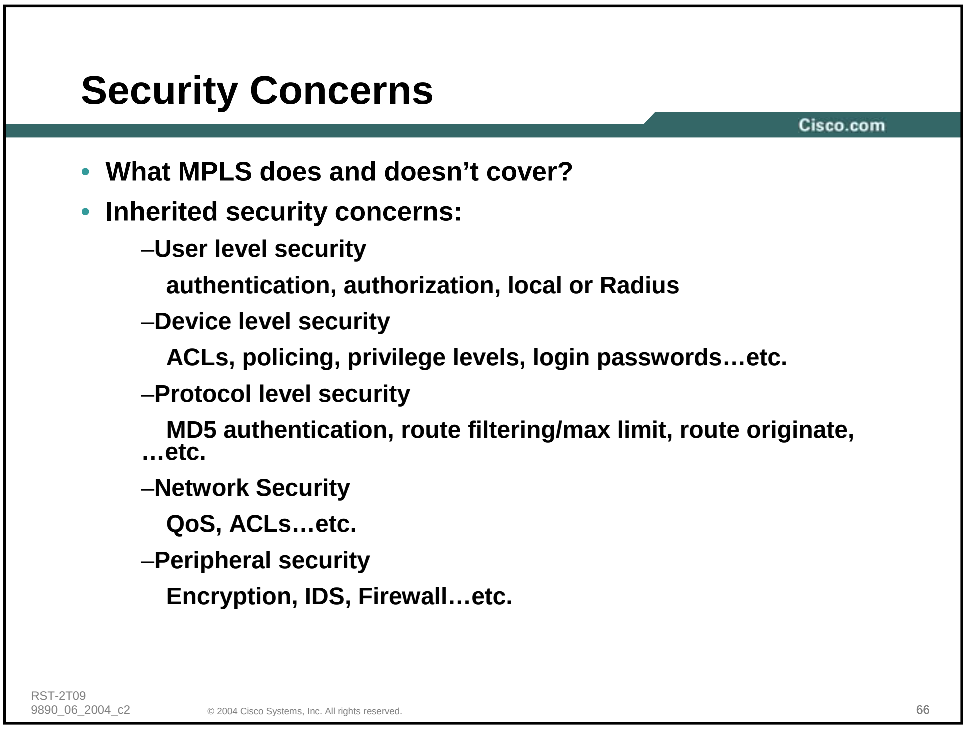
Interface

```
interface ethernet0
 ip nat outside
interface serial1
 ip nat inside
interface serial2
 ip nat inside
```



Agenda

- Managed IP Services
- Managed Security Services
- SP Edge Redundancy
- Shared Management



Security Concerns

- What MPLS does and doesn't cover?
- Inherited security concerns:
  - –User level security
    authentication, authorization, local or Radius
  - –Device level security
    ACLs, policing, privilege levels, login passwords…etc.
  - –Protocol level security
    MD5 authentication, route filtering/max limit, route originate, …etc.
  - –Network Security
    QoS, ACLs…etc.
  - –Peripheral security
    Encryption, IDS, Firewall…etc.

Security Services for MPLS VPNs

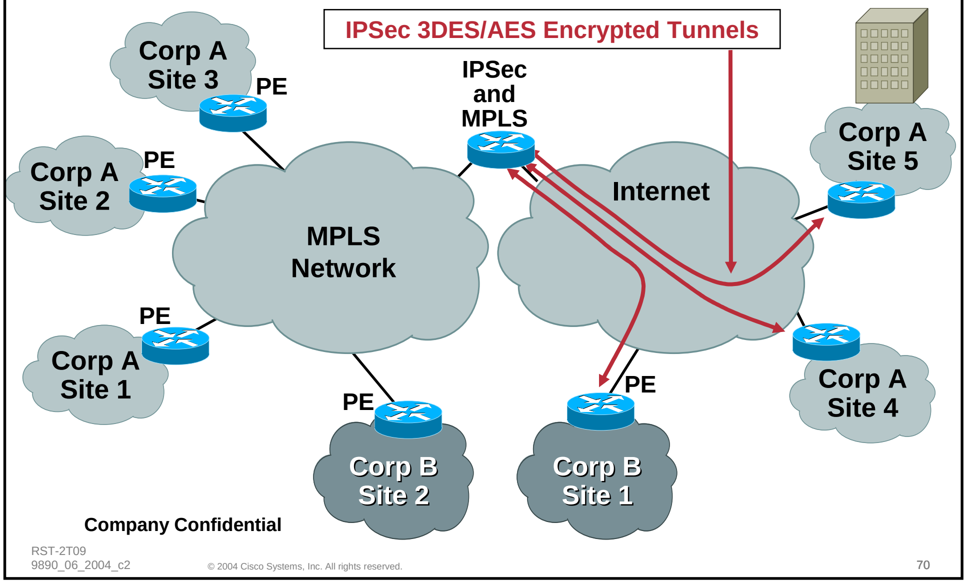
- VRF aware IPSec
- VRF aware Cisco IOS firewall

VRF-AWARE IPSEC

Why VRF-Aware IPSec?

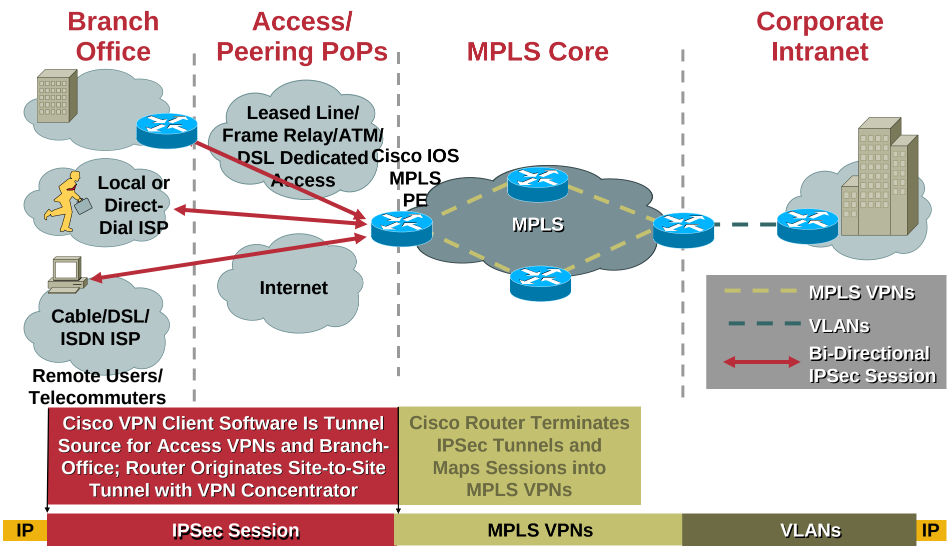
- Enterprises are looking to expand their IPSec VPNs to geographically separate locations for internal or outsourced services
- Reduces two box solution to one box solution
- Provide additional security to MPLS VPN traffic
    1. Protect critical data
    2. Selected VPN sites that might be crossing multiple Service Providers
    3. Support off-net remote access over the Internet
        - -Site to site
        - -Broadband user connections
        - -Dial-In, mobile user connections

IPSec Off-Net Service for Multiple MPLS VPNs


IPSec 3DES/AES Encrypted Tunnels
Corp A
Site 3
PE
IPSec
and
MPLS
Corp A
Site 5
Corp A
Site 2
PE
MPLS
Network
Internet
PE
Corp A
Site 1
PE
PE
Corp A
Site 4
Corp B
Site 2
Corp B
Site 1

Company Confidential

Cisco IOS IPSec + MPLS PE Single box Solution



Branch Office

Access/ Peering PoPs

MPLS Core

Corporate Intranet

Local or Direct-Dial ISP

Cable/DSL/ ISDN ISP

Remote Users/ Telecommuters

Leased Line/ Frame Relay/ATM/ DSL Dedicated Access

Internet

Cisco IOS MPLS PE

MPLS

MPLS VPNs

VLANs

Bi-Directional IPSec Session

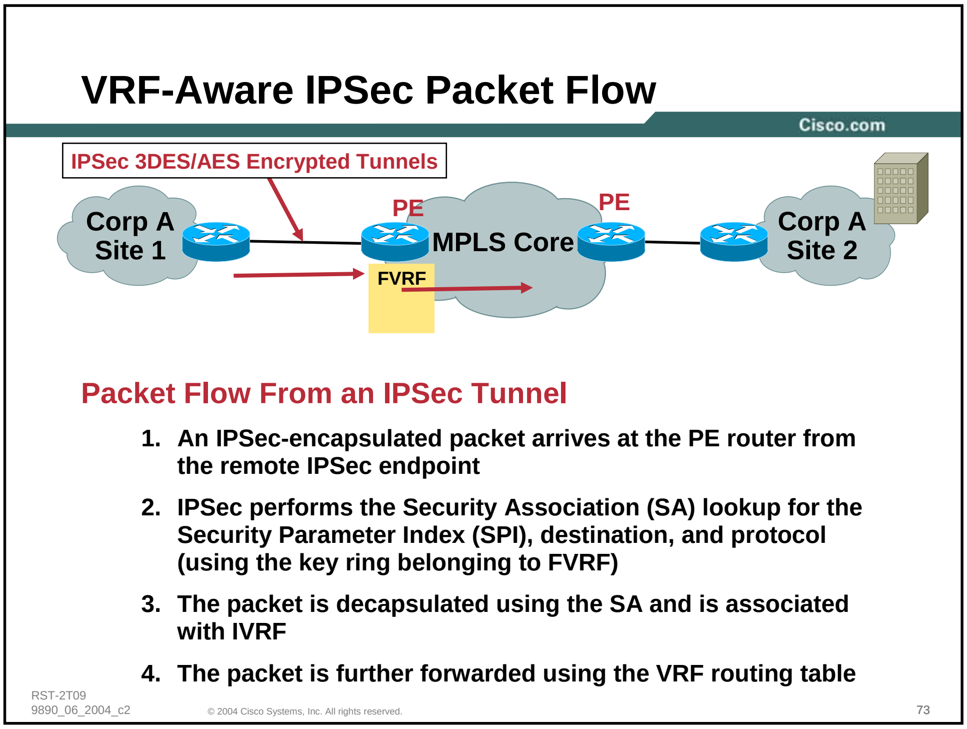
Cisco VPN Client Software Is Tunnel Source for Access VPNs and Branch-Office; Router Originates Site-to-Site Tunnel with VPN Concentrator

Cisco Router Terminates IPSec Tunnels and Maps Sessions into MPLS VPNs

IP | IPSec Session | MPLS VPNs | VLANs | IP

VRF-Aware IPSec Key Elements

- **VRF instance**
- **MPLS distribution**
- **Key rings:**
  - Are required
  - They store keys belonging to different VRFs
  - IKE exchange is authenticated if the peer key is present in the keyring belonging to the FVRF of the IKE SA
- **Front door VRF**
  - Local endpoint (or outer IKE source/destination) of the IPSec tunnel belongs to the FVRF
- **Inside VRF**
  - The source and destination addresses of the inside packet belong to the IVRF

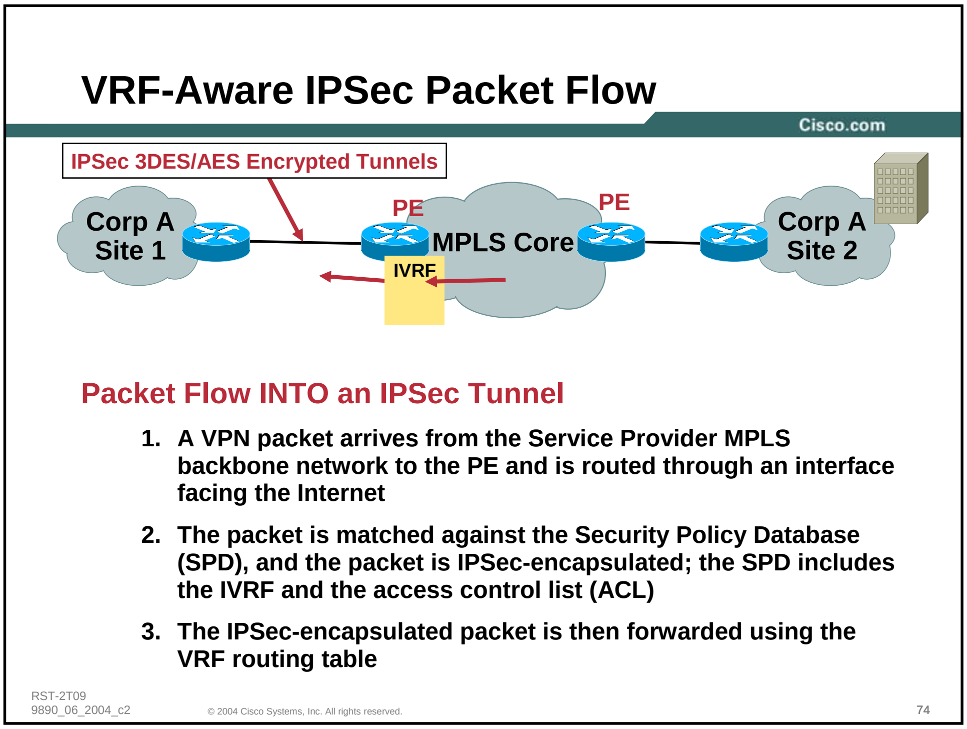


VRF-Aware IPSec Packet Flow

Cisco.com
IPSec 3DES/AES Encrypted Tunnels
Corp A
Site 1
PE
MPLS Core
PE
Corp A
Site 2
FVRF

Packet Flow From an IPSec Tunnel

1. An IPSec-encapsulated packet arrives at the PE router from the remote IPSec endpoint
2. IPSec performs the Security Association (SA) lookup for the Security Parameter Index (SPI), destination, and protocol (using the key ring belonging to FVRF)
3. The packet is decapsulated using the SA and is associated with IVRF
4. The packet is further forwarded using the VRF routing table

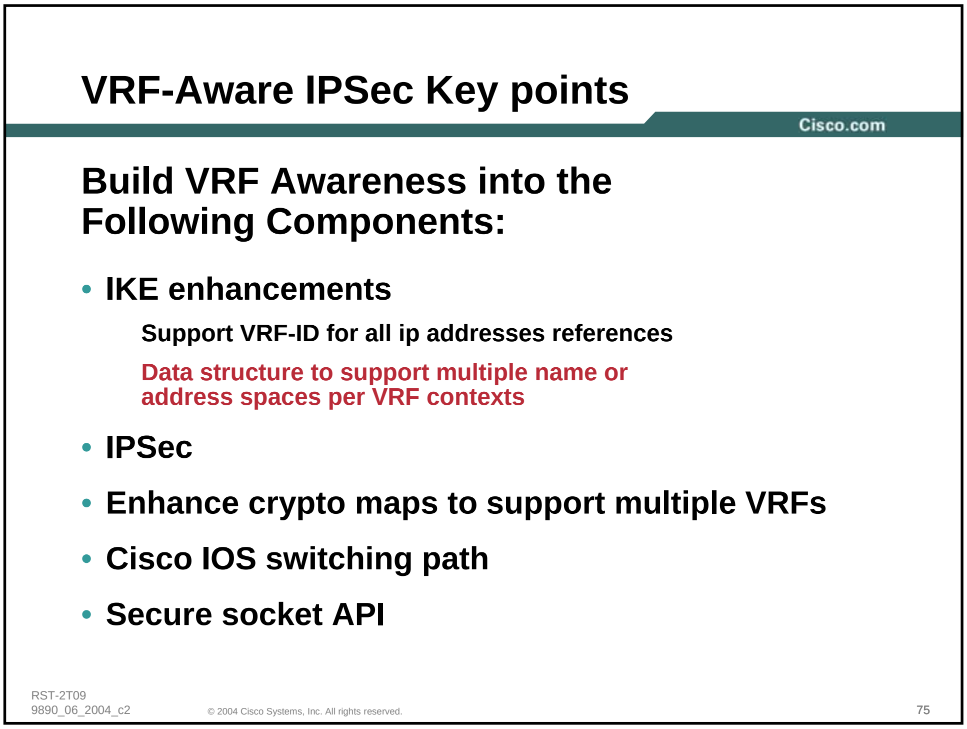


VRF-Aware IPSec Packet Flow

Cisco.com
IPSec 3DES/AES Encrypted Tunnels
Corp A
Site 1
PE
MPLS Core
PE
Corp A
Site 2
IVRF

Packet Flow INTO an IPSec Tunnel

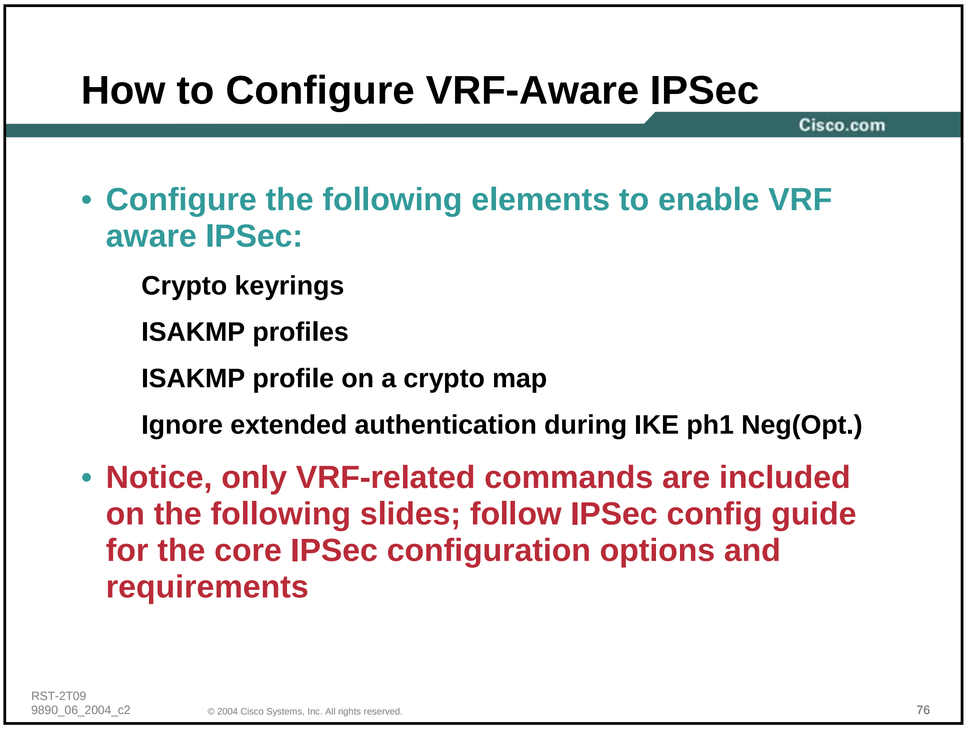
1. A VPN packet arrives from the Service Provider MPLS backbone network to the PE and is routed through an interface facing the Internet
2. The packet is matched against the Security Policy Database (SPD), and the packet is IPSec-encapsulated; the SPD includes the IVRF and the access control list (ACL)
3. The IPSec-encapsulated packet is then forwarded using the VRF routing table



VRF-Aware IPSec Key points

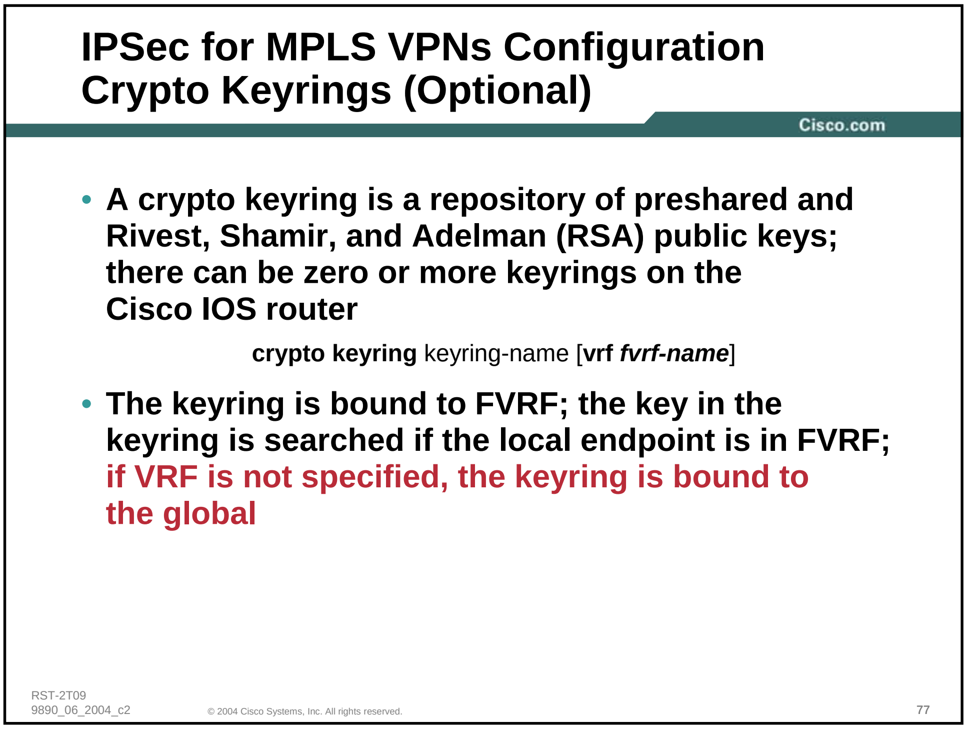
Build VRF Awareness into the Following Components:

- IKE enhancements
  - Support VRF-ID for all ip addresses references
  - Data structure to support multiple name or address spaces per VRF contexts
- IPSec
- Enhance crypto maps to support multiple VRFs
- Cisco IOS switching path
- Secure socket API



How to Configure VRF-Aware IPSec

- Configure the following elements to enable VRF aware IPSec:
  - Crypto keyrings
  - ISAKMP profiles
  - ISAKMP profile on a crypto map
  - Ignore extended authentication during IKE ph1 Neg(Opt.)
- Notice, only VRF-related commands are included on the following slides; follow IPSec config guide for the core IPSec configuration options and requirements



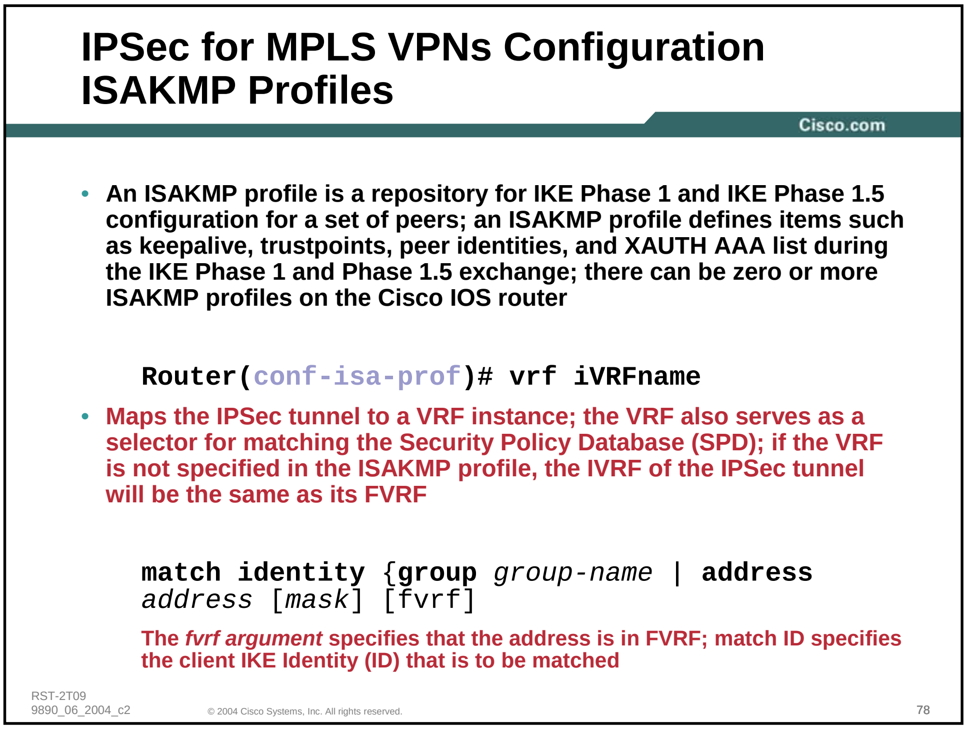
IPSec for MPLS VPNs Configuration Crypto Keyrings (Optional)

Cisco.com

- **A crypto keyring is a repository of preshared and Rivest, Shamir, and Adelman (RSA) public keys; there can be zero or more keyrings on the Cisco IOS router**

crypto keyring keyring-name [**vrf** ***fvrf-name***]

- **The keyring is bound to FVRF; the key in the keyring is searched if the local endpoint is in FVRF; if VRF is not specified, the keyring is bound to the global**



IPSec for MPLS VPNs Configuration ISAKMP Profiles

Cisco.com

- **An ISAKMP profile is a repository for IKE Phase 1 and IKE Phase 1.5 configuration for a set of peers; an ISAKMP profile defines items such as keepalive, trustpoints, peer identities, and XAUTH AAA list during the IKE Phase 1 and Phase 1.5 exchange; there can be zero or more ISAKMP profiles on the Cisco IOS router**

```
Router(conf-isa-prof)# vrf iVRFname
```

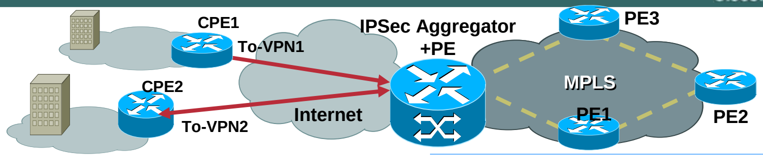
- **Maps the IPSec tunnel to a VRF instance; the VRF also serves as a selector for matching the Security Policy Database (SPD); if the VRF is not specified in the ISAKMP profile, the IVRF of the IPSec tunnel will be the same as its FVRF**

```
match identity {group group-name | address
address [mask] [fvrf]
```

The *fvrf argument* specifies that the address is in FVRF; match ID specifies the client IKE Identity (ID) that is to be matched

---

Static IPSec tunnels to MPLS VPN Sample Config



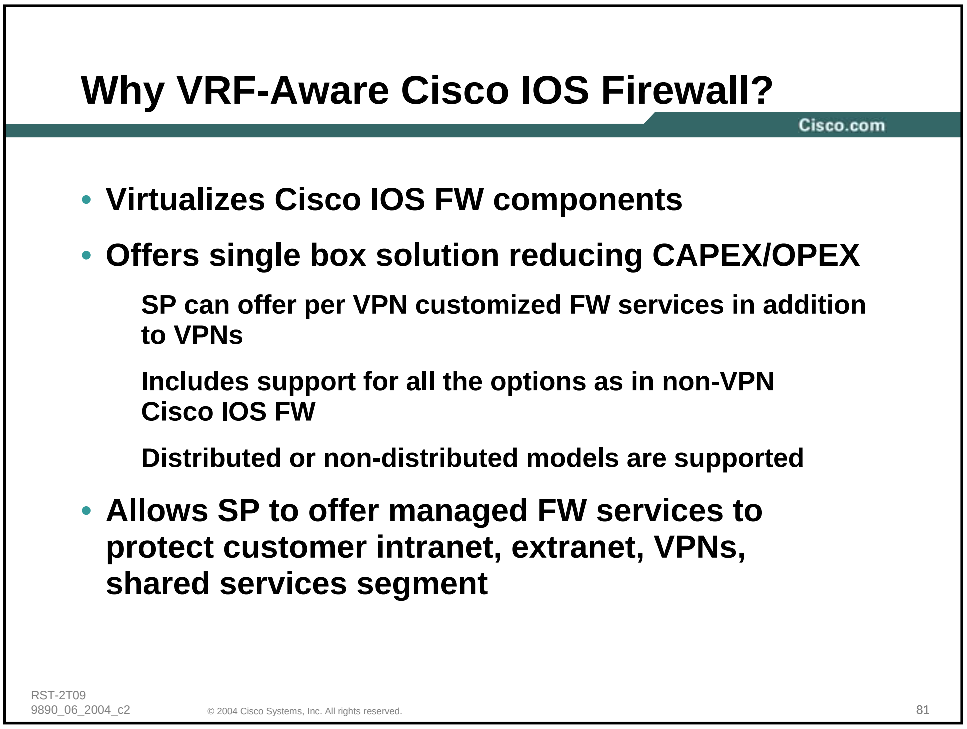
CPE1:

```
crypto isakmp policy 1
  encr 3des
  authentication pre-share
  group 2
crypto isakmp key vpn1 address 172.18.1.1
!
crypto ipsec transform-set vpn1 esp-3des esp-sha-hmac
!
crypto map vpn1 1 ipsec-isakmp
 set peer 172.18.1.1
 set transform-set vpn1
 match address 101
!
interface FastEthernet1/0
  ip address 172.16.1.1 255.255.255.0
  crypto map vpn1
!
interface FastEthernet1/1
ip address 10.2.1.1 255.255.0.0
!
access-list 101 permit ip 10.2.0.0 0.0.255.255 10.1.0.0 0.0.255.255
!
```

Aggregator-PE:

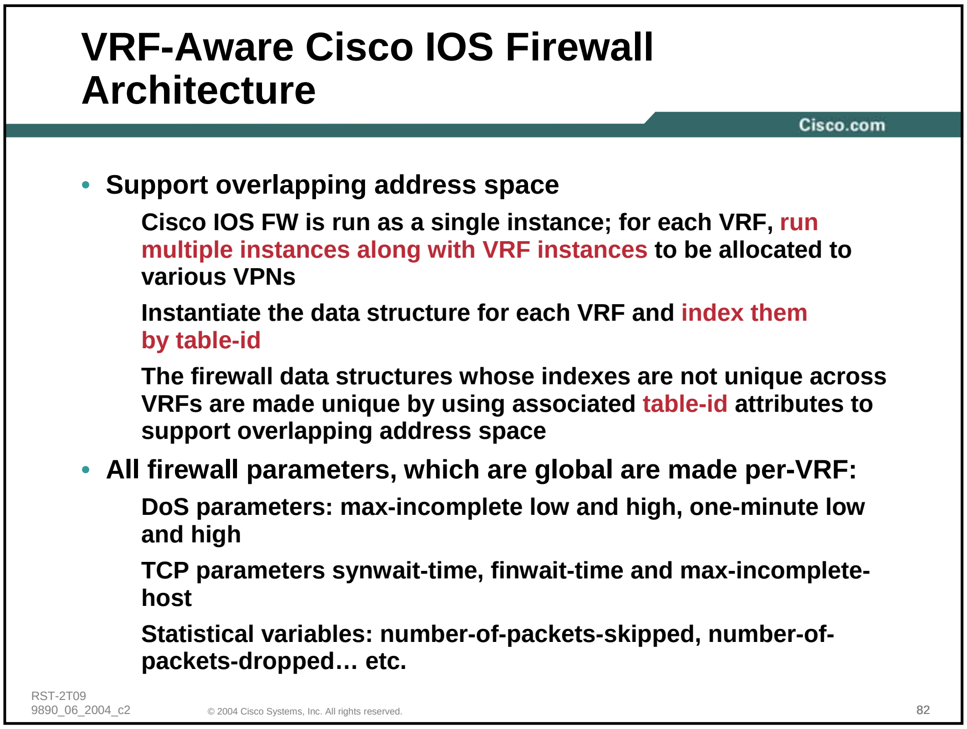
```
crypto keyring vpn1
pre-shared-key address 172.16.1.1 key vpn1
!
crypto isakmp policy 1
  encr 3des
  authentication pre-share
  group 2
!
crypto isakmp profile vpn1
   vrf vpn1
   keyring vpn1
   match identity address 172.16.1.1 255.255.255.255
!
crypto ipsec transform-set vpn1 esp-3des esp-sha-hmac
!
crypto map crypmap 1 ipsec-isakmp
  set peer 172.16.1.1
  set transform-set vpn1
  set isakmp-profile vpn1
  match address 101
!
interface Ethernet1/2
  ip address 172.18.1.1 255.255.255.0
  crypto map crypmap
!
ip route vrf vpn1 10.2.0.0 255.255.0.0 172.18.1.2 global
!
access-list 101 permit ip 10.1.0.0 0.0.255.255 10.2.0.0 0.0.255.255
```

VRF-AWARE CISCO IOS FIREWALL



Why VRF-Aware Cisco IOS Firewall?

- Virtualizes Cisco IOS FW components
- Offers single box solution reducing CAPEX/OPEX
    - SP can offer per VPN customized FW services in addition to VPNs
    - Includes support for all the options as in non-VPN Cisco IOS FW
    - Distributed or non-distributed models are supported
- Allows SP to offer managed FW services to protect customer intranet, extranet, VPNs, shared services segment



VRF-Aware Cisco IOS Firewall Architecture

- Support overlapping address space
    - Cisco IOS FW is run as a single instance; for each VRF, run multiple instances along with VRF instances to be allocated to various VPNs
    - Instantiate the data structure for each VRF and index them by table-id
    - The firewall data structures whose indexes are not unique across VRFs are made unique by using associated table-id attributes to support overlapping address space
- All firewall parameters, which are global are made per-VRF:
    - DoS parameters: max-incomplete low and high, one-minute low and high
    - TCP parameters synwait-time, finwait-time and max-incomplete-host
    - Statistical variables: number-of-packets-skipped, number-of-packets-dropped… etc.

VRF Aware Cisco IOS Firewall Architecture (Cont.)

- Firewall alert and audit-trail syslog messages

  Useful for the network administrators to manage the firewall: adjust FW parameters, add additional security policies, detect malicious sources and attacks, etc...

  Isolate the syslog messages per VRF

  VRF name corresponding to the VPN will be tagged along with each syslog message being logged

  VPN customers can examine logs for their own VPNs

- Per VRF URL Filtering

  URL Filter Server per VRF; server could be located in the services segment (could affect performance caused by large # of per-process TCP connections for large # of VRFs)

  Virtualize URL Filtering server; support multiple URL instances per VRF on a single server

  Send VPN-ID along with URL filter query request and forward the request to the corresponding URL Filter server instance based on the VPN-ID

VRF-Aware Cisco IOS Firewall Architecture (Cont.)

- Session control mechanism

  To limit one VRF hogging all the resources

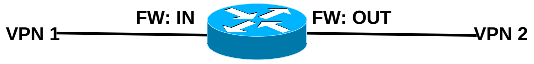
  Session counter will be maintained for each firewall

  Send Alert Message to syslog server when session exceeds max session limit; no new sessions allowed for this VRF

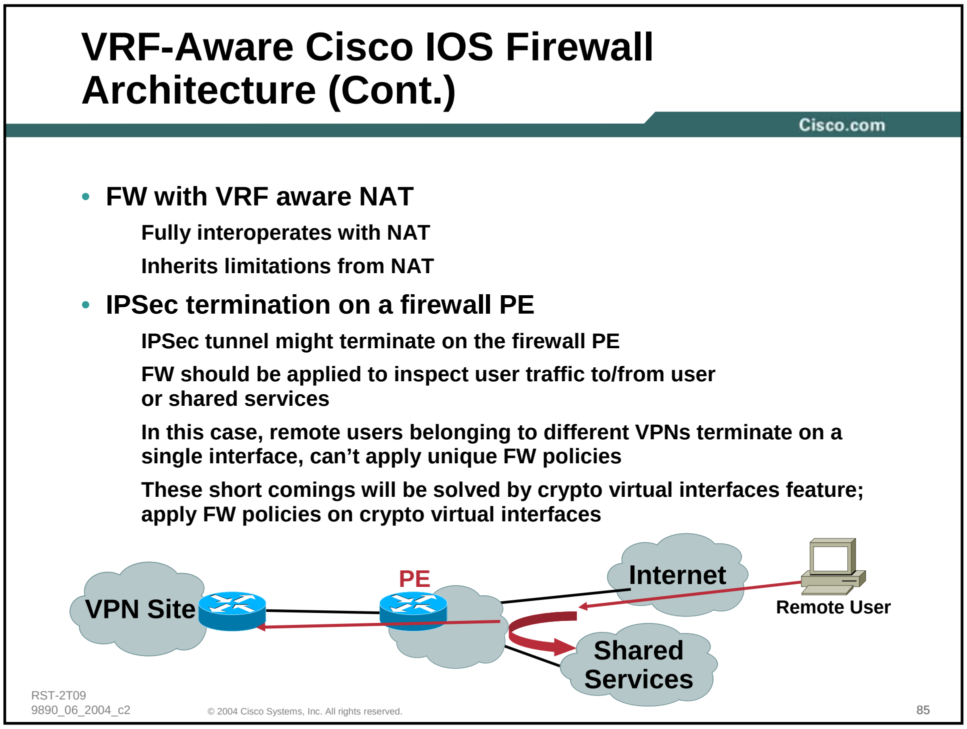
- VRF-VRF

  Firewall policies applied on both inbound and outbound interfaces

  The firewall on the inbound takes precedence over the FW on the outbound interface if there is conflict



  Firewall rule on the outbound will be applied on a packet if none of the rules match on inbound



VRF-Aware Cisco IOS Firewall Architecture (Cont.)

- FW with VRF aware NAT
  - Fully interoperates with NAT
  - Inherits limitations from NAT
- IPSec termination on a firewall PE
  - IPSec tunnel might terminate on the firewall PE
  - FW should be applied to inspect user traffic to/from user or shared services
  - In this case, remote users belonging to different VPNs terminate on a single interface, can't apply unique FW policies
  - These short comings will be solved by crypto virtual interfaces feature; apply FW policies on crypto virtual interfaces

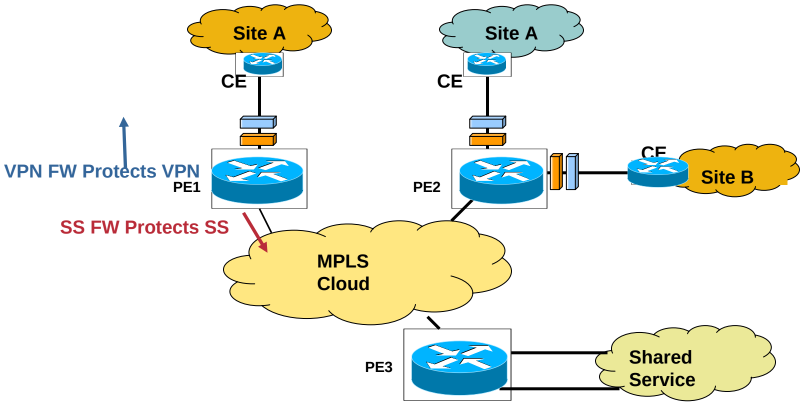
RST-2T09
9890_06_2004_c2

Cisco IOS Firewall for MPLS VPNs Deployment Scenarios

- Distributed model
- Hub-and-spoke model
- Virtual interface for VRF firewall

RST-2T09
9890_06_2004_c2

VRF-Aware Cisco IOS Firewall Distributed Model


Site A
CE
Site A
CE
CE
Site B
VPN FW Protects VPN
PE1
PE2
SS FW Protects SS
MPLS Cloud
PE3
Shared Service

🟦 VPN Firewall (VPN1-FW, VPN2-FW)

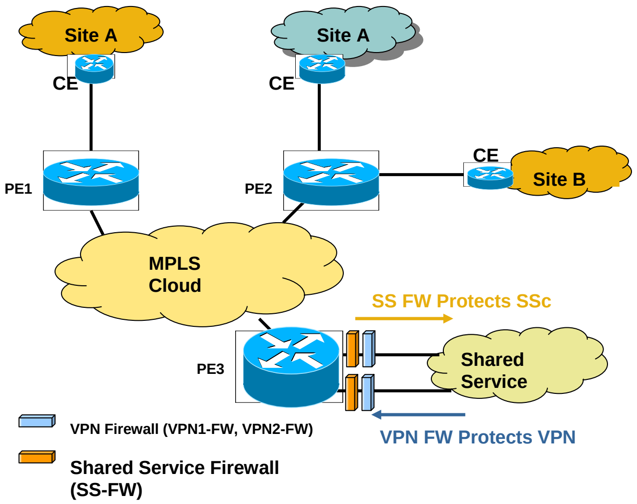
🟧 Shared Service Firewall (SS- FW)

Cisco Confidential

VRF-Aware Cisco IOS Firewall Distributed Model

- Pros
  - Firewall processing load is distributed to participating PEs
  - Firewall features can be deployed in the inbound direction
  - Shared services is protected @ ingress, malicious packets from VPNs don’t travel through the network
- Cons
  - MPLS cloud is open to the shared services; malicious packets from shared services travel through MPLS NW
  - Shared service FW features can’t be deployed in the inbound direction
  - Distributed provisioning, management, and troubleshooting

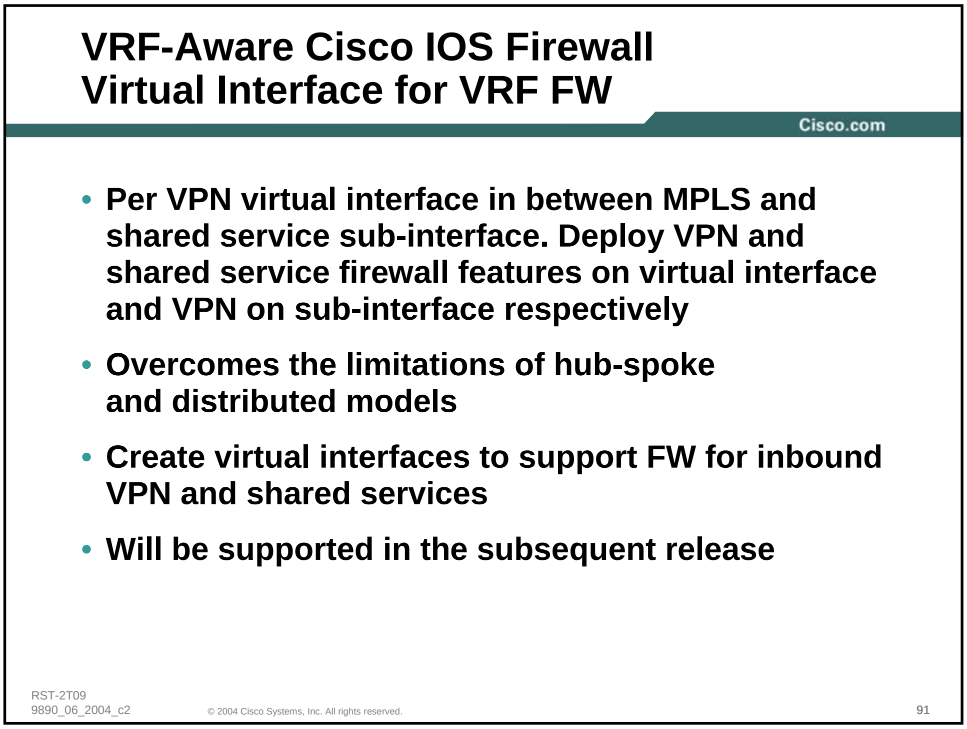
VRF-Aware Cisco IOS Firewall Hub-and-Spoke Model


Site A
CE
Site A
CE
CE
Site B
PE1
PE2
MPLS Cloud
SS FW Protects SSc
PE3
Shared Service
VPN FW Protects VPN
VPN Firewall (VPN1-FW, VPN2-FW)
Shared Service Firewall (SS-FW)

Cisco Confidential

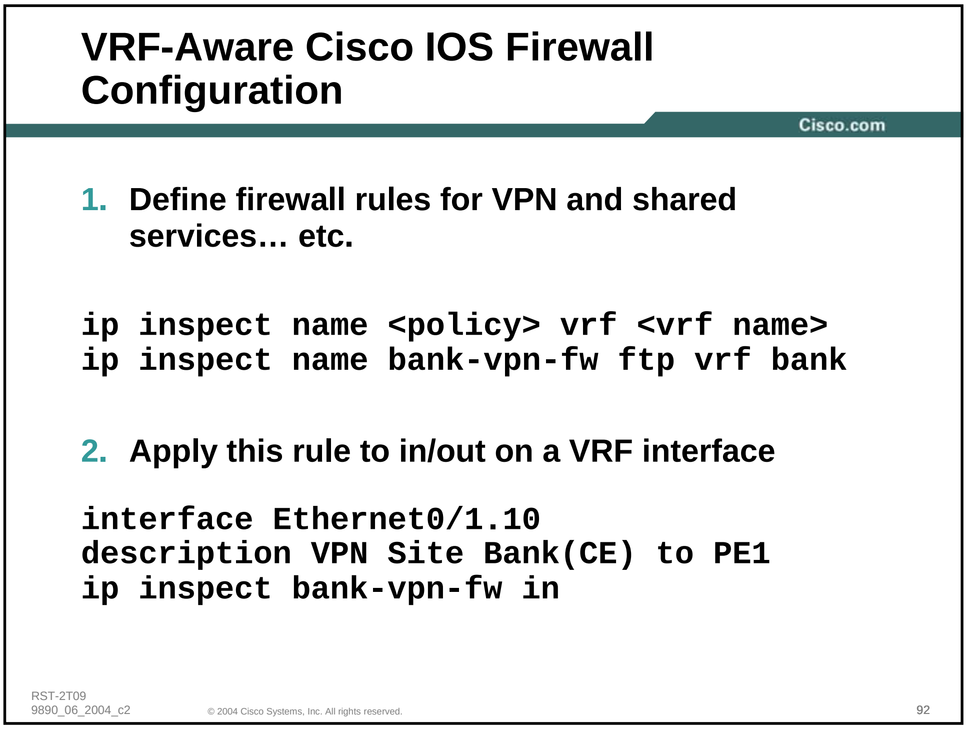
VRF-Aware Cisco IOS Firewall Hub-and-Spoke Model

- Pros
  - Firewall deployed centrally, easier prov., mgmt, troubleshooting
  - Shared services FW can be applied in the inbound direction
  - VPN site is protected from shared service at egress PE itself; malicious packets from shared service will be filtered at this PE before they enter into MPLS cloud
- Cons
  - VPN firewall features can not be deployed in the inbound direction
  - MPLS cloud is open to the VPN site; the malicious packets from VPN sites will be filtered only at egress PE—after traveling through all core routers



VRF-Aware Cisco IOS Firewall Virtual Interface for VRF FW

- Per VPN virtual interface in between MPLS and shared service sub-interface. Deploy VPN and shared service firewall features on virtual interface and VPN on sub-interface respectively
- Overcomes the limitations of hub-spoke and distributed models
- Create virtual interfaces to support FW for inbound VPN and shared services
- Will be supported in the subsequent release



VRF-Aware Cisco IOS Firewall Configuration

1. Define firewall rules for VPN and shared services… etc.

```
ip inspect name <policy> vrf <vrf name>
ip inspect name bank-vpn-fw ftp vrf bank
```

2. Apply this rule to in/out on a VRF interface

```
interface Ethernet0/1.10
description VPN Site Bank(CE) to PE1
ip inspect bank-vpn-fw in
```

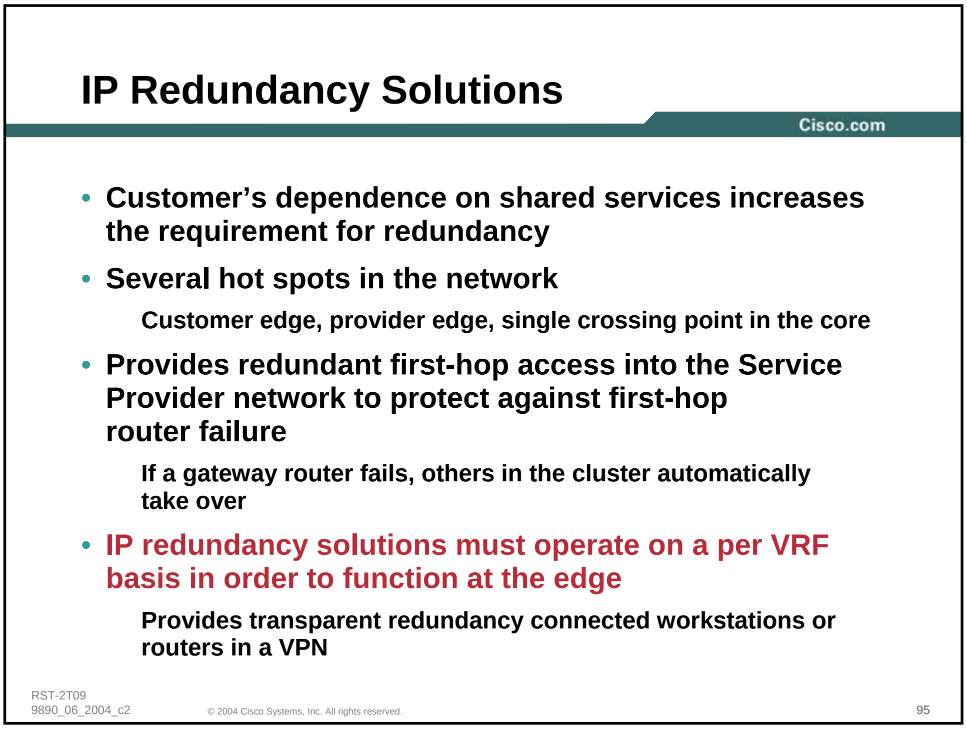
VRF-Aware Cisco IOS Firewall Restrictions

Firewall on the MPLS Interface: It Is Not Recommended to Apply Firewall Policies on the MPLS Interfaces:

1. Firewall on MPLS interface will be treated as global firewall hence VPN-specific instantiation will not happen
2. ACL is not supported on the MPLS interface (not VRF aware, thus configuring firewall without an ACL makes less sense)

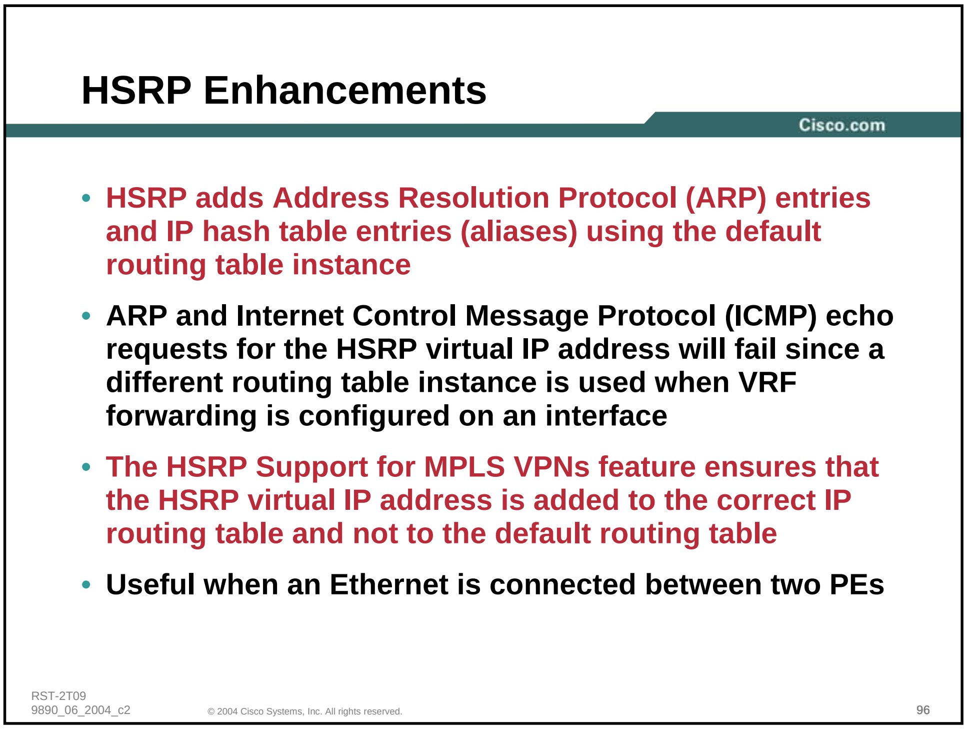
Agenda

- Managed IP Services
- Managed Security Services
- SP Edge Redundancy
- Shared Management



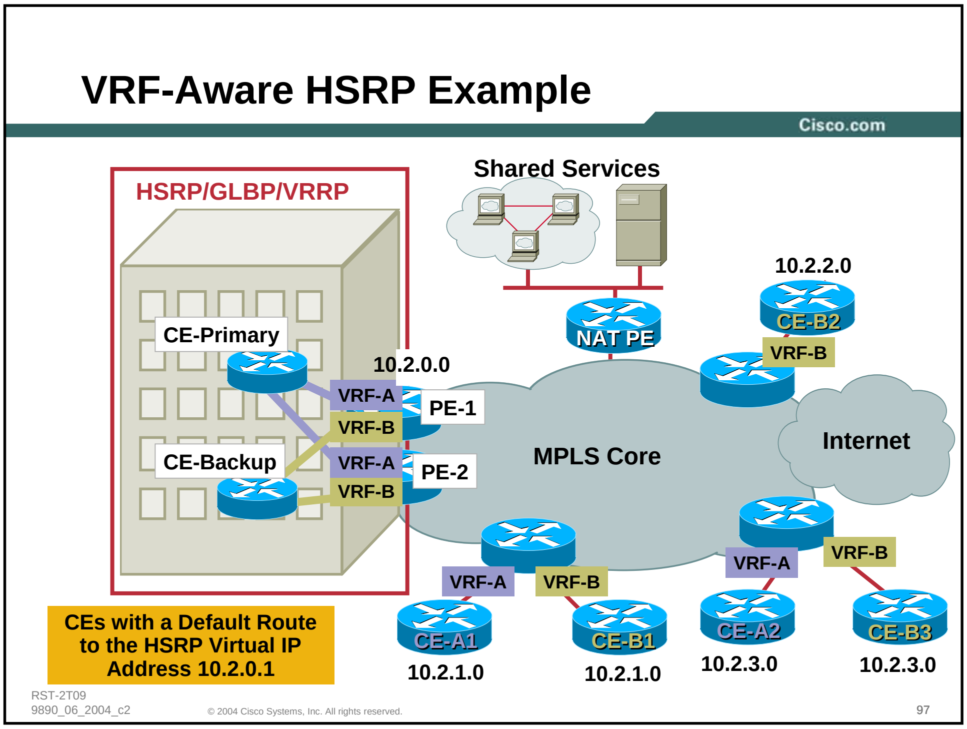
IP Redundancy Solutions

- Customer's dependence on shared services increases the requirement for redundancy
- Several hot spots in the network
  - Customer edge, provider edge, single crossing point in the core
- Provides redundant first-hop access into the Service Provider network to protect against first-hop router failure
  - If a gateway router fails, others in the cluster automatically take over
- IP redundancy solutions must operate on a per VRF basis in order to function at the edge
  - Provides transparent redundancy connected workstations or routers in a VPN



HSRP Enhancements

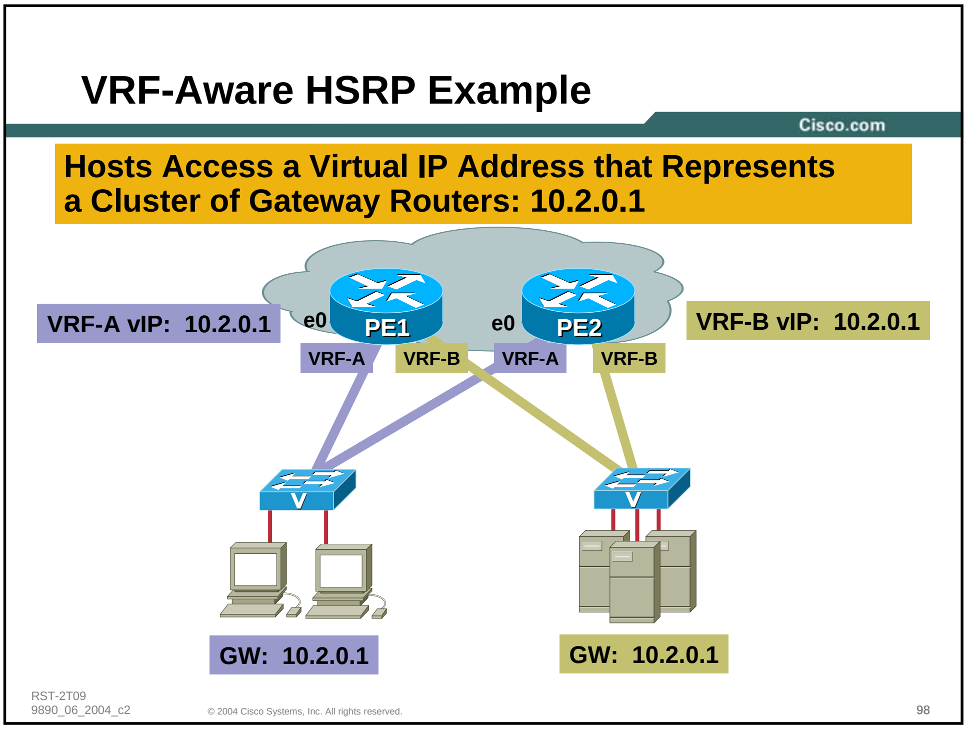
- HSRP adds Address Resolution Protocol (ARP) entries and IP hash table entries (aliases) using the default routing table instance
- ARP and Internet Control Message Protocol (ICMP) echo requests for the HSRP virtual IP address will fail since a different routing table instance is used when VRF forwarding is configured on an interface
- The HSRP Support for MPLS VPNs feature ensures that the HSRP virtual IP address is added to the correct IP routing table and not to the default routing table
- Useful when an Ethernet is connected between two PEs



VRF-Aware HSRP Example

Cisco.com
HSRP/GLBP/VRRP
CE-Primary
CE-Backup
10.2.0.0
VRF-A
VRF-B
PE-1
VRF-A
VRF-B
PE-2
Shared Services
NAT PE
MPLS Core
10.2.2.0
CE-B2
VRF-B
Internet
VRF-A
VRF-B
CE-A1
10.2.1.0
CE-B1
10.2.1.0
VRF-A
VRF-B
CE-A2
10.2.3.0
CE-B3
10.2.3.0

CEs with a Default Route to the HSRP Virtual IP Address 10.2.0.1



VRF-Aware HSRP Example

Cisco.com

Hosts Access a Virtual IP Address that Represents a Cluster of Gateway Routers: 10.2.0.1

VRF-A vIP: 10.2.0.1
e0
PE1
e0
PE2
VRF-B vIP: 10.2.0.1
VRF-A
VRF-B
VRF-A
VRF-B
V
V
GW: 10.2.0.1
GW: 10.2.0.1

MPLS VPN HSRP Configuration

PEs with HSRP Running between Their VRF Interfaces
CEs Configured with HSRP Virtual IP Address as Its Default Route

Router PE1 Configuration

```
!
 ip vrf vrf-A
  rd 100:1
  route-target export 100:1
  route-target import 100:1
 !
 interface ethernet0
  ip vrf forwarding vrf-A
  ip address 10.2.0.3 255.255.0.0
  standby 1 ip 10.2.0.1
  standby 1 priority 105 preempt delay 10
  standby 1 timers 1 3
  standby 1 track ethernet1 10
  standby 1 track ethernet2 10
```

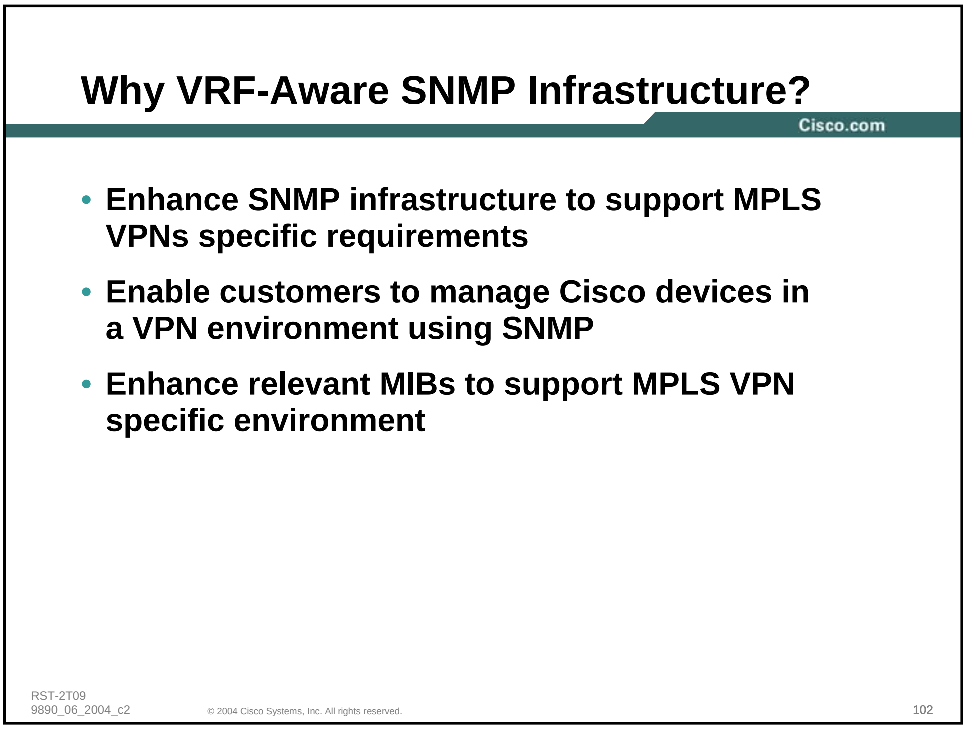
Router PE2 Configuration

```
!
 ip vrf vrf-A
  rd 100:1
  route-target export 100:1
  route-target import 100:1
 !
 interface ethernet0
  ip vrf forwarding vrf-A
  ip address 10.2.0.2 255.255.0.0
  standby 1 ip 10.2.0.1
  standby 1 priority 100 preempt delay 10
  standby 1 timers 1 3
  standby 1 track ethernet1 10
  standby 1 track ethernet2 10
```

Agenda

- Managed IP Services
- Managed Security Services
- SP Edge Redundancy
- Shared Management

VRF-AWARE SNMP INFRASTRUCTURE



Why VRF-Aware SNMP Infrastructure?

- Enhance SNMP infrastructure to support MPLS VPNs specific requirements
- Enable customers to manage Cisco devices in a VPN environment using SNMP
- Enhance relevant MIBs to support MPLS VPN specific environment

VRF-Aware SNMP Infrastructure Enhancements

- This framework requires an agreement between manager and agent entities on a mapping between SNMP SecurityNames and the VPN Ids
- IP SNMP (UDP) transport process has been modified to allow SNMP receive and respond to request from VPN interfaces (VRF tables)
- Support multiple (instead of a single) context (per VRF) at the agent initialization time
- MIB instrumentation has been enhanced so it can maintain MIB data of different VPNs and provide a secure way to access the MIB data within the framework provided by SNMPv3
- MIBs which are made VPN aware, must use context-name as parameter to retrieve the instances of objects in different contexts, store or modify data in different (relevant) contexts
- Community string should include VRF name since SNMP v1/v2C doesn't have security; if a request is coming through a VRF interface, it must include VRF name, otherwise the request will not be processed

---

Summary

- Cisco to continue enhancing feature sets to support innovative services that work over a single infrastructure

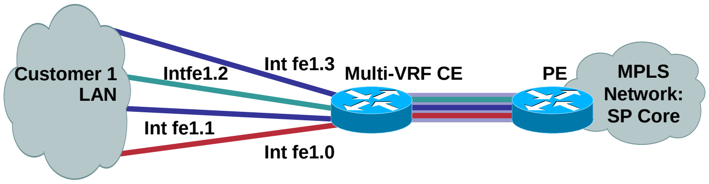
NETWORK INTEGRATION WITH MPLS: ADVANCED L3 VPN SERVICES MODELS AND MECHANISMS

Agenda

- Multi-VRF
- Remote Access to MPLS Solutions
  - VRF Select
  - Half Duplex VRFs
  - VPNID
- InterProvider Solutions
  - Inter-AS Service Models
  - RT Rewrite, ipV4BGP Label Distribution
  - Carrier Supporting Carrier Service Models
  - VRF-Aware MPLS Static Labels
- Multipath Load Balancing

Multi-VRF Architecture

Extend VPN Functionality to Customer Edge Router to Offer Privacy and Separation without Full PE Functionality

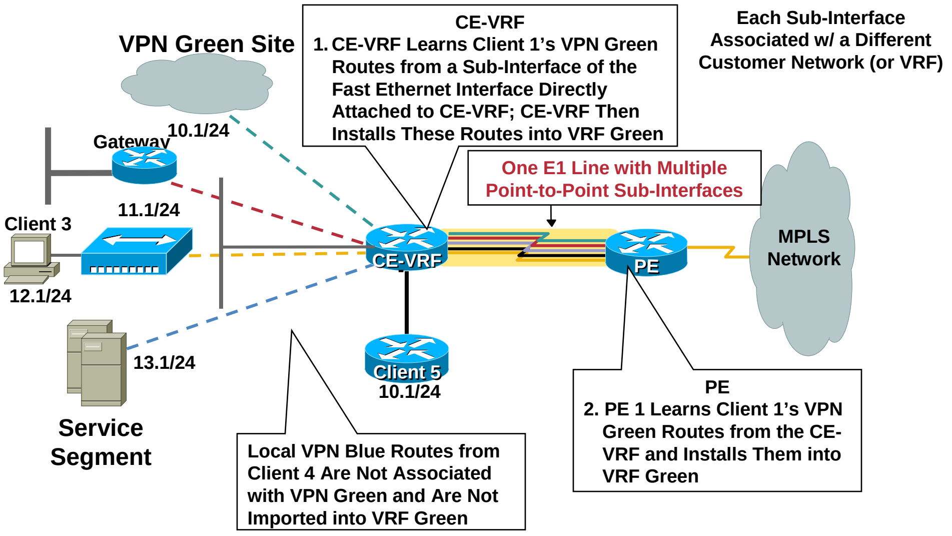

Customer 1
LAN
Intfe1.2
Int fe1.3
Multi-VRF CE
PE
MPLS
Network:
SP Core
Int fe1.1
Int fe1.0

- Each ingress interface is bound to a VRF
- Each sub-interface associated is bound to different VRF
- Pair of ingress/egress interface can be mapped to the same VRF
- Separate dedicated VRF instances + a global routing table

Multi-VRF Benefits

- Reduces number of edge devices per VPN (metro area, multi-tenant bldg, multi-VLAN support… etc.)
- No MPLS functionality on the CE, no CE-PE label exchange
- Overlapping address space is supported
- No VRF ID or labels are created and/or used in the control or Forwarding Plane
- No MP-BGP meshing required with remote PEs, less # of routing tables to manage
- Same routing protocol support as in normal VRF
- Local inter-VRF routing is supported
- CE supports int. that support sub-ints (FE, FR, ATM,…)

Multi-VRF Operational Model



CE-VRF
1. CE-VRF Learns Client 1's VPN Green Routes from a Sub-Interface of the Fast Ethernet Interface Directly Attached to CE-VRF; CE-VRF Then Installs These Routes into VRF Green

Each Sub-Interface Associated w/ a Different Customer Network (or VRF)

One E1 Line with Multiple Point-to-Point Sub-Interfaces

Local VPN Blue Routes from Client 4 Are Not Associated with VPN Green and Are Not Imported into VRF Green

PE
2. PE 1 Learns Client 1's VPN Green Routes from the CE-VRF and Installs Them into VRF Green

MULTI-VRF APPLICATIONS

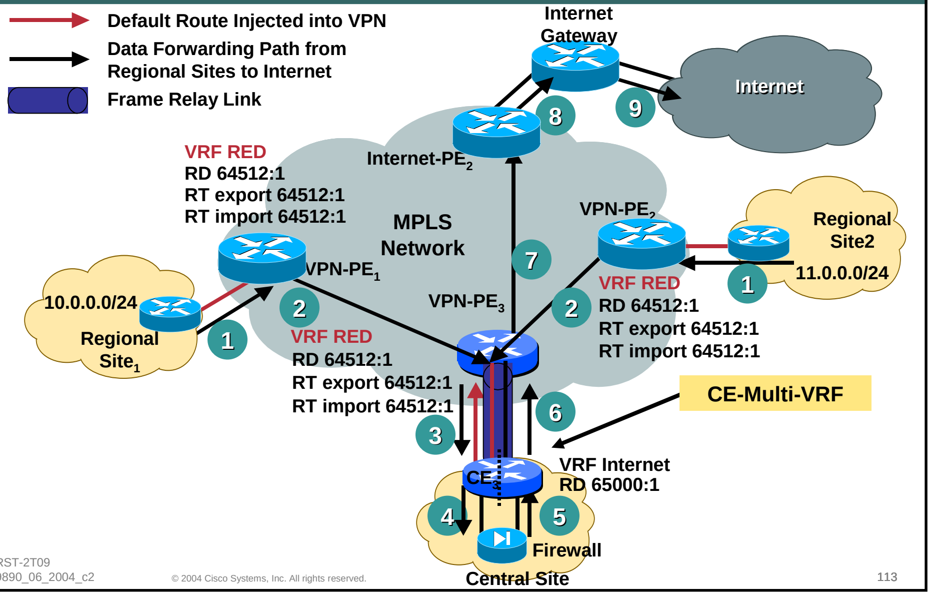
Multi-VRF Applications

Cisco.com

1. Provide Internet and VPN services using the same CE
   - Traffic separation on VRF lite CE instead of on a PE; public and private traffic are kept separate
   - There is an option to put Internet table in a VRF on a CE
2. Provide VLAN-based VPN services
3. Implement multiple VPNs in a customer site using single router in a multi-tenant environment
4. L3 VLAN Service Model

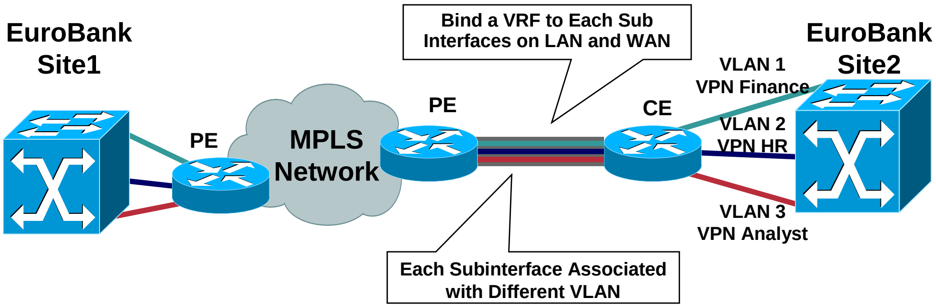
Application 1: Internet and VPN Services Using a Single CE

Cisco.com


Default Route Injected into VPN
Data Forwarding Path from Regional Sites to Internet
Frame Relay Link
Internet Gateway
Internet
8
9
Internet-PE$_2$
VRF RED
RD 64512:1
RT export 64512:1
RT import 64512:1
MPLS Network
VPN-PE$_2$
Regional Site2
11.0.0.0/24
VPN-PE$_1$
7
1
10.0.0.0/24
Regional Site$_1$
2
1
VPN-PE$_3$
2
VRF RED
RD 64512:1
RT export 64512:1
RT import 64512:1
VRF RED
RD 64512:1
RT export 64512:1
RT import 64512:1
CE-Multi-VRF
6
3
CE$_3$
VRF Internet
RD 65000:1
4
5
Firewall
Central Site

Application 2: VLAN-Based VPN Services

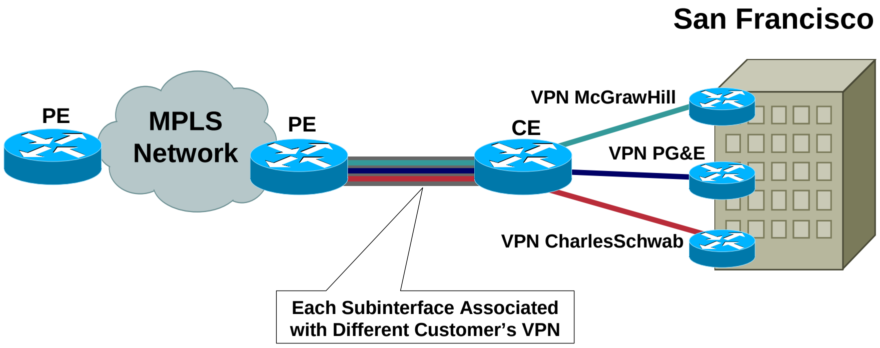
Cisco.com


EuroBank Site1
PE
MPLS Network
PE
Bind a VRF to Each Sub Interfaces on LAN and WAN
CE
EuroBank Site2
VLAN 1
VPN Finance
VLAN 2
VPN HR
VLAN 3
VPN Analyst
Each Subinterface Associated with Different VLAN

Alternate Method Would Be to Use Inbound/Outbound ACLs or Different Routing Protocols (Do Not Redistribute)! Neither Are Scalable!

Application 3: MetroEther: Multi-Tenant Building

Cisco.com

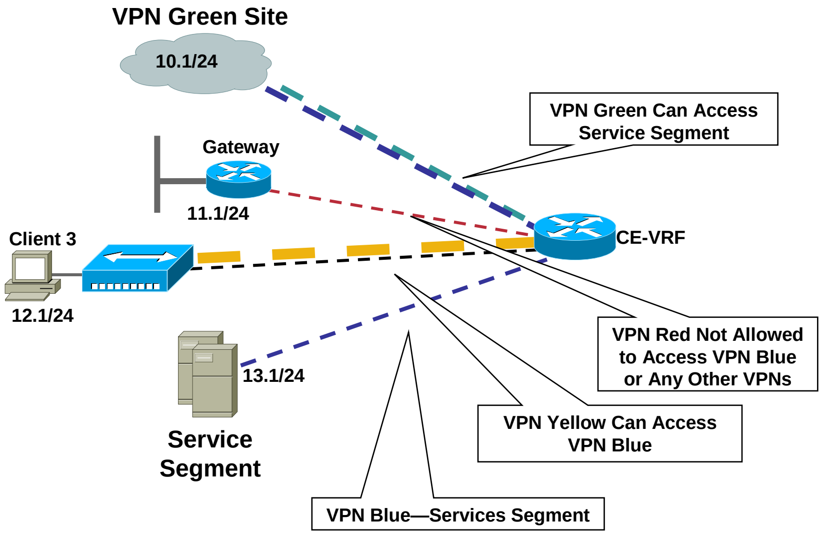

San Francisco
PE
MPLS Network
PE
CE
VPN McGrawHill
VPN PG&E
VPN CharlesSchwab

Each Subinterface Associated with Different Customer's VPN

Alternate Method Would Be to Use Inbound/Outbound ACLs or Different Routing Protocols (Do Not Redistribute)! Neither Are Scalable!

Application 4: L3 VLAN Service Model

Cisco.com

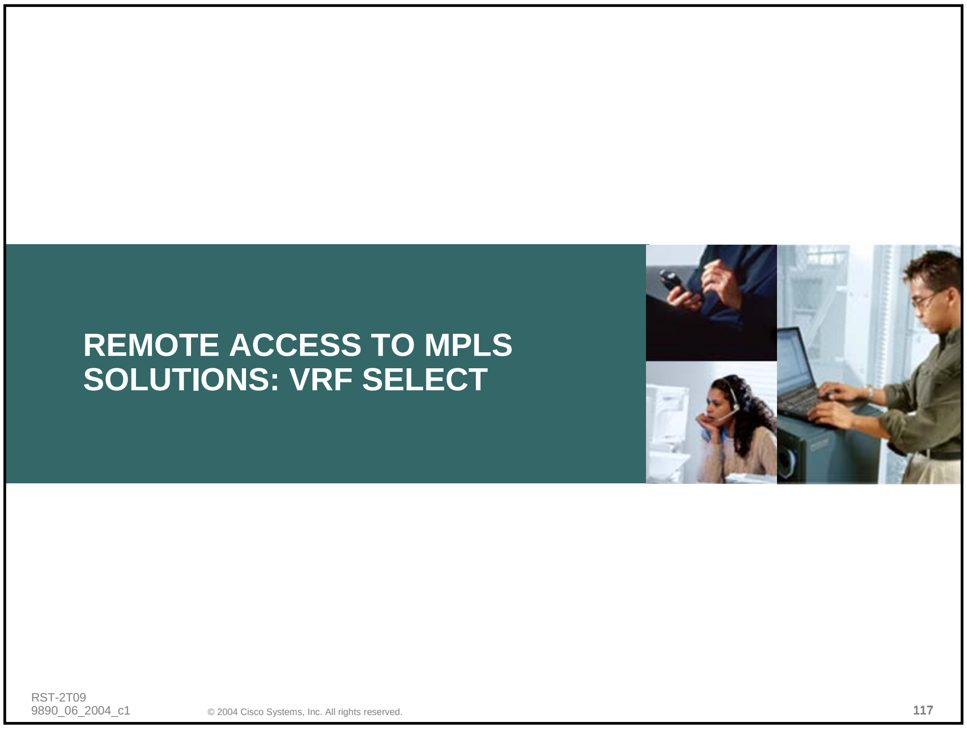

VPN Green Site
10.1/24
Gateway
11.1/24
Client 3
12.1/24
CE-VRF
13.1/24
Service Segment

VPN Green Can Access Service Segment

VPN Red Not Allowed to Access VPN Blue or Any Other VPNs

VPN Yellow Can Access VPN Blue

VPN Blue—Services Segment



REMOTE ACCESS TO MPLS SOLUTIONS: VRF SELECT

Why VRF Select?

- **Flexible solution for access providers**
- **Allows access providers to map DSL/cable customers to any ISP that provides VPN capabilities**
- **Allows remote users to connect to VPNs, irrespective of access provider**
- **Eliminates the dependence on physical interface to VRF binding**
- **Better alternative to policy-based routing**

How VRF Select Works

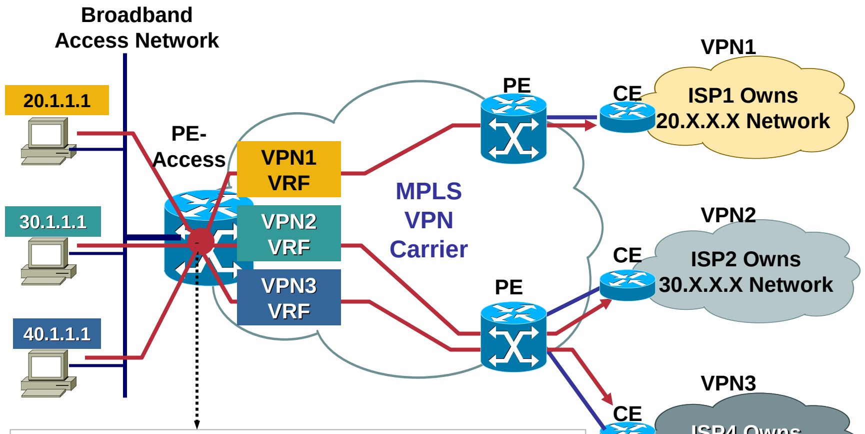
Cisco.com

- De-couple the association between VRF and an interface and populate a source IP address table used to select VRF
- VRF selection is performed at the ingress interface on the PE router
- Use a two-table lookup mechanism at the ingress interface of the PE router; perform:
  1. ‘Criteria Selection’ table look up to select a VRF table
  2. Look up the destination IP address of the packet on the selected VRF table to determine the output int. and adjacency

---

VRF Select: Deployment Scenario

Cisco.com


Broadband
Access Network
20.1.1.1
30.1.1.1
40.1.1.1
PE-
Access
VPN1
VRF
VPN2
VRF
VPN3
VRF
MPLS
VPN
Carrier
PE
PE
CE
CE
CE
VPN1
ISP1 Owns
20.X.X.X Network
VPN2
ISP2 Owns
30.X.X.X Network
VPN3
ISP4 Owns
40.X.X.X Network

- VRF select decouples the interface with a VRF
- The VRF selection will be based on the source address of the incoming traffic

VRF Select Configuration Process on an Access PE

Cisco.com

- Create a VRF Routing Table

```
Access-PE(config)# ip vrf VPN1
Access-PE(config-vrf)# rd 100:1
Access-PE(config-vrf)# route-target export 100:1
Access-PE(config-vrf)# route-target export 100:1
```

- Define VRF Selection Entries

```
Access-PE(config)# vrf selection source 20.0.0.0 255.0.0.0 vrf vpn1
Access-PE(config)# vrf selection source 30.0.0.0 255.0.0.0 vrf vpn2
```

- Define Static Routes for a VRF

```
Access-PE(config)# ip route vrf  vpn1 20.0.0.0 255.0.0.0 pos1/0
Access-PE(config)# ip route vrf  vpn1 30.0.0.0 255.0.0.0 pos1/0
```

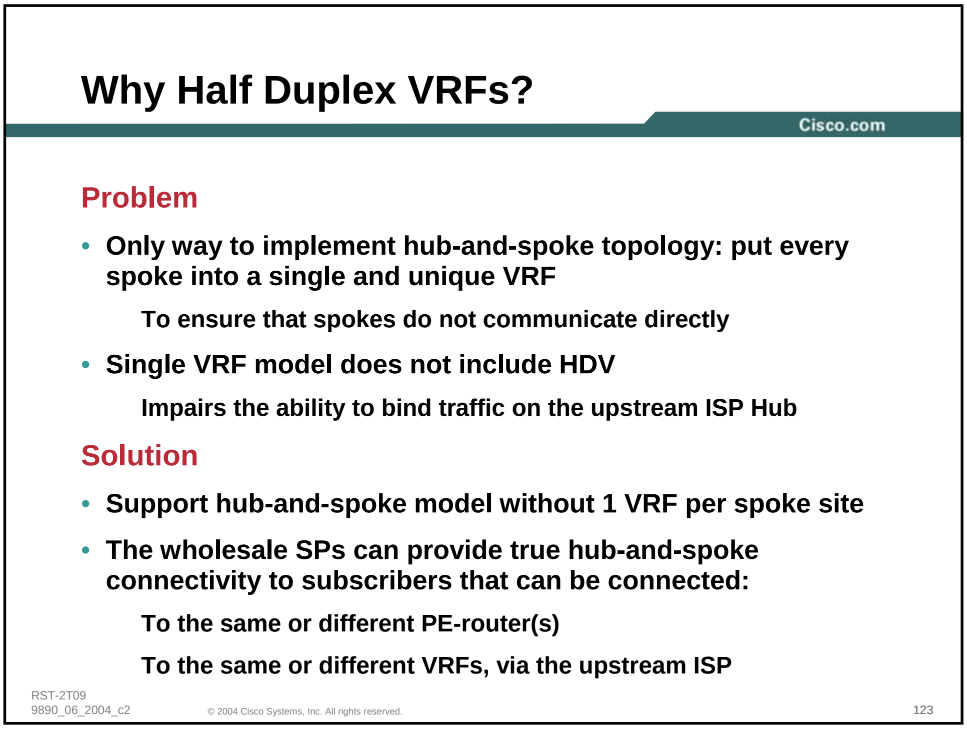
- Configure VRF Selection on the interface

```
Access-PE(config)# interface pos1/0
Access-PE(config-if)# ip vrf select source
Access-PE(config-if)# ip vrf receive vpn1
Access-PE(config-if)# ip vrf receive vpn2
```

- Configure BGP for VRF Selection

```
Access-PE(config)# address-family ipv4 vrf vpn2
Access-PE(config)#  redistribute static
```

REMOTE ACCESS TO MPLS SOLUTIONS: HALF DUPLEX VRFS (HDV)



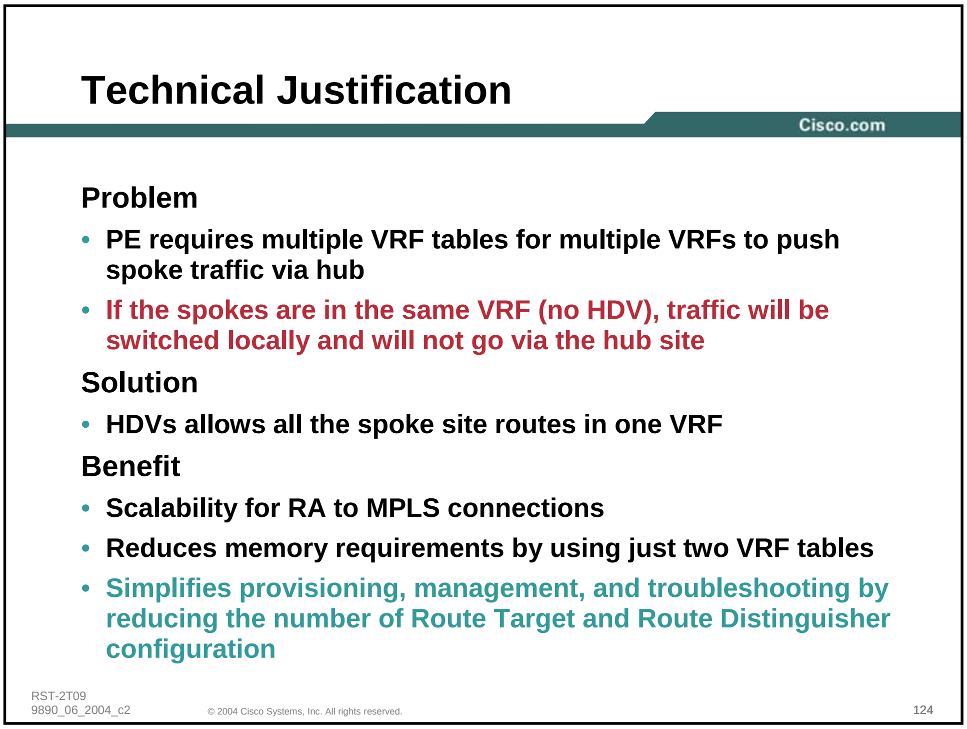
Why Half Duplex VRFs?

Problem

- Only way to implement hub-and-spoke topology: put every spoke into a single and unique VRF
  - To ensure that spokes do not communicate directly
- Single VRF model does not include HDV
  - Impairs the ability to bind traffic on the upstream ISP Hub

Solution

- Support hub-and-spoke model without 1 VRF per spoke site
- The wholesale SPs can provide true hub-and-spoke connectivity to subscribers that can be connected:
  - To the same or different PE-router(s)
  - To the same or different VRFs, via the upstream ISP



Technical Justification

Problem

- PE requires multiple VRF tables for multiple VRFs to push spoke traffic via hub
- If the spokes are in the same VRF (no HDV), traffic will be switched locally and will not go via the hub site

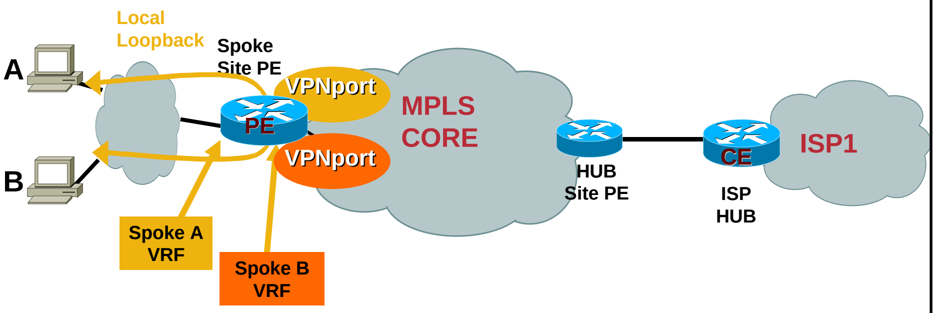
Solution

- HDVs allows all the spoke site routes in one VRF

Benefit

- Scalability for RA to MPLS connections
- Reduces memory requirements by using just two VRF tables
- Simplifies provisioning, management, and troubleshooting by reducing the number of Route Target and Route Distinguisher configuration

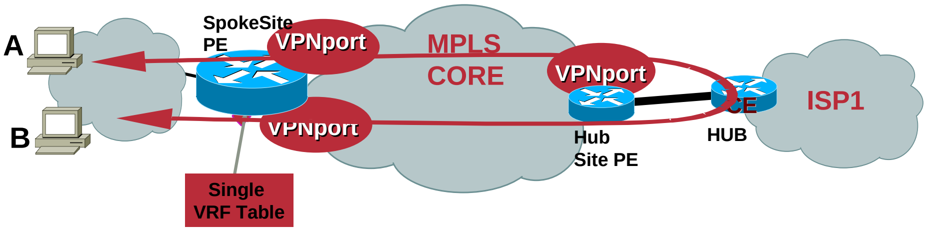
Hub-and-Spoke Connectivity without HDV Requires Dedicated VRF Tables per Spoke


Local
Loopback
Spoke
Site PE
A
B
PE
VPNport
VPNport
MPLS
CORE
HUB
Site PE
CE
ISP
HUB
ISP1
Spoke A
VRF
Spoke B
VRF

- All the spokes in the same VPN (yellow)
- Traffic will get switched locally
- Dedicated (separate) VRF per spoke is needed to push all traffic through upstream ISP Hub

---

Hub-and-Spoke Connectivity with HDV Using a Single VRF


A
B
SpokeSite
PE
VPNport
VPNport
MPLS
CORE
VPNport
Hub
Site PE
CE
HUB
ISP1
Single
VRF Table

- If two subscribers of the same service terminate on the same PE-router, then traffic between them can be switched locally at the PE-router (as shown), which is undesirable
- All inter-subscriber traffic needs to follow the default route via the Home Gateway (located at upstream ISP)
- Traffic doesn't get switched locally

HDV Supported Features

- IP unnumbered any point-to-point interfaces
  - Virtual access/template interfaces
- Spokes connected to Spoke PE or the Hub PE
- Subscriber using single or multiple ISPs
- Reverse Path Forwarding (RPF)
  - Used by Service Provider determine the source IP address of an incoming IP packet and ascertain whether it entered the router via the correct inbound interface
  - Concern: HDV populates a different VRF than the one used for “upstream” forwarding
  - Solution: Extend the RPF mechanism so the “downstream” VRF is checked; enable RPF extension using:

```
ip verify unicast reverse-path <downstream vrfname>
```

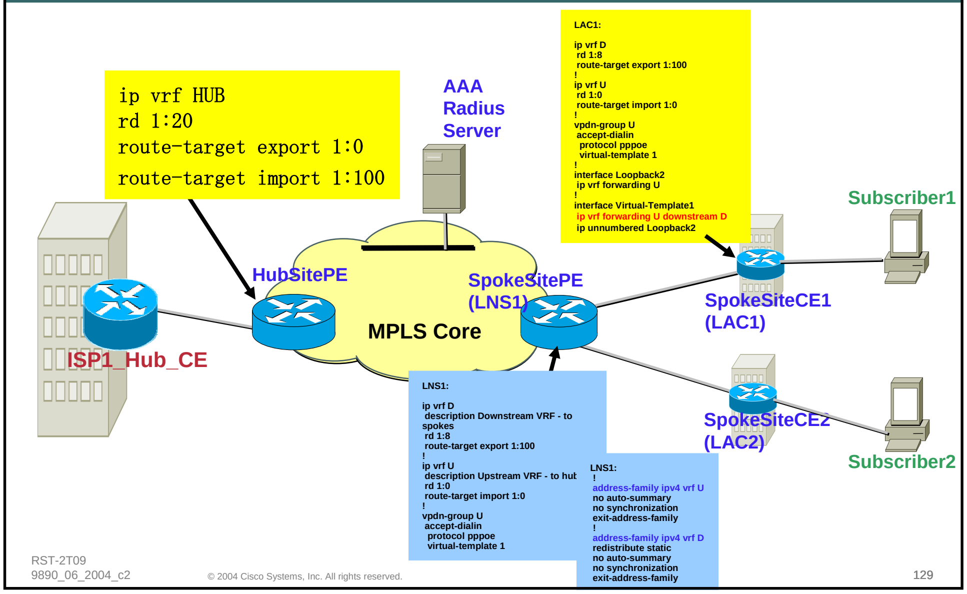
Sample Configuration

- Each VRF is created on the Spoke PE-router (LAC) before PPPoA or PPPoE client connections are established

```
ip vrf U
 rd 10:26
 route-target import 10:26
!
ip vrf D
 rd 10:27
 route-target export 10:27
```

- Upstream VRF:
  - Used for forwarding packets from spokes to hub & contains a static default route
  - Imports the default route from the Hub PE router (@WholeSale Provider)
  - Only requires a `route-target import` statement
- Downstream VRF:
  - Used for forwarding packets from hub to spoke
  - Contains a /32 route to a subscriber (installed from PPP)
  - Used to export all of the /32 (virtual-access ints) addresses toward the Hub PE-router; only requires a `route-target export` command

HDVRF Deployment Topology



```
ip vrf HUB
rd 1:20
route-target export 1:0
route-target import 1:100
```

```
LAC1:

ip vrf D
 rd 1:8
 route-target export 1:100
!
ip vrf U
 rd 1:0
 route-target import 1:0
!
vpdn-group U
 accept-dialin
  protocol pppoe
  virtual-template 1
!
interface Loopback2
 ip vrf forwarding U
!
interface Virtual-Template1
 ip vrf forwarding U downstream D
 ip unnumbered Loopback2
```

```
LNS1:

ip vrf D
 description Downstream VRF - to
spokes
 rd 1:8
 route-target export 1:100
!
ip vrf U
 description Upstream VRF - to hub
 rd 1:0
 route-target import 1:0
!
vpdn-group U
 accept-dialin
  protocol pppoe
  virtual-template 1
```

```
LNS1:
 !
 address-family ipv4 vrf U
 no auto-summary
 no synchronization
 exit-address-family
 !
 address-family ipv4 vrf D
 redistribute static
 no auto-summary
 no synchronization
 exit-address-family
```

Half Duplex VRF Functionality

1. HDVs are used in only one direction by incoming traffic

   Ex: upstream toward the MPLS VPN backbone or downstream toward the attached subscriber

2. PPP client dial, and is authenticated, authorized, and assigned an IP address

3. Peer route is installed in the downstream VRF table

   One single downstream VRF for all spokes in the single VRF

4. To forward the traffic among spokes (users), upstream VRF is consulted at the Spoke PE and traffic is forwarded from a Hub PE to Hub CE

   Return path: downstream VRF is consulted on the Hub PE before forwarding traffic to appropriate spoke PE and to the spoke (user)

5. Source address look up occurs in the downstream VRF, if unicast RPF check is configured on the interface on which HDV is enabled

---

VPNID

VPNID

- Standard specification as described in RFC 2685
- VPNID identifies MPLS VPNs for simpler and consistent management purposes
- Not used in any control plane or forwarding plane functionality
- VRF name or VPNID can be used to reference a VPN
- Use a unique VPNID for each VPN, configure this consistently on all the associated PEs
- Remote access applications, such as RADIUS and DHCP can use the MPLS VPN ID to identify a VPN. RADIUS can use the VPN ID to assign dial-in users to the proper VPN, based on each user's authentication information
- Can be used in any MPLS VPN topology: inter-provider or single service provider implementations

VPNID Configuration

Cisco.com

- **VPNID components:**

  vpn id oui:vpn-index

  oui (universal LAN MAC addresses assigned by IEEE Registration authority-3octets hex #)

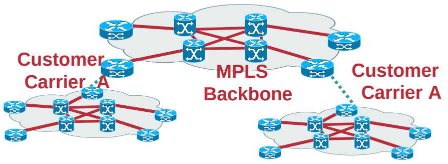
  vpn-index (4-octet hex #, identifies each VPN within a company

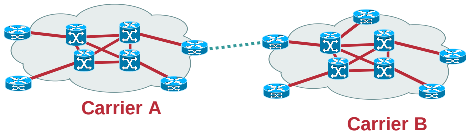
- **Sample config:**

```
!
ip vrf Cisco
 vpn id 36B:10
!
```

INTER-PROVIDER SOLUTIONS:
INTER-AS MPLS-VPN

CsC vs. Inter-AS





CSC

- Client-server topologies
- ISP or MPLS VPN provider is a customer of another MPLS VPN backbone provider
- MPLS VPN backbone services needed between the same carrier POPs
- Subscribing Service Provider may or may not have been MPLS-enabled
- Customer's sites do not distribute reachability information to the backbone carrier

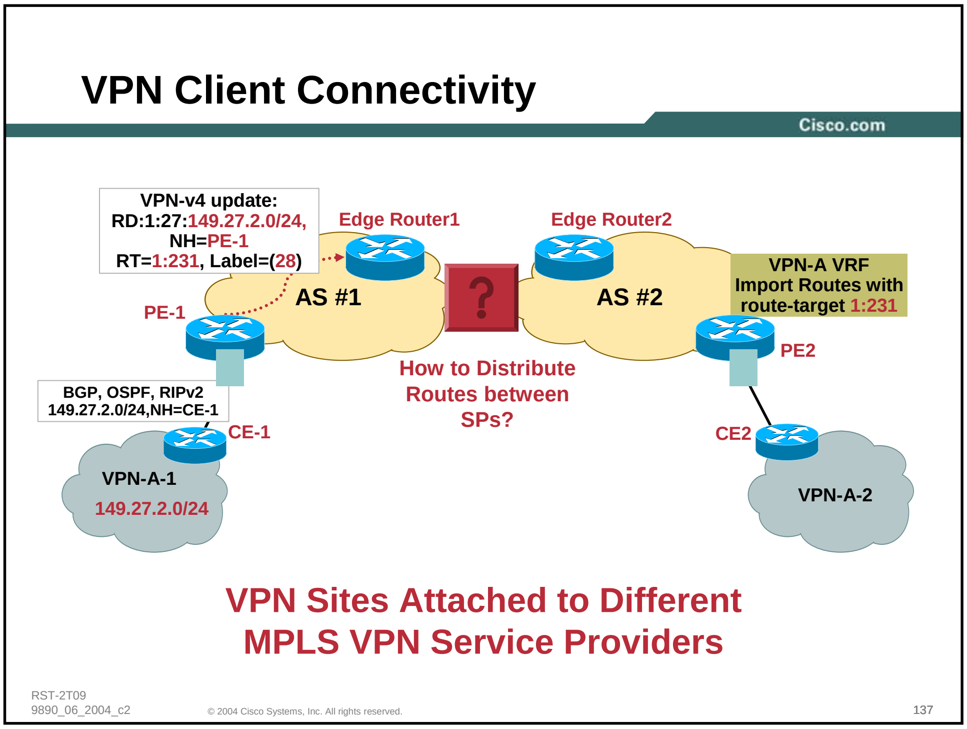
Inter-AS

- Peer-to-peer topologies
- Two ISPs peer up providing services to some of the common customer base
- Single SP POPs not available in all geographical areas required by their customers
- Both SPs must support MPLS VPNs
- Customer's sites distribute reachability information directly to the participating Service Providers

---

Why Inter-AS?

- Extends MPLS-VPN services across geographical boundaries allowing service providers to support customer base in geographical locations where their POPs are not available
- Allows multiple Service Providers to build common services
  - Allows separate ASs to communicate
  - Implies exchange of VPN routing information between providers
- Provides traffic separation and maintains privacy end-to-end
- Allows a single service provider to partition their network into multiple domains for scalability and inter-departmental privacy

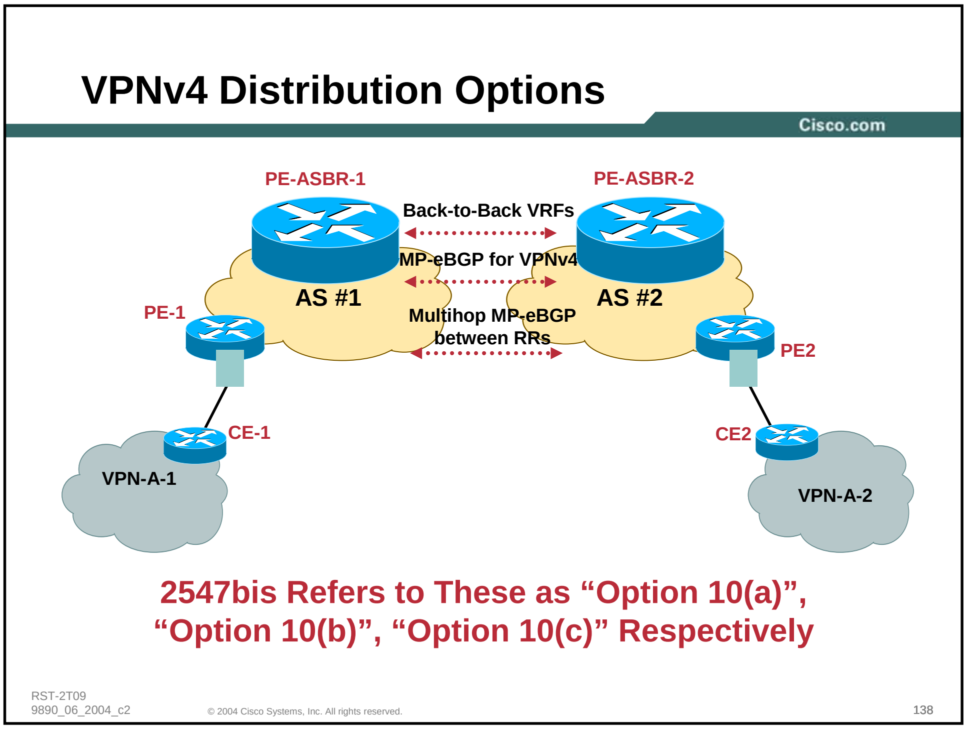


VPN Client Connectivity

Cisco.com

VPN-v4 update:
RD:1:27:149.27.2.0/24,
NH=PE-1
RT=1:231, Label=(28)
Edge Router1
Edge Router2
AS #1
AS #2
VPN-A VRF
Import Routes with
route-target 1:231
PE-1
PE2
How to Distribute
Routes between
SPs?
BGP, OSPF, RIPv2
149.27.2.0/24,NH=CE-1
CE-1
CE2
VPN-A-1
149.27.2.0/24
VPN-A-2

VPN Sites Attached to Different MPLS VPN Service Providers



VPNv4 Distribution Options

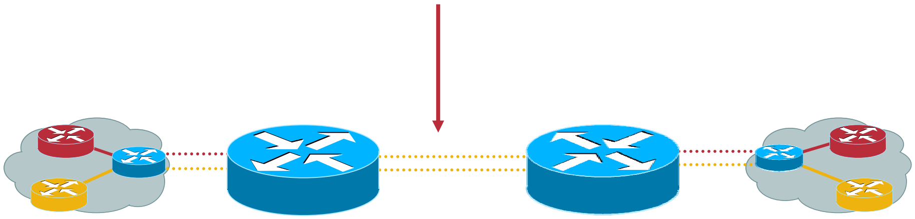
Cisco.com

PE-ASBR-1
PE-ASBR-2
Back-to-Back VRFs
MP-eBGP for VPNv4
Multihop MP-eBGP
between RRs
AS #1
AS #2
PE-1
PE2
CE-1
CE2
VPN-A-1
VPN-A-2

2547bis Refers to These as “Option 10(a)”, “Option 10(b)”, “Option 10(c)” Respectively

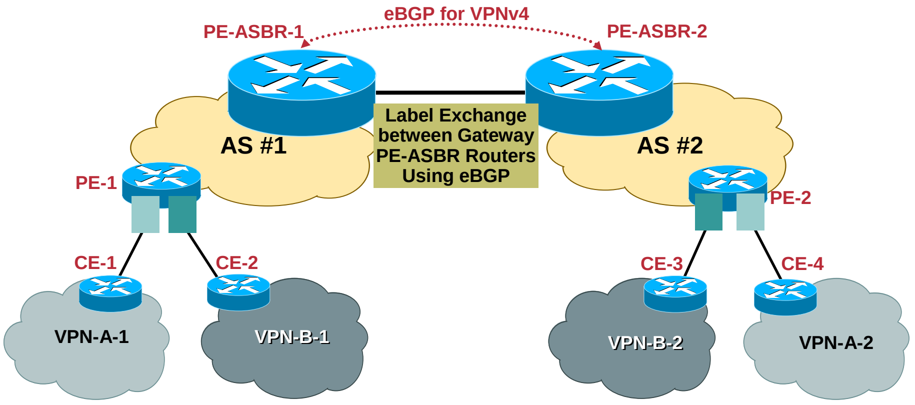
Back-to-Back VRFs (Option 10(a))

Each PE-ASBR Thinks the Other Is a CE



- 10(a) is the most popular Inter-AS tool today
- Directly connects ASBRs, over a sub-interface per VRF
- Packet is forwarded as an IP packet between the ASBRs
- Link may use any supported PE-CE routing protocol
- 10(a) is the most secure and easiest to provision
- May not be easy to manage as it grows

EBGP VPNv4 (Option 10(b))


eBGP for VPNv4
PE-ASBR-1
PE-ASBR-2
AS #1
AS #2
PE-1
PE-2
CE-1
CE-2
CE-3
CE-4
VPN-A-1
VPN-B-1
VPN-B-2
VPN-A-2

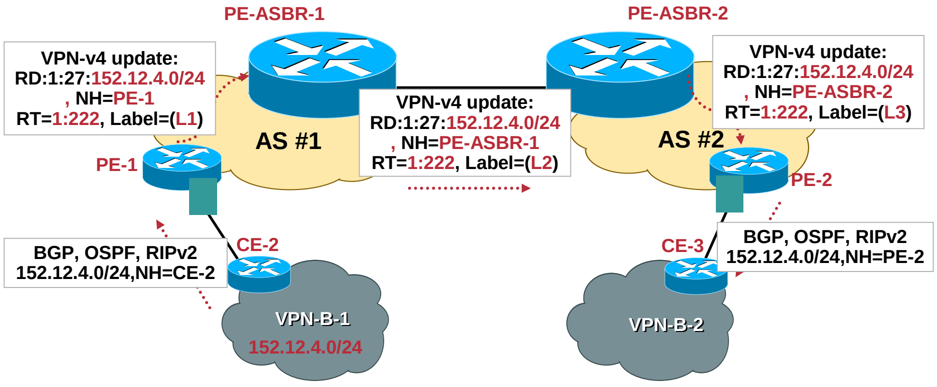
Label Exchange between Gateway PE-ASBR Routers Using eBGP

MP-BGP VPNv4 Prefix Exchange Between Gateway PE-ASBRs

EBGP VPNv4 (Option 10(b))

- **PE-ASBRs exchange routes directly using BGP**
  - External MP-BGP for VPNv4 prefix exchange; no LDP or IGP
- **MP-BGP session with NH to advertising PE-ASBR**
  - Next-hop and labels are rewritten when advertised across the inter-provider MP-BGP session
- **Receiving PE-ASBR automatically creates a /32 host route to a peer ASBR**
  - Which must be advertised into receiving IGP if next-hop-self is not in operation to maintain the LSP
- **PE-ASBR stores all VPN routes that need to be exchanged**
  - But only within the BGP table
  - No VRFs; labels are populated into the LFIB of the PE-ASBR
- **Receiving PE-ASBRs may allocate new label**
  - Controlled by configuration of next-hop-self (default is off)

EBGP VPNv4 Control Plane (Option 10(b))



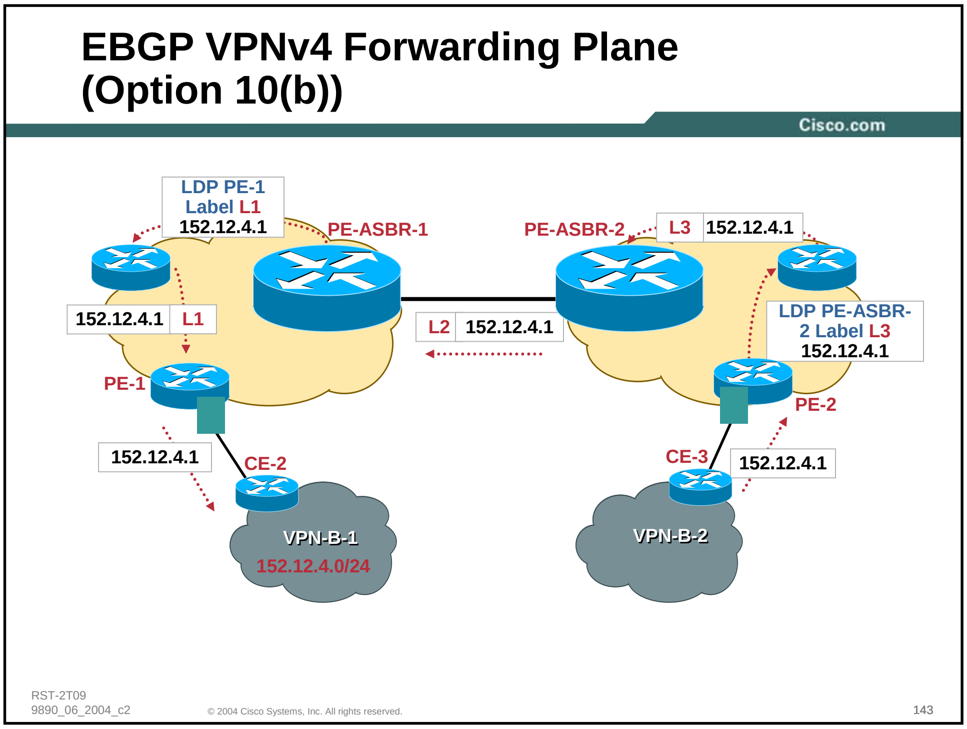
VPN-v4 update:
RD:1:27:152.12.4.0/24
, NH=PE-1
RT=1:222, Label=(L1)

VPN-v4 update:
RD:1:27:152.12.4.0/24
, NH=PE-ASBR-1
RT=1:222, Label=(L2)

VPN-v4 update:
RD:1:27:152.12.4.0/24
, NH=PE-ASBR-2
RT=1:222, Label=(L3)

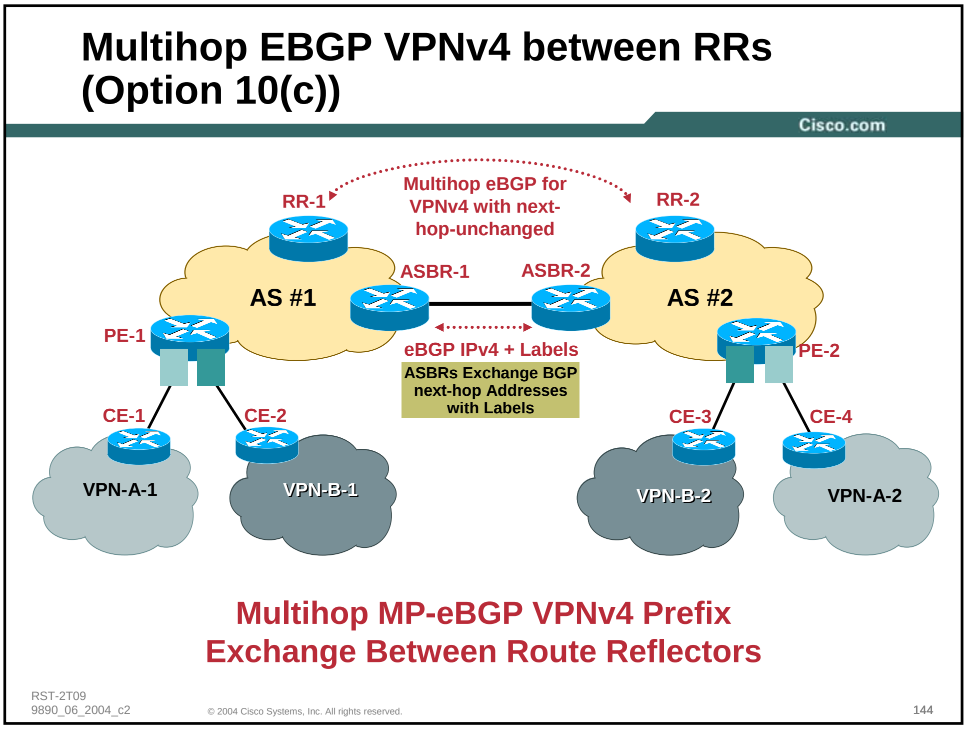
BGP, OSPF, RIPv2
152.12.4.0/24,NH=CE-2

BGP, OSPF, RIPv2
152.12.4.0/24,NH=PE-2



EBGP VPNv4 Forwarding Plane (Option 10(b))

Cisco.com
LDP PE-1
Label L1
152.12.4.1
PE-ASBR-1
PE-ASBR-2
L3
152.12.4.1
152.12.4.1
L1
L2
152.12.4.1
LDP PE-ASBR-2 Label L3
152.12.4.1
PE-1
PE-2
152.12.4.1
CE-2
CE-3
152.12.4.1
VPN-B-1
152.12.4.0/24
VPN-B-2



Multihop EBGP VPNv4 between RRs (Option 10(c))

Cisco.com
RR-1
Multihop eBGP for
VPNv4 with next-
hop-unchanged
RR-2
ASBR-1
ASBR-2
AS #1
AS #2
PE-1
PE-2
eBGP IPv4 + Labels
CE-1
CE-2
CE-3
CE-4
VPN-A-1
VPN-B-1
VPN-B-2
VPN-A-2

ASBRs Exchange BGP next-hop Addresses with Labels

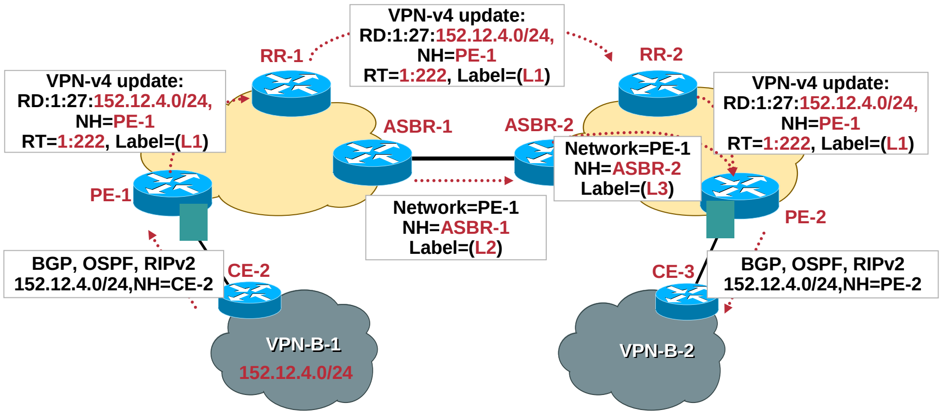
Multihop MP-eBGP VPNv4 Prefix Exchange Between Route Reflectors

Multihop EBGP VPNv4 between RRs (Option 10(c))

- MPLS VPN providers exchange VPNv4 prefixes via their route reflectors
  - Requires multihop MP-eBGP (VPNv4 routes)
- Next-hop-self must be disabled on route reflector
  - Preserves next-hop and label as allocated by the originating PE router
- Providers exchange IPv4 routes with labels between directly connected ASBRs using eBGP
  - Only PE loopback addresses exchanged as these are BGP next-hop addresses

---

Multihop EBGP VPNv4 between RRs (Option 10(c)): Control Plane



VPN-v4 update:
RD:1:27:152.12.4.0/24,
NH=PE-1
RT=1:222, Label=(L1)

VPN-v4 update:
RD:1:27:152.12.4.0/24,
NH=PE-1
RT=1:222, Label=(L1)

VPN-v4 update:
RD:1:27:152.12.4.0/24,
NH=PE-1
RT=1:222, Label=(L1)

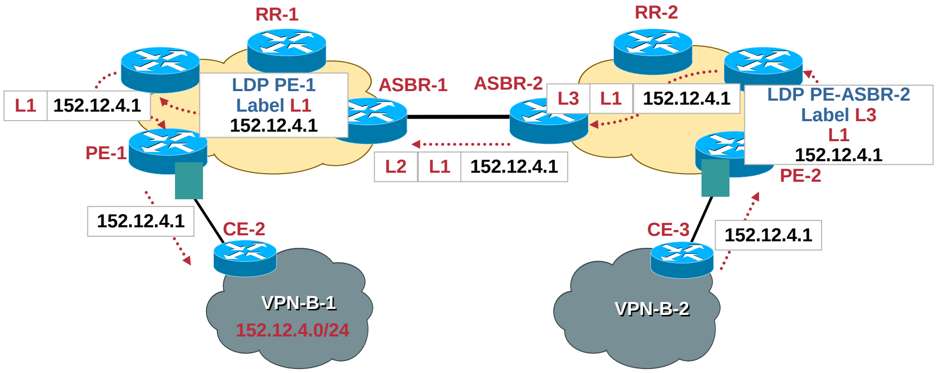
Network=PE-1
NH=ASBR-2
Label=(L3)

Network=PE-1
NH=ASBR-1
Label=(L2)

BGP, OSPF, RIPv2
152.12.4.0/24,NH=CE-2

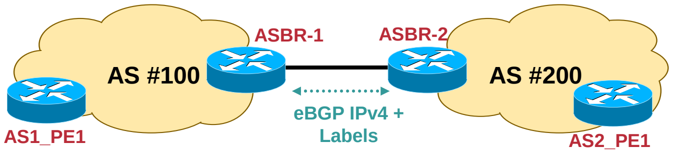
BGP, OSPF, RIPv2
152.12.4.0/24,NH=PE-2

Multihop EBGP VPNv4 between RRs (Option 10(c)): Forwarding Plane


RR-1
RR-2
L1 | 152.12.4.1
LDP PE-1
Label L1
152.12.4.1
ASBR-1
ASBR-2
L3 | L1 | 152.12.4.1
LDP PE-ASBR-2
Label L3
L1
152.12.4.1
PE-1
L2 | L1 | 152.12.4.1
PE-2
152.12.4.1
CE-2
CE-3
152.12.4.1
VPN-B-1
152.12.4.0/24
VPN-B-2

Why IPV4 BGP Label Distribution?

- Allows a VPN Service Provider network to exchange IPv4 BGP routes with MPLS labels
- Use BGP to distribute labels associated with the routes at the same time it distributes the routes


ASBR-1
ASBR-2
AS #100
AS #200
eBGP IPv4 +
Labels
AS1_PE1
AS2_PE1

Benefits:

- Eliminates the need for any other label distribution protocol between the two ASBRs
- Allows a non-VPN core network to act as a transit network for VPN traffic

IPV4 BGP Label Distribution Architecture

- Described in RFC 3107
- Subsequent Address Family Identifier (value 4) field is used to indicate that the NLRI contains a label
- If a BGP peer indicates, through BGP Capability Advertisement, that it can process update messages with the specified SAFI field, a BGP speaker can use BGP to send labels
- No specific procedures are enforced in RFC when the BGP peers are non-adjacent
- Accept labels from only trusted source to assure proper security

---

IPV4 BGP Label Distribution Config

On ASBRs(and RR if in use)

```
address-family ipv4
 neighbor <peer's loopback add> send-label
```

On AS1_PE1

```
  neighbor <RR> send-label
  neighbor <ASBR-1> send-label
```

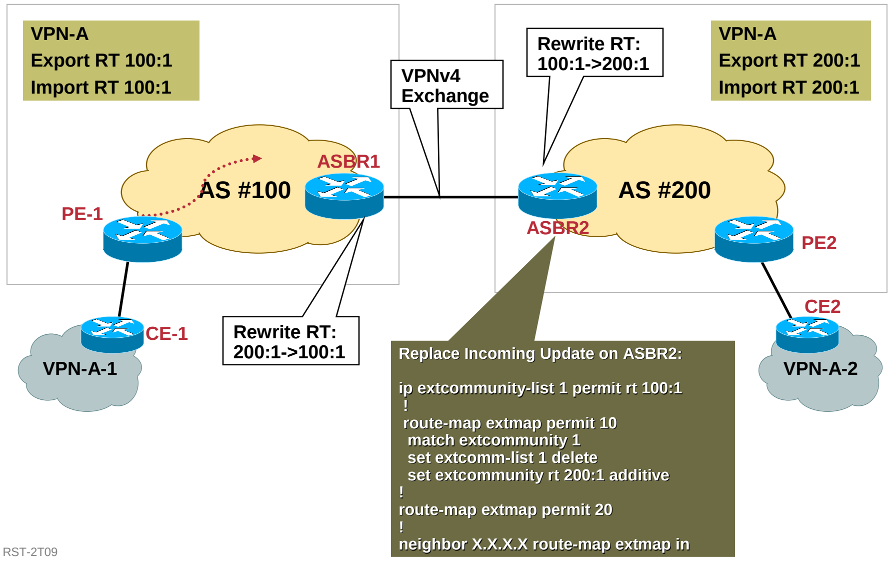
On RR

```
  neighbor <ASBR-1> send-label
  neighbor <AS1_PE1> send-label
```

Route Target Rewrite

- RTs are carried as extended community attributes in bgp-vpnv4 updates across ASBRs
- For ease of management, keeps the administration of routing policy local to the autonomous system
- Can replace RTs with incoming or outgoing VPNv4 updates
- Can replace RTs on RR or PEs, typically done on ASBRs
- RT rewrite configured via BGP in/out bound route-map

Route Target Rewrite


VPN-A
Export RT 100:1
Import RT 100:1
VPNv4
Exchange
Rewrite RT:
100:1->200:1
VPN-A
Export RT 200:1
Import RT 200:1
ASBR1
AS #100
PE-1
AS #200
ASBR2
PE2
CE2
CE-1
VPN-A-1
Rewrite RT:
200:1->100:1
VPN-A-2

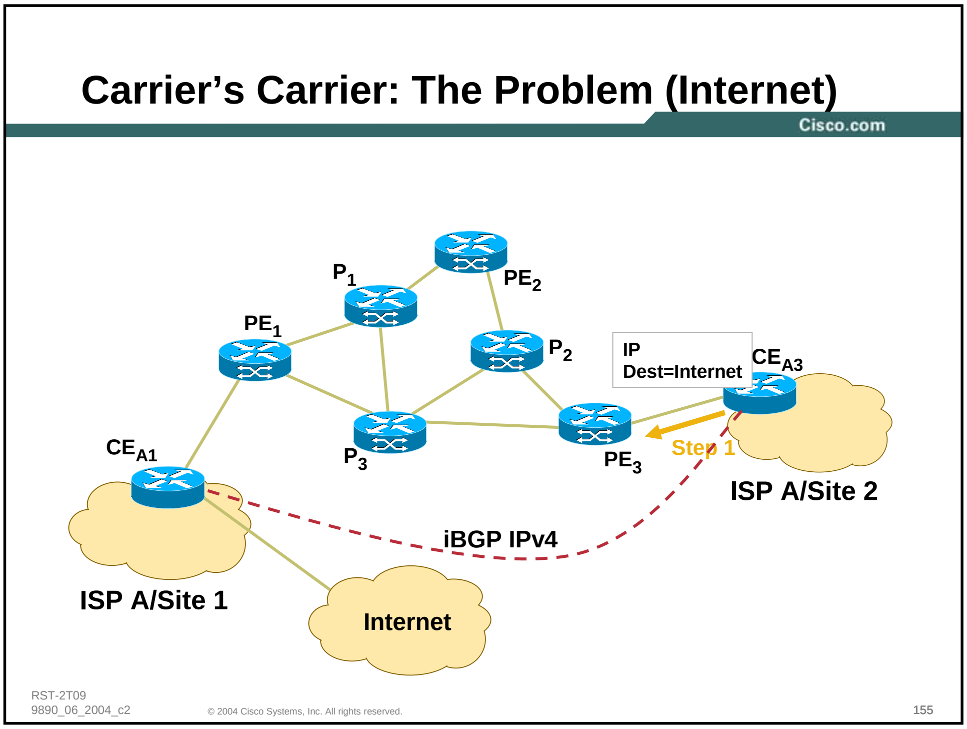
Replace Incoming Update on ASBR2:

```
ip extcommunity-list 1 permit rt 100:1
 !
 route-map extmap permit 10
  match extcommunity 1
  set extcomm-list 1 delete
  set extcommunity rt 200:1 additive
!
route-map extmap permit 20
!
neighbor X.X.X.X route-map extmap in
```

INTER-PROVIDER SOLUTIONS: CARRIER SUPPORTING CARRIER

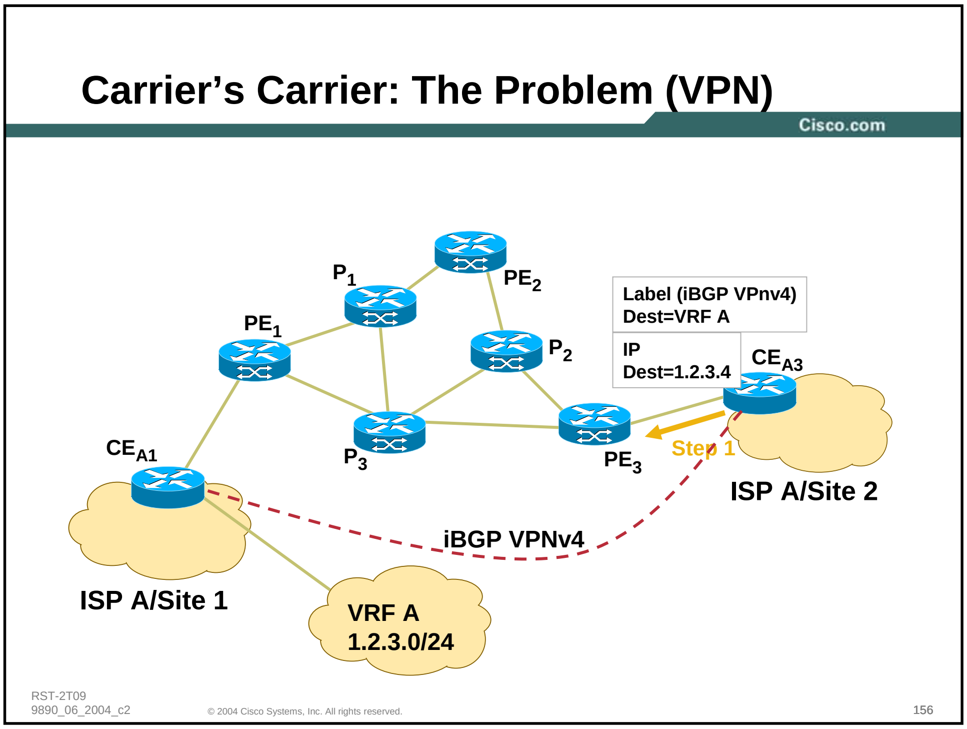
The Problem

- MPLS-VPN works well for carrying customer IGPs
- Platforms, network scale to N*O(IGP) routes
- What if the CE wants the PE to carry all their BGP routes?
- Or if CE wants to run their own VPN service?
- Or if the CE just has a lot of route? (example: 30 sites, 20k prefixes)



Carrier's Carrier: The Problem (Internet)

P$_1$
PE$_2$
PE$_1$
P$_2$
IP
Dest=Internet
CE$_{A3}$
CE$_{A1}$
P$_3$
PE$_3$
Step 1
ISP A/Site 2
iBGP IPv4
ISP A/Site 1
Internet



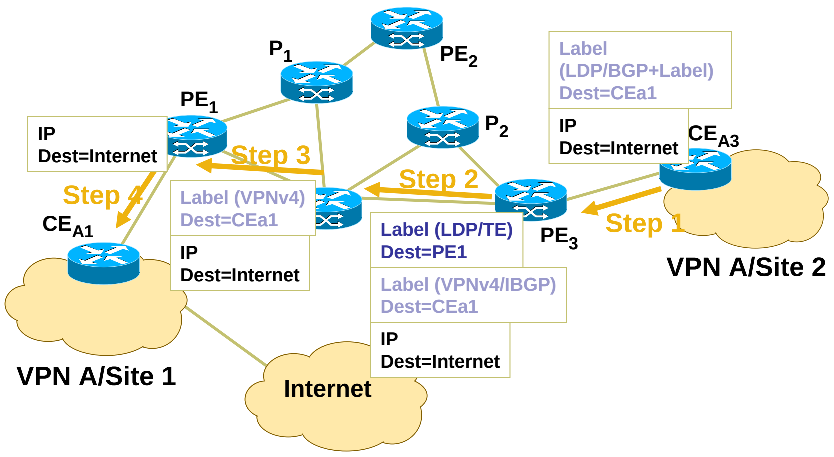
Carrier's Carrier: The Problem (VPN)

P$_1$
PE$_2$
PE$_1$
P$_2$
Label (iBGP VPnv4)
Dest=VRF A
IP
Dest=1.2.3.4
CE$_{A3}$
CE$_{A1}$
P$_3$
PE$_3$
Step 1
ISP A/Site 2
iBGP VPNv4
ISP A/Site 1
VRF A
1.2.3.0/24

Carrier's Carrier: The Solution

- **MPLS between PE and CE**
  - **Either IGP+LDP or BGP+Label**
- **CEs exchange labels for their IGP routes with the PEs**
- **CEs iBGP peer with each other**
- **PEs are back to O(IGP) information**

Carrier's Carrier: The Solution (Internet)



P$_1$
PE$_2$
PE$_1$
P$_2$
PE$_3$
CE$_{A1}$
CE$_{A3}$
Step 1
Step 2
Step 3
Step 4
VPN A/Site 1
VPN A/Site 2
Internet

Label (LDP/BGP+Label)
Dest=CEa1

IP
Dest=Internet

Label (LDP/TE)
Dest=PE1

Label (VPNv4/IBGP)
Dest=CEa1

IP
Dest=Internet

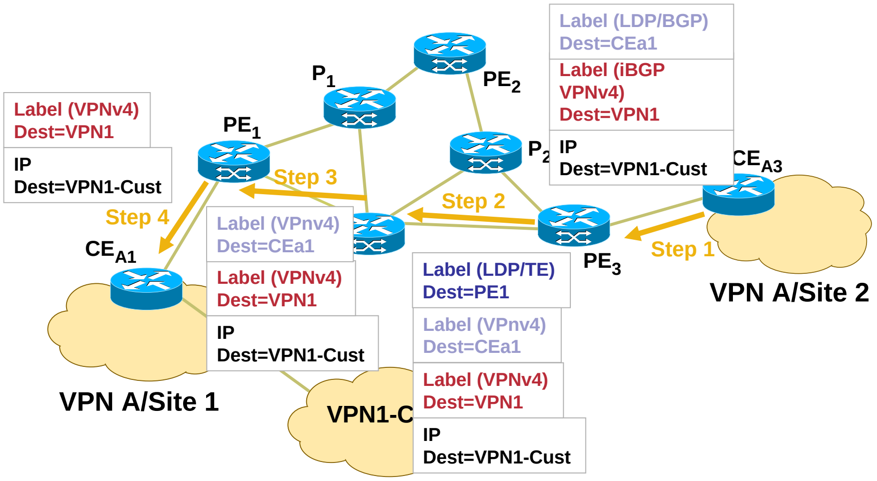
Label (VPNv4)
Dest=CEa1

IP
Dest=Internet

IP
Dest=Internet

Carrier's Carrier: The Solution (VPN)

Cisco.com


P1
PE2
PE1
P2
CEA3
CEA1
PE3
Step 1
Step 2
Step 3
Step 4
VPN A/Site 1
VPN A/Site 2
VPN1-C

Label (VPNv4)
Dest=VPN1

IP
Dest=VPN1-Cust

Label (LDP/BGP)
Dest=CEa1

Label (iBGP
VPNv4)
Dest=VPN1

IP
Dest=VPN1-Cust

Label (VPnv4)
Dest=CEa1

Label (VPNv4)
Dest=VPN1

IP
Dest=VPN1-Cust

Label (LDP/TE)
Dest=PE1

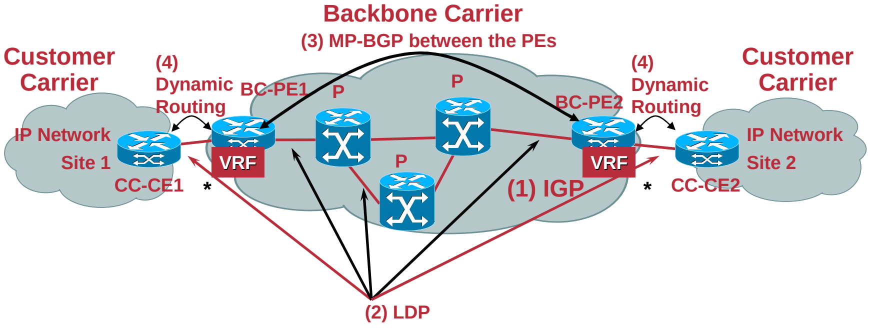
Label (VPnv4)
Dest=CEa1

Label (VPNv4)
Dest=VPN1

IP
Dest=VPN1-Cust

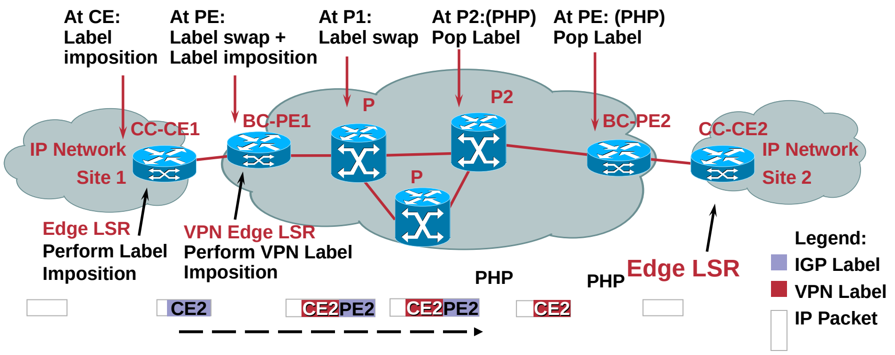
Carrier's Carrier Service Model I Customer Carrier Is Running IP Only

Cisco.com


Backbone Carrier
(3) MP-BGP between the PEs
Customer Carrier
(4) Dynamic Routing
BC-PE1
P
P
BC-PE2
(4) Dynamic Routing
Customer Carrier
IP Network Site 1
VRF
VRF
IP Network Site 2
CC-CE1
*
P
(1) IGP
*
CC-CE2
(2) LDP

- Control Plane configuration is similar to MPLS VPN
- BGP-4 used by customer to distribute external routing information between all sites
- BGP next-hop addresses exchanged between customer and carrier PE routers and are placed into VRFs and distributed using MP-BGP
- The only addition is (*) LDP at the PE-CE links

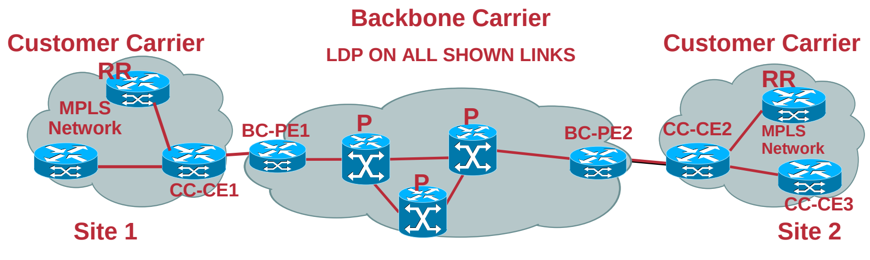
Carrier's Carrier Service Model I
Customer Carrier Is Running IP Only


At CE:
Label
imposition
At PE:
Label swap +
Label imposition
At P1:
Label swap
At P2:(PHP)
Pop Label
At PE: (PHP)
Pop Label
CC-CE1
BC-PE1
P
P2
P
BC-PE2
CC-CE2
IP Network
Site 1
IP Network
Site 2
Edge LSR
Perform Label
Imposition
VPN Edge LSR
Perform VPN Label
Imposition
PHP
PHP
Edge LSR
CE2
CE2 PE2
CE2 PE2
CE2
Legend:
IGP Label
VPN Label
IP Packet

- Forward using labels (as opposed to IP add) starting at the CE
- Label Switch Path is extended to the CEs
- Penultimate hop popping at PE2, requested by the Edge LSR (CE2) during label distribution

---

Carrier's Carrier Service Model II
Customer Carrier Is Running MPLS

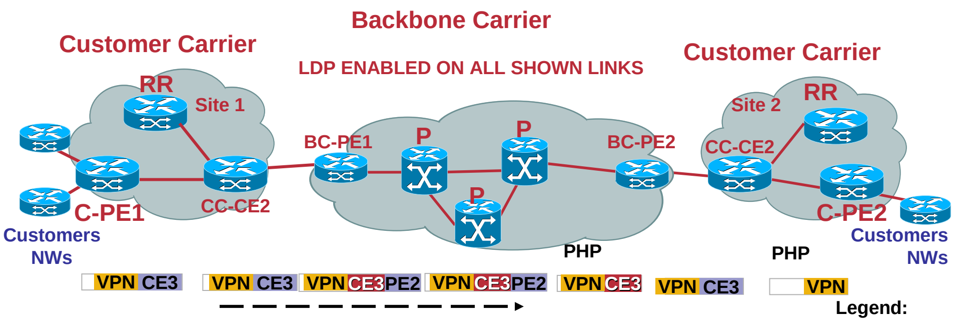

Backbone Carrier
LDP ON ALL SHOWN LINKS
Customer Carrier
Customer Carrier
RR
MPLS
Network
CC-CE1
Site 1
BC-PE1
P
P
P
BC-PE2
CC-CE2
RR
MPLS
Network
CC-CE3
Site 2
CE3
CE3
CE3 PE2
CE3 PE2
CE3
CE3
CE3

- LSP is extended into the sites, CC-CE advertises labels to PE internal routes
- CE1 perform label swap for the site IGP label, PE swaps the site IGP label with a VPN label and pushes IGP label
- PHP is now extended to inside of site 2
- External routes are carried by BGP between sites, same as before
- The same BGP peering from before can be used here

Legend:
🟪 IGP Label
🟥 VPN Label
☐ IP Packet

Carrier’s Carrier Service Model III
Customer Carrier Supports MPLS VPNs

Cisco.com


Backbone Carrier
Customer Carrier
LDP ENABLED ON ALL SHOWN LINKS
Customer Carrier
RR
Site 1
Site 2
RR
BC-PE1
P
P
BC-PE2
CC-CE2
C-PE1
CC-CE2
P
C-PE2
Customers NWs
PHP
PHP
Customers NWs
VPN CE3
VPN CE3
VPN CE3 PE2
VPN CE3 PE2
VPN CE3
VPN CE3
VPN

Legend:

- 🟪 IGP Label
- 🟥 VPN Label
- 🟨 Site VPN Label
- ☐ IP Packet

- LSP is extended to C-PE, CE advertises labels for internal routes to PE; C-PE1 perform imposition for site VPN label and IGP label
- PE swaps the site IGP label with a BB VPN label and push IGP label; PHP is now extended to inside of site 2
- External and VPNv4 routes are carried by MP-BGP between customer carrier sites

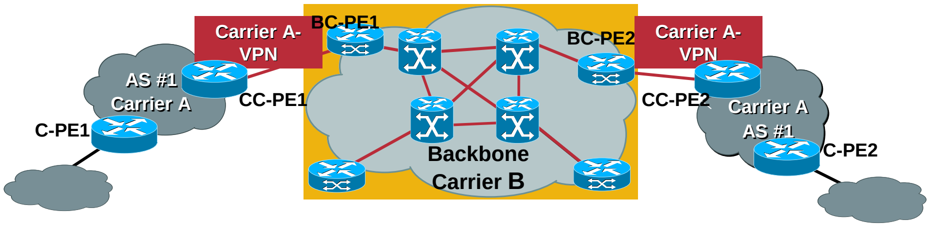
---

VRF-Aware MPLS Static Labels

Cisco.com

- Static label bindings to IPv4 prefixes on hop-by-hop forwarding through neighbor router where LDP is not used:
  - Eliminates the need to enable LDP between customer and backbone provider
  - Easier management, monitoring customer NWs, +security
- MPLS static labels introduced in 12.0.23S, but only supported global routing tables (not VRF aware):
  - Config. only MPLS forwarding table entries for the global table
  - Assign label values to FECs learned by the LDP for the global table
  - Limits usage to the provider core only
- Feature is enhanced so static labels can be used for VRF traffic at the VPN edge for CSC networks:
  - Static labels can be used at the VPN edge
  - Static bindings between labels and IPv4 prefixes can be configured statically

VRF-Aware MPLS Static Labels: Config.


Carrier A-VPN
BC-PE1
BC-PE2
Carrier A-VPN
AS #1
Carrier A
CC-PE1
CC-PE2
Carrier A
AS #1
C-PE1
C-PE2
Backbone
Carrier B

I. Define label range

mpls label range *min-label max-label* [static *min-static-label max-static-label*]

```
Router(config)# mpls label range 200 100000 static 16 199
```

II. Bind a prefix to local labels

mpls static binding ipv4 vrf *vpn-name prefix mask* **[[input]** label**]**

```
Router(config)# mpls static binding ipv4 vrf vpnRed
       13.0.66.0.0 255.255.0.0 input 17
```

Bindings specified are installed automatically in the MPLS forwarding table as routing demands

---

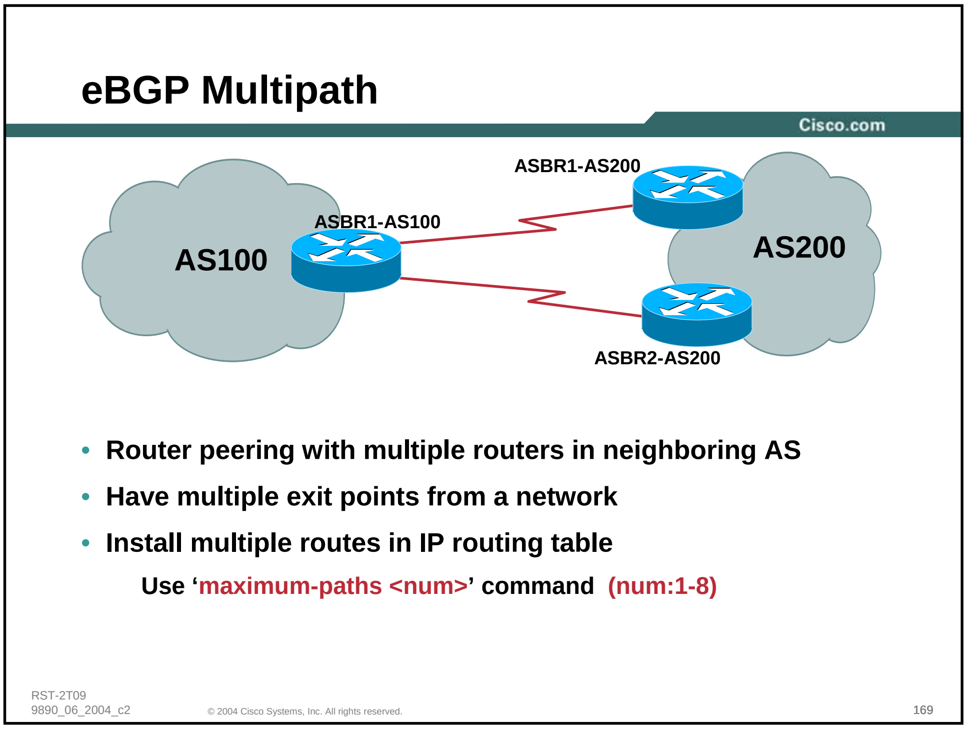
eiBGP MULTIPATH LOAD BALANCING

eiBGP Multipath Load Balancing

- **Improves load balancing deployment and service offering capabilities**
- **Useful for multihomed AS systems and PE routers that import both eBGP and iBGP paths from multihomed and stub networks**
- **Supported for CSC and inter-AS as well**

eiBGP Multipath Load Balancing Topics

- **eBGP Multipath**
- **iBGP Multipath**
- **eiBGP Multipath (MPLS/VPN)**

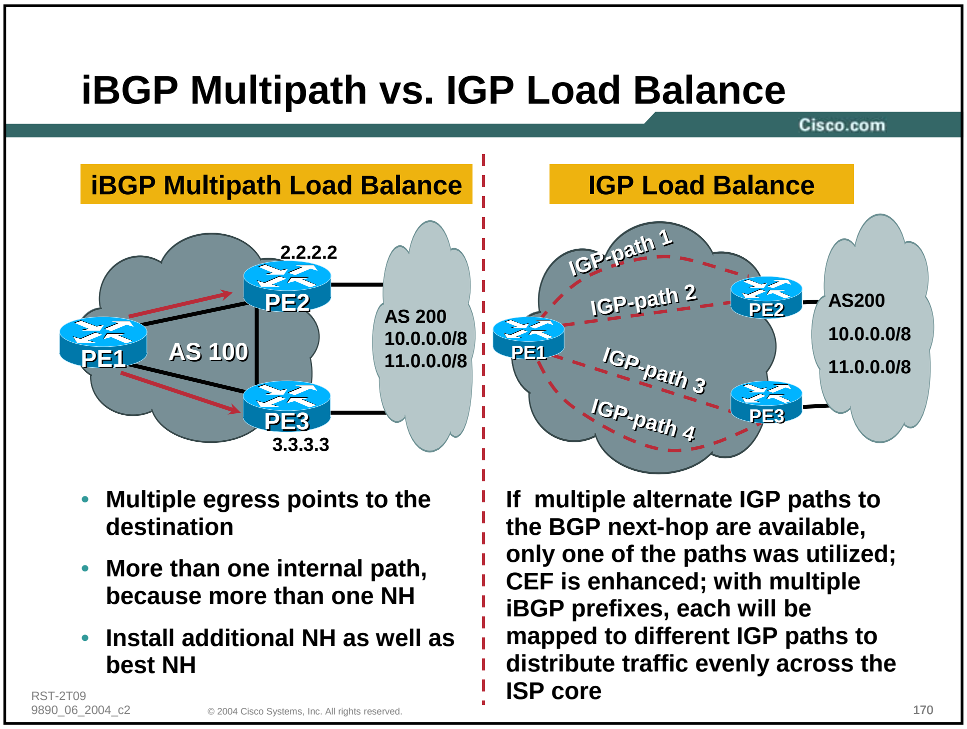


eBGP Multipath

Cisco.com
ASBR1-AS200
ASBR1-AS100
AS100
AS200
ASBR2-AS200

- Router peering with multiple routers in neighboring AS
- Have multiple exit points from a network
- Install multiple routes in IP routing table

  Use ‘maximum-paths <num>’ command (num:1-8)



iBGP Multipath vs. IGP Load Balance

Cisco.com

iBGP Multipath Load Balance

2.2.2.2
PE2
PE1
AS 100
PE3
3.3.3.3
AS 200
10.0.0.0/8
11.0.0.0/8

- Multiple egress points to the destination
- More than one internal path, because more than one NH
- Install additional NH as well as best NH

IGP Load Balance

IGP-path 1
IGP-path 2
PE2
AS200
10.0.0.0/8
11.0.0.0/8
PE1
IGP-path 3
IGP-path 4
PE3

If multiple alternate IGP paths to the BGP next-hop are available, only one of the paths was utilized; CEF is enhanced; with multiple iBGP prefixes, each will be mapped to different IGP paths to distribute traffic evenly across the ISP core

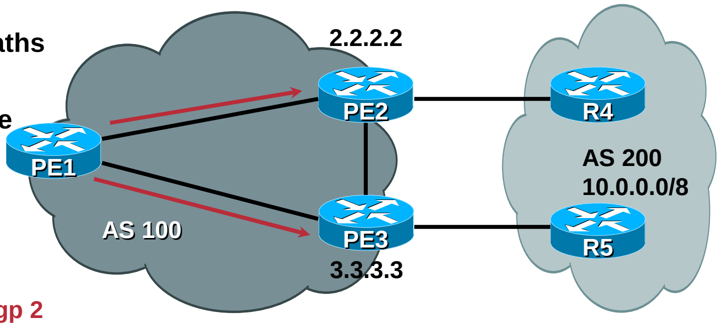
iBGP Multipath: Concept and Requirements

- **Allows BGP to install more than one internal path to a destination**
  - **Useful for load sharing in the core for multiple egress PEs**
- **The paths MUST be equivalent: all the absolute attributes MUST tie during the best path selection process**
  - **router-id, cluster-id, peer-address are not absolute attributes**
- **The best path (as determined by the selection process) is advertised**
  - **All eligible multipaths are installed in the RIB/FIB**
  - **Each path has a unique NEXT_HOP**
- **Use 'maximum-paths ibgp <num>' command (num:1-8)**

---

iBGP Multipath: Example

- **R1 has two paths for 10.0.0.0/8**
- **Both paths are flagged as "multipath"**



maximum-paths ibgp 2

```
PE1#sh ip bgp 10.0.0.0
  200
    20.20.20.3 from 20.20.20.3 (3.3.3.3)
    Origin IGP, metric 0, localpref 100, valid, internal, multipath
  200
    20.20.20.2 from 20.20.20.2 (2.2.2.2)
    Origin IGP, metric 0, localpref 100, valid, internal, multipath, best
```

eiBGP MULTIPATH LOAD BALANCING

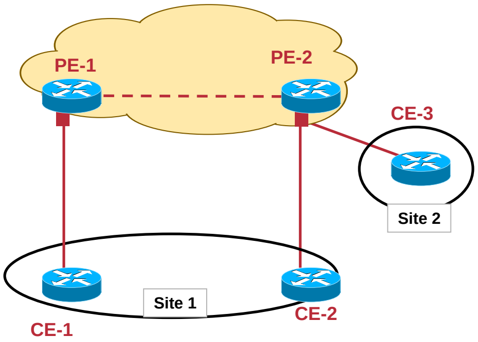
eiBGP Multipath Load Balancing

- The traffic destined to a site may be load shared between all entry points
  - From the MPLS/VPNs provider's point of view, these entry points may not all correspond to internal or external peers
  - The intent is for the MPLS/VPN network to be transparent to the customers
- The ability to consider both iBGP and eBGP paths, when using multipath, is needed
  - Applies only to the MPLS/VPN case

eiBGP Multipath Example

- **PE-2 has two possible paths into site 1**
  - **eiBGP multipath allows both paths to be used**

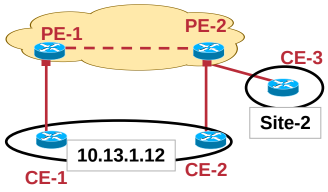
maximum-paths eibgp <num>
(num is <1- 8>)



---

eiBGP Multipath PE2 Configuration

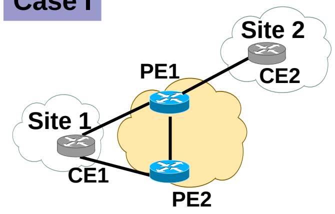
```
PE2# show  run
router bgp 1
 no synchronization
 bgp log-neighbor-changes
 neighbor 13.1.1.5 remote-as 1
 neighbor 13.1.1.5 update-source Loopback0
 no auto-summary
 !
 address-family vpnv4
 neighbor 13.1.1.5 activate
 neighbor 13.1.1.5 send-community extended
 exit-address-family
 !
address-family ipv4 vrf red
 neighbor 100.1.1.1 remote-as 2
 neighbor 100.1.1.1 activate
 neighbor 100.1.1.1 as-override
 neighbor 200.1.1.1 remote-as 2
 neighbor 200.1.1.1 activate
 neighbor 200.1.1.1 as-override
 maximum-paths eibgp 2
 no auto-summary
 no synchronization
 exit-address-family
```



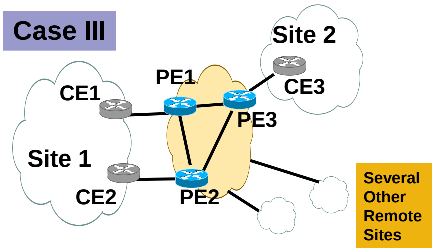
eiBGP Multipath: Cases

Cisco.com

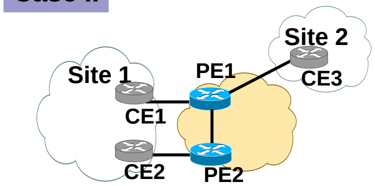
Case I


Site 2
CE2
PE1
Site 1
CE1
PE2

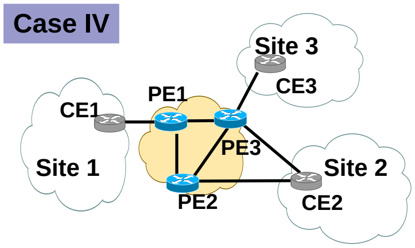
Case III


Site 2
CE3
PE1
CE1
PE3
Site 1
CE2
PE2
Several Other Remote Sites

Case II


Site 2
Site 1
PE1
CE3
CE1
CE2
PE2

Case IV


Site 3
CE3
PE1
CE1
PE3
Site 1
Site 2
PE2
CE2

Q & A

CISCO SYSTEMS

NETWORK INTEGRATION WITH MPLS: ADVANCED TE APPLICATIONS

Cool Things You Can Do with MPLS-TE

- Providing services with MPLS-TE
- Inter-AS TE
- Fast reroute
- Automation
- Miscellaneous yet highly cool stuff

Providing Services with MPLS-TE

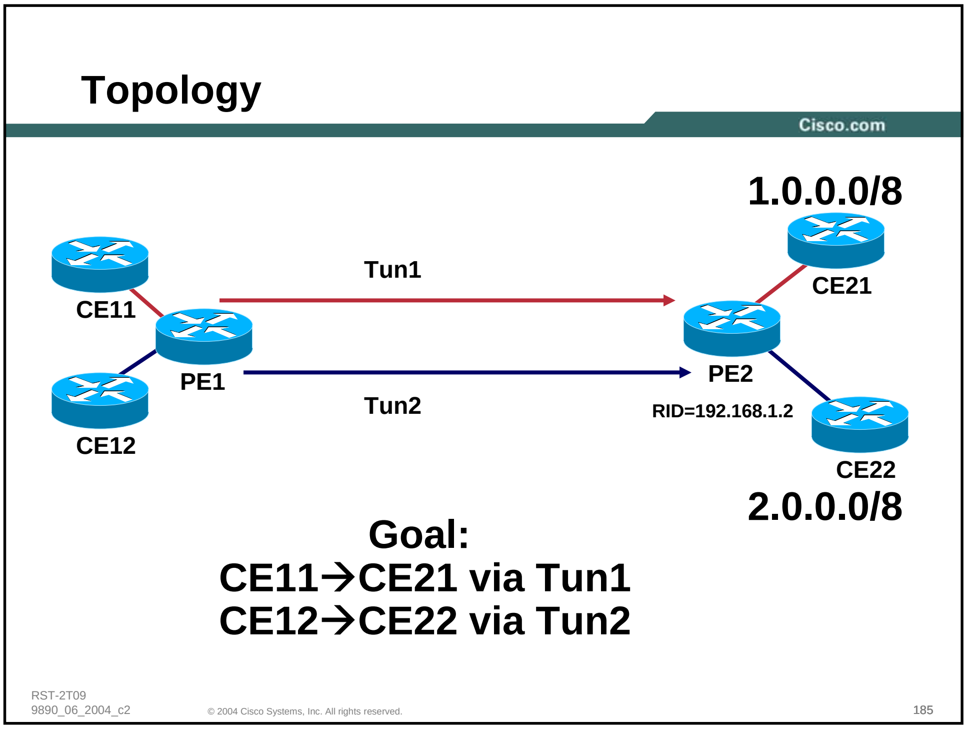
- Integrating TE and L3VPNs
- TE+QoS with CBTS
- Integrating TE and L2VPNs

Integrating TE and L3VPNs

- How do I get traffic for my L3VPNs down my TE tunnels?
- A few possible answers, depending on the granularity you want:
  - All traffic to a given destination PE
    - Answer is the same as for L2—see RST2603
  - All traffic for a set of VPNs

All Traffic for a Set of VPNs

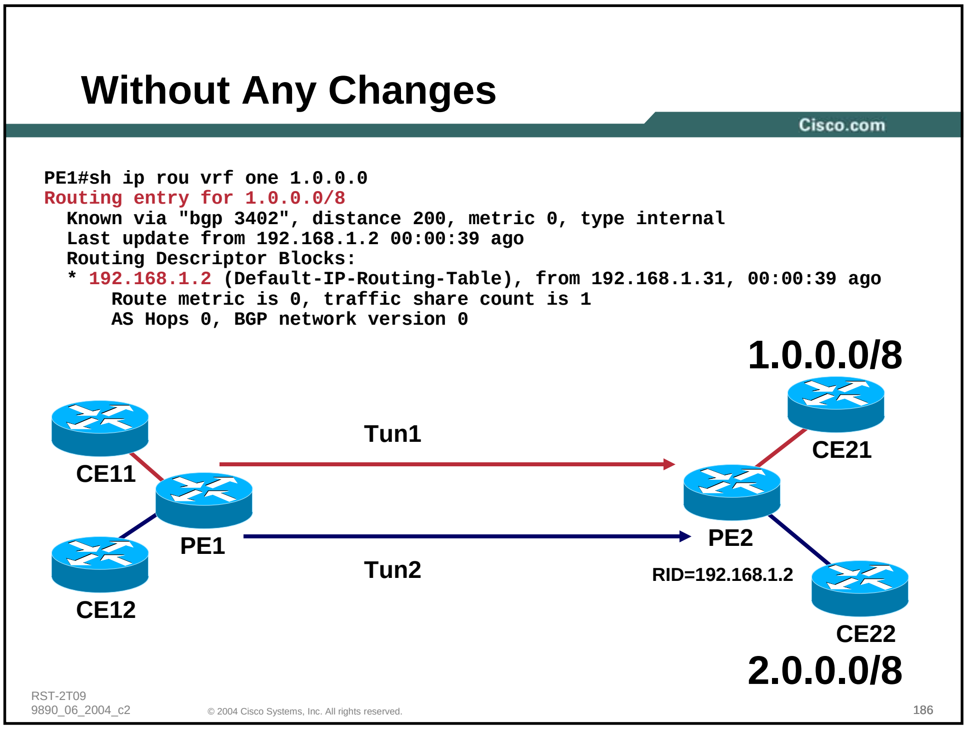
- L3VPN routes via other PEs are all learned via BGP
- BGP carries a next-hop
- All L3VPN routes recourse through their next-hop
- To attract traffic down a certain set of tunnels
  1. Set the next-hop to an address X
  2. Route address X down a given tunnel



Topology

Cisco.com
1.0.0.0/8
CE21
Tun1
CE11
PE1
PE2
CE12
Tun2
RID=192.168.1.2
CE22
2.0.0.0/8

Goal:
CE11→CE21 via Tun1
CE12→CE22 via Tun2

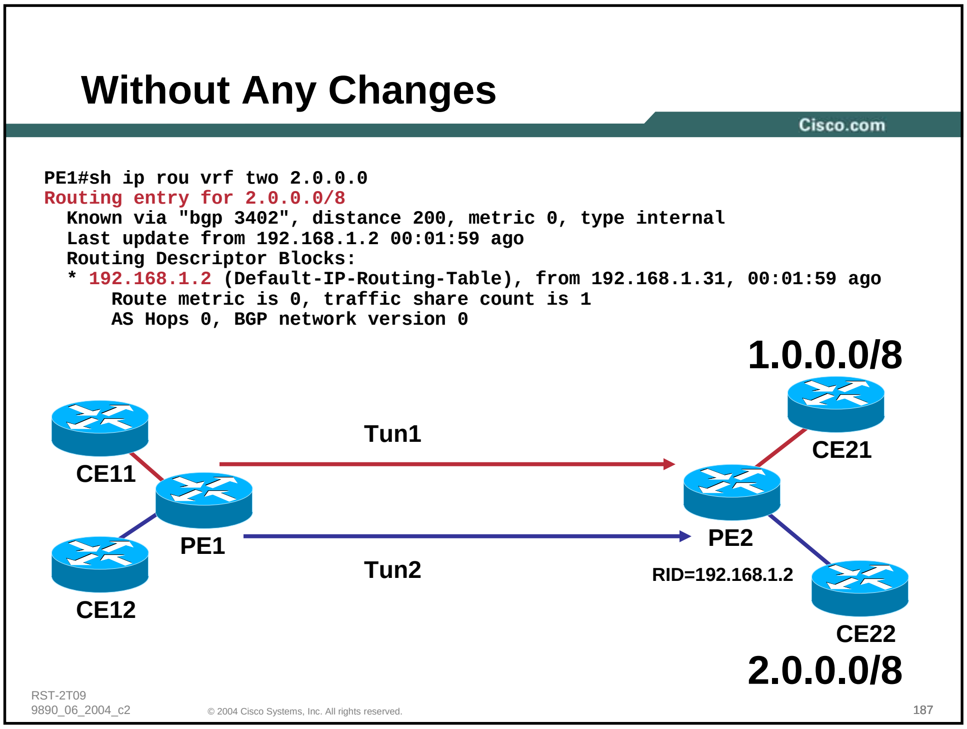


Without Any Changes

Cisco.com

```
PE1#sh ip rou vrf one 1.0.0.0
Routing entry for 1.0.0.0/8
  Known via "bgp 3402", distance 200, metric 0, type internal
  Last update from 192.168.1.2 00:00:39 ago
  Routing Descriptor Blocks:
  * 192.168.1.2 (Default-IP-Routing-Table), from 192.168.1.31, 00:00:39 ago
      Route metric is 0, traffic share count is 1
      AS Hops 0, BGP network version 0
```

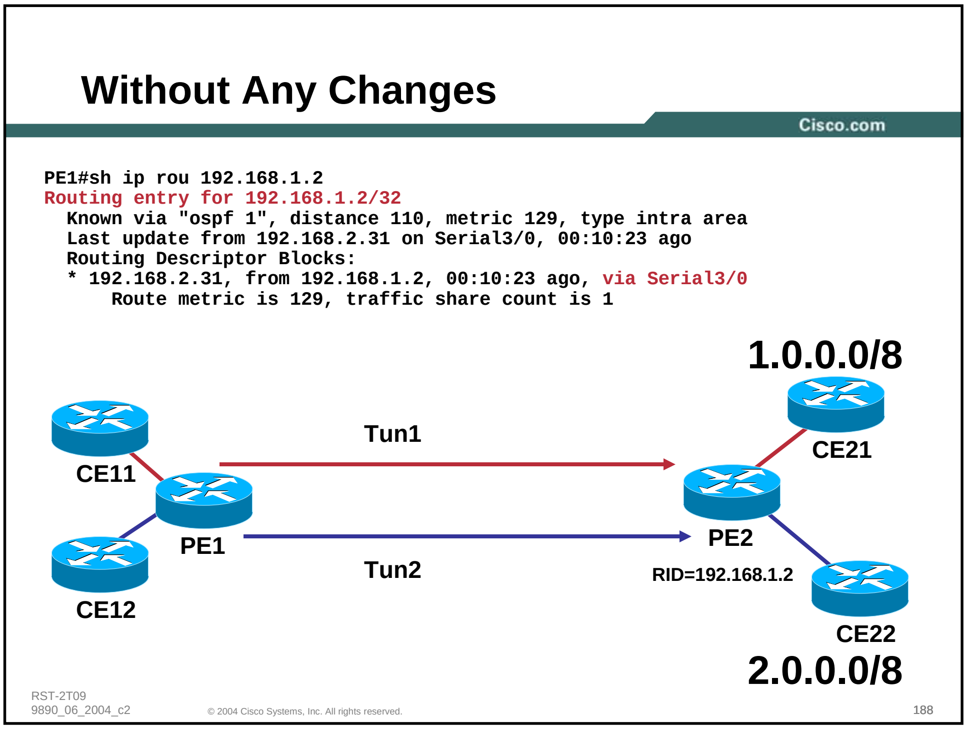
1.0.0.0/8
CE21
Tun1
CE11
PE1
PE2
CE12
Tun2
RID=192.168.1.2
CE22
2.0.0.0/8



Without Any Changes

```
PE1#sh ip rou vrf two 2.0.0.0
Routing entry for 2.0.0.0/8
  Known via "bgp 3402", distance 200, metric 0, type internal
  Last update from 192.168.1.2 00:01:59 ago
  Routing Descriptor Blocks:
  * 192.168.1.2 (Default-IP-Routing-Table), from 192.168.1.31, 00:01:59 ago
      Route metric is 0, traffic share count is 1
      AS Hops 0, BGP network version 0
```

1.0.0.0/8
Tun1
CE21
CE11
PE1
PE2
Tun2
RID=192.168.1.2
CE12
CE22
2.0.0.0/8

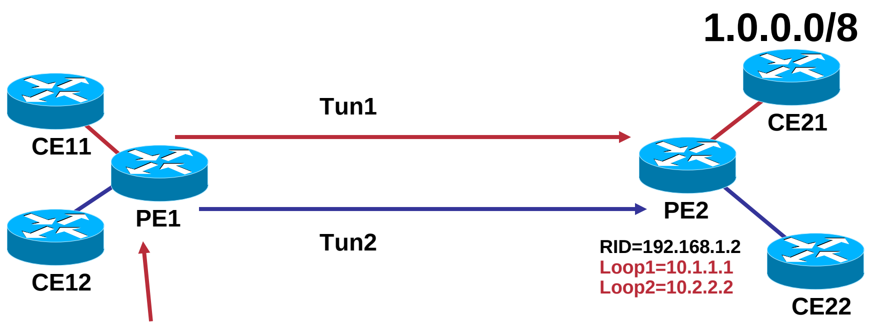


Without Any Changes

```
PE1#sh ip rou 192.168.1.2
Routing entry for 192.168.1.2/32
  Known via "ospf 1", distance 110, metric 129, type intra area
  Last update from 192.168.2.31 on Serial3/0, 00:10:23 ago
  Routing Descriptor Blocks:
  * 192.168.2.31, from 192.168.1.2, 00:10:23 ago, via Serial3/0
      Route metric is 129, traffic share count is 1
```

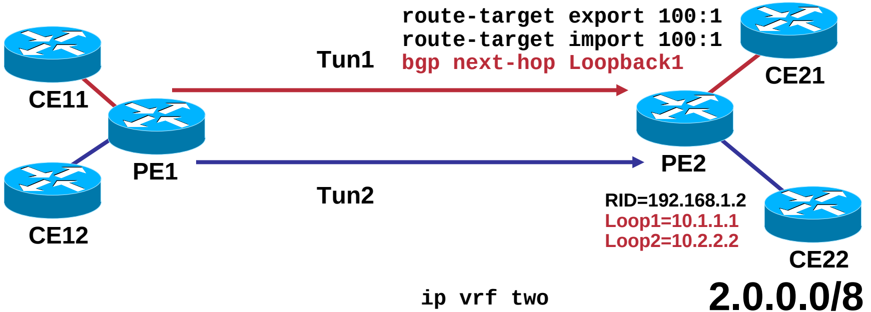
1.0.0.0/8
Tun1
CE21
CE11
PE1
PE2
Tun2
RID=192.168.1.2
CE12
CE22
2.0.0.0/8

Add Loopbacks on PE2 and Routes on PE1


1.0.0.0/8
CE11
PE1
CE12
Tun1
Tun2
PE2
CE21
CE22
RID=192.168.1.2
Loop1=10.1.1.1
Loop2=10.2.2.2
2.0.0.0/8

```
ip route 10.1.1.1 255.255.255.255 Tunnel1
ip route 10.2.2.2 255.255.255.255 Tunnel2
```

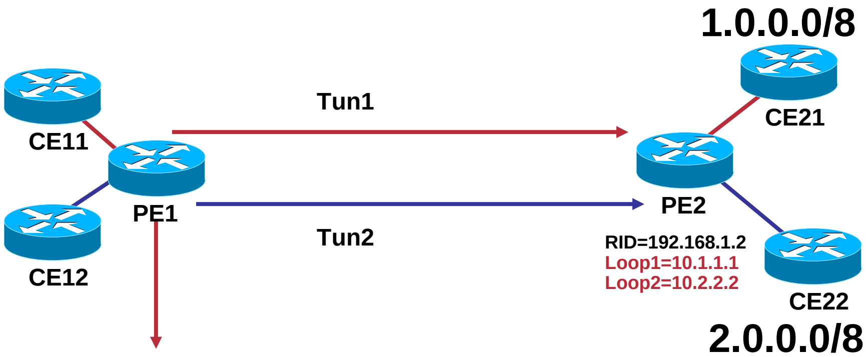
Set Next-Hops on PE2 or PE1


1.0.0.0/8
CE11
PE1
CE12
Tun1
Tun2
PE2
CE21
CE22
RID=192.168.1.2
Loop1=10.1.1.1
Loop2=10.2.2.2
2.0.0.0/8

```
ip vrf one
 rd 100:1
 route-target export 100:1
 route-target import 100:1
 bgp next-hop Loopback1
```

```
ip vrf two
 rd 100:2
 route-target export 100:2
 route-target import 100:2
 bgp next-hop Loopback2
```

Voila!


1.0.0.0/8
CE21
Tun1
CE11
PE2
PE1
CE12
Tun2
RID=192.168.1.2
Loop1=10.1.1.1
Loop2=10.2.2.2
CE22
2.0.0.0/8

```
PE1#sh ip cef vrf one 1.0.0.0
1.0.0.0/8, version 12, epoch 0, cached adjacency to Tunnel1
…
PE1#sh ip cef vrf two 2.0.0.0 int
2.0.0.0/8, version 9, epoch 0, cached adjacency to Tunnel2
```

TE+QoS with CBTS

- Currently, traffic is sent down a TE tunnel based on destination address
  - Exception here is PBR, which isn't a great idea large-scale
- Class-Based Tunnel Selection (CBTS) adds flexibility here
- Multiple tunnels to the same destination, which can carry different classes of traffic
  - …i.e. VoIP goes over a shorter path, which is a tunnel that's requested FRR, but bulk data takes a different path and may not request FRR
- A lot like VC selection for ATM

CBTS Config

- On a TE tunnel, just add

```
R3(config-if)#tunnel mpls traffic-eng exp ?
  <0-7>   exp
  default default all unconfigured exp to this interface
```

- You can have up to 8 EXP or ‘default’ on a tunnel
- The rest of the tunnel config proceeds as before
- Must enable autoroute or otherwise have routing on the TE tunnel

Integrating TE and L2VPNs

- How do I get traffic for my L2VPNs down my TE tunnels?
- A few possible answers, depending on the granularity you want:
  - All traffic to a given destination PE
  - All traffic for a set of pseudowires

All Traffic to a Given Destination PE

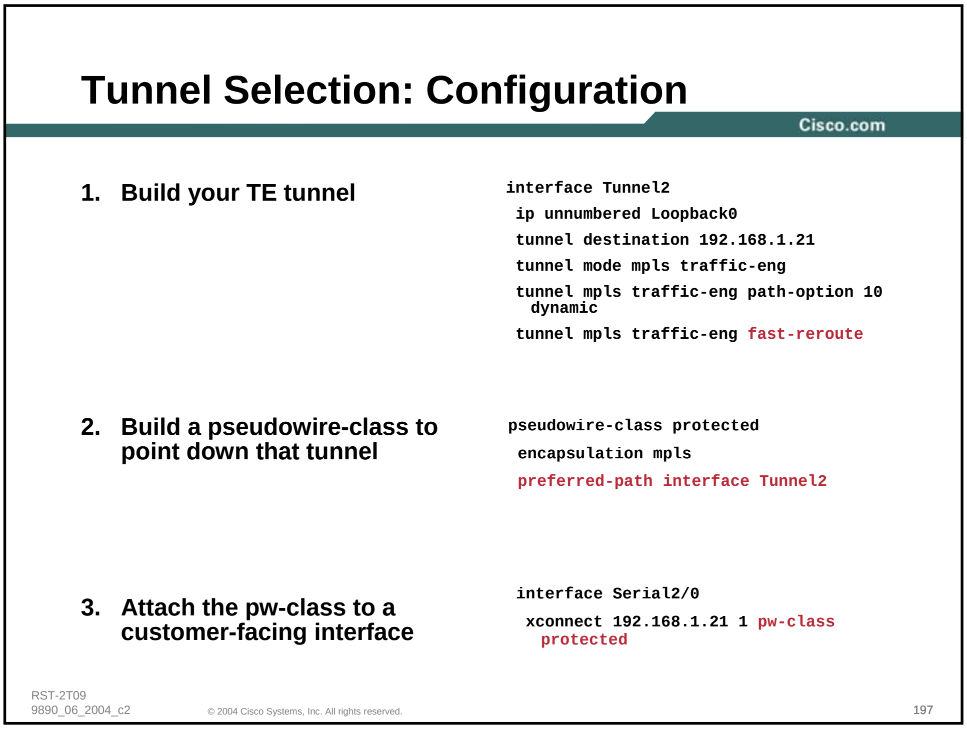
Cisco.com

- Autoroute
- Forwarding-adjacency
- Static routes
- … just like any other traffic down a TE tunnel
- Covered in RST-2603, “Deploying MPLS-TE”

All Traffic for a Set of Pseudowires

Cisco.com

- Use tunnel selection
- Set of PW can be 1 or more to the same destination
- Shipped in 12.0(25)S
- Easy way to connect a given AToM pseudowire to a specific TE tunnel
- Enables you to sell pseudowires attached to TE tunnels with different services
- Could offer a BW guarantee vs. best-effort PW
- Could offer protected vs. unprotected PW
- Could offer both



Tunnel Selection: Configuration

1. Build your TE tunnel

```
interface Tunnel2
 ip unnumbered Loopback0
 tunnel destination 192.168.1.21
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng path-option 10
   dynamic
 tunnel mpls traffic-eng fast-reroute
```

2. Build a pseudowire-class to point down that tunnel

```
pseudowire-class protected
 encapsulation mpls
 preferred-path interface Tunnel2
```

3. Attach the pw-class to a customer-facing interface

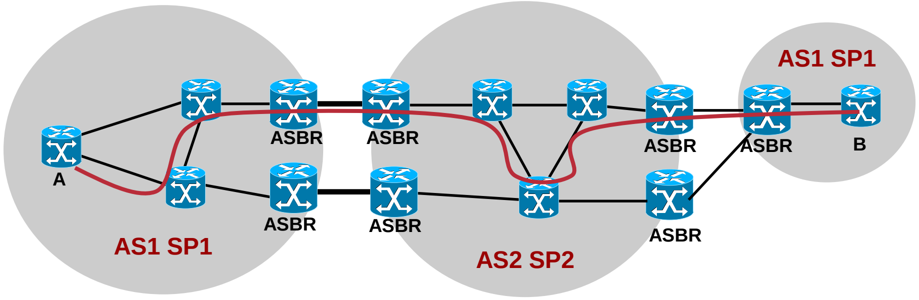
```
interface Serial2/0
 xconnect 192.168.1.21 1 pw-class
   protected
```

INTER-AS TE

Inter-AS TE

Cisco.com

- Goal: Build a TE LSP from A to B
- Why: connect separate parts of an AS across another's backbone
- How: inter-AS TE
  - Looks a lot like inter-area TE


AS1 SP1
A
ASBR
ASBR
ASBR
ASBR
ASBR
ASBR
ASBR
B
ASBR
AS1 SP1
AS2 SP2

Inter-AS TE

Cisco.com

Extends Existing TE Capabilities with

- ASBR node protection
- Loose path
- ASBR forced link flooding
- Cisco IOS RSVP local policy extensions for inter-AS
- Per-neighbor RSVP keys

- Will ship in 12.0(29)S (fall '04)

FAST REROUTE

Fast Reroute

Cisco.com

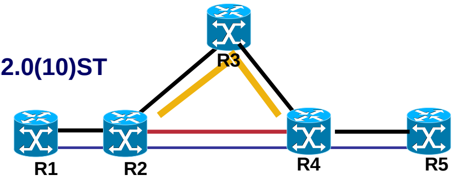
- **Link and node protection are shipping today, have been for a while**
- **Path protection is on its way (12.0(30)S)**
- **Goal of all protection schemes: use MPLS-TE to build loop-free paths to send traffic down in the event of a failure**

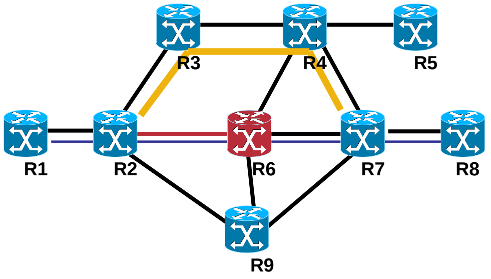
Types of FRR: Local Repair

MPLS Fast Reroute Local Repair

- **Link protection**: the backup tunnel tail-head (MP) is one hop away from the PLR

12.0(10)ST





- **Node protection + ENHANCEMENTS**: the backup tunnel tail-end (MP) is two hops away from the PLR

12.0(22)S

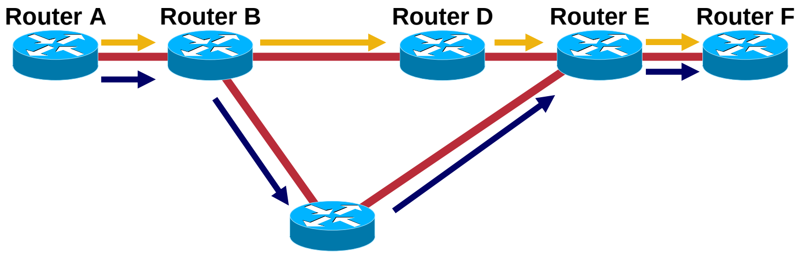
Types of FRR: Local Repair

- Link protection: pre-establish a TE tunnel to the Next-Hop (NHOP)
- Node protection: pre-establish a TE tunnel to the Next-Next-Hop (NNHOP)
- In both cases, when a failure is detected, send traffic that used to pass through the NHop/NNHop down the pre-established backup tunnel
- Loss is O(msec)—depends on design, HW, SW

Types of FRR: Path Protection

Cisco.com

- Link, Node protection: pre-establish 1 TE tunnel per {NHop/NNHop}
- Path protection: pre-establish 2 LSPs for 1 TE tunnel, taking two diverse paths



Types of FRR: Path Protection

Cisco.com

- Path protection: pre-establishes 2 LSPs for 1 TE tunnel, taking 2 diverse paths
- When a failure is detected anywhere on the tunnel, the headend switches over to the other LSP
- Failure detection takes longer because we're not just protecting against failure of a directly connected resource
- Loss is msec to sec, depending on size of the network and location of the failure

FRR Scalability

- N = Number of nodes involved in FRR
- D = Degree of connectivity per N
- Link protection scales O(N*D)
- Node protection scales O(N*D^2)
- Path protection scales O(N^2)

FRR Scalability

- In the real world, 10 < N < 200, and 4 < D < 6
- How many LSPs does this translate to?

N	D	Link	Node	Path
10	4	40	160	100
100	5	500	2500	10000
200	6	1200	7200	40000

FRR Scalability

- Moral of the story: use whatever fits your needs
- Link protection is easiest, scales best
- Node protection is probably best overall
- Path protection solves specific problems (rectangles), fits some people’s models better
- Autotunnel makes link and node protection easier (see next section)

TE AUTOMATION

Automation

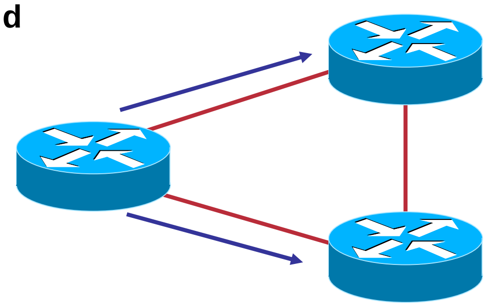
Cisco.com

- Autotunnel
  - Primary 1hop
  - Backup
  - Meshgroup
- Autobandwidth

Autotunnel Primary Onehop

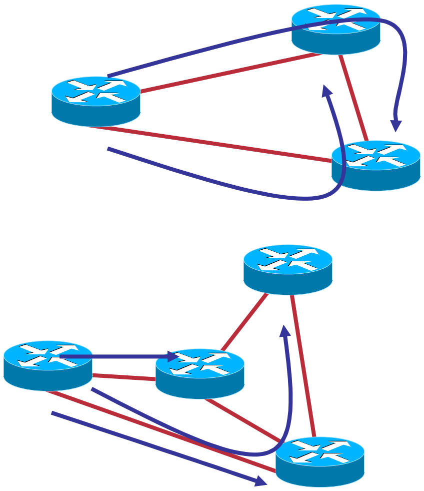
Cisco.com

- Primary onehop—build a TE tunnel to each directly connected neighbor
- This is so we have something to protect
- Not terribly useful by itself, designed to be used in conjunction with NHop tunnels (next slide)



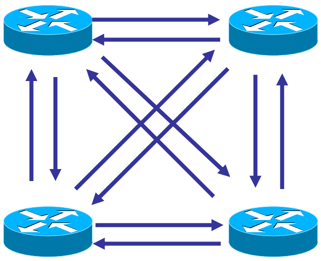
Autotunnel Backup

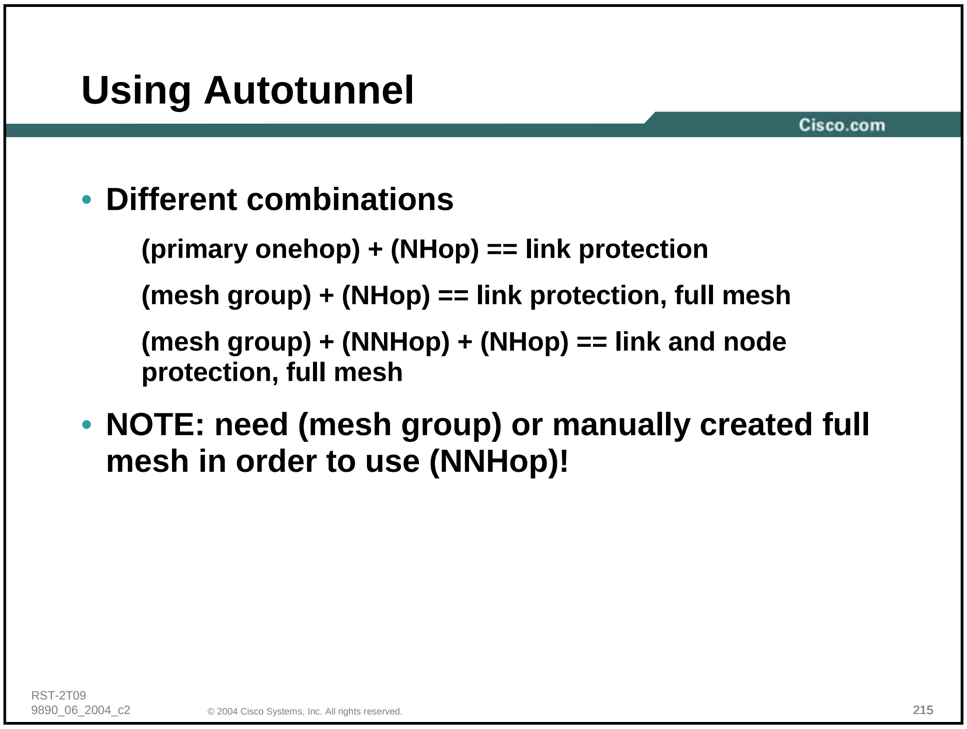
- Build NHop or NNHop (aka link or node) protection tunnels automatically
- Automatically (link or node) disjoint
- Lets you provide connectivity protection with minimal config



Autotunnel Mesh Group

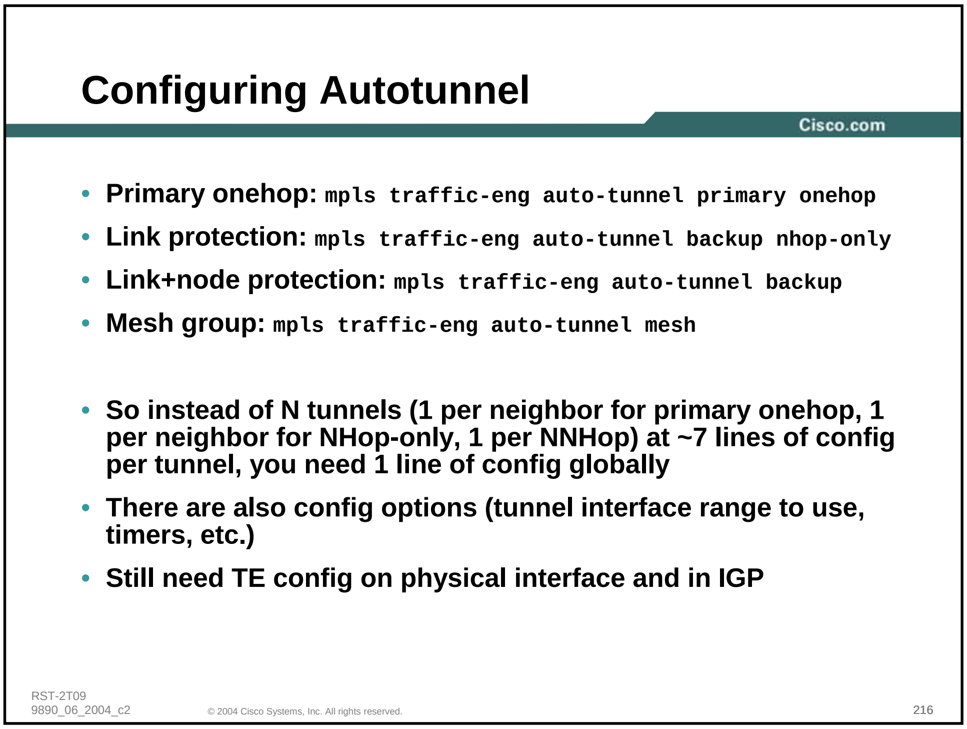
- Automatically build a full mesh of TE tunnels between a set of routers
- Lets you take advantage of NNHop protection
- Also can use the full mesh for a traffic matrix





Using Autotunnel

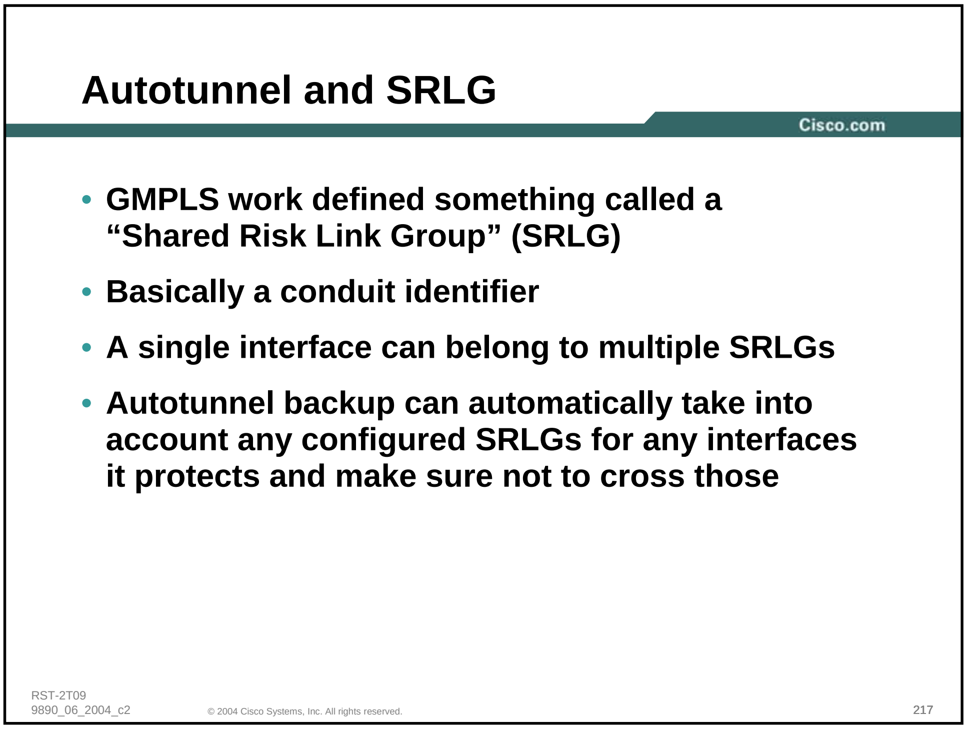
- Different combinations
  - (primary onehop) + (NHop) == link protection
  - (mesh group) + (NHop) == link protection, full mesh
  - (mesh group) + (NNHop) + (NHop) == link and node protection, full mesh
- NOTE: need (mesh group) or manually created full mesh in order to use (NNHop)!



Configuring Autotunnel

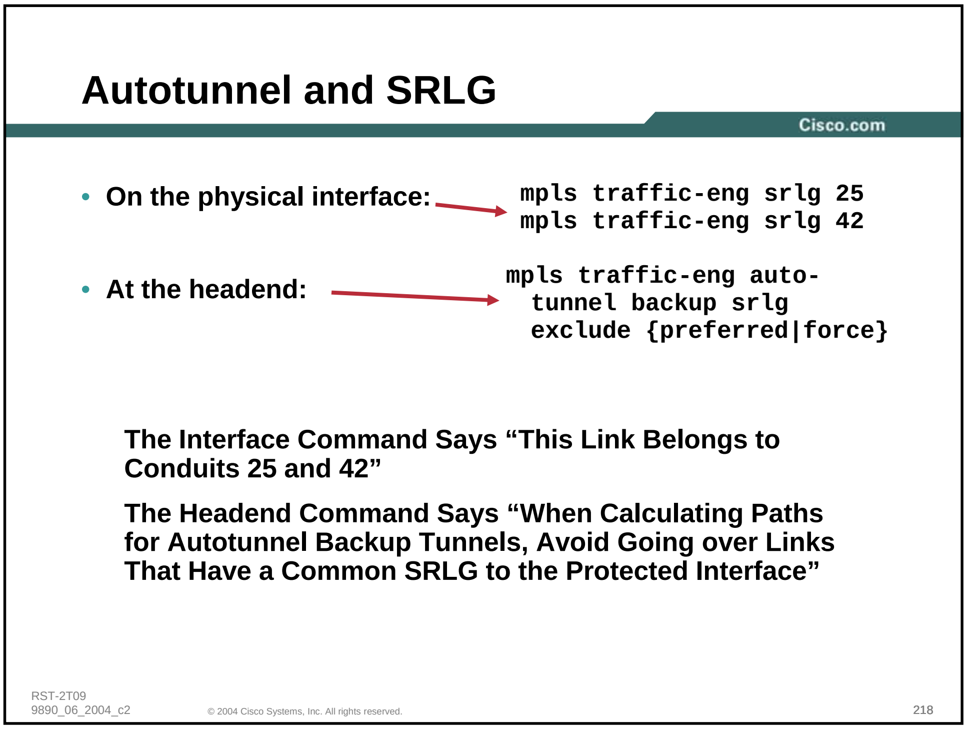
- Primary onehop: `mpls traffic-eng auto-tunnel primary onehop`
- Link protection: `mpls traffic-eng auto-tunnel backup nhop-only`
- Link+node protection: `mpls traffic-eng auto-tunnel backup`
- Mesh group: `mpls traffic-eng auto-tunnel mesh`

- So instead of N tunnels (1 per neighbor for primary onehop, 1 per neighbor for NHop-only, 1 per NNHop) at ~7 lines of config per tunnel, you need 1 line of config globally
- There are also config options (tunnel interface range to use, timers, etc.)
- Still need TE config on physical interface and in IGP



Autotunnel and SRLG

- GMPLS work defined something called a "Shared Risk Link Group" (SRLG)
- Basically a conduit identifier
- A single interface can belong to multiple SRLGs
- Autotunnel backup can automatically take into account any configured SRLGs for any interfaces it protects and make sure not to cross those



Autotunnel and SRLG

- On the physical interface:

```
mpls traffic-eng srlg 25
mpls traffic-eng srlg 42
```

- At the headend:

```
mpls traffic-eng auto-
  tunnel backup srlg
  exclude {preferred|force}
```

The Interface Command Says "This Link Belongs to Conduits 25 and 42"

The Headend Command Says "When Calculating Paths for Autotunnel Backup Tunnels, Avoid Going over Links That Have a Common SRLG to the Protected Interface"

Autotunnel and SRLG

- SRLG support is currently only for autotunnel backup
  - No support for autotunnel mesh group
  - No support for manually created tunnels
- Not terribly useful for primary tunnels
- If you’re going to create a manual tunnel that protects an interface, you need to explicitly place it to avoid any common SRLGs

Autobandwidth

- Autobandwidth has been around for a while
- Growing in popularity in conjunction with autotunnel
- Autobandwidth: periodically resize a TE tunnel based on the traffic that goes down it
- On a TE tunnel: `tunnel mpls traffic-eng auto-bw`
- Knobs exist for changing the period and setting minimum and maximum bandwidths

MORE COOL STUFF

Miscellaneous Yet Highly Cool Stuff

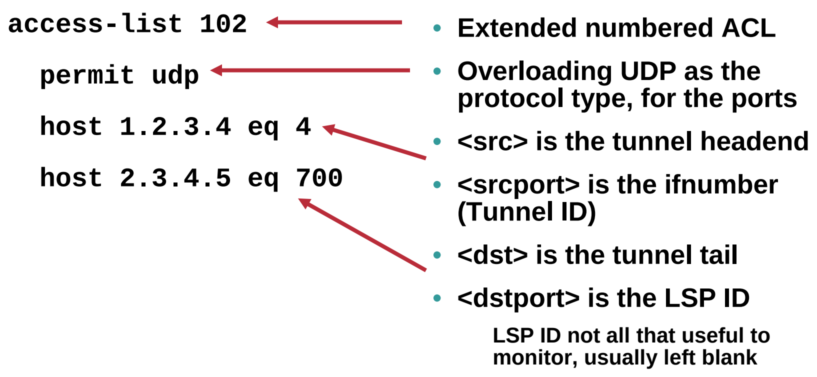
Cisco.com

- Scalable Debugs
- Refresh Reduction
- Pacing

Scalable Debugs

- Debugs are a fact of life
- In my experience, the single most useful TE-related debug is of RSVP signaling
- How do you debug a single LSP at a midpoint that may have 10,000 LSPs through it?
- Answer: ACLs!

Scalable Debugs



```
access-list 102
   permit udp
   host 1.2.3.4 eq 4
   host 2.3.4.5 eq 700
```

- Extended numbered ACL
- Overloading UDP as the protocol type, for the ports
- <src> is the tunnel headend
- <srcport> is the ifnumber (Tunnel ID)
- <dst> is the tunnel tail
- <dstport> is the LSP ID
  - LSP ID not all that useful to monitor, usually left blank

Scalable Debugs

1. `debug ip rsvp filter 102`
2. `debug ip rsvp {signalling|dump-messages}`

- **Lots more RSVP debugs available, these are the big two**
- **Debugs changed a bit in 12.0(24)S; for older software and for more details see the techtip:**
  - http://www/en/US/tech/tk436/tk428/technologies_tech_not
    09186a008019d842.shtml

Refresh Reduction

- **Normal RSVP state maintenance is 1 PATH and 1 RESV per LSP per link every 30s +/- 50%**
- **For 1,000 LSPs this is 2,000 PATH and 2,000 RESV per minute, on average**
- **RFC2961, “RSVP Refresh Overhead Reduction Extensions” greatly reduces this**
- **In practice, haven’t seen the number of messages be a problem**

```
ip rsvp signalling refresh reduction globally
```

Message Pacing

- Early on, we discovered that we can send RSVP messages so fast, we fill up the input queue on the other side
- Full input queue == Bad
- Need to pace messages, but still send them out quickly
- `ip rsvp signalling rate-limit` globally
- Best practice (soon to be default):

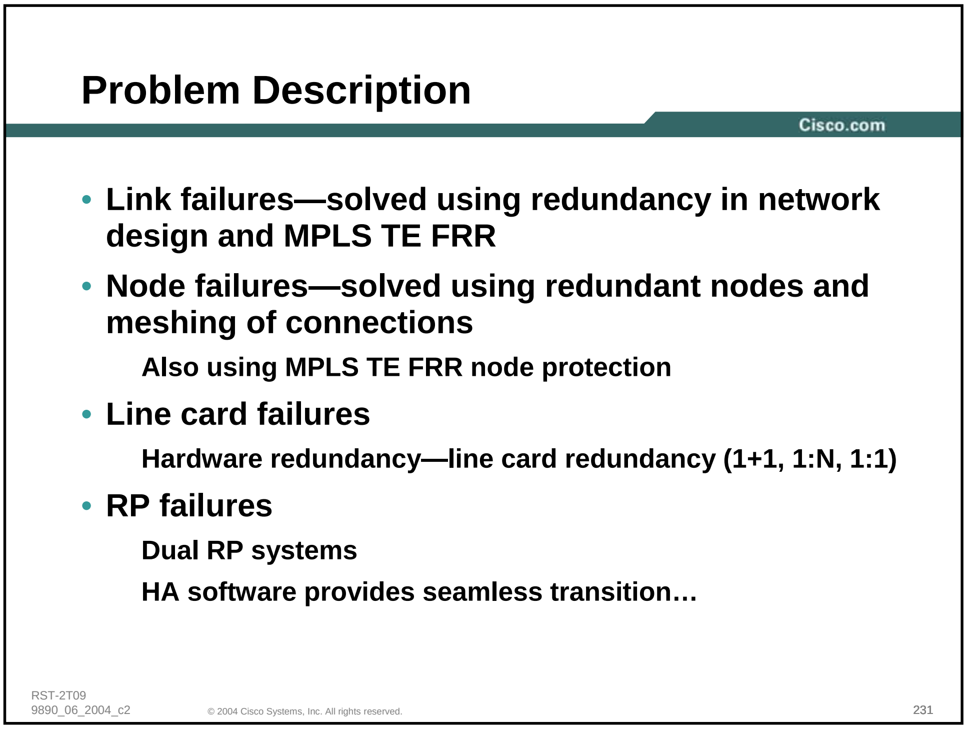
```
ip rsvp signalling rate-limit burst 8
 maxsize 2000 period 20
```

CISCO SYSTEMS

NETWORK INTEGRATION WITH MPLS: MPLS HIGH AVAILABILITY

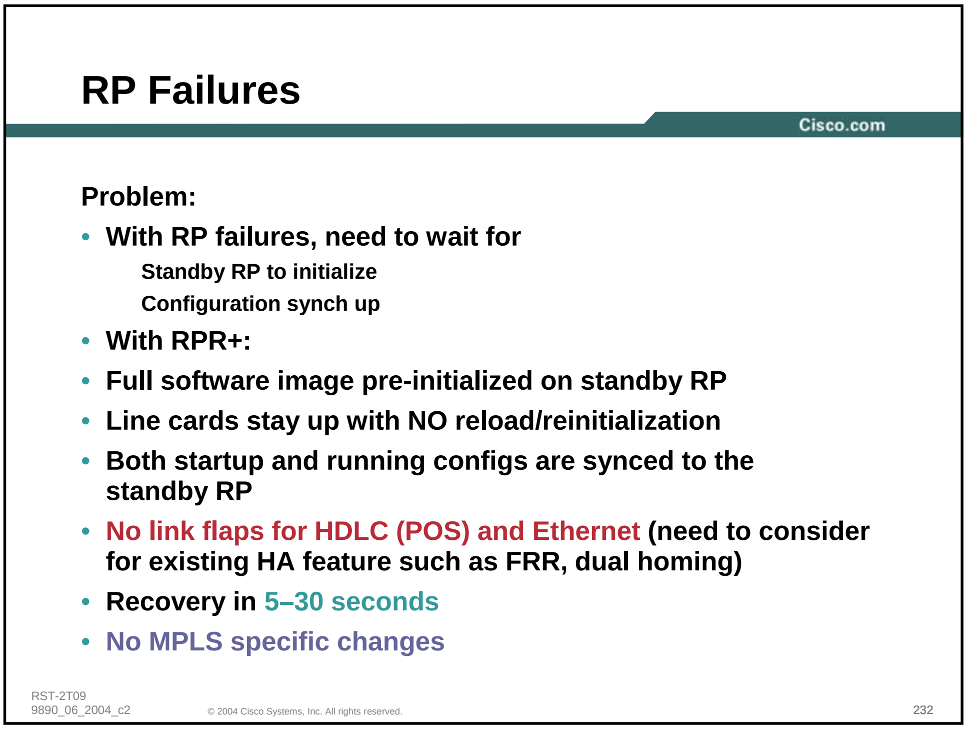
Agenda

- MPLS High Availability (HA) Overview
- Coexistence with IP Non-Stop Forwarding/Stateful Switchover (NSF/SSO)
- MPLS HA Components
    - MPLS HA—LDP NSF/SSO
    - MPLS HA—BGP VPNv4 NSF/SSO
    - MPLS HA—Management
- Summary



Problem Description

- Link failures—solved using redundancy in network design and MPLS TE FRR
- Node failures—solved using redundant nodes and meshing of connections
    - Also using MPLS TE FRR node protection
- Line card failures
    - Hardware redundancy—line card redundancy (1+1, 1:N, 1:1)
- RP failures
    - Dual RP systems
    - HA software provides seamless transition…



RP Failures

Problem:

- With RP failures, need to wait for
    - Standby RP to initialize
    - Configuration synch up
- With RPR+:
- Full software image pre-initialized on standby RP
- Line cards stay up with NO reload/reinitialization
- Both startup and running configs are synced to the standby RP
- No link flaps for HDLC (POS) and Ethernet (need to consider for existing HA feature such as FRR, dual homing)
- Recovery in 5–30 seconds
- No MPLS specific changes

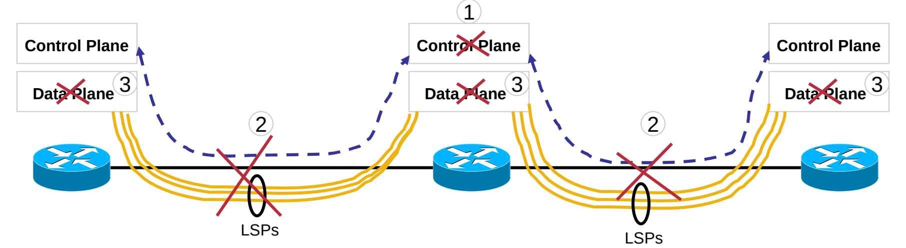
Problem Description

- The TCP session between two LDP/BGP peers may go down for several reasons such as RP switchover due to HW/SW failures
- LDP and BGP use TCP as a reliable transport mechanism for its protocol messages
- On detection of TCP session failure (if HW/SW failure), existing LDP and BGP control plane components would disrupt their forwarding state

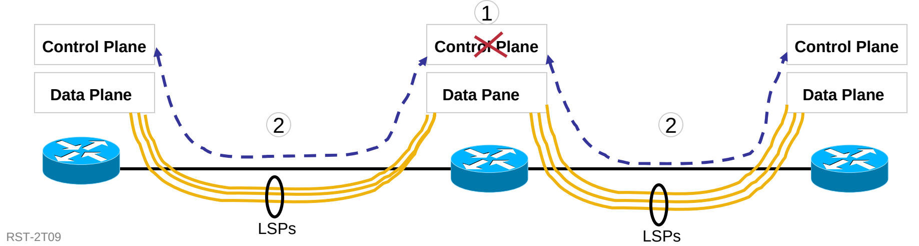
---

Continuous Forwarding of Traffic During Control Plane Service Disruption

Before Cisco NSF/SSO Support and Graceful Restart

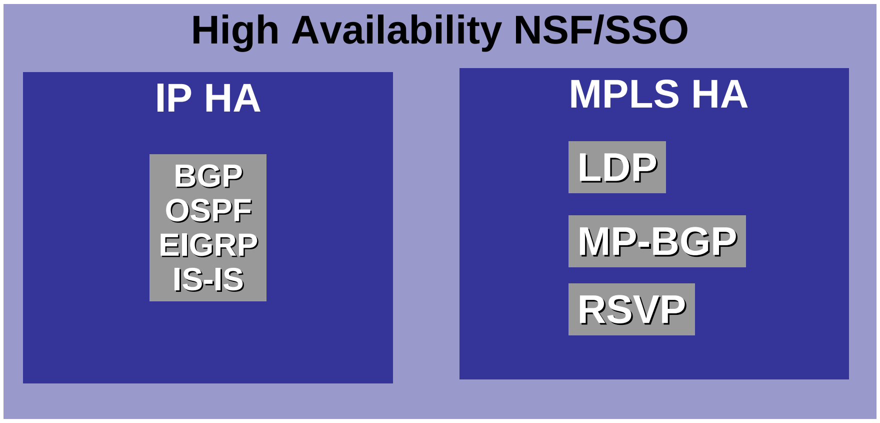

1
Control Plane
Data Plane
3
Control Plane
Data Plane
3
Control Plane
Data Plane
3
2
2
LSPs
LSPs

After Cisco NSF/SSO Support and Graceful Restart


1
Control Plane
Data Plane
Control Plane
Data Pane
Control Plane
Data Plane
2
2
LSPs
LSPs

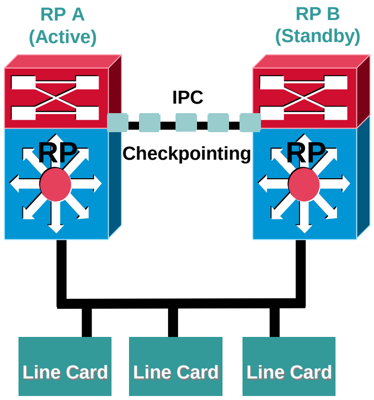
Solution: MPLS HA Umbrella

Enhance the Key Protocols Used in MPLS Control Plane to Minimize the Disruption in MPLS Forwarding Plane


High Availability NSF/SSO
IP HA
BGP
OSPF
EIGRP
IS-IS
MPLS HA
LDP
MP-BGP
RSVP

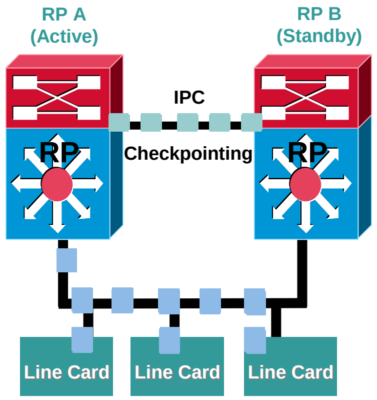
NSF/SSO Architecture Phase I

1. RPs initialize, assign roles (Active/Standby), negotiate mode (SSO), begin checkpointing state


RP A
(Active)
RP B
(Standby)
IPC
Checkpointing
RP
RP
Line Card
Line Card
Line Card

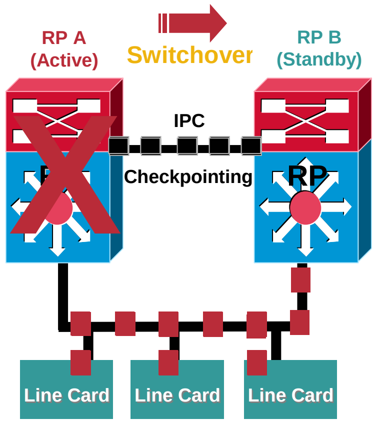
NSF/SSO Architecture Phase II

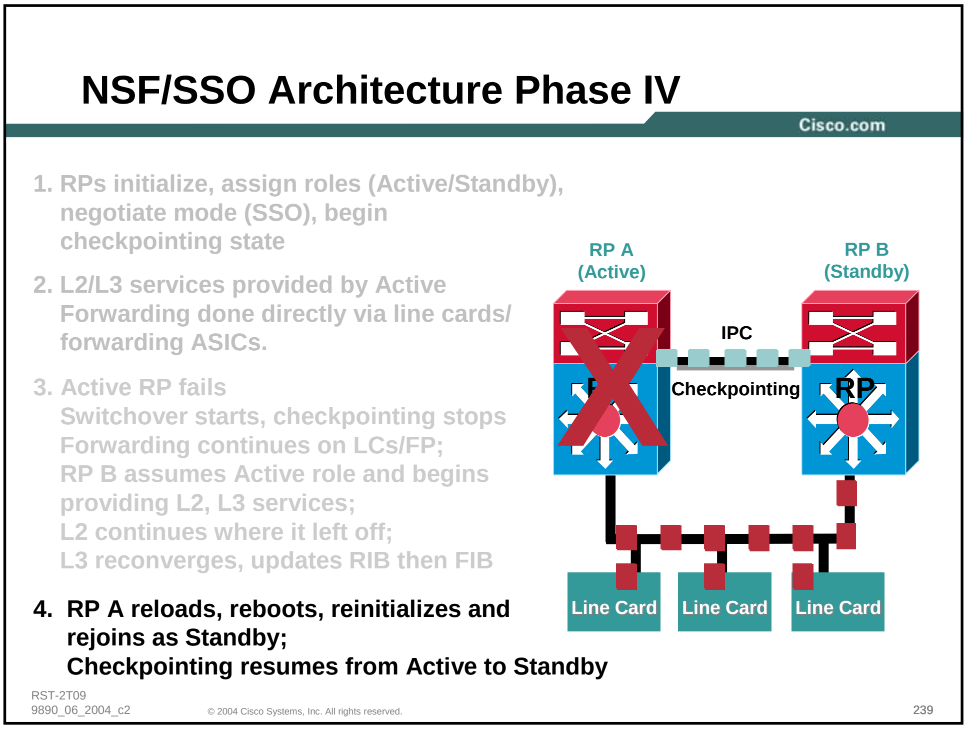
1. RPs initialize, assign roles (Active/Standby), negotiate mode (SSO), begin checkpointing state
2. L2/L3 services provided by Active
   Forwarding done directly via line cards/ forwarding ASICs


RP A
(Active)
RP B
(Standby)
IPC
Checkpointing
RP
RP
Line Card
Line Card
Line Card

NSF/SSO Architecture Phase III

1. RPs initialize, assign roles (Active/Standby), negotiate mode (SSO), begin checkpointing state
2. L2/L3 services provided by Active
   Forwarding done directly via line cards/ forwarding ASICs
3. Active RP fails
   Switchover starts, checkpointing stops
   Forwarding continues on LCs/FP;
   RP B assumes Active role and begins providing L2, L3 services;
   L2 continues where it left off;
   L3 reconverges, updates RIB then FIB

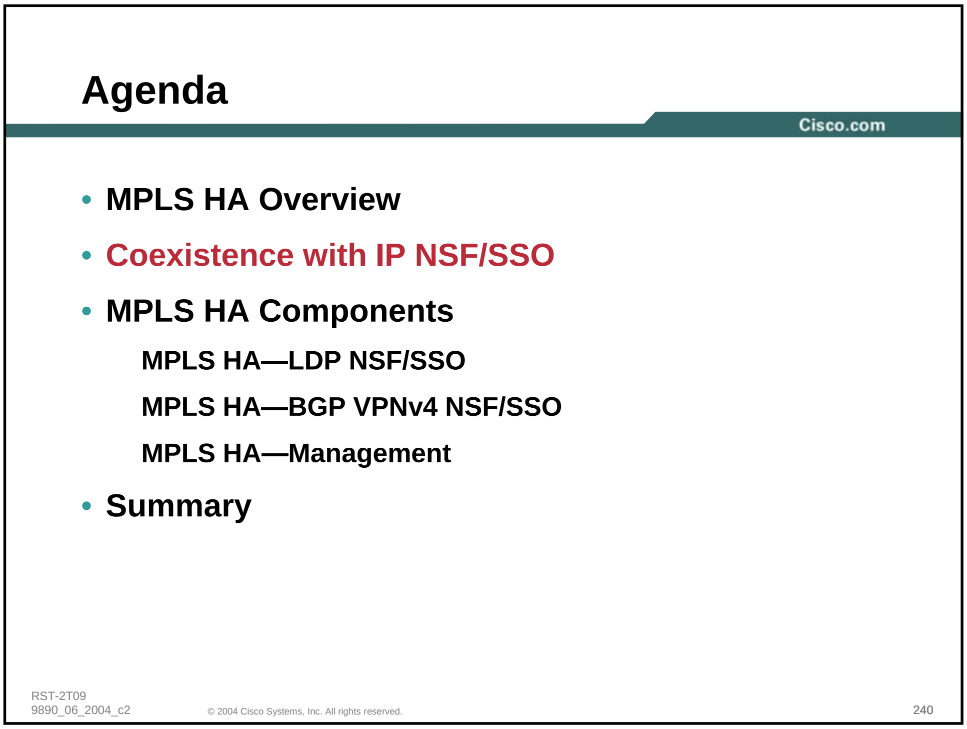

RP A
(Active)
Switchover
RP B
(Standby)
IPC
Checkpointing
RP
Line Card
Line Card
Line Card



NSF/SSO Architecture Phase IV

1. RPs initialize, assign roles (Active/Standby), negotiate mode (SSO), begin checkpointing state
2. L2/L3 services provided by Active Forwarding done directly via line cards/ forwarding ASICs.
3. Active RP fails
   Switchover starts, checkpointing stops
   Forwarding continues on LCs/FP;
   RP B assumes Active role and begins providing L2, L3 services;
   L2 continues where it left off;
   L3 reconverges, updates RIB then FIB
4. **RP A reloads, reboots, reinitializes and rejoins as Standby;**
   Checkpointing resumes from Active to Standby

RP A
(Active)
RP B
(Standby)
IPC
Checkpointing
RP
RP
Line Card
Line Card
Line Card



Agenda

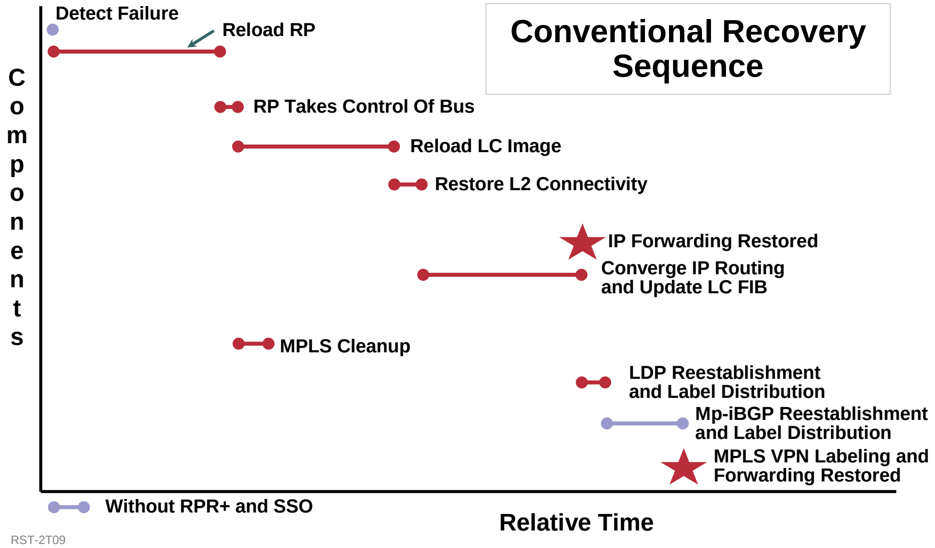
- **MPLS HA Overview**
- **Coexistence with IP NSF/SSO**
- **MPLS HA Components**
  - **MPLS HA—LDP NSF/SSO**
  - **MPLS HA—BGP VPNv4 NSF/SSO**
  - **MPLS HA—Management**
- **Summary**

MPLS NSF/SSO-Coexistence

- IP HA speeds up MPLS recovery
  - No waiting for the route processor
  - No loss of Layer 2 connectivity, so it does not need to be re-established
  - MPLS with IP SSO begin rebuilding more quickly after a switchover to the standby RP
- SSO coexistence feature allows the mix of SSO and non-SSO features at the same time
- During the switchover, MPLS forwarding entries are removed from the linecards and MPLS forwarding is stopped

---

Restoring MPLS Following an RP Failure



Conventional Recovery Sequence

Detect Failure

Reload RP

RP Takes Control Of Bus

Reload LC Image

Restore L2 Connectivity

IP Forwarding Restored

Converge IP Routing and Update LC FIB

MPLS Cleanup

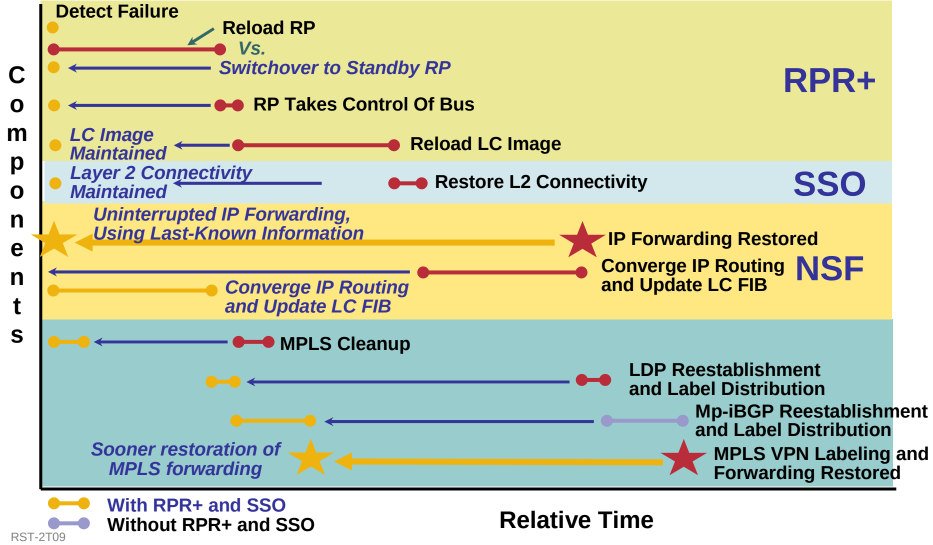
LDP Reestablishment and Label Distribution

Mp-iBGP Reestablishment and Label Distribution

MPLS VPN Labeling and Forwarding Restored

Without RPR+ and SSO

Faster MPLS Recovery Coexists with NSF



Detect Failure

Reload RP

Vs.

Switchover to Standby RP

RP Takes Control Of Bus

LC Image Maintained

Reload LC Image

Layer 2 Connectivity Maintained

Restore L2 Connectivity

Uninterrupted IP Forwarding, Using Last-Known Information

IP Forwarding Restored

Converge IP Routing and Update LC FIB

Converge IP Routing and Update LC FIB

MPLS Cleanup

LDP Reestablishment and Label Distribution

Mp-iBGP Reestablishment and Label Distribution

Sooner restoration of MPLS forwarding

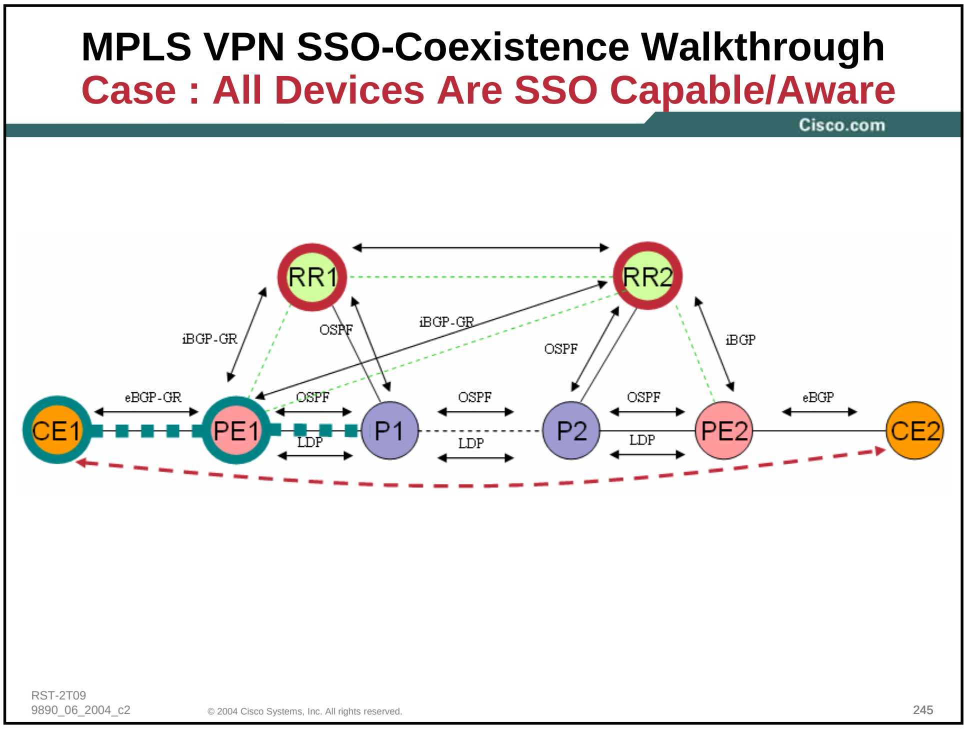
MPLS VPN Labeling and Forwarding Restored

●—● With RPR+ and SSO

●—● Without RPR+ and SSO

Various MPLS SSO-Coexistence Scenarios

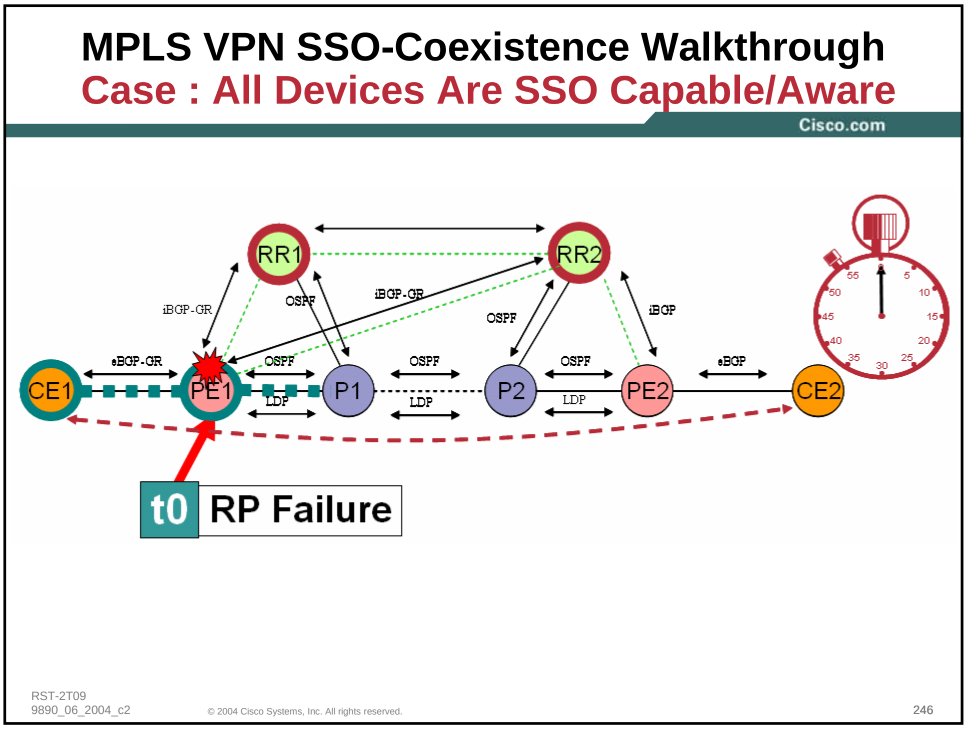
- All peers are SSO capable/aware:
  - IP over MPLS (LDP)
  - MPLS VPN—best case
- All PEs are SSO capable/aware:
  - IP over MPLS (LDP)
  - MPLS VPN—worst case



MPLS VPN SSO-Coexistence Walkthrough
Case : All Devices Are SSO Capable/Aware

Cisco.com

RR1
RR2
iBGP-GR
OSPF
iBGP-GR
OSPF
iBGP
eBGP-GR
OSPF
OSPF
OSPF
eBGP
CE1
PE1
P1
P2
PE2
CE2
LDP
LDP
LDP



MPLS VPN SSO-Coexistence Walkthrough
Case : All Devices Are SSO Capable/Aware

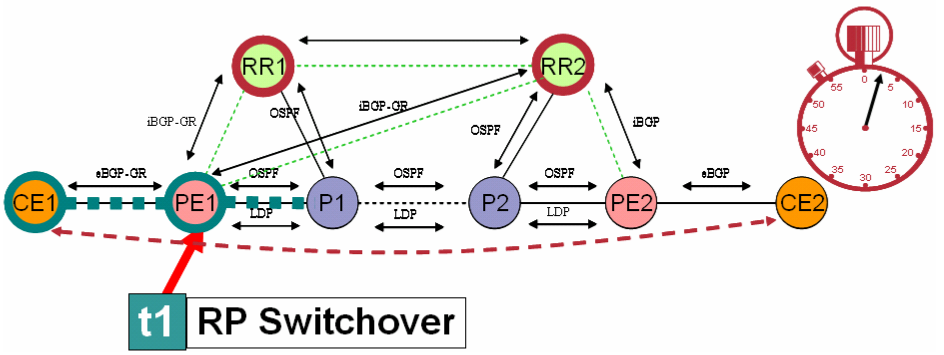
Cisco.com

RR1
RR2
iBGP-GR
OSPF
iBGP-GR
OSPF
iBGP
eBGP-GR
OSPF
OSPF
OSPF
eBGP
CE1
PE1
P1
P2
PE2
CE2
LDP
LDP
LDP
55
5
50
10
45
15
40
20
35
30
25

t0 RP Failure

MPLS VPN SSO-Coexistence Walkthrough
Case : All Devices Are SSO Capable/Aware

Cisco.com

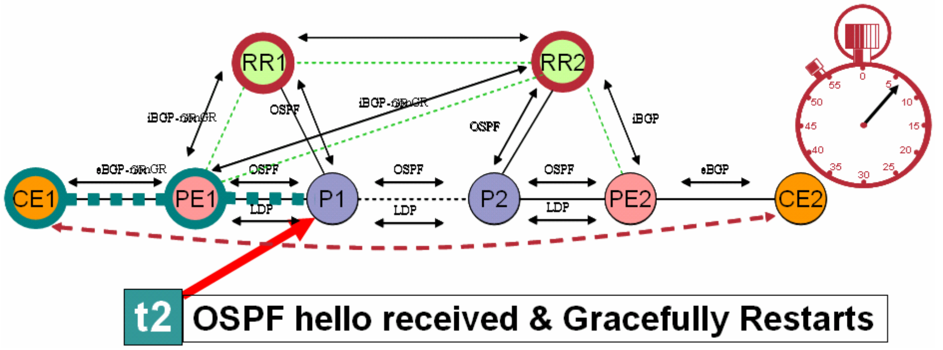

RR1
RR2
iBGP-GR
OSPF
iBGP-GR
OSPF
iBGP
eBGP-GR
OSPF
OSPF
OSPF
eBGP
CE1
PE1
P1
P2
PE2
CE2
LDP
LDP
LDP
0
5
10
15
20
25
30
35
40
45
50
55

t1 RP Switchover

RST-2T09
9890_06_2004_c2

MPLS VPN SSO-Coexistence Walkthrough
Case : All Devices Are SSO Capable/Aware

Cisco.com

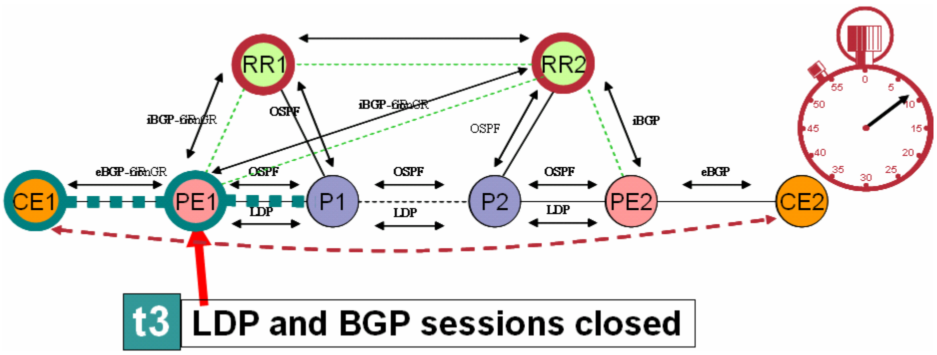

RR1
RR2
iBGP-GRnGR
OSPF
iBGP-GRnGR
OSPF
iBGP
eBGP-GRnGR
OSPF
OSPF
OSPF
eBGP
CE1
PE1
P1
P2
PE2
CE2
LDP
LDP
LDP
0
5
10
15
20
25
30
35
40
45
50
55

t2 OSPF hello received & Gracefully Restarts

RST-2T09
9890_06_2004_c2

MPLS VPN SSO-Coexistence Walkthrough
Case : All Devices Are SSO Capable/Aware

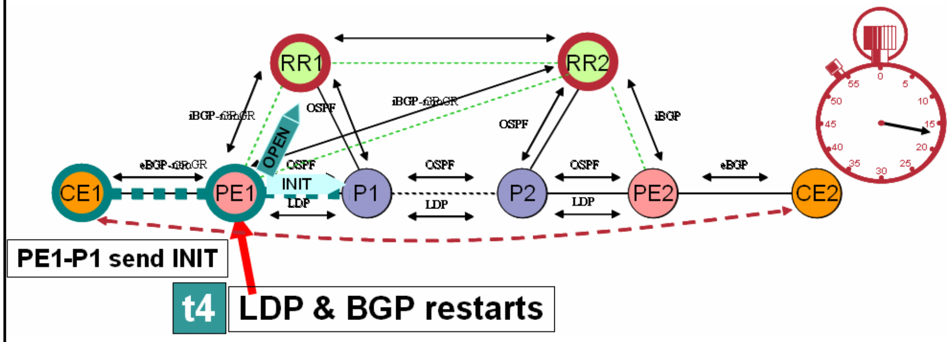
Cisco.com


RR1
RR2
iBGP-GR
OSPF
iBGP-GR
OSPF
iBGP
eBGP-GR
OSPF
OSPF
OSPF
eBGP
CE1
PE1
P1
P2
PE2
CE2
LDP
LDP
LDP
0
5
10
15
20
25
30
35
40
45
50
55

t3 LDP and BGP sessions closed

MPLS VPN SSO-Coexistence Walkthrough
Case : All Devices Are SSO Capable/Aware

Cisco.com

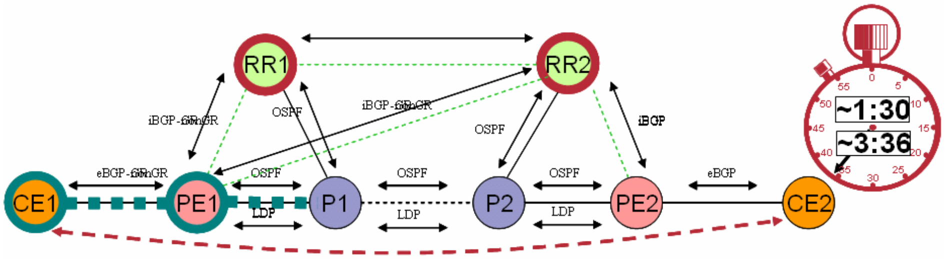

RR1
RR2
iBGP-GR
OSPF
OPEN
iBGP-GR
OSPF
iBGP
eBGP-GR
OSPF
INIT
OSPF
OSPF
eBGP
CE1
PE1
P1
P2
PE2
CE2
LDP
LDP
LDP
0
5
10
15
20
25
30
35
40
45
50
55

PE1-P1 send INIT

t4 LDP & BGP restarts

MPLS VPN SSO-Coexistence Walkthrough

Case : All Devices Are SSO Capable/Aware


RR1
RR2
iBGP-GR
OSPF
iBGP-GR
OSPF
iBGP
eBGP-GR
OSPF
OSPF
OSPF
eBGP
CE1
PE1
P1
P2
PE2
CE2
LDP
LDP
LDP
55
0
5
50
~1:30
10
45
15
40
~3:36
20
35
30
25

t5 End to End Traffic Resumed

Agenda

- MPLS HA Overview
- Coexistence with IP NSF/SSO
- MPLS HA Components
  - MPLS HA—LDP NSF/SSO
  - MPLS HA—BGP VPNv4 NSF/SSO
  - MPLS HA—Management
- Summary

LDP HA Key Elements

1. Checkpointing local label bindings

   On devices with route processor redundancy

2. LDP graceful restart capability

   On participating PEs, RRs, and P routers

Why LDP Checkpointing?

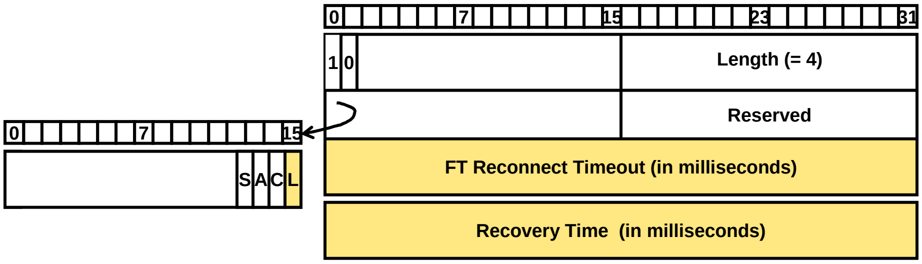
- The LSRs that support LDP GR, require that the restarting control plane be able to query the forwarding plane for local label binding information; the MPLS forwarding plane will not support such queries
- LDP checkpointing makes local label bindings available to the restarting(standby) LDP control plane

MPLS LDP Local Checkpointing Key Points

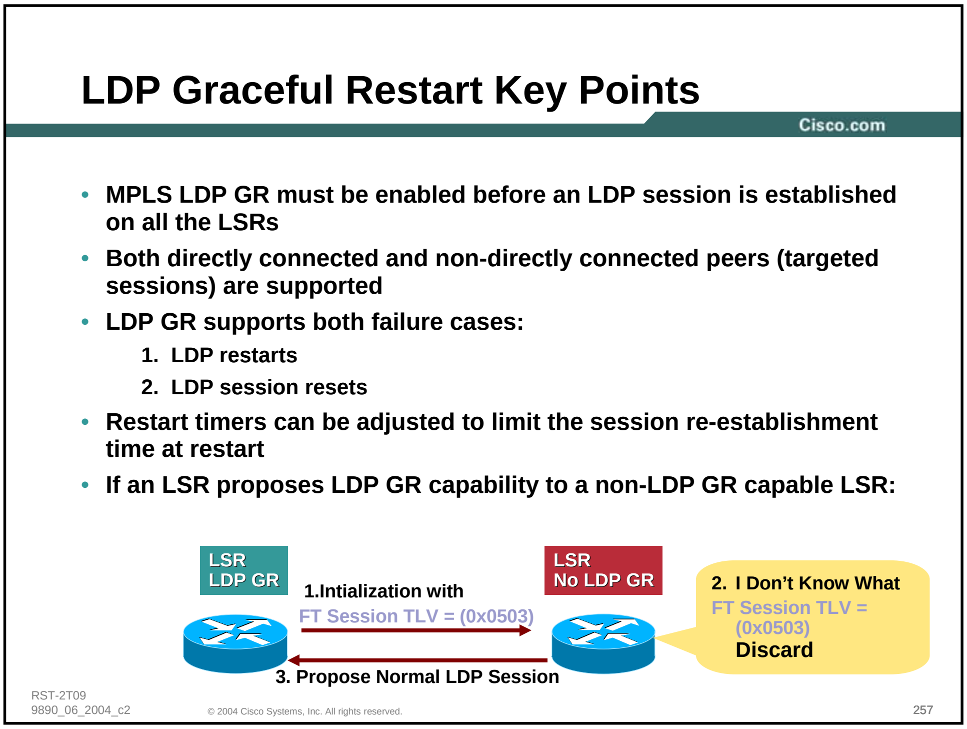
- The checkpointing function is enabled by default and done using a separate process
- The checkpointing function copies active RP's LDP local label bindings to the backup RP; for the first round, all the labels are copied from active to back up RP
- Periodic incremental updates are done to reflect new routes that have been learned or routes that have been removed and/or when labels are allocated or freed
- Checkpointing stops during control plane disruptions, GR, and recovery
- Label bindings on backup RP are marked checkpointed
- This marking is removed when it becomes active RP

LDP Graceful Restart Mechanism

- Described in RFC3478
- The LSR sends the LDP initialization message to a neighbor to establish an LDP session
- The Fault Tolerant (FT) session type length value (TLV) is included in the LDP initialization message


0
7
15
23
31
1
0
Length (= 4)
Reserved
FT Reconnect Timeout (in milliseconds)
Recovery Time (in milliseconds)
0
7
15
S
A
C
L

L=1 Means GR-LDP Is Selected
FT Reconnect = 0 Means LSR Is Not NSF Capable
FT Recovery Time = 0 Means LSR Was Unable to Preserve MPLS Forwarding State Across Restart



LDP Graceful Restart Key Points

- MPLS LDP GR must be enabled before an LDP session is established on all the LSRs
- Both directly connected and non-directly connected peers (targeted sessions) are supported
- LDP GR supports both failure cases:
  1. LDP restarts
  2. LDP session resets
- Restart timers can be adjusted to limit the session re-establishment time at restart
- If an LSR proposes LDP GR capability to a non-LDP GR capable LSR:

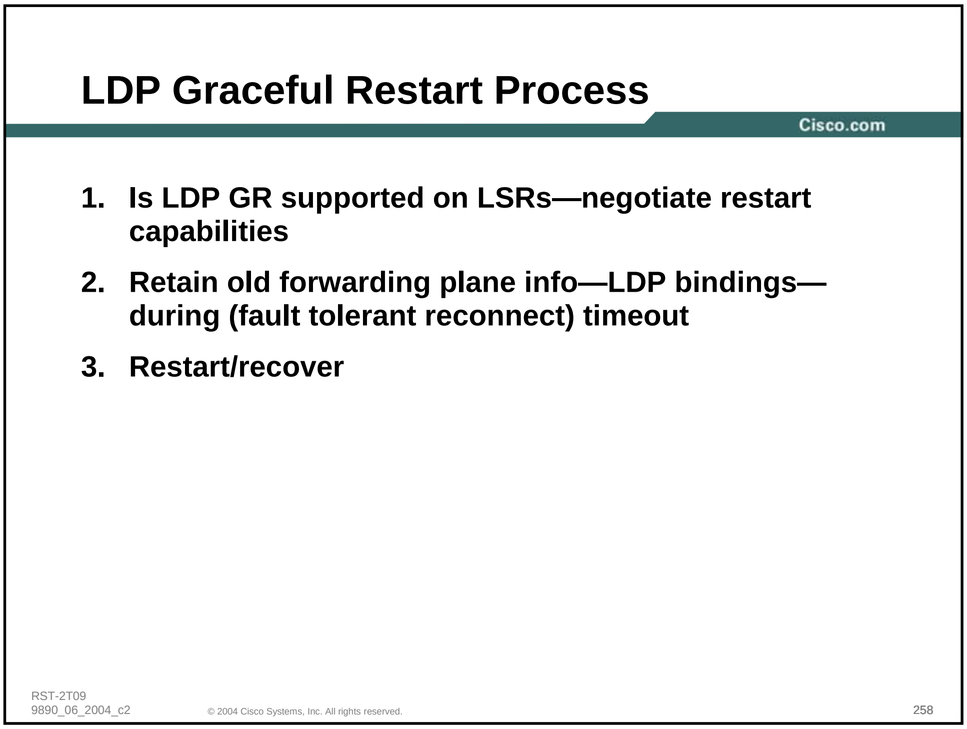
LSR
LDP GR

LSR
No LDP GR

1.Intialization with
FT Session TLV = (0x0503)

2. I Don't Know What
FT Session TLV =
(0x0503)
Discard

3. Propose Normal LDP Session



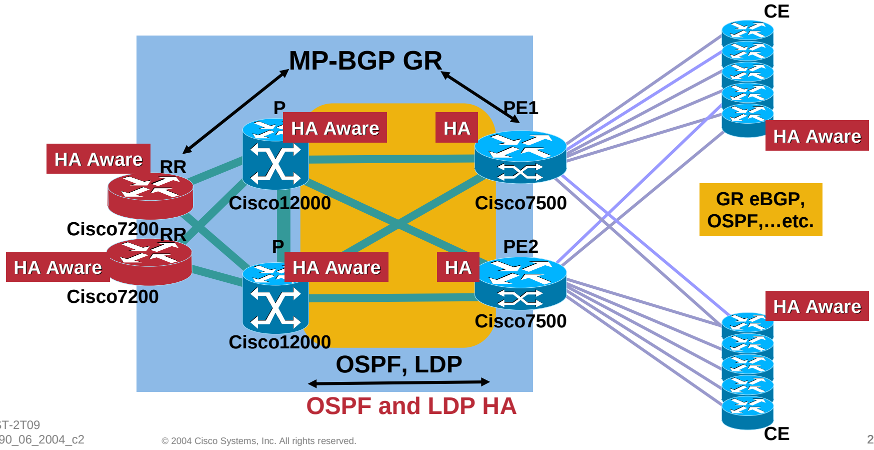
LDP Graceful Restart Process

1. Is LDP GR supported on LSRs—negotiate restart capabilities
2. Retain old forwarding plane info—LDP bindings—during (fault tolerant reconnect) timeout
3. Restart/recover

Deploying MPLS HA Example

HA Aware Devices:
Graceful Restart Functionality for All the Related Protocols CEs, Ps, and RRs

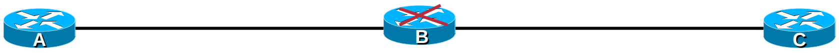
HA Capable Devices
Full NSF/SSO Functionality (PE1 and PE2)


MP-BGP GR
HA Aware
RR
Cisco7200
RR
HA Aware
Cisco7200
P
HA Aware
Cisco12000
P
HA Aware
Cisco12000
HA
PE1
Cisco7500
HA
PE2
Cisco7500
CE
HA Aware
GR eBGP, OSPF,...etc.
HA Aware
CE
OSPF, LDP
OSPF and LDP HA

LDP GR on SSO-LSR and SSO-Aware Peer

- **SSO capable LSR failed:**
  - Active RP failed
  - Continue forwarding using the stale state
  - Standby RP becomes Active
  - Mark the forwarding state as stale, and retain it
  - Reestablish LDP session
- **SSO aware neighbor LSR:**
  - LDP failure detected
  - Mark the forwarding state as stale, and retain it
  - Reestablish LDP session

LDP Graceful Restart Operation



- LDP paths established, LDP GR negotiated
- When RP fails on LSRb, communication between peers is lost; LSRb encounters a LDP restart, while LSRa and LSRc encounter an LDP session reset
- LSRa and LSRc mark all the label bindings from LSRb as stale, but continue to use the same bindings for MPLS forwarding
- LSRa and LSRc attempt to reestablish an LDP session with Rb
- LSRb restarts and marks all of its forwarding entries as stale; at this point, LDP state is in restarting mode
- LSRa and LSRc reestablish LDP sessions with Rb, but keep their stale label bindings; at this point, the LDP state is in recovery mode
- All routers re-advertise their label binding info; stale flags are removed if a label has been relearned; new LFIB table is ready with new local label, outgoing label or VC, prefix or tunnel ID, label-switched bytes, outgoing interface and Next Hop

LDP Graceful Restart Configuration

```
mpls ldp graceful-restart
mpls ldp graceful-restart timers forwarding-holding
mpls ldp graceful-restart timers neighbor-liveness
mpls ldp graceful-restart timers max-recovery
```

If LDP Session Don't Reestablish Or If LSR Doesn't Restart

Configuration Options:

- If enable globally, don't need to enable per interface
- Don't need to enable globally, can enable on selected interfaces

LDP Graceful Restart Timers

- mpls ldp graceful-restart timers forwarding-holding

  The amount of time the MPLS forwarding state should be preserved after the Control Plane restarts (in seconds); if the timer expires, all the entries marked stale are deleted

- mpls ldp graceful-restart timers neighbor-liveness

  The amount of time an LSR should wait for an LDP session to reconnect (in seconds); if the LSR cannot recreate an LDP session with the neighbor in the time allotted, the LSR deletes the stale label-FEC bindings received from that neighbor (def: 5sec)

- mpls ldp graceful-restart timers max-recovery

  The amount of time an LSR should hold stale label-FEC bindings after an LDP session has been re-established; after the timer expires, all prefixes are removed from the binding and forwarding table; set this timer to a value that allows the neighboring LSRs to re-sync the LSPs without creating congestion in the LDP control plan (def: 120sec)

Agenda

- MPLS HA Overview
- Coexistence with IP NSF/SSO
- MPLS HA Components
  - MPLS HA––LDP NSF/SSO
  - MPLS HA—BGP VPNv4 NSF/SSO
  - MPLS HA—Management
- Summary

MPLS VPNv4 HA Elements

- MPLS VPN checkpointing capability
- BGP graceful restart capability

MPLS VPN Checkpointing

- Active RP checkpoints the following information to the back up RP after the MPLS forwarding is updated:

  {<VRFID>, <prefix>, <mask>, <local label>}

MPLS BGP VPNv4 GR Mechanism

Cisco.com

- MPLS BGP GR mechanism defines preservation of forwarding state across BGP restart

  draft-ietf-mpls-bgp-mpls-restart

- A new graceful restart capability is carried in BGP open message

  A BGP update message with no reachable NLRI and empty withdrawn NLRI is specified as an End-of-RIB marker

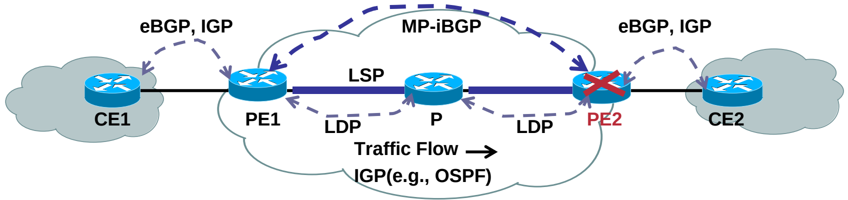
  AFI/SAFI (Subsequent Address Family Identifier) pairs in the graceful restart cap is used by an LSR for indicating its ability to preserve MPLS forwarding state across BGP restart

  SAFI in the advertise capability indicates that NLRI field carries addressing information as well as labels

---

MPLS VPN: BGP Graceful Restart Procedure

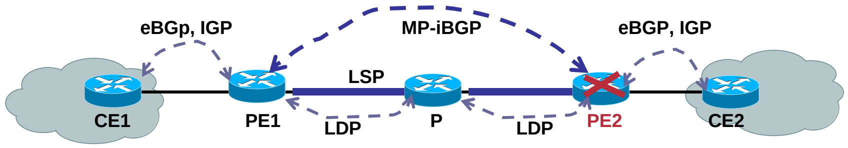
Cisco.com



Traffic Is Forwarded Continuously

- PE1-PE2 exchange Graceful Restart Cap (restart time, AFI/SAFI, etc.)
- PE1 and PE2 exchange routing information via BGP update messages
- Assume LSP has been established from PE1 to PE2
- PE2 restarts (active RP fails, switchover occurs)

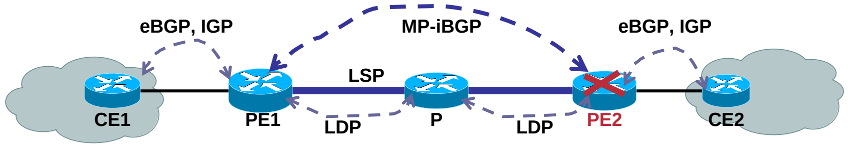
MPLS VPN: BGP Graceful Restart Procedure (Cont.)



- PE1 detects PE2's failure, retains its last Adj-RIB-In and forwarding state learned from PE2; PE1 will delete this state if session is not re-established within restart time
- The BGP session between PE1 and PE2 goes down
- PE1 marks the entries in its BGP table which it learned from PE2 as stale but does not delete the routes from its VRF tables; hence it continues to forward traffic to PE2
- PE2 switches over to back up RP; continues to forward traffic using the backed up info

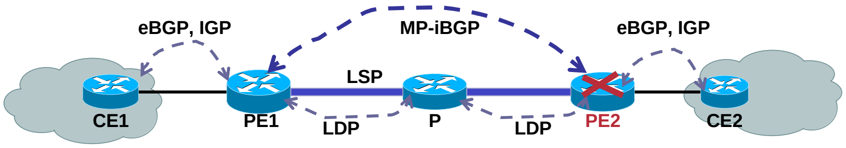
---

MPLS VPN: BGP Graceful Restart Procedure (Cont.)



- PE2 re-establishes TCP session with PE1 (hopefully within Restart Time-180seconds default)
- PE1 sends BGP Updates from Adj-RIBs-Out to PE2 along with the label mapping; on completion, sends End-of-RIB marker
- PE2 runs decision process after receiving End-of-RIB markers from all BGP peers, updates its Loc-RIB, FIB, Adj_RIBs-Out and advertise its routes to PE1; on completion sends End-of-RIB marker

MPLS VPN: BGP Graceful Restart Procedure (Cont.)



- Suppose PE2 had bound an in label L1 to a FEC, it picks the same label (if checkpointed) or allocates a new one and advertises it to PE1 for this route
- PE2 updates its Adj-RIBs-In, on receipt of End-of-RIB marker, deletes stale entries, runs its decision process, updates its Loc-RIB, FIB, and Adj-RIBs-Out; removes stale state if the labels are the same (Routing info and Out_Label are recovered from peer; In_Label is either checkpointed or a new one allocated)
- Back to normal operation

How to Enable BGP VPNv4 SSO/NSF

The Following Components Must Be Enabled to Support BGP VPNv4 NSF

- Enabling SSO
  http://www.cisco.com/univercd/cc/td/doc/product/software/ios122s/122snwft/release/122s18/sso18s.htm
- Enabling LDP Graceful Restart
  http://www.cisco.com/univercd/cc/td/doc/product//product/software/ios122s/122snwft/release/122srls4
- Enabling Nonstop Forwarding for routing protocols
  http://www.cisco.com/univercd/cc/td/doc/product/software/ios122s/122snwft/release/122s18/nsf18s.htm
- Enabling NSF support for basic VPNs
- Configuring NSF support for VPN interfaces that use BGP for label distribution
- Verifying VPN NSF (optional)

MPLS L3 VPN Features

Cisco.com

Supported Features:

- MPLS VPN BGP GR for VPNv4
- MPLS VPN BGP checkpointing
- MPLS VPN SSO/NSF support for VRF
- MPLS VPN SSO/NSF support for I-AS
- MPLS VPN SSO/NSF support for CSC

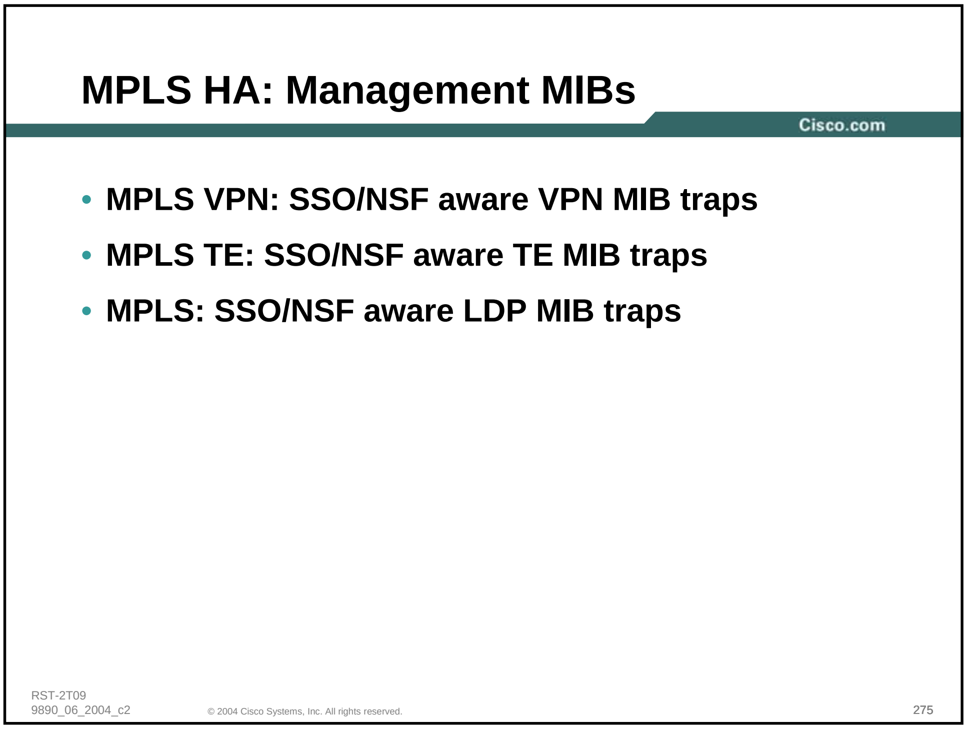
  Additional features will be supported in subsequent releases; please refer to Cisco Feature Navigator for the latest information

  http://www.cisco.com/go/fn

Agenda

Cisco.com

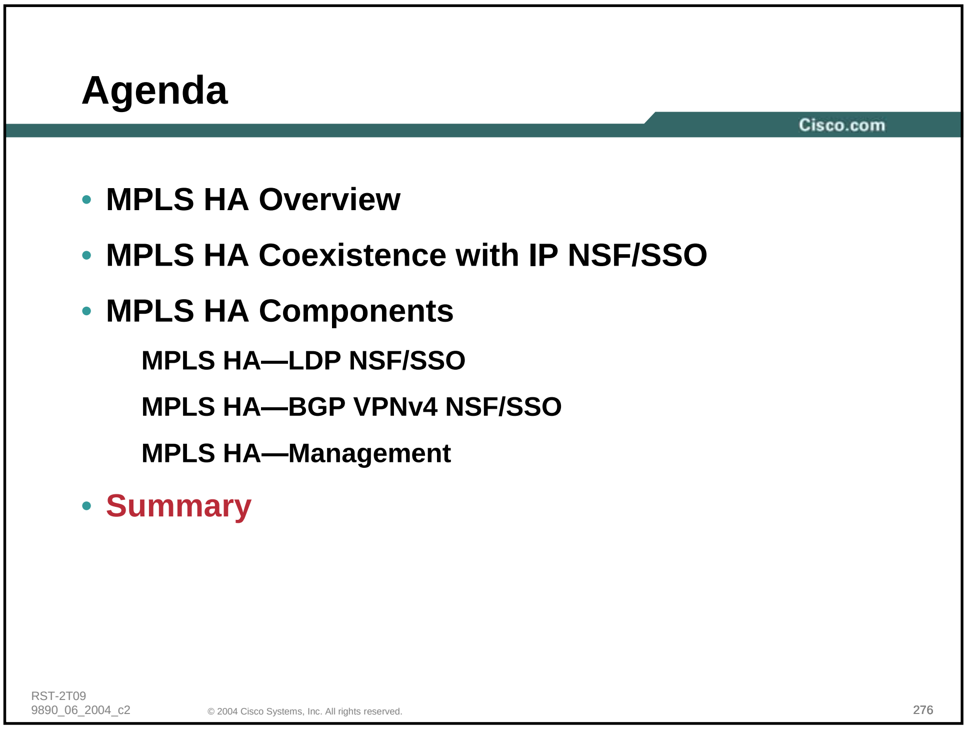
- MPLS HA Overview
- Coexistence with IP NSF/SSO
- MPLS HA Components
  - MPLS HA—LDP NSF/SSO
  - MPLS HA—BGP VPNv4 NSF/SSO
  - MPLS HA—Management
- Summary



MPLS HA: Management MIBs

Cisco.com

- MPLS VPN: SSO/NSF aware VPN MIB traps
- MPLS TE: SSO/NSF aware TE MIB traps
- MPLS: SSO/NSF aware LDP MIB traps



Agenda

Cisco.com

- MPLS HA Overview
- MPLS HA Coexistence with IP NSF/SSO
- MPLS HA Components
    - MPLS HA—LDP NSF/SSO
    - MPLS HA—BGP VPNv4 NSF/SSO
    - MPLS HA—Management
- Summary

Summary

- Cisco is enhancing its portfolio to add features for improved full HA solution; MPLS HA features provide stateful switchover and NSF capability for VPN, LDP, TE, etc.
- Need IP HA enabled to support MPLS HA; GR must be enabled for all participating RPs (OSPF, BGP, IS-IS) on P, PE, and CE routers
- For MPLS VPN HA: MFI HA, LDP HA, BGP HA is required
- AToM NSF/SSO is exactly the same as directed LDP
- AToM application will do checkpointing for local labels only
- TE NSF/SSO is defined in GMPLS-RSVP(TE) RFC 3473
- FRR NSF/SSO is same as NSF/SSO for TE
- LC-ATM NSF/SSO is DoD LDP GR
- Some CLI changes

MPLS AND LAYER 1: GENERALIZED MPLS AND OPTICAL USER NETWORK INTERFACE (OUNI) GMPLS AND OUNI

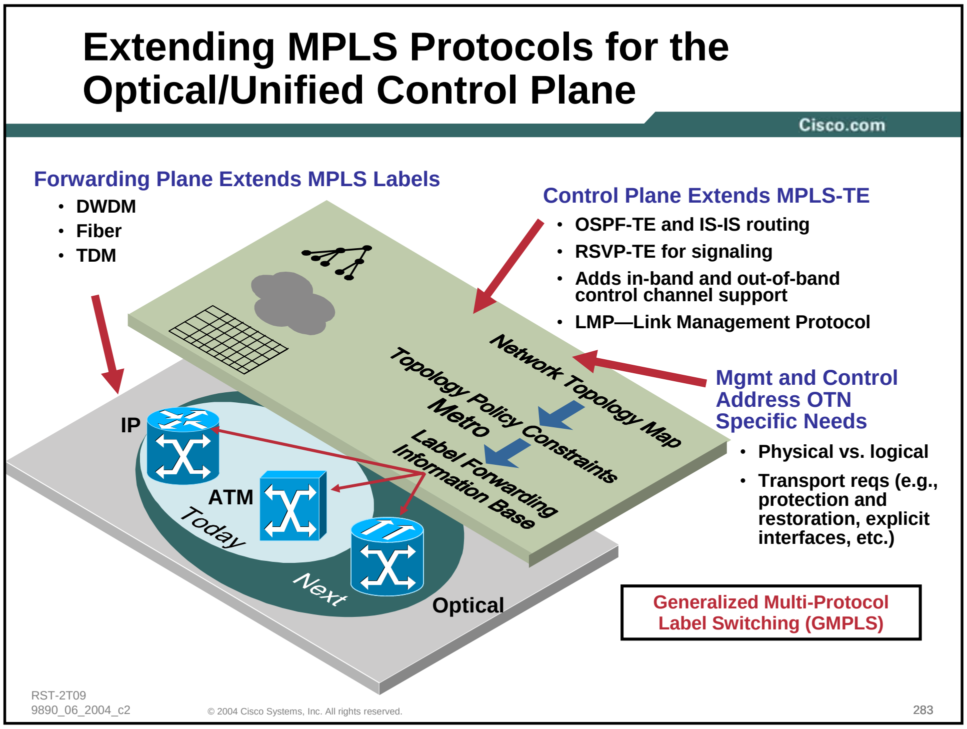
GMPLS AND OUNI

Terminology

- GMPLS—Generalized MPLS
  - Extensions to MPLS to integrate IP and optical networks
- OUNI—Optical UNI
- UCP—Unified Control Plane
  - Same control plane for MPLS and GMPLS

UCP Overview

- GMPLS provides
  - Connection protection/restoration capabilities
  - Separation b/w transmission, control and management plane
  - Network management using SNMP (dedicated MIB)
- Optical UNI
  - Signaling interface (demarcation) between the optical user equipment and Service Provider transport network
  - Signaling and routing interface between optical networking elements



Extending MPLS Protocols for the Optical/Unified Control Plane

Forwarding Plane Extends MPLS Labels

- DWDM
- Fiber
- TDM

Control Plane Extends MPLS-TE

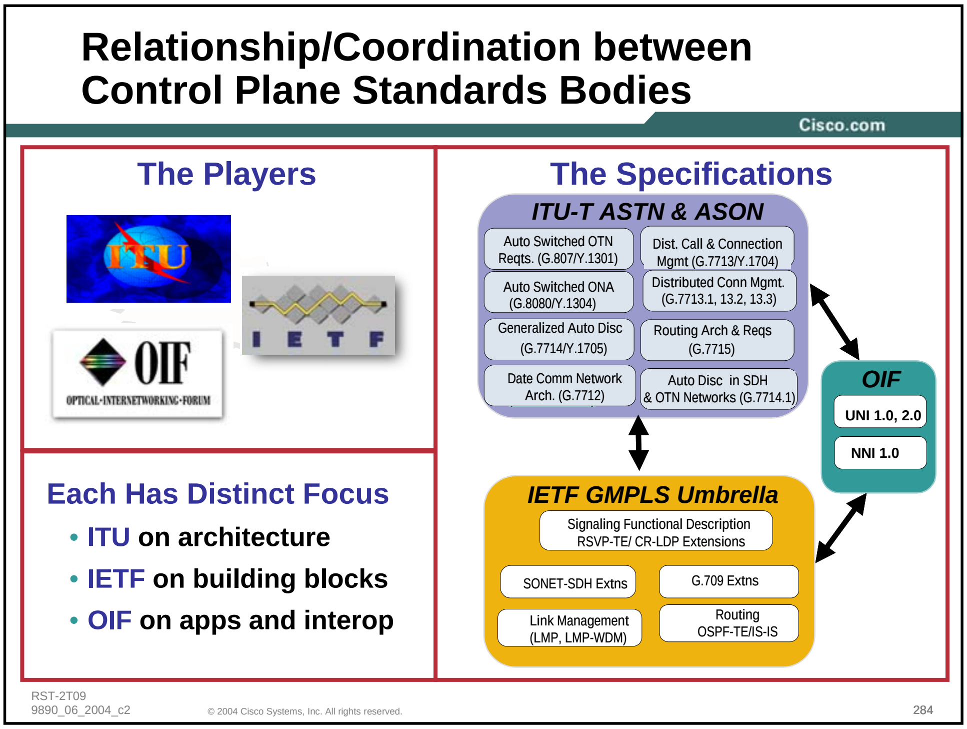
- OSPF-TE and IS-IS routing
- RSVP-TE for signaling
- Adds in-band and out-of-band control channel support
- LMP—Link Management Protocol

Mgmt and Control Address OTN Specific Needs

- Physical vs. logical
- Transport reqs (e.g., protection and restoration, explicit interfaces, etc.)

Generalized Multi-Protocol Label Switching (GMPLS)

Network Topology Map
Topology Policy Constraints
Metro
Label Forwarding Information Base
IP
ATM
Today
Next
Optical



Relationship/Coordination between Control Plane Standards Bodies

The Players

ITU
OIF
OPTICAL-INTERNETWORKING-FORUM
IETF

Each Has Distinct Focus

- **ITU** on architecture
- **IETF** on building blocks
- **OIF** on apps and interop

The Specifications

ITU-T ASTN & ASON

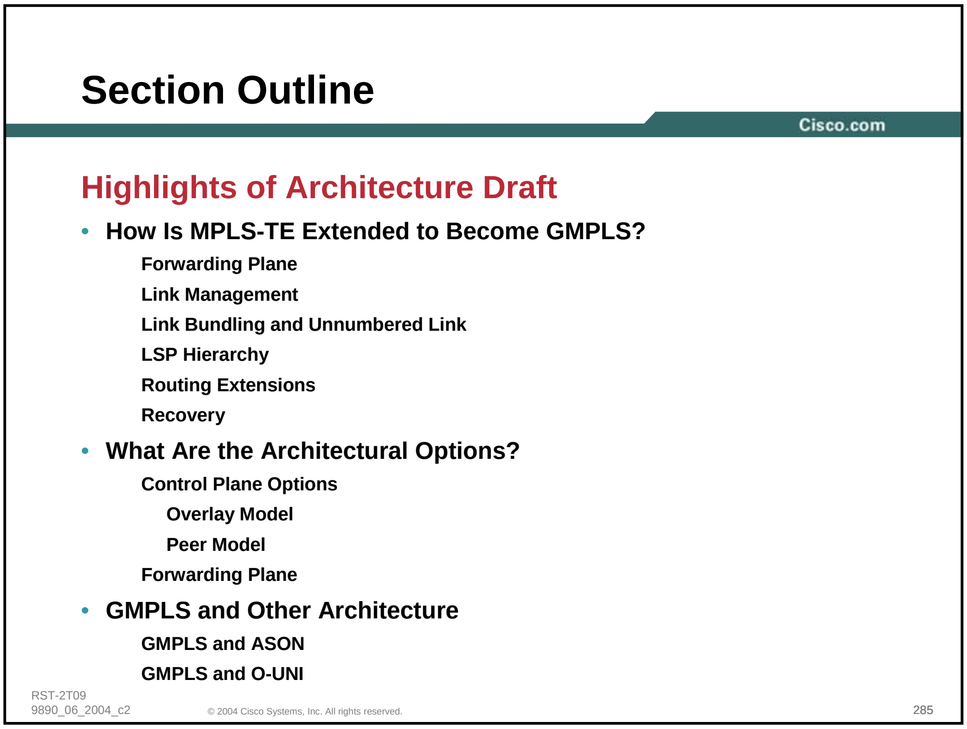
- Auto Switched OTN Reqts. (G.807/Y.1301)
- Dist. Call & Connection Mgmt (G.7713/Y.1704)
- Auto Switched ONA (G.8080/Y.1304)
- Distributed Conn Mgmt. (G.7713.1, 13.2, 13.3)
- Generalized Auto Disc (G.7714/Y.1705)
- Routing Arch & Reqs (G.7715)
- Date Comm Network Arch. (G.7712)
- Auto Disc in SDH & OTN Networks (G.7714.1)

OIF

- UNI 1.0, 2.0
- NNI 1.0

IETF GMPLS Umbrella

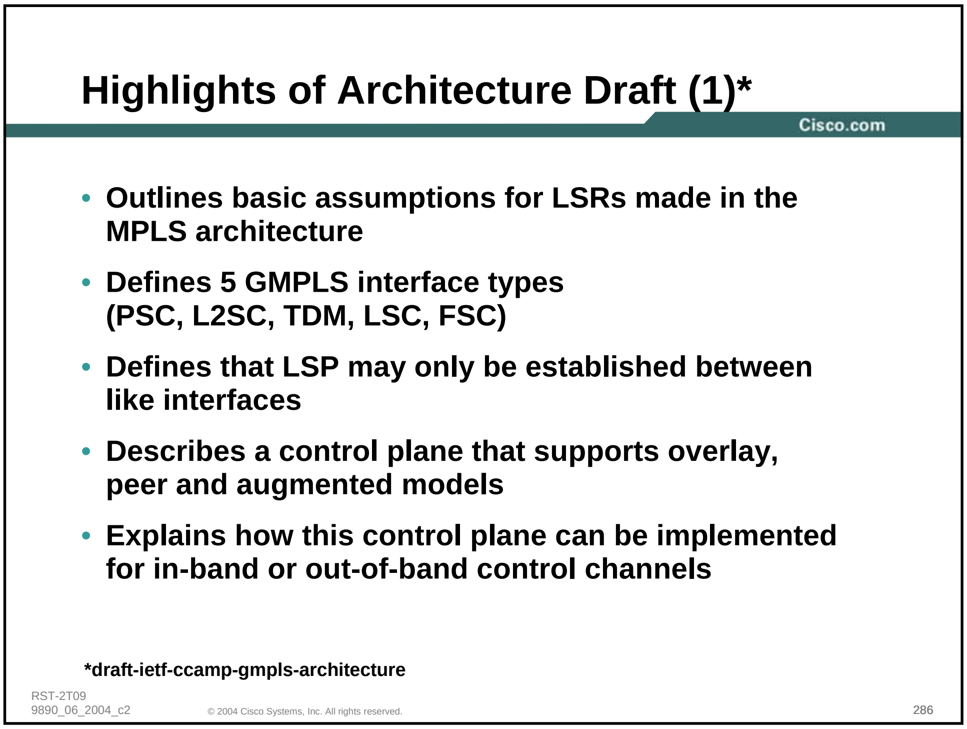
- Signaling Functional Description RSVP-TE/ CR-LDP Extensions
- SONET-SDH Extns
- G.709 Extns
- Link Management (LMP, LMP-WDM)
- Routing OSPF-TE/IS-IS



Section Outline

Highlights of Architecture Draft

- How Is MPLS-TE Extended to Become GMPLS?
    - Forwarding Plane
    - Link Management
    - Link Bundling and Unnumbered Link
    - LSP Hierarchy
    - Routing Extensions
    - Recovery
- What Are the Architectural Options?
    - Control Plane Options
        - Overlay Model
        - Peer Model
    - Forwarding Plane
- GMPLS and Other Architecture
    - GMPLS and ASON
    - GMPLS and O-UNI



Highlights of Architecture Draft (1)*

- Outlines basic assumptions for LSRs made in the MPLS architecture
- Defines 5 GMPLS interface types (PSC, L2SC, TDM, LSC, FSC)
- Defines that LSP may only be established between like interfaces
- Describes a control plane that supports overlay, peer and augmented models
- Explains how this control plane can be implemented for in-band or out-of-band control channels

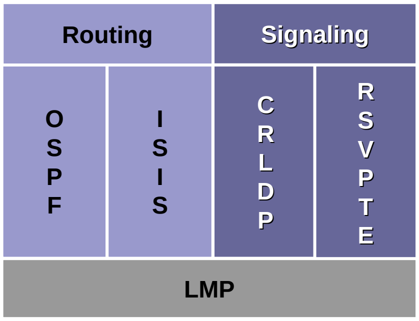
*draft-ietf-ccamp-gmpls-architecture

Highlights of Architecture Draft (2)*

- Extends the concept of a label to include implicit labelling via timeslots, wavelengths and fibres
- Describes the extensions to existing routing and signaling protocols
- Highlights that only one new protocol (LMP) is needed
- Describes the need for un-numbered links
- Describes the need for link bundling as an aid to scalability

*draft-ietf-ccamp-gmpls-architecture

Basic Concepts and Components



- Topology discovery
  - Running an IGP (OSPF or IS-IS) with extensions
- Route computation
  - Route computation done by NEs
  - Link state aggregation and lack of lightpath related information affects efficiency
- Neighbor discovery
  - Link Management Protocol like LMP/NDP run in distributed way
- Lightpath setup
  - Done by ingress NE using signaling protocol like RSVP-TE

RFC 3472 GMPLS Signaling CR-LDP Extensions

RFC 3473 GMPLS Signaling RSVP-TE Extensions

GMPLS Mechanisms

- GMPLS control plane supports
    - Domain and unified Service Model
    - Overlay, augmented, and peer Control Plane interconnection model (known as overlay and Peer Models)
- GMPLS control plane architecture includes several extended MPLS-TE building blocks
    - Signalling protocols: RSVP-TE and CR-LDP
    - Routing protocol extension: OSPF-TE, ISIS-TE
    - Link Management Protocol (LMP): new protocol
    - Bi-directional LSP

GMPLS Mechanisms

TE-Routing Enhanced Scalability and Flexibility

- Link bundling (TE-links)
- Forwarding adjacencies (generalized virtual TE-links)
- Generalized unnumbered interfaces
- Extended explicit routing

Section Outline

Highlights of Architecture Draft

- How Is MPLS-TE Extended to Become GMPLS?
  - Forwarding Plane
  - Link Management
  - Link Bundling and Unnumbered Link
  - LSP Hierarchy
  - Routing Extensions
  - Recovery
- What Are the Architectural Options?
  - Control Plane Options
    - Overlay Model
    - Peer Model
  - Forwarding Plane
- GMPLS and Other Architecture
  - GMPLS and ASON
  - GMPLS and O-UNI

Extending MPLS-TE Model
Basic Building Blocks

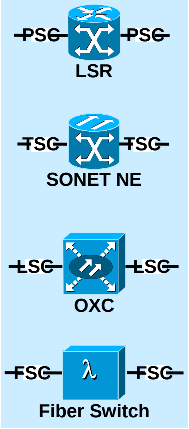
1. A new generic label request format
2. Labels for TDM, LSC and FSC interfaces, known as Generalised Label
3. Waveband switching support
4. Support of unnumbered and bundled link
5. Label restriction by the upstream to support some optical constraints
6. Bi-directional LSP establishment
7. In-band and out-of-band control channel
8. Rapid failure notification to ingress node
9. Protection information currently focusing on link protection, plus primary and secondary LSP indication
10. Explicit routing with explicit label control for a fine degree of control
11. Specific traffic parameters per technology

Highlighted = Mandatory

Forwarding Plane

Must Match Expectation with Product Capability, AND Work Within Limitations of Legacy Equipment

- Packet-Switch Capable Interfaces (PSC)
- Time-Division Multiplex Capable Interfaces (TDM)
- Layer 2 Switch Capable Interfaces (L2SC)
- Lambda Switch Capable Interfaces (LSC)
- Fibre-Switch Capable Interfaces (FSC)



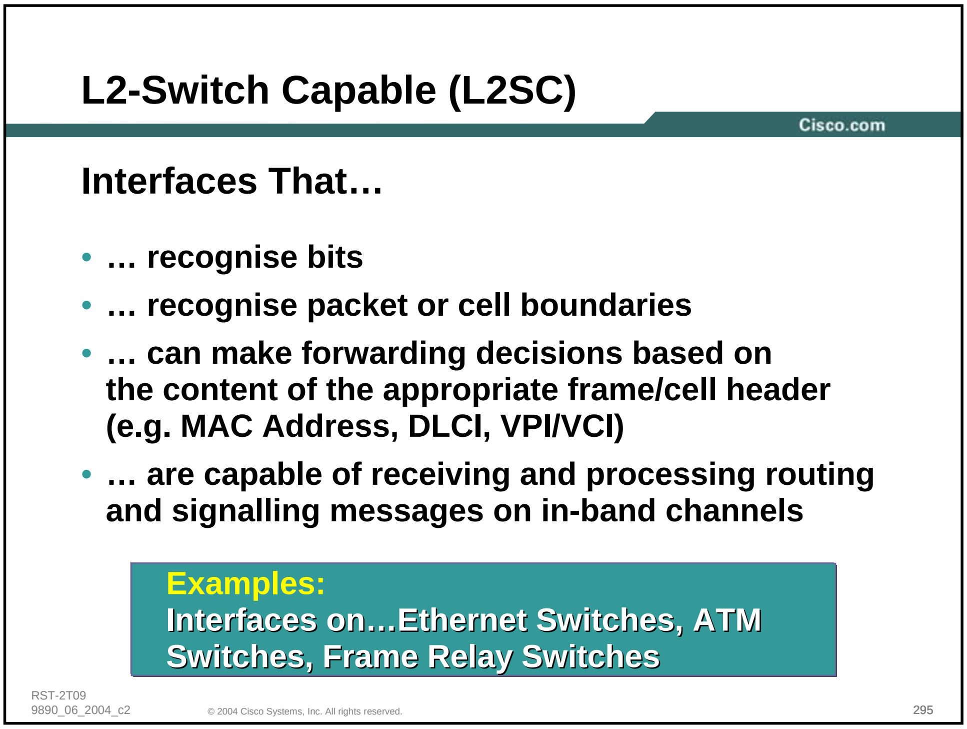
RFC 3471 GMPLS Signaling Functional Description

Packet-Switch Capable (PSC)

Interfaces That…

- … recognise bits
- … recognise packet or cell boundaries
- … can make forwarding decisions based on the content of the appropriate packet or MPLS header
- … are capable of receiving and processing routing and signalling messages on in-band channels

Examples:
Interfaces on…Routers, MPLS LSRs

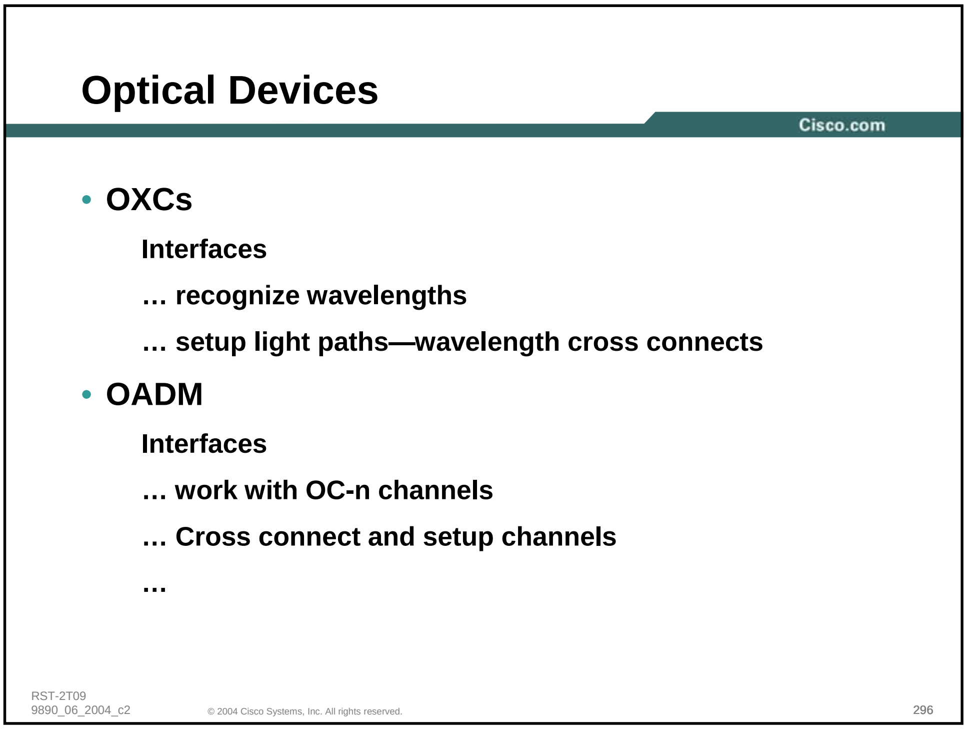


L2-Switch Capable (L2SC)

Interfaces That…

- … recognise bits
- … recognise packet or cell boundaries
- … can make forwarding decisions based on the content of the appropriate frame/cell header (e.g. MAC Address, DLCI, VPI/VCI)
- … are capable of receiving and processing routing and signalling messages on in-band channels

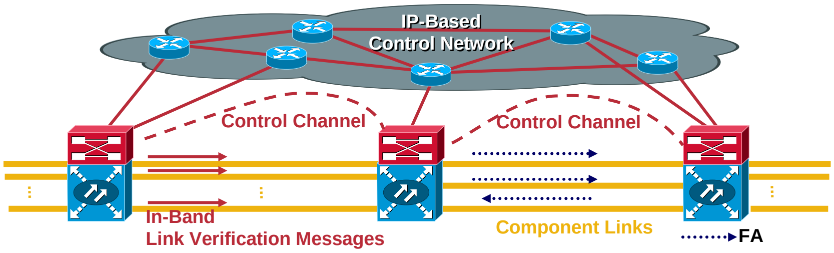
Examples:
Interfaces on…Ethernet Switches, ATM Switches, Frame Relay Switches



Optical Devices

- OXCs
  - Interfaces
  - … recognize wavelengths
  - … setup light paths—wavelength cross connects
- OADM
  - Interfaces
  - … work with OC-n channels
  - … Cross connect and setup channels
  - …

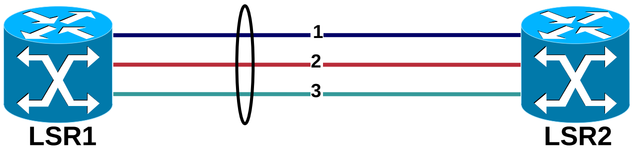
LMP and Link Management


IP-Based
Control Network
Control Channel
Control Channel
In-Band
Link Verification Messages
Component Links
FA

- LMP Functionality
  - Most LMP messages sent out-of-band through CC
  - In-band messages sent for Component Link Verification
  - Once allocated, Component Link is not assumed to be opaque
  - Port ID mapping
  - One CC per one or more Component Link Bundles
  - Fault isolation
  - End-system and service discovery (UNI related)
- Flooding Adjacencies are maintained over CC (via control network)
- Forwarding Adjacencies (FA) are maintained over Component Links and announced as links into the IGP

draft-ietf-ccamp-lmp
draft-ietf-ccamp-lmp-wdm

Link Bundling and Unnumbered Links

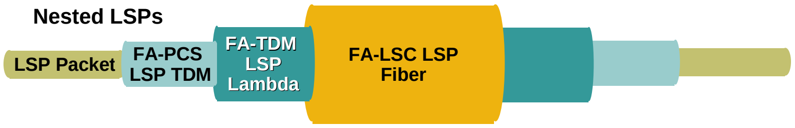

1
2
3
LSR1
LSR2

- Issue
  - Neighboring LSRs connected by multiple parallel links
  - Each link is addressed at each end and advertised into routing database… lots of links!!!
- Solution
  - Aggregate multiple components links into a single abstract link
  - Use (router ID, interface #) for link identifiers
- Reduces number of links in routing database and amount of per-link configuration

draft-kompella-mpls-bundle
draft-kompella-mpls-unnum

LSP Hierarchy

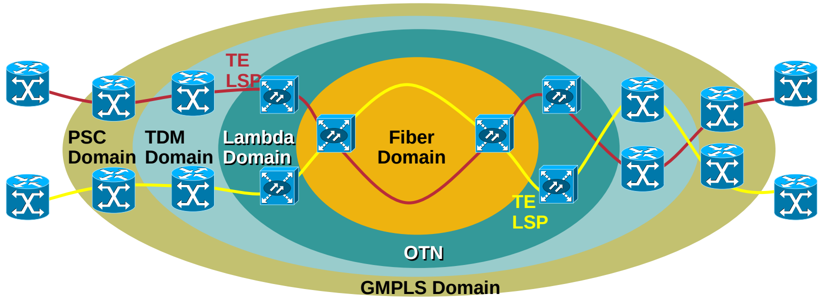
FA-LSP…FORWARDING ADJACENCY LSP


Nested LSPs
LSP Packet
FA-PCS LSP TDM
FA-TDM LSP Lambda
FA-LSC LSP Fiber

- Enables aggregation of GMPLS LSP tunnels
- Accomplished by
  - Inter-LSR LSP tunnel (FA-LSP) link is created
  - Ingress LSR injects link (FA-LSP) into IGP database
  - Other routers use the link in path calculation/setup
  - Other LSP tunnels are nested inside FA-LSP
- Advantages
  - Fewer high-order labels (e.g. lambdas) consumed
  - Nested LSPs can be of non-discrete bandwidth
  - FA-LSP can “hide” topology

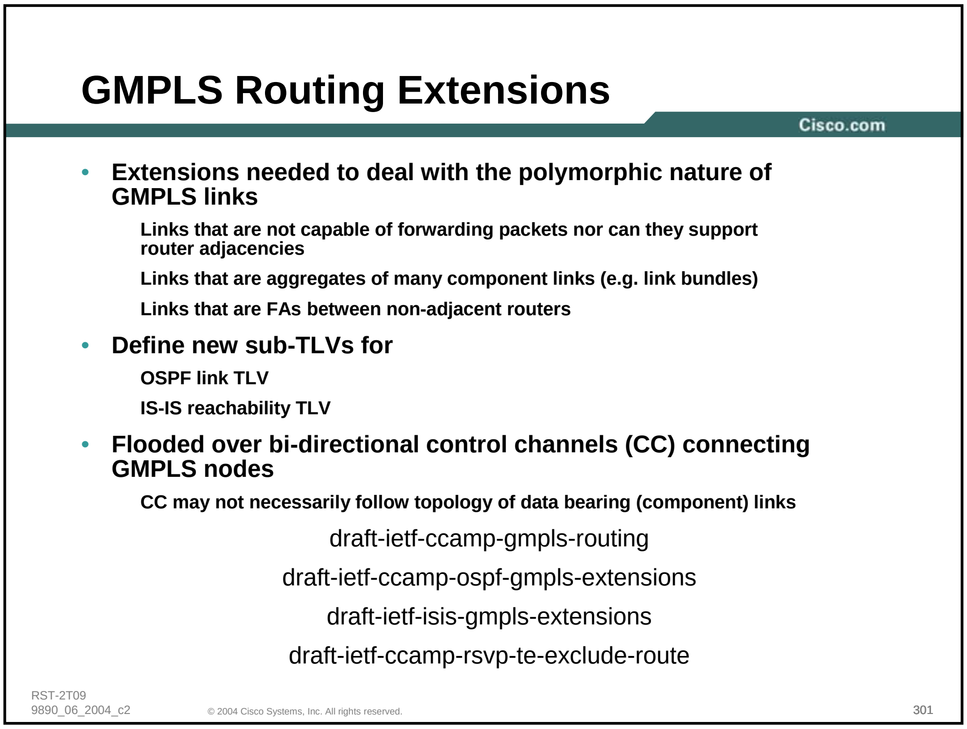
draft-ietf-mpls-lsp-hierarchy

GMPLS Signaling


TE LSP
PSC Domain
TDM Domain
Lambda Domain
Fiber Domain
TE LSP
OTN
GMPLS Domain

- Extended label semantics for Fiber, Waveband, Lambda, TDM and PSC LSP setup
- Extend RSVP-TE/CR-LDP to support new label objects over explicit/non explicit path
- Suggested label—conveyed by upstream LSR to downstream LSR to speed up configuration (on upstream)
- Label set—limits choice of labels that downstream LSR can choose from
  - If no wavelength conversion available then same lambdas must be used ete
- Bidirectional LSP setup

RFC 3471 Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description



GMPLS Routing Extensions

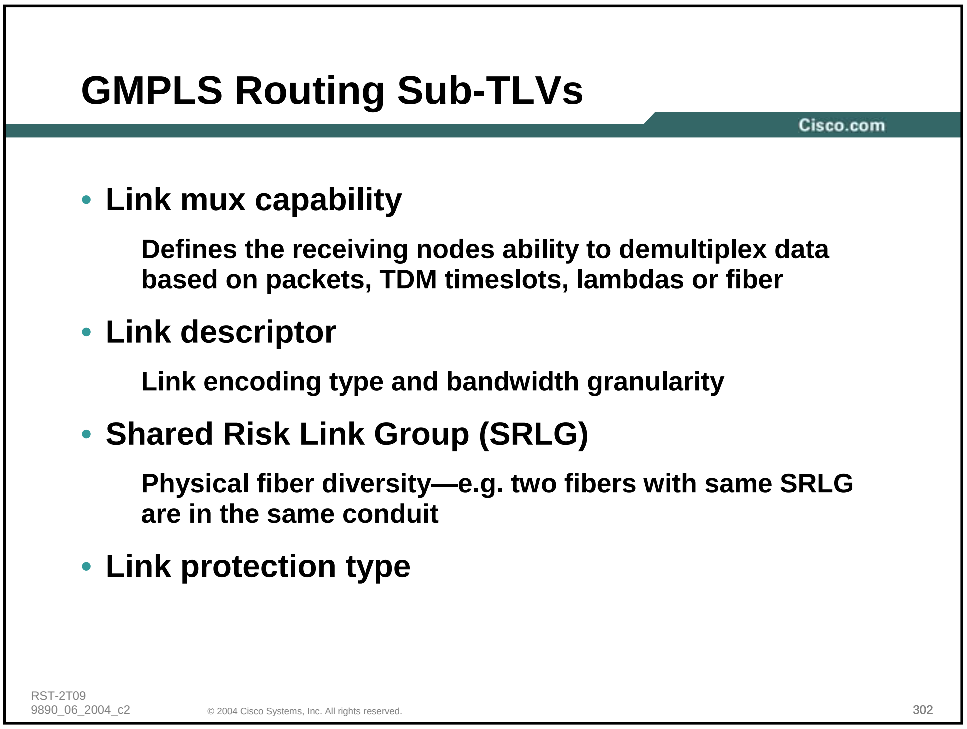
- Extensions needed to deal with the polymorphic nature of GMPLS links
  - Links that are not capable of forwarding packets nor can they support router adjacencies
  - Links that are aggregates of many component links (e.g. link bundles)
  - Links that are FAs between non-adjacent routers
- Define new sub-TLVs for
  - OSPF link TLV
  - IS-IS reachability TLV
- Flooded over bi-directional control channels (CC) connecting GMPLS nodes
  - CC may not necessarily follow topology of data bearing (component) links

draft-ietf-ccamp-gmpls-routing

draft-ietf-ccamp-ospf-gmpls-extensions

draft-ietf-isis-gmpls-extensions

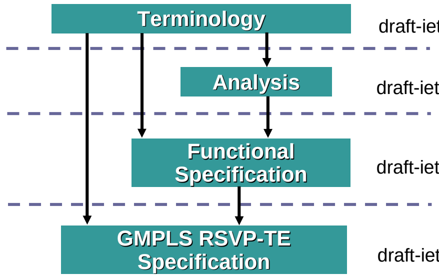
draft-ietf-ccamp-rsvp-te-exclude-route



GMPLS Routing Sub-TLVs

- Link mux capability
  - Defines the receiving nodes ability to demultiplex data based on packets, TDM timeslots, lambdas or fiber
- Link descriptor
  - Link encoding type and bandwidth granularity
- Shared Risk Link Group (SRLG)
  - Physical fiber diversity—e.g. two fibers with same SRLG are in the same conduit
- Link protection type

GMPLS-Based Recovery


Terminology
Analysis
Functional Specification
GMPLS RSVP-TE Specification

draft-ietf-ccamp-gmpls-recovery-terminology

draft-ietf-ccamp-gmpls-recovery-analysis

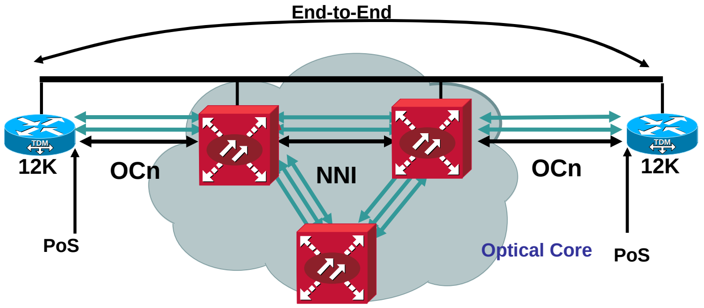
draft-ietf-ccamp-gmpls-recovery-functional

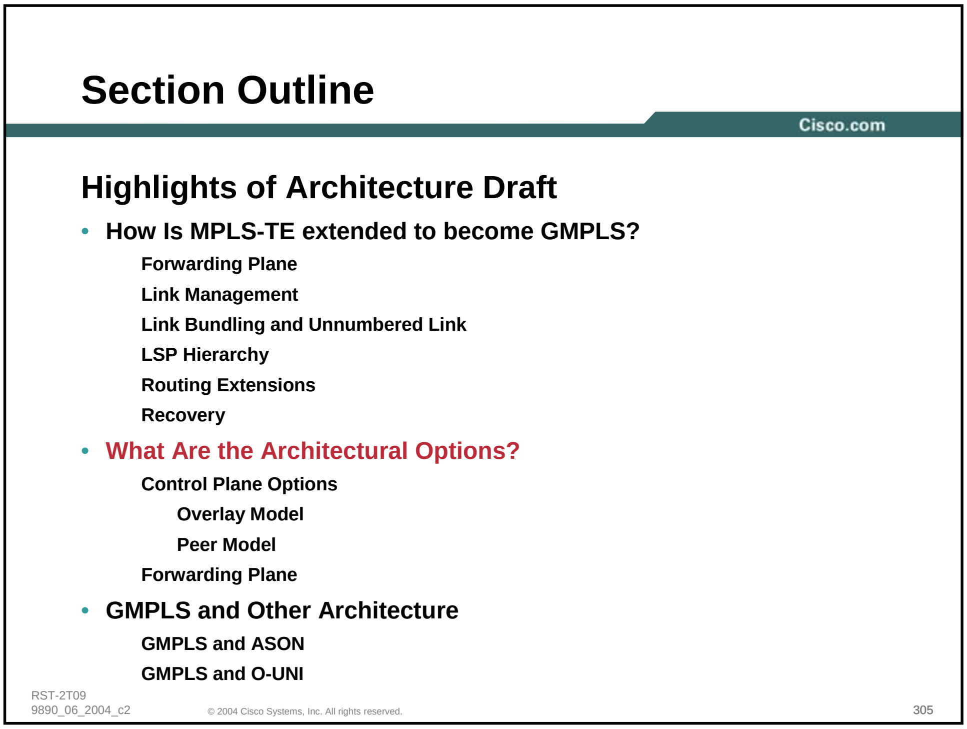
draft-ietf-ccamp-gmpls-recovery-e2e-signaling

- End-to-end path protection
  - 1+1 uni or bi-directional dedicated (traffic duplicated on the protection path)
  - “Head-end re-write”: back up path established, on stand-by
- Link/node protection
  - Similar to MPLS TE FRR…
- LSP dynamic rerouting (restoration)
  - Full LSP signalling after failure occurrence

Protection and Restoration

- Interface standby
  - Similar to link protection of MPLS FRR
- End-to-end protection
  - Similar to path protection of MPLS FRR

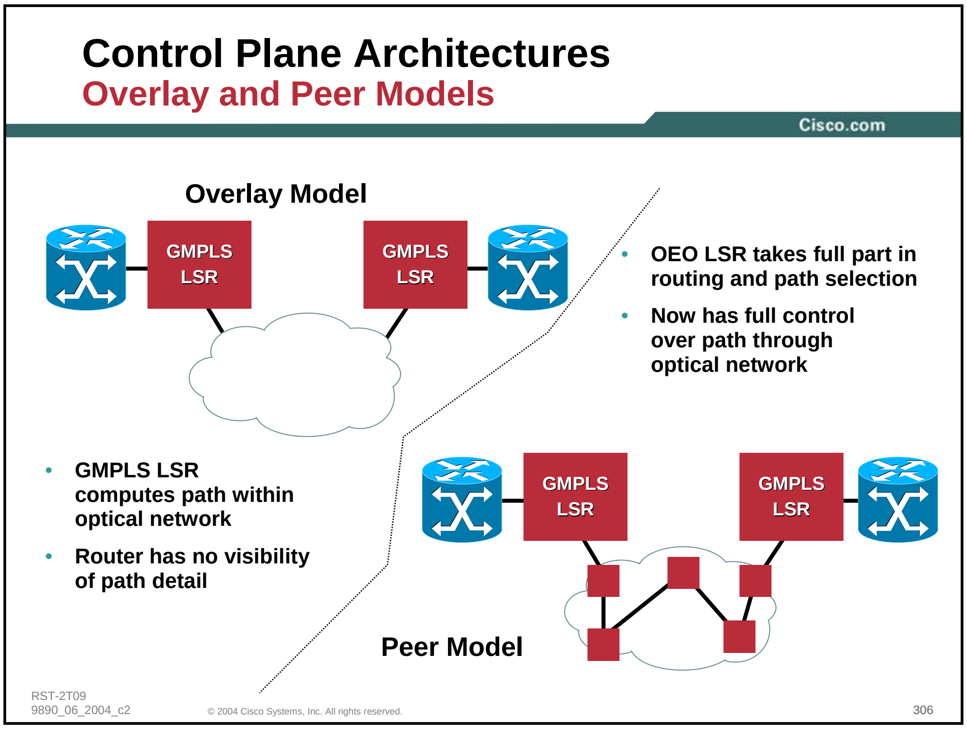

End-to-End
12K
OCn
NNI
OCn
12K
PoS
Optical Core
PoS



Section Outline

Highlights of Architecture Draft

- How Is MPLS-TE extended to become GMPLS?
  - Forwarding Plane
  - Link Management
  - Link Bundling and Unnumbered Link
  - LSP Hierarchy
  - Routing Extensions
  - Recovery
- What Are the Architectural Options?
  - Control Plane Options
    - Overlay Model
    - Peer Model
  - Forwarding Plane
- GMPLS and Other Architecture
  - GMPLS and ASON
  - GMPLS and O-UNI



Control Plane Architectures
Overlay and Peer Models

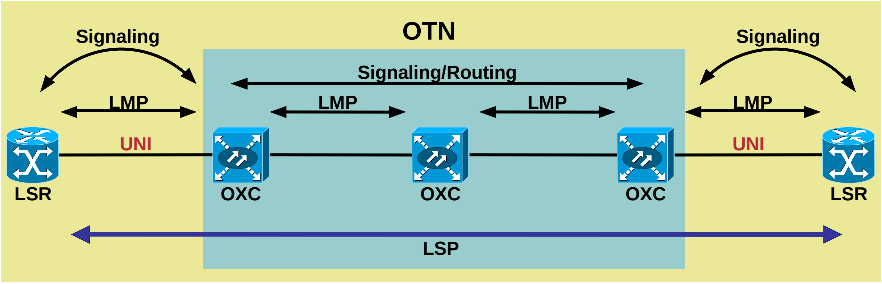
- OEO LSR takes full part in routing and path selection
- Now has full control over path through optical network

- GMPLS LSR computes path within optical network
- Router has no visibility of path detail

Overlay vs. Peer Models

- Overlay Model with UNI is preferred in environments when clients and optical domain operated by different entities (with limited, or unknown trust)
  - Limit routing structure transfer
  - Apply policies and authentication
  - Use between IP autonomous systems
  - Similar to “BGP model”
- Peer Model is preferred in environments with full trust
  - Full routing information transfer
  - More efficient use of resources
  - Use inside IP autonomous systems
  - Follows IGP model

GMPLS Overlay Routing Model

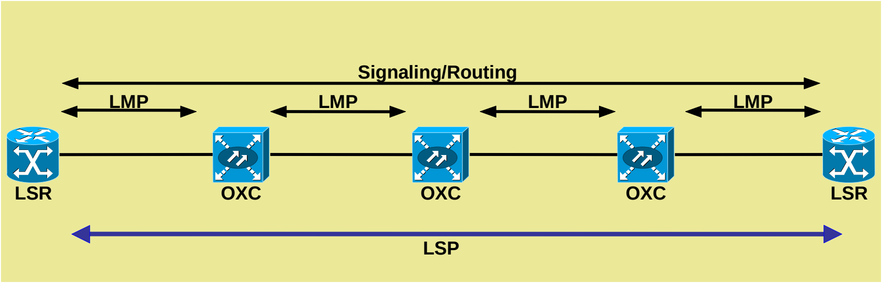

Signaling
OTN
Signaling
Signaling/Routing
LMP
LMP
LMP
LMP
UNI
UNI
LSR
OXC
OXC
OXC
LSR
LSP

- UNI interactions—GMPLS signaling, LMP
- OTN interactions—GMPLS signaling, routing, and LMP

draft-ietf-ccamp-gmpls-overlay-xx.txt
(RSVP Support for Overlay Model)

GMPLS Peer Routing Model

Cisco.com



- OTN interactions—GMPLS signaling, routing, and LMP
- GMPLS protocol machinery can support overlay or peer routing models

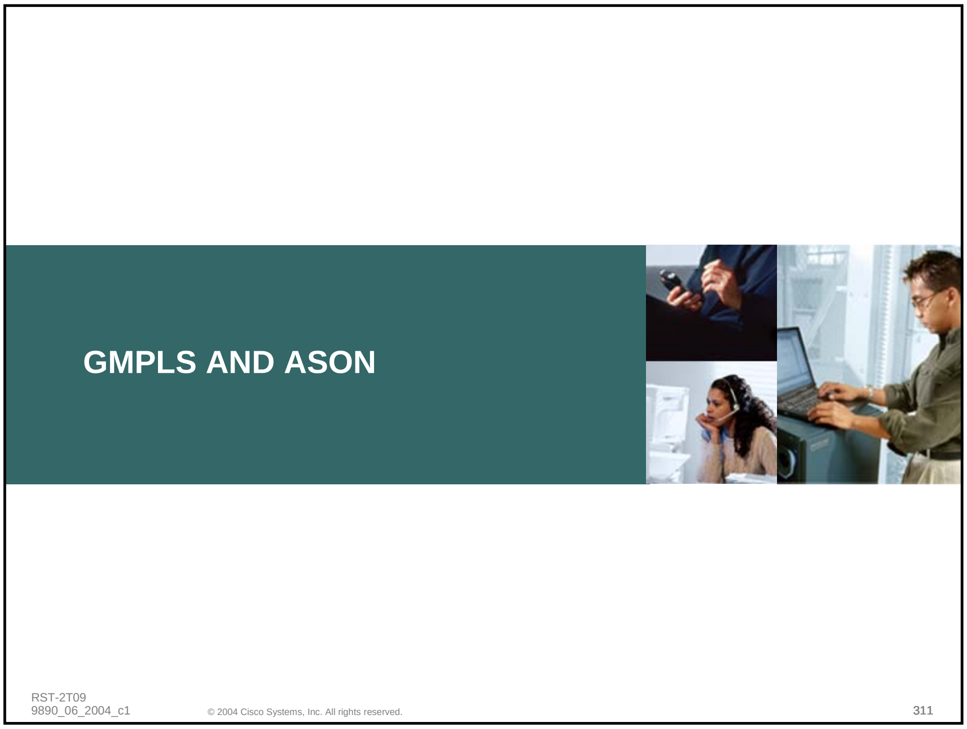
RFC 3473 GMPLS Signaling RSVP-TE Extensions

Section Outline

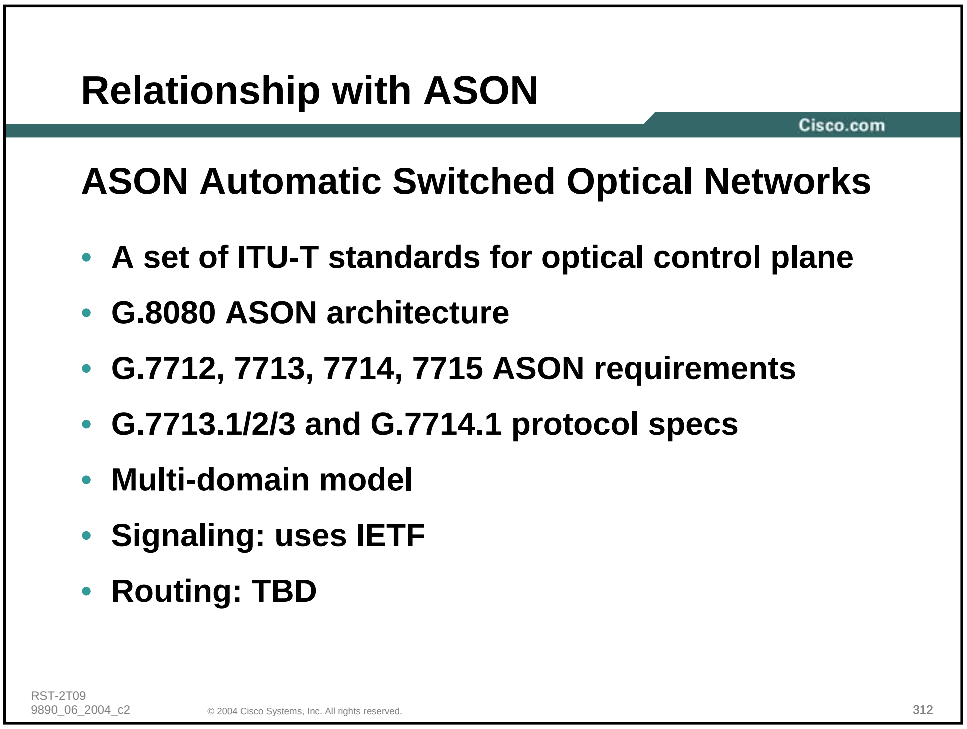
Cisco.com

Highlights of Architecture Draft

- How Is MPLS-TE extended to become GMPLS?
  - Forwarding Plane
  - Link Management
  - Link Bundling and Unnumbered Link
  - LSP Hierarchy
  - Routing Extensions
  - Recovery
- What Are the Architectural Options?
  - Control Plane Options
    - Overlay Model
    - Peer Model
  - Forwarding Plane
- GMPLS and Other Architecture
  - GMPLS and ASON
  - GMPLS and O-UNI



GMPLS AND ASON



Relationship with ASON

ASON Automatic Switched Optical Networks

- A set of ITU-T standards for optical control plane
- G.8080 ASON architecture
- G.7712, 7713, 7714, 7715 ASON requirements
- G.7713.1/2/3 and G.7714.1 protocol specs
- Multi-domain model
- Signaling: uses IETF
- Routing: TBD

GMPLS RSVP TE Signaling in Support of ASON

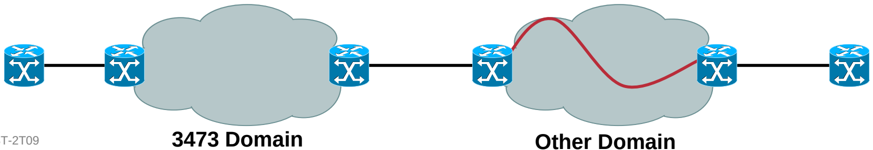
- Backward/forward compatible with GMPLS RFCs (RFC 3471/73)
- Independence between UNI and E-NNI (agnosticism)
- Interworking (at UNI and/or E-NNI) must be impact free on GMPLS RFCs
- Intra-Domain and Inter-Domain signaling
- Only define new object and procedures when strictly needed (max re-use principle)

Requirements	Info RFC 3474/76	Proposal
Soft Permanent Connection	Yes (SPC Label)	Yes (RFC 3473)
E2e Capability Negotiation	No	Yes
Call w/o Connection Setup	No	Yes
Call w/ (Single) Connection Setup	Yes (Limited to Single Hop Sessions)	Yes
Multiple Connections Per Call (Add/Remove)	No	Yes

draft-dimitri-ccamp-gmpls-rsvp-te-ason-xx.txt

G.7713.2/RFC3474: RFC3473 Interworking

- RFCs 3473 and 3474 interworking explained in
  - draft-ong-ccamp-3473-3474-iw-xx.txt
  - Specifics are in the draft
  - More details and clarifications to be added
- RFC 3474 key concepts
  - Overlay or multiple domain model
    - Client interface (overlay)
    - ENNI (between domains)
  - Client address space (TNA)
    - Separate address space and format
  - Call-ID and related information
    - Carried transparently across intermediate nodes
  - Multi-session RSVP
    - E2E connection stitched together from multiple tunnels



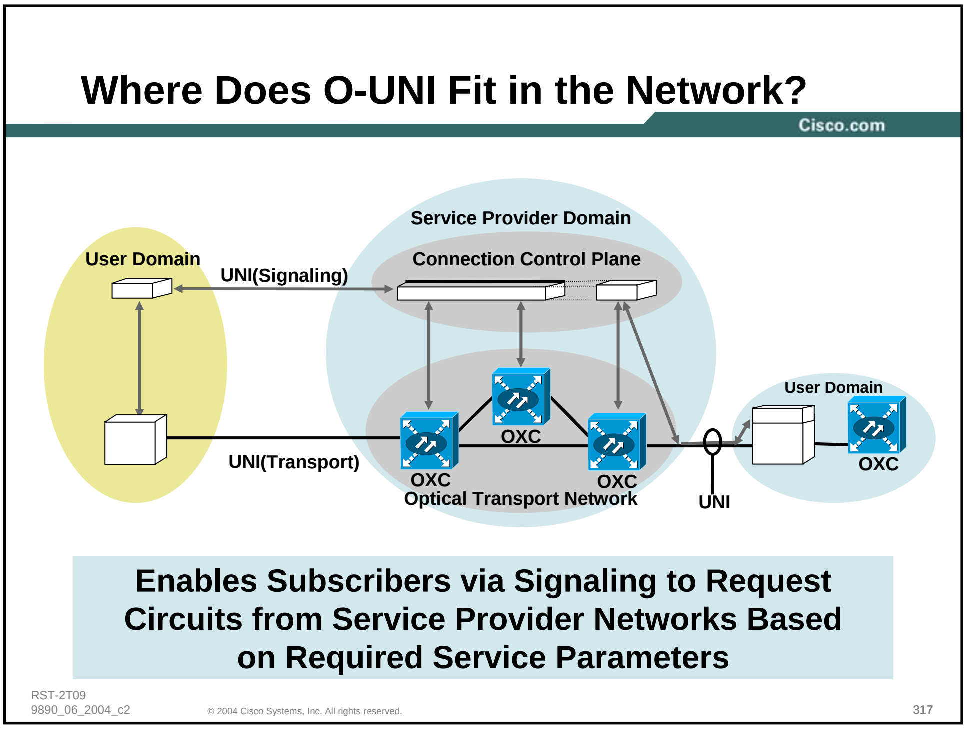
OPTICAL UNI

What Is O-UNI?

A Signaling Interface (Demarcation) between the Optical User Equipment and the Service Provider Transport Network!

Optical User Equipment (Client)

- **Service provider, enterprise, organization**
- **IP router, SONET/SDH, ATM NEs**

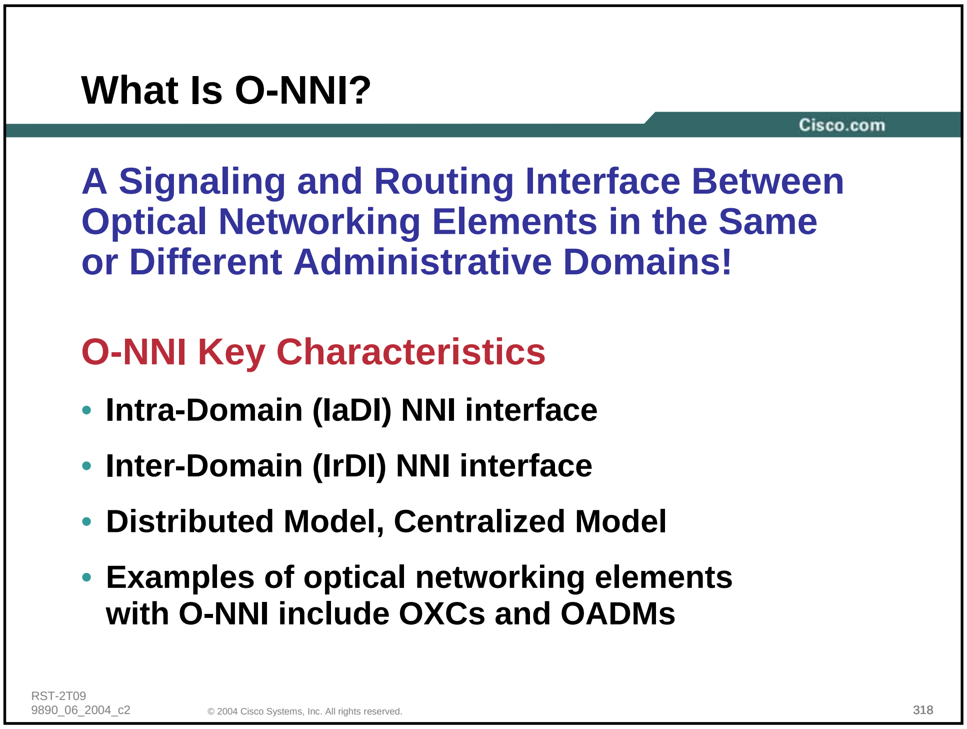


Where Does O-UNI Fit in the Network?

Cisco.com

Service Provider Domain
User Domain
Connection Control Plane
UNI(Signaling)
OXC
UNI(Transport)
OXC
OXC
Optical Transport Network
UNI
User Domain
OXC

Enables Subscribers via Signaling to Request Circuits from Service Provider Networks Based on Required Service Parameters



What Is O-NNI?

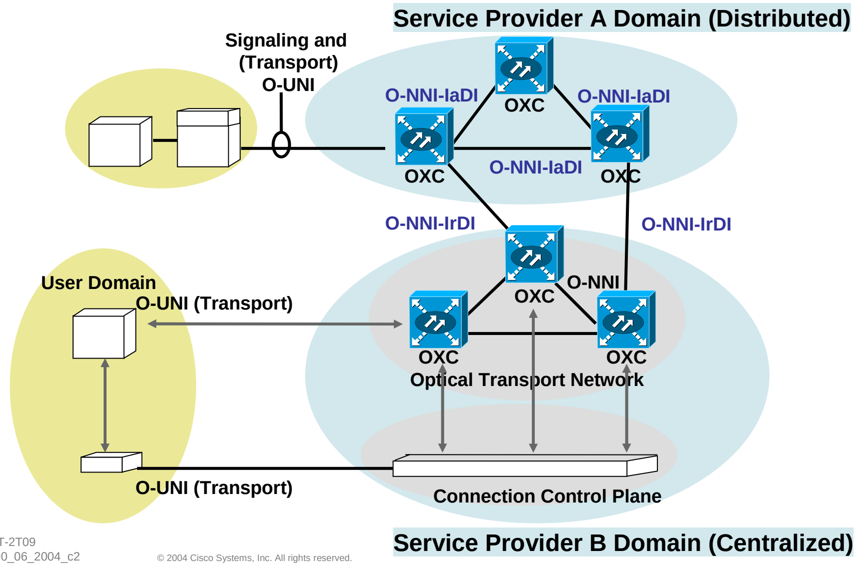
Cisco.com

A Signaling and Routing Interface Between Optical Networking Elements in the Same or Different Administrative Domains!

O-NNI Key Characteristics

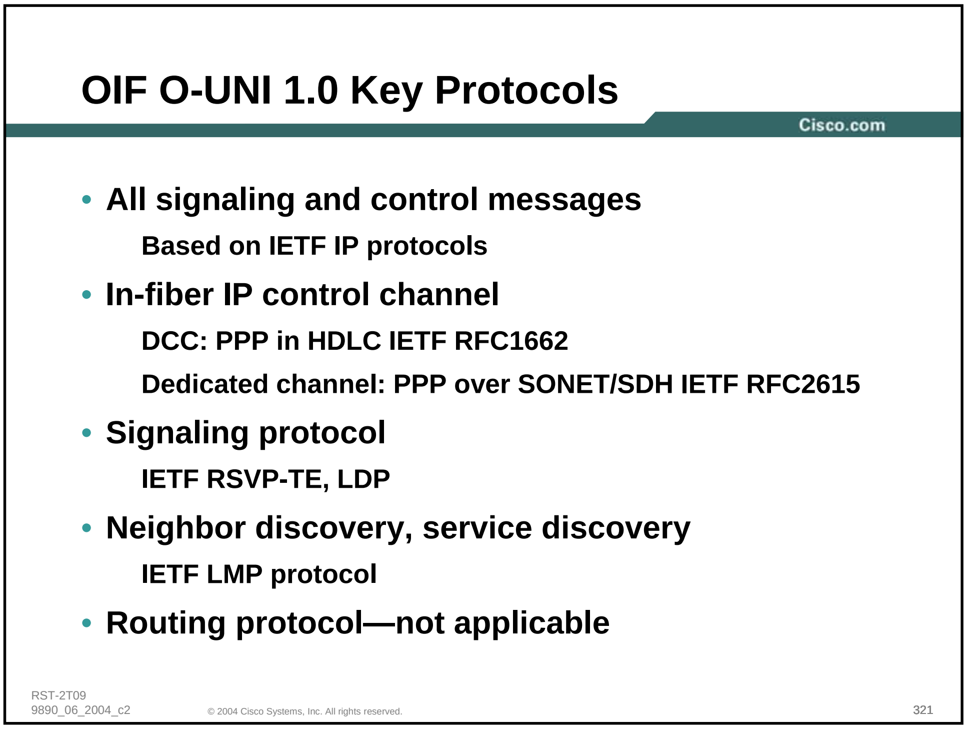
- Intra-Domain (IaDI) NNI interface
- Inter-Domain (IrDI) NNI interface
- Distributed Model, Centralized Model
- Examples of optical networking elements with O-NNI include OXCs and OADMs

OIF UNI/NNI Network Reference Model


Service Provider A Domain (Distributed)
Signaling and
(Transport)
O-UNI
O-NNI-IaDI
OXC
O-NNI-IaDI
OXC
O-NNI-IaDI
OXC
O-NNI-IrDI
O-NNI-IrDI
User Domain
O-UNI (Transport)
OXC
O-NNI
OXC
OXC
Optical Transport Network
O-UNI (Transport)
Connection Control Plane
Service Provider B Domain (Centralized)

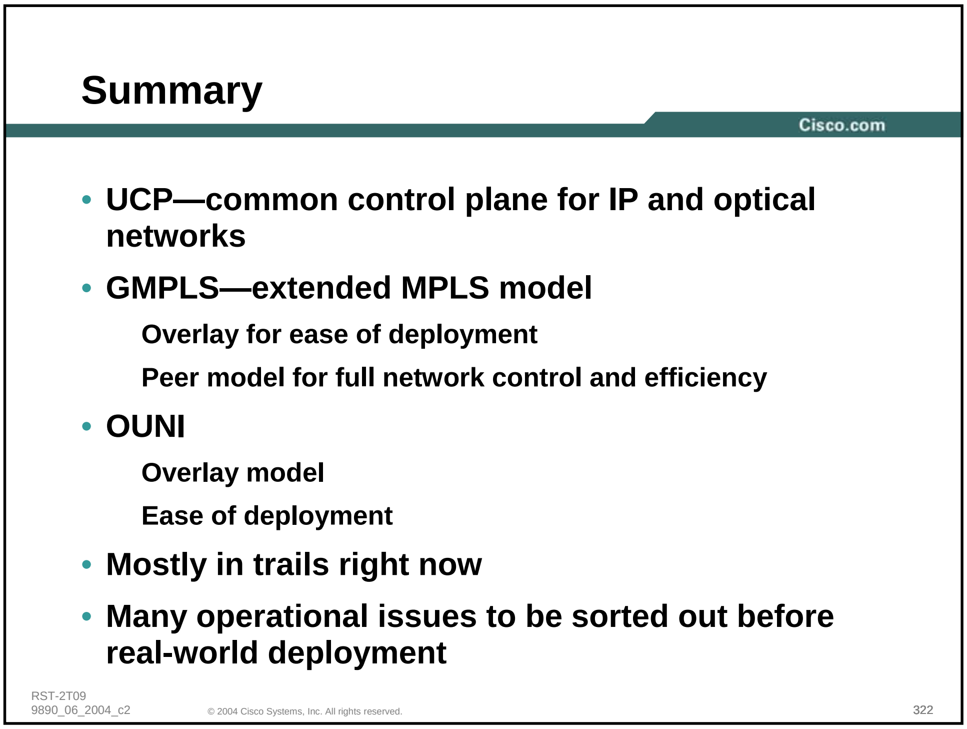
O-UNI Key Features

- Signaling Interface between optical network and clients
  - IP routers, ATM switches, SONET ADMs
- UNI functional components
  - Neighbor discovery and control channel maintenance
    - Control channel configuration
    - Hello initiation and link verification (up/down status)
    - Neighbor discovery information retrieval
  - Service discovery and address registration
    - Discovery of service attributes
    - Service granularity (min, max bandwidth)
    - Signaling protocols (RSVP-TE/LDP)
  - Signaling message exchange
    - Connection create, delete, status inquiry



OIF O-UNI 1.0 Key Protocols

- All signaling and control messages
  - Based on IETF IP protocols
- In-fiber IP control channel
  - DCC: PPP in HDLC IETF RFC1662
  - Dedicated channel: PPP over SONET/SDH IETF RFC2615
- Signaling protocol
  - IETF RSVP-TE, LDP
- Neighbor discovery, service discovery
  - IETF LMP protocol
- Routing protocol—not applicable



Summary

- UCP—common control plane for IP and optical networks
- GMPLS—extended MPLS model
  - Overlay for ease of deployment
  - Peer model for full network control and efficiency
- OUNI
  - Overlay model
  - Ease of deployment
- Mostly in trails right now
- Many operational issues to be sorted out before real-world deployment

Q & A

References: GMPLS

- **E. Mannie (Editor) et al., ‘Generalized MPLS Architecture’, Informational Draft, draft-ietf-ccamp-gmpls-architecture-07.txt, May 2003**
- **Lou Berger (Editor), et al., ‘Generalized MPLS Signaling—Signaling Functional Description,’ Internet Draft, RFC 3471, January 2003**
- **Lou Berger (Editor) et al., ‘Generalized MPLS Signaling—RSVP-TE Extensions,’ Internet Draft, RFC 3473, January 2003**
- **Lou Berger (Editor) et al., ‘Generalized MPLS Signaling—CR-LDP Extensions,’ Internet Draft, RFC 3472, January 2003**
- **E. Mannie (Editor) et al., ‘Generalized MPLS Extensions for SONET and SDH Control’, Internet Draft, Work in progress, draft-ietf-ccamp-gmpls-sonet-sdh-08.txt, February 2003**
- **D. Papadimitriou (Editor) et al., ‘Generalized MPLS Extensions for G.079 Optical Transport Networks Control’, Internet Draft, Work in progress, draft-fontana-ccamp-gmpls-g709-07.txt, March 2004**

References: GMPLS-TE

- K. Kompella et al., “Routing Extensions in Support of Generalized MPLS”, Internet Draft, Work in progress, draft-ietf-ccamp-gmpls-routing-09.txt, October 2003
- K. Kompella et al., “IS-IS Extensions in Support of Generalized MPLS”, Internet Draft, Work in progress, draft-ietf-isis-gmpls-extensions-19.txt, October 2003
- K. Kompella et al. “OSPF Extensions in Support of Generalized MPLS”, Internet Draft, Work in progress, draft-ietf-ccamp-ospf-gmpls-extensions-12.txt, October 2003
- G. Swallow et al., ‘GMPLS UNI: RSVP Support for the Overlay Model ’, Internet Draft, Work in progress, draft-ietf-ccamp-gmpls-overlay-04.txt, April 2004.
- P. Lang (Editor) et al., ‘GMPLS Recovery Functional Specification’, Internet Draft, Work in progress, draft-ietf-ccamp-gmpls-recovery-functional-02.txt, April 2004

References: (G)MPLS-TE

- J. Lang (Editor) et al., “Link Management Protocol”, Internet Draft, Work in progress, draft-ietf-ccamp-lmp-10.txt, October 2003
- K. Kompella, Y. Rekhter, “Signalling Unnumbered Links in RSVP-TE”, Internet Draft, RFC 3477, January 2003
- K. Kompella, Y. Rekhter, “Signalling Unnumbered Links in CR-LDP”, Internet Draft, RFC 3480, February 2003
- K. Kompella and Y. Rekhter, LSP Hierarchy with Generalized MPLS TE, Internet Draft, Work in progress, draft-ietf-mpls-lsp-hierarchy-08.txt, March 2002
- K. Kompella, Y. Rekhter and L. Berger, “Link Bundling in MPLS Traffic Engineering”, Internet Draft, Work in progress, draft-ietf-mpls-bundle-04.txt, January 2003

MPLS DEPLOYMENT EXPERIENCE

Agenda

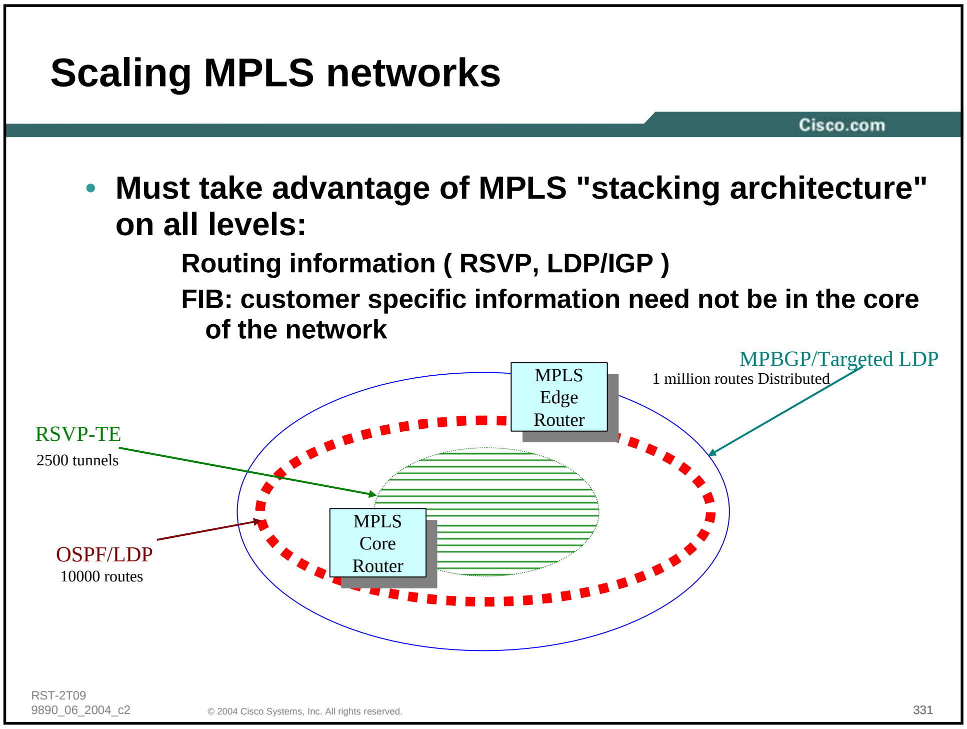
Cisco.com

- **Scalability**
- **Core vs Edge**
- **Management**
- **Migrations L3 and L2**

Assumption:

Cisco.com

Existing IP only network -> MPLS multi-service network



Scaling MPLS networks

- Must take advantage of MPLS "stacking architecture" on all levels:
  - Routing information (RSVP, LDP/IGP)
  - FIB: customer specific information need not be in the core of the network

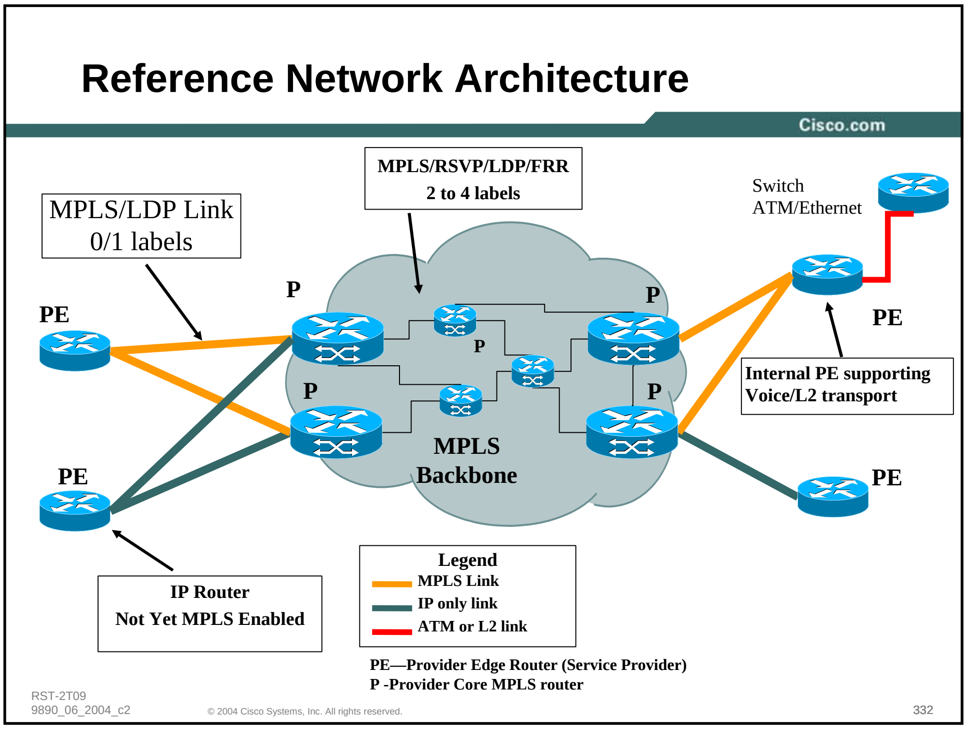
MPBGP/Targeted LDP
1 million routes Distributed

MPLS Edge Router

RSVP-TE
2500 tunnels

MPLS Core Router

OSPF/LDP
10000 routes



Reference Network Architecture

MPLS/RSVP/LDP/FRR
2 to 4 labels

MPLS/LDP Link
0/1 labels

Switch
ATM/Ethernet

PE

P

P

P

P

P

PE

PE

PE

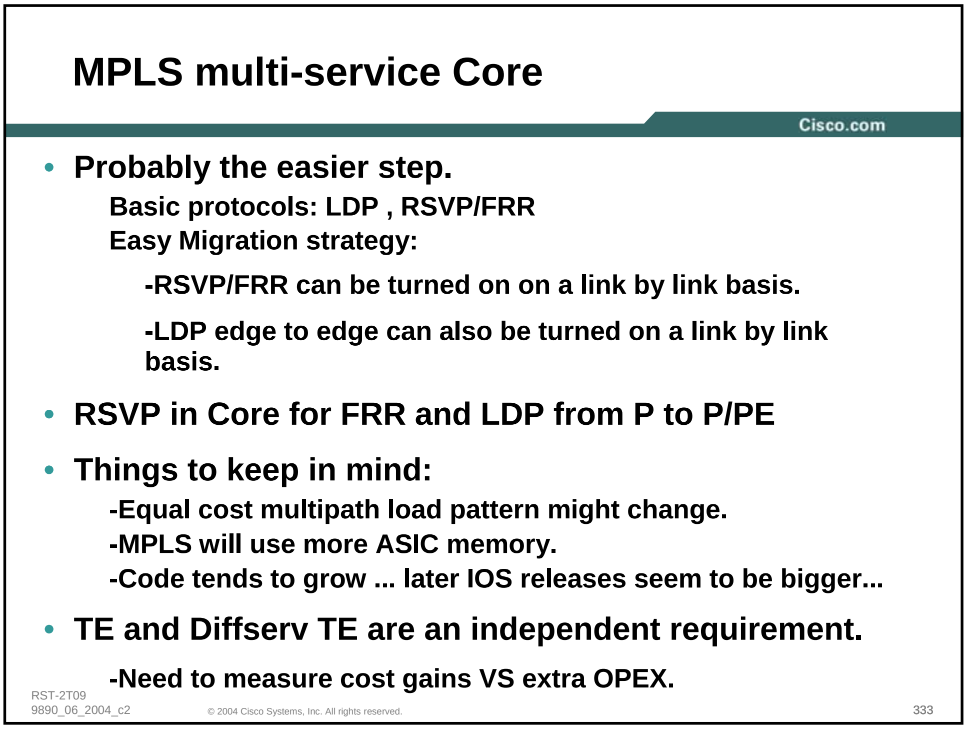
Internal PE supporting Voice/L2 transport

MPLS Backbone

IP Router
Not Yet MPLS Enabled

Legend
- MPLS Link
- IP only link
- ATM or L2 link

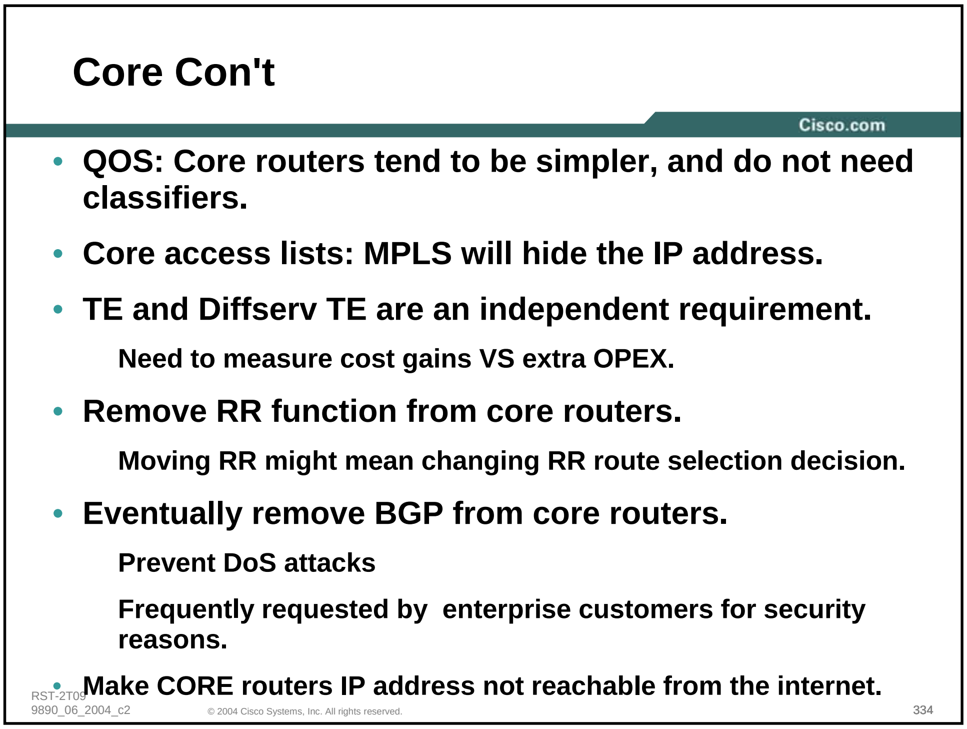
PE—Provider Edge Router (Service Provider)
P -Provider Core MPLS router



MPLS multi-service Core

Cisco.com

- Probably the easier step.
  - Basic protocols: LDP , RSVP/FRR
  - Easy Migration strategy:
    - -RSVP/FRR can be turned on on a link by link basis.
    - -LDP edge to edge can also be turned on a link by link basis.
- RSVP in Core for FRR and LDP from P to P/PE
- Things to keep in mind:
  - -Equal cost multipath load pattern might change.
  - -MPLS will use more ASIC memory.
  - -Code tends to grow ... later IOS releases seem to be bigger...
- TE and Diffserv TE are an independent requirement.
  - -Need to measure cost gains VS extra OPEX.



Core Con't

Cisco.com

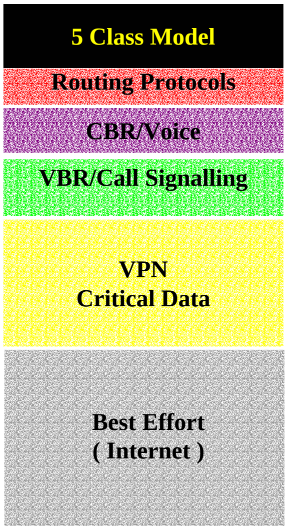
- QOS: Core routers tend to be simpler, and do not need classifiers.
- Core access lists: MPLS will hide the IP address.
- TE and Diffserv TE are an independent requirement.
  - Need to measure cost gains VS extra OPEX.
- Remove RR function from core routers.
  - Moving RR might mean changing RR route selection decision.
- Eventually remove BGP from core routers.
  - Prevent DoS attacks
  - Frequently requested by enterprise customers for security reasons.
- Make CORE routers IP address not reachable from the internet.

QoS and L2 services

- QoS is a necessary evil....
  - A must for failure, and unexpected traffic situations.
- Customer expectation is that the service works!
- Constant Bit Rate (CBR), Variable Bit Rate (VBR), Unspecified Bit Rate (UBR) class a minimum.
- Typically a 10Gb backbone vs classical OC3 ATM network.
  - Behaves better then OC3/OC12 ATM network.

QoS: Core

- Choose services: IP/Internet, L3 VPN, L2 VPN
- Edge OR Edge and Core ?
- Core queuing based on MPLS exp Bits.
- Keep the Core simple with at most 5 classes
- VoIP hardened , Video/ATM is not.
- Classify/Police at the input, queue in the core:
  - L2 QoS is the simplest, per VC , very controlled traffic flow.
  - L3 IP Internet is the most difficult/costly (traffic flow completely unpredictable)
- Capacity planing -> Per LSP data



5 Class Model

Routing Protocols

CBR/Voice

VBR/Call Signalling

VPN
Critical Data

Best Effort
(Internet)

MPLS Edge

- Easy migration: each edge element can be migrated independently.
- Memory consumption needs to be considered.
  - Any new IP based services like MPLS-VPN will require more memory.
  - Limit LDP label advertisement to loopbacks only.
- Disable propagate-ttl:
  - Customer traceroute no longer shows the core network hops.
- Before removing BGP for Core routers use accounting to check if all IP traffic is using MPLS LSPs.

Monitoring

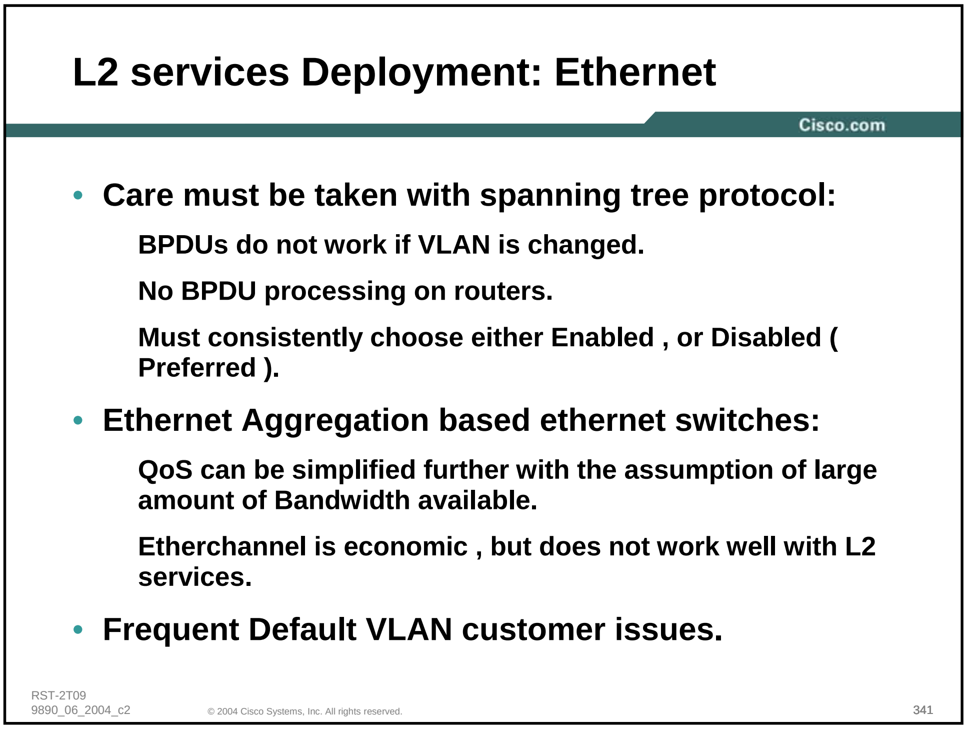
- MPLS LDP session Flaps.
- RSVP LSPs flaps
- Internal IP tools Like SAA might not report the correct status, as they run IP, not MPLS.
  - Use of MPLS VPN can help.
  - Or L2 mesh of PVCs
- IP is very resilient, other services will not be.
  - A network interruption of less then 10 seconds will not be detected by IP customers.
  - Packet loss of 0.01% will not be detected by IP customers, but will render video or voice unusable.

L2 point-to-point services

- Point-to-point links - operationally similar to classical frame/ATM.
- MPLS based Multi-service network.
- Service management granularity/scalability compromise
- Classes of Service:
  - Multiple CoS on same logical interface.
  - Multiple CoS on same physical port.
- Billing models:
  - PVC charge/ KB of CIR
  - Usage based (Port aggregate/PVC)

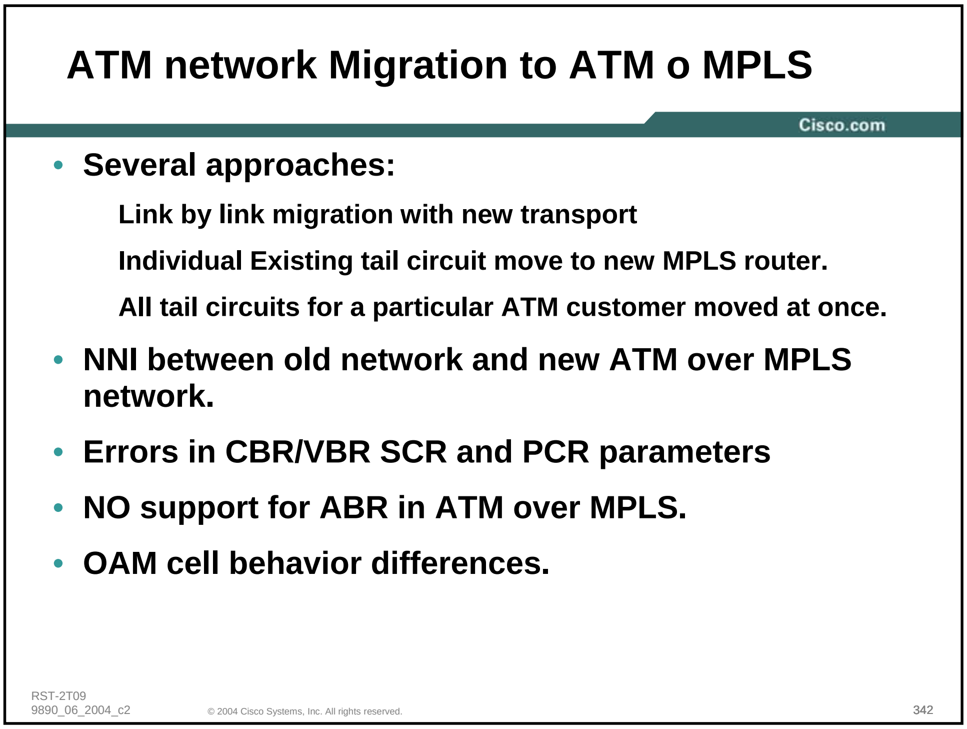
L2 Deployment OAM:

- Example - Ethernet:
  - No Management LMI/Ping
  - Monitor Error counters
  - Interface flaps (Syslog/Traps)
- Monitor PW status changes/MPLS RSVP events.
- Future Enhancements: IETF Virtual circuit connection Verification Draft
- SLA measurement: SAA or probe.
  - Qos Classes will be identical without network failure condition
  - One probe infrastructure to measure all services.



L2 services Deployment: Ethernet

- Care must be taken with spanning tree protocol:
  - BPDUs do not work if VLAN is changed.
  - No BPDU processing on routers.
  - Must consistently choose either Enabled , or Disabled (Preferred).
- Ethernet Aggregation based ethernet switches:
  - QoS can be simplified further with the assumption of large amount of Bandwidth available.
  - Etherchannel is economic , but does not work well with L2 services.
- Frequent Default VLAN customer issues.



ATM network Migration to ATM o MPLS

- Several approaches:
  - Link by link migration with new transport
  - Individual Existing tail circuit move to new MPLS router.
  - All tail circuits for a particular ATM customer moved at once.
- NNI between old network and new ATM over MPLS network.
- Errors in CBR/VBR SCR and PCR parameters
- NO support for ABR in ATM over MPLS.
- OAM cell behavior differences.

Complete Your Online Session Evaluation!

WHAT: Complete an online session evaluation and your name will be entered into a daily drawing

WHY: Win fabulous prizes! Give us your feedback!

WHERE: Go to the Internet stations located throughout the Convention Center

HOW: Winners will be posted on the onsite Networkers Website; four winners per day

Additional MPLS Sessions

- RST-1601 Introduction to MPLS
- RST-1607 QoS in MPLS Networks
- RST-2602 Deploying MPLS VPNs
- RST-2603 Deploying MPLS Traffic Engineering
- RST-2606 Understanding Convergence in MPLS VPN Networks
- RST-3605 Troubleshooting MPLS Networks
- RST-3606 Troubleshooting MPLS VPNs
- RST-4607 Advanced Topics and Future Directions in MPLS
- NMS-4012 MPLS Embedded Management Tools