

Implementing Cisco Service Provider Next-Generation Core Network Services

Volume 1

Version 1.01

Student Guide

Text Part Number: 97-3153-02



Americas Headquarters

Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters

Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters

Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

DISCLAIMER WARRANTY: THIS CONTENT IS BEING PROVIDED "AS IS" AND AS SUCH MAY INCLUDE TYPOGRAPHICAL, GRAPHICS, OR FORMATTING ERRORS. CISCO MAKES AND YOU RECEIVE NO WARRANTIES IN CONNECTION WITH THE CONTENT PROVIDED HEREUNDER, EXPRESS, IMPLIED, STATUTORY OR IN ANY OTHER PROVISION OF THIS CONTENT OR COMMUNICATION BETWEEN CISCO AND YOU. CISCO SPECIFICALLY DISCLAIMS ALL IMPLIED WARRANTIES, INCLUDING WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT AND FITNESS FOR A PARTICULAR PURPOSE, OR ARISING FROM A COURSE OF DEALING, USAGE OR TRADE PRACTICE. This learning product may contain early release content, and while Cisco believes it to be accurate, it falls subject to the disclaimer above.



Students, this letter describes important course evaluation access information!

Welcome to Cisco Systems Learning. Through the Cisco Learning Partner Program, Cisco Systems is committed to bringing you the highest-quality training in the industry. Cisco learning products are designed to advance your professional goals and give you the expertise you need to build and maintain strategic networks.

Cisco relies on customer feedback to guide business decisions; therefore, your valuable input will help shape future Cisco course curricula, products, and training offerings. We would appreciate a few minutes of your time to complete a brief Cisco online course evaluation of your instructor and the course materials in this student kit. On the final day of class, your instructor will provide you with a URL directing you to a short post-course evaluation. If there is no Internet access in the classroom, please complete the evaluation within the next 48 hours or as soon as you can access the web.

On behalf of Cisco, thank you for choosing Cisco Learning Partners for your Internet technology training.

Sincerely,

Cisco Systems Learning

Table of Contents

Volume 1

<i>Course Introduction</i>	<i>1</i>
Overview	1
Learner Skills and Knowledge	2
Course Goal and Objectives	3
Course Flow	4
Additional References	5
Cisco Glossary of Terms	5
Your Training Curriculum	6
Your Training Curriculum	7
<i>Multiprotocol Label Switching</i>	<i>1-1</i>
Overview	1-1
Module Objectives	1-1
<i>Introducing MPLS</i>	<i>1-3</i>
Overview	1-3
Objectives	1-3
Traditional ISP vs Traditional Telco	1-5
Modern Service Provider	1-7
Cisco IP NGN Architecture	1-8
SONET/SDH	1-10
DWDM and ROADM	1-11
IP over DWDM (IPoDWDM)	1-12
10/40/100 Gigabit Ethernet Standards	1-13
Transformation to IP	1-15
Traditional IP Routing	1-16
MPLS Introduction	1-17
MPLS Features	1-18
MPLS Benefits	1-19
MPLS Terminology	1-20
MPLS Architecture: Control Plane	1-22
MPLS Architecture: Data Plane	1-23
Forwarding Structures	1-24
MPLS Architecture Example	1-25
MPLS Labels	1-26
MPLS Packet Flow Basic Example	1-28
MPLS Label Stack	1-32
MPLS Applications	1-34
MPLS Unicast IP Routing	1-35
MPLS Multicast IP Routing	1-36
MPLS VPNs	1-37
Layer 3 MPLS VPNs	1-38
Layer 2 MPLS VPNs	1-41
MPLS Traffic Engineering	1-44
MPLS QoS	1-46
Interaction between MPLS Applications	1-47
Summary	1-48
<i>Label Distribution Protocol</i>	<i>1-51</i>
Overview	1-51
Objectives	1-51
Label Distribution Protocol (LDP)	1-53
LDP Neighbor Adjacency Establishment	1-54
LDP Link Hello Message	1-55
LDP Session Negotiation	1-57
LDP Discovery of Nonadjacent Neighbors	1-58
LDP Session Protection	1-59

LDP Graceful Restart and NonStop Routing (NSR)	1-60
MPLS Forwarding Structures	1-62
Label Switched Path (LSP)	1-63
Label Allocation and Distribution	1-66
Packet Propagation across an MPLS Domain	1-71
MPLS Steady State Condition	1-73
MPLS Label Control Methods	1-74
Impact of IP Aggregation on LSPs	1-75
Loop Detection using the MPLS TTL field	1-77
Disabling TTL Propagation	1-79
Steady State Condition	1-81
Link Failure MPLS Convergence Process	1-82
Link Recovery MPLS Convergence Process	1-86
IP Switching Mechanisms	1-89
Standard IP Switching Example	1-90
CEF Switching Example	1-91
CEF in IOS XE and IOS XR	1-92
Monitoring IPv4 Cisco Express Forwarding	1-95
Summary	1-96

Implementing MPLS in the Service Provider Core 1-99

Overview	1-99
Objectives	1-99
MPLS Configuraton on Cisco IOS XR vs Cisco IOS/IOS XE	1-101
MPLS Configuration Tasks	1-102
Basic MPLS Configuration	1-103
MTU Requirements	1-104
MPLS MTU Configuration	1-105
IP TTL Propagation	1-106
Disabling IP TTL Propagation	1-108
LDP Session Protection Configuration	1-110
LDP Graceful Restart and NSR Configuration	1-111
LDP IGP Synchronization Configuration	1-113
LDP Autoconfiguration	1-114
Label Advertisement Control Configuration	1-115
Monitor MPLS	1-120
Debugging MPLS and LDP	1-133
Classic Ping and Traceroute	1-134
MPLS Ping and Traceroute	1-136
Troubleshoot MPLS	1-141
Summary	1-147
Module Summary	1-149
Module Self-Check	1-151
Module Self-Check Answer Key	1-155

MPLS Traffic Engineering 2-1

Overview	2-1
Module Objectives	2-1

Introducing MPLS Traffic Engineering Components 2-3

Overview	2-3
Objectives	2-3
Traffic Engineering Concepts	2-4
Traffic Engineering with a Layer 2 Overlay Model	2-9
Layer 3 routing model without Traffic Engineering	2-12
Traffic Engineering with a layer 3 routing model	2-13
Traffic Engineering with the MPLS TE Model	2-14
MPLS TE Traffic Tunnels	2-16
Traffic Tunnels Attributes	2-19
Link Resource Attributes	2-21
Constraint-Based Path Computation	2-22

MPLS TE Process	2-26
Role of RSVP in Path Setup Procedures	2-28
Path Setup and Admission Control with RSVP	2-29
Forwarding Traffic to a Tunnel	2-31
Autoroute	2-32
Summary	2-34
MPLS Traffic Engineering Operations	2-35
Overview	2-35
Objectives	2-35
Attributes used by Constraint-Based Path Computation	2-37
MPLS TE Link Resource Attributes	2-38
MPLS TE Link Resource Attributes: Maximum Bandwidth and Maximum Reservable Bandwidth	2-39
MPLS TE Link Resource Attributes: Link Resource Class	2-40
MPLS TE Link Resource Attributes: Constraint-Based Specific Link Metric (Administrative Weight)	2-41
MPLS TE Tunnel Attributes	2-42
MPLS TE Tunnel Attributes: Traffic Parameter and Path Selection and Management	2-43
MPLS TE Tunnel Attributes: Tunnel Resource Class Affinity	2-44
MPLS TE Tunnel Attributes: Adaptability, Priority, Preemption	2-45
MPLS TE Tunnel Attributes: Resilience	2-46
Implementing TE Policies with Affinity Bits	2-47
Propagating MPLS TE Link Attributes with Link-State Routing Protocol	2-51
Constraint-Based Path Computation	2-55
Path Setup	2-62
RSVP usage in Path Setup	2-64
Tunnel and Link Admission Control	2-69
Path Rerouting	2-71
Assigning Traffic to Traffic Tunnels	2-76
Using Static Routing to Assign Traffic to Traffic Tunnel	2-77
Autoroute	2-78
Autoroute: Default Metric	2-80
Autoroute: Relative and Absolute Metric	2-83
Forwarding Adjacency	2-85
Summary	2-89
Implementing MPLS TE	2-91
Overview	2-91
Objectives	2-91
MPLS TE Configuration Tasks	2-92
MPLS TE Configuration	2-93
RSVP Configuration	2-94
OSPF Configuration	2-95
IS-IS Configuration	2-97
MPLS TE Tunnels Configuration	2-99
Static Route and Autoroute Configurations	2-101
Monitoring MPLS TE Operations	2-102
MPLS TE Case Study: Dynamic MPLS TE Tunnel	2-108
MPLS TE Case Study Continue: Explicit MPLS TE Tunnel	2-115
MPLS TE Case Study Continue: Periodic Tunnel Optimization	2-117
MPLS TE Case Study Continue: Path Selection Restrictions	2-119
MPLS TE Case Study Continue: Modifying the Administrative Weight	2-123
MPLS TE Case Study Continue: Autoroute and Forwarding Adjacency	2-125
Summary	2-128
Protecting MPLS TE Traffic	2-129
Overview	2-129
Objectives	2-129
Improving MPLS TE Convergence Time	2-130
Configuring Backup MPLS TE tunnels	2-131
Drawbacks of Backup MPLS TE tunnels	2-133
Fast Reroute Case Study	2-134

Fast Reroute Case Study Continue: Link Protection	2-136
Fast Reroute Case Study Continue: Node Protection	2-138
Fast Reroute Case Study Continue: Fast Reroute Link Protection Configurations	2-140
MPLS TE Bandwidth Control	2-143
DiffServ-Aware MPLS TE Tunnels	2-148
Summary	2-155
Module Summary	2-157
Module Self-Check	2-159
Module Self-Check Answer Key	2-164

QoS in the Service Provider Network **3-1**

Overview	3-1
Module Objectives	3-1

Understanding QoS **3-3**

Overview	3-3
Objectives	3-3
Cisco IP NGN Architecture	3-4
QoS Issues in Converged Networks	3-6
QoS and Traffic Classes	3-8
Applying QoS Policies on Traffic Classes	3-10
Service Level Agreement	3-12
Service Level Agreement Measuring Points	3-13
Models for Implementing QoS	3-14
IntServ Model and RSVP	3-15
Differentiated Services Model	3-17
DSCP Field	3-18
QoS Actions on Interfaces	3-22
MQC Introduction	3-23
Summary	3-24

Implementing QoS in the SP Network **3-25**

Overview	3-25
Objectives	3-25
QoS Mechanisms	3-27
Classification	3-28
Marking	3-29
Congestion Management	3-30
Congestion Avoidance	3-31
Policing	3-32
Shaping	3-33
Shaping vs. Policing	3-34
Implementing QoS	3-35
MQC	3-38
QoS in Service Provider Environment	3-44
Service Provider Trust Boundary	3-46
PE router QoS Requirements	3-47
P Router QoS Requirements	3-48
Hierarchical QoS Policies	3-49
Summary	3-52

Implementing MPLS Support for QoS **3-53**

Overview	3-53
Objectives	3-53
MPLS QoS	3-54
MPLS EXP	3-55
QoS Group	3-56
Configuring MPLS QoS on a PE Router	3-58
Configuring MPLS QoS on a P Router	3-60
Monitoring MPLS QoS	3-62
QoS-Enabled MPLS VPNs: Point-to-Cloud Service Model	3-63

QoS-Enabled MPLS VPNs: Point-to-Point Service Model	3-64
MPLS DiffServ QoS Models	3-65
MPLS DiffServ Pipe Mode	3-67
MPLS DiffServ Short-Pipe Mode	3-68
MPLS DiffServ Uniform Mode	3-70
MPLS DS-TE	3-71
Summary	3-75
Module Summary	3-77
Module Self-Check	3-79
Module Self-Check Answer Key	3-81

Course Introduction

Overview

Implementing Cisco Service Provider Next-Generation Core Network Services (SPCORE) 1.01 is an instructor-led course presented by Cisco Learning Partners to their end-user customers. This five-day course provides network engineers and technicians with the knowledge and skills necessary to implement and support a service provider network.

The course is designed to provide service provider network professionals with the information that they need to use technologies in a service provider core network. The goal is to provide network professionals with the knowledge, skills, and techniques that are required to plan, implement, and monitor a service provider core network.

The course also features classroom activities, including remote labs, to teach practical skills on deploying Cisco IOS, IOS XE, and IOS XR features to operate and support a service provider network.

Learner Skills and Knowledge

This subtopic lists the skills and knowledge that learners must possess to benefit fully from the course. The subtopic also includes recommended Cisco learning offerings that learners should first complete to benefit fully from this course.

Learner Skills and Knowledge

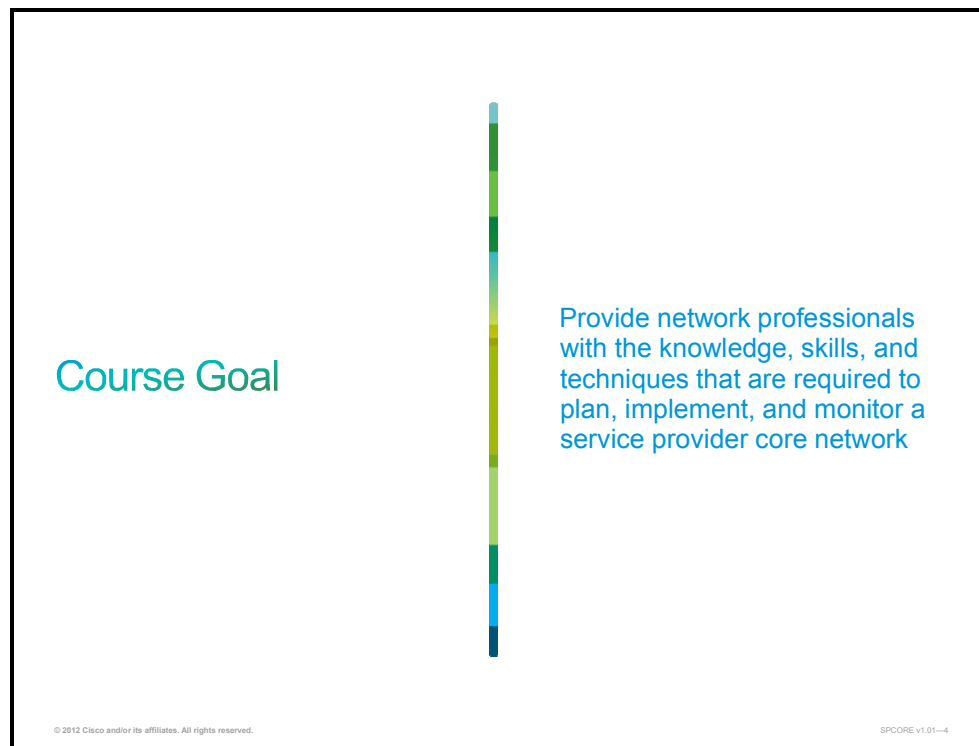
- Students considered for this training will have attended the following courses or obtained equivalent level training:
 - *Building Cisco Service Provider Next-Generation Networks, Part 1* (SPNGN1)
 - *Building Cisco Service Provider Next-Generation Networks, Part 2* (SPNGN2)
 - *Deploying Cisco Service Provider Network Routing* (SPROUTE)
 - *Deploying Cisco Service Provider Advanced Network Routing* (SPADVROUTE)

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—3

Course Goal and Objectives

This topic describes the course goal and objectives.

A graphic with a vertical bar composed of many small colored segments (blue, green, yellow, orange, red) in the center. To the left of the bar is the text 'Course Goal' in a teal color. To the right of the bar is a block of text in blue: 'Provide network professionals with the knowledge, skills, and techniques that are required to plan, implement, and monitor a service provider core network'. At the bottom left of the graphic is the copyright notice '© 2012 Cisco and/or its affiliates. All rights reserved.' and at the bottom right is 'SPCORE v1.01--4'.

Course Goal

Provide network professionals with the knowledge, skills, and techniques that are required to plan, implement, and monitor a service provider core network

© 2012 Cisco and/or its affiliates. All rights reserved. SPCORE v1.01--4

Upon completing this course, you will be able to meet these objectives:

- Describe the features of MPLS, and how MPLS labels are assigned and distributed
- Discuss the requirement for traffic engineering in modern networks that must attain optimal resource utilization
- Describe the concept of QoS and explain the need to implement QoS
- Classify and mark network traffic to implement an administrative policy requiring QoS
- Compare the different Cisco QoS queuing mechanisms that are used to manage network congestion
- Explain the concept of traffic policing and shaping, including token bucket, dual token bucket, and dual-rate policing

Course Flow

This topic presents the suggested flow of the course materials.

Course Flow

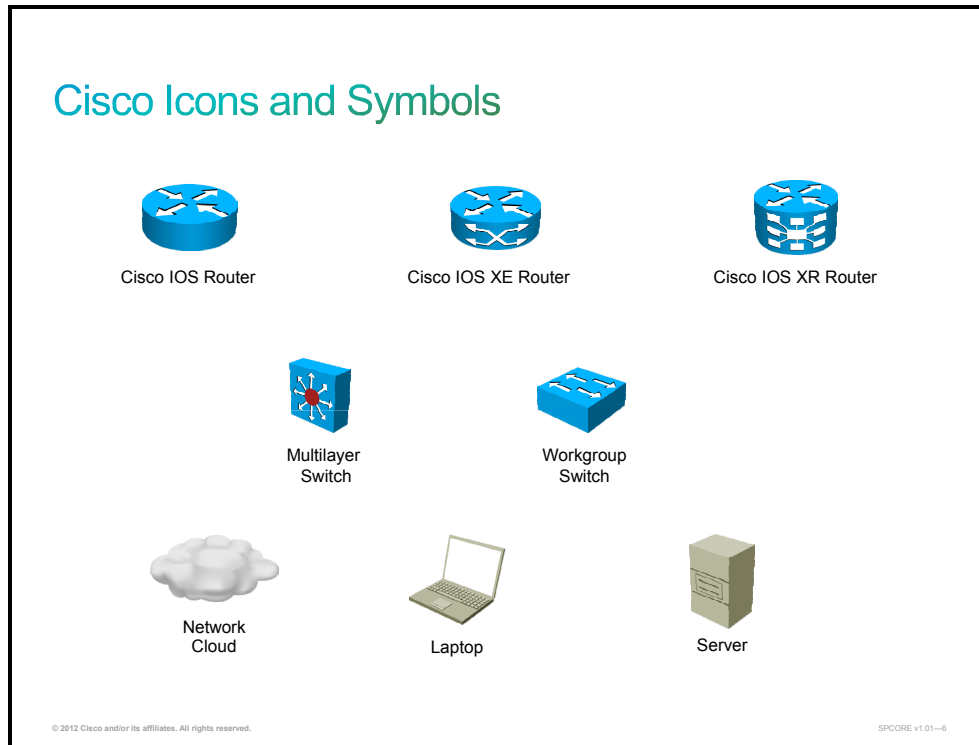
	Day 1	Day 2	Day 3	Day 4	Day 5
A M	Course Introduction	Module 1 (Cont.)	Module 2 (Cont.)	Module 3 (Cont.)	Module 5 (Cont.)
	Module 1: Multiprotocol Label Switching	Module 2: MPLS Traffic Engineering	Module 3: QoS in the Service Provider Network	Module 4: QoS Classification and Marking	Module 6: QoS Traffic Policing and Shaping
	Lunch				
P M	Module 1 (Cont.)	Module 2: (Cont.)	Module 3 (Cont.)	Module 5: QoS Congestion Management and Avoidance	Module 6 (Cont.)

© 2012 Cisco and/or its affiliates. All rights reserved. SPCORE v1.01—5

The schedule reflects the recommended structure for this course. This structure allows enough time for the instructor to present the course information and for you to work through the lab activities. The exact timing of the subject materials and labs depends on the pace of your specific class.

Additional References

This topic presents the Cisco icons and symbols that are used in this course, as well as information on where to find additional technical references.



Cisco Glossary of Terms

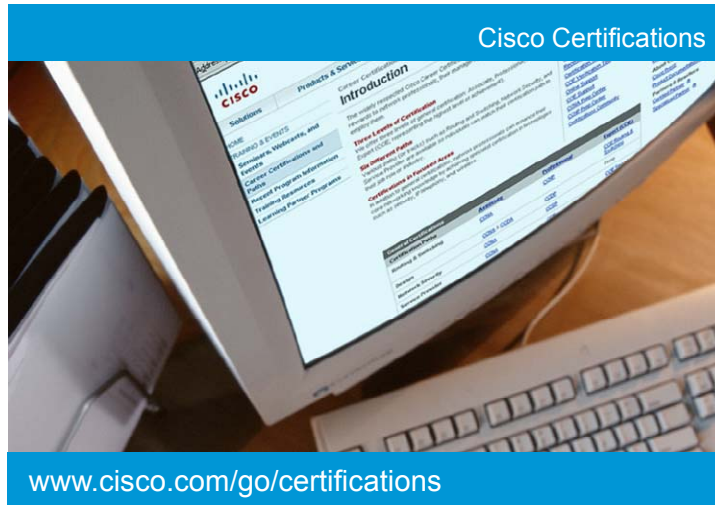
For additional information on Cisco terminology, refer to the *Cisco Internetworking Terms and Acronyms* glossary of terms at

http://docwiki.cisco.com/wiki/Internetworking_Terms_and_Acronyms_%28ITA%29_Guide.

Your Training Curriculum

This topic presents the training curriculum for this course.

Cisco Certifications



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-7

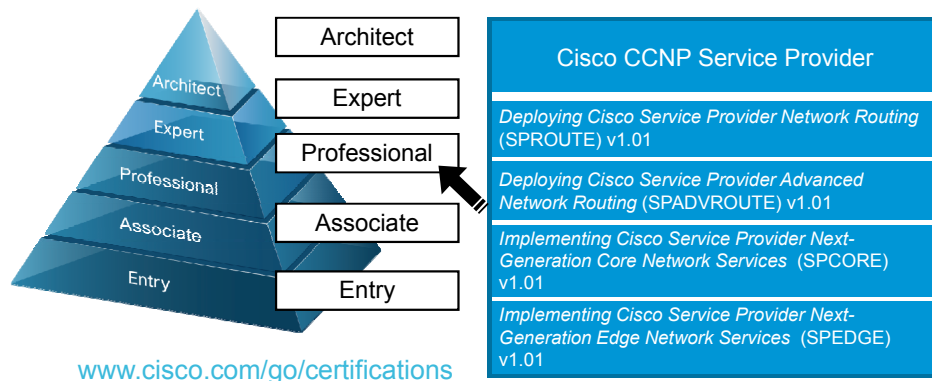
You are encouraged to join the Cisco Certification Community, a discussion forum open to anyone holding a valid Cisco Career Certification (such as Cisco CCIE®, CCNA®, CCDA®, CCNP®, CCDP®, CCIP®, CCVP®, or CCSP®). It provides a gathering place for Cisco certified professionals to share questions, suggestions, and information about Cisco Career Certification programs and other certification-related topics. For more information, visit www.cisco.com/go/certifications.

Your Training Curriculum

This topic presents the training curriculum for this course.

Cisco Career Certifications: Cisco CCNP Service Provider

Expand Your Professional Options and Advance Your Career



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-8

Multiprotocol Label Switching

Overview

This module explains the features of Multiprotocol Label Switching (MPLS) compared with those of traditional hop-by-hop IP routing. MPLS concepts and terminology are explained in this module, along with MPLS label format and label switch router (LSR) architecture and operations. This module describes the assignment and distribution of labels in an MPLS network, including neighbor discovery and session establishment procedures. Label distribution, control, and retention modes will also be explained. This module also covers the functions and benefits of penultimate hop popping (PHP) and provides a review of switching implementations, focusing on Cisco Express Forwarding.

The module also covers the details of implementing MPLS on Cisco IOS, IOS XE, and IOS XR platforms, giving detailed configuration, monitoring, and debugging guidelines for a typical service provider network. In addition, this module includes the advanced topics of controlling Time-to-Live (TTL) propagation and label distribution.

Module Objectives

Upon completing this module, you will be able to explain and configure the features of MPLS, and describe how MPLS labels are assigned and distributed. This ability includes being able to meet these objectives:

- Discuss the basic concepts and architecture of MPLS
- Discuss the label allocation and distribution function and describe the LDP neighbor discovery process via hello messages and by the type of information that is exchanged
- Configure MPLS on Cisco IOS, IOS XE, and Cisco IOS XR platforms

Introducing MPLS

Overview

Multiprotocol Label Switching (MPLS) is a switching mechanism that is often found in service provider environments. MPLS leverages traditional IP routing and supports several services that are required in next-generation IP networks.

This lesson discusses the basic concept and architecture of MPLS. The lesson also describes, at a high level, some of the various types of applications with which you can use MPLS. It is important to have a clear understanding of the role of MPLS and the makeup of the devices and components. This understanding will help you have a clear picture of how to differentiate between the roles of certain devices and to understand how information is transferred across an MPLS domain.

Objectives

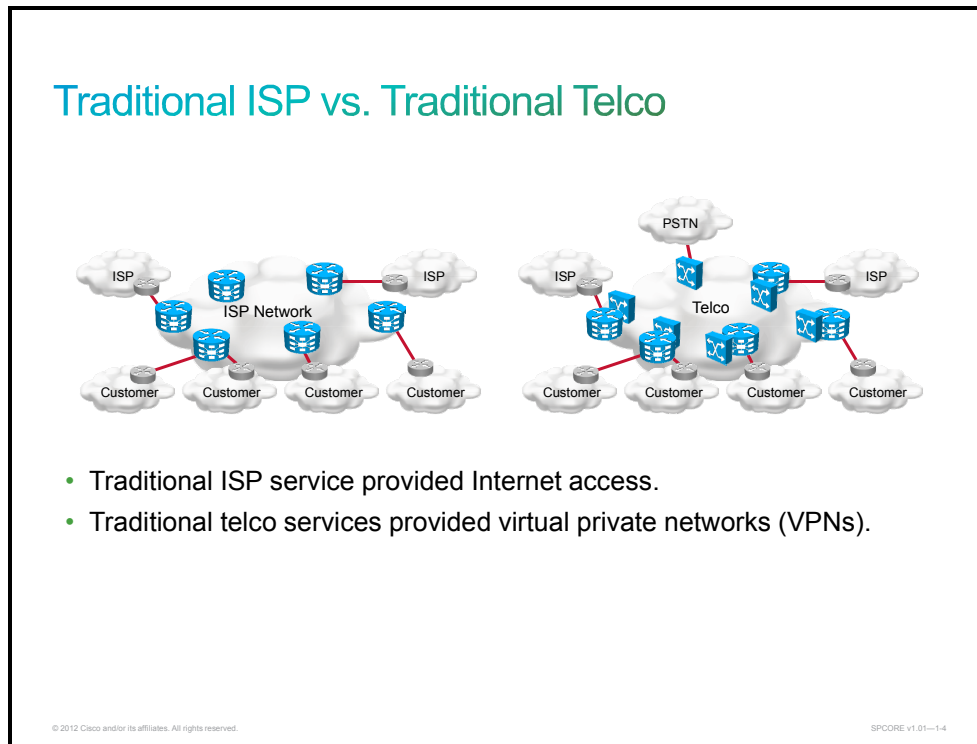
Upon completing this lesson, you will be able to describe the basic MPLS process in a service provider network. You will be able to meet these objectives:

- Describe a traditional Telco and a traditional ISP
- Describe a modern Service Provider
- Show the Cisco IP NGN Architecture
- Describe SONET/SDH
- Describe DWDM and ROADM
- Describe IPoDWDM
- Describe the 10/40/100 Gigabit Ethernet Standards
- Describe traditional Service Providers Transformation to IP
- Describe traditional IP routing where packet forwarding discussions are based on the IP address
- Describe a MPLS at a high level
- Describe MPLS forwarding based on the MPLS label
- Describe the benefits of MPLS
- Describe the LSR, Edge LSR and LSP terminologies
- Describe the MPLS Control Plane

- Describe the MPLS Data Plane
- Describe the FIB and LFIB
- Show an example of the protocols used in the MPLS Control Plane and the LFIB in the Data Plane
- Describe the MPLS label
- Explain how an IP packet is forwarded using MPLS label switching
- Describe the MPLS label stack
- List the MPLS applications in a service provider environment
- Describe MPLS support for Unicast IP routing
- Describe MPLS support for Multicast IP routing
- Describe MPLS support for VPNs
- Describe MPLS support for Layer 3 VPNs
- Describe MPLS support for Layer 2 VPNs
- Describe MPLS support for Traffic Engineering
- Describe MPLS support for QoS
- Describe the interaction Between MPLS Applications

Traditional ISP vs Traditional Telco

This topic describes a traditional Telco and a traditional ISP.



The figure illustrates two types of networks:

- On the right is a traditional telco network that comprises many different devices to offer various services. ATM switches were used to provide VPNs to customers. Time-division multiplexing (TDM) switches were used to provide circuits or telephony to customers. Synchronous Digital Hierarchy (SONET/SDH) was used to carry ATM and TDM across an optical network. Routers were used to provide Internet access to customers.
- On the left is a traditional ISP whose initial focus was only to provide Internet access to their customers. No other services were offered because there was limited capability to offer anything comparable to what telcos could offer through their extensive range of equipment and technologies.

There were advantages with traditional ISPs:

- Employed one technology—IP (cost-effective)
- Had competitive pricing

There were limitations to traditional ISPs:

- Provided only Internet access
- Had no VPNs

There were advantages with traditional telcos:

- Employed multiple technologies (IP, ATM, FR, SONET/SDH)
- Used good QoS mechanisms
- Offered Internet access, VPNs, and telephony

There were limitations with traditional telcos:

- Employed multiple technologies (costly)
- Provided costly services

Modern Service Provider

This topic describes a modern Service Provider.

Modern Service Provider

- Uses IP-based infrastructure
- Maintains only one technology aided by **Multi-Protocol Label Switching (MPLS)** to support additional services:
 - Internet access
 - Virtual private networks
 - Telephony
 - Quality of service
- Retains competitive pricing
- Includes additional developments:
 - WDM in the core (more core bandwidth)
 - DSL, Cable, Ethernet on the edge (more edge bandwidth)
 - Various wireless technologies

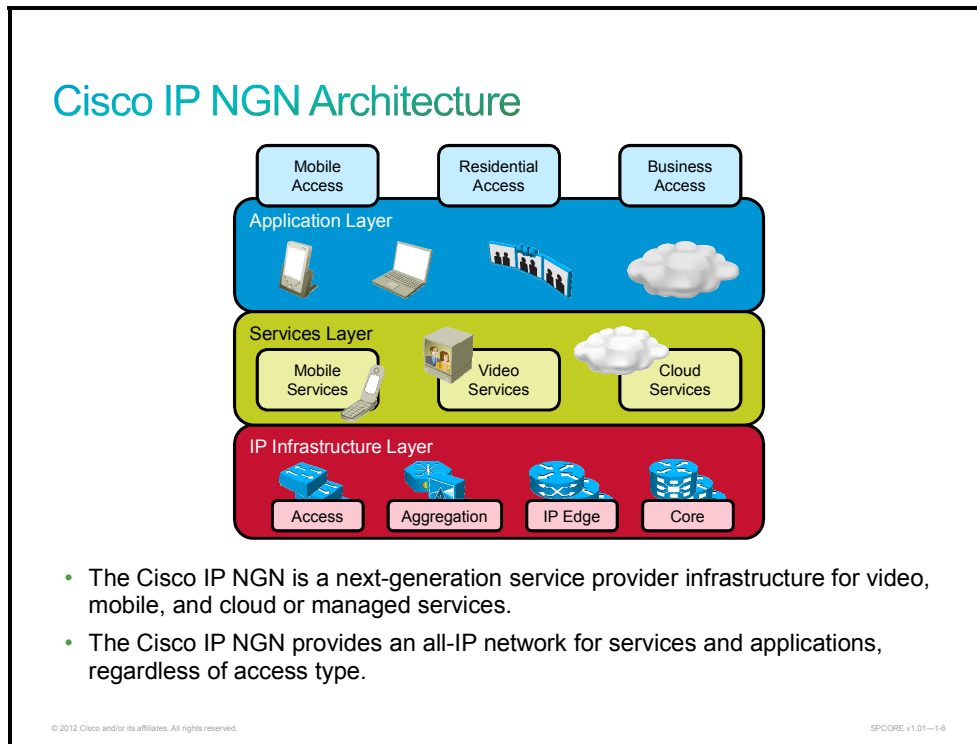
© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—1-6

A modern service provider network, whether it evolved from a traditional telco or from a greenfield ISP, can accommodate the same customer requirements as traditional telcos did, without having to use different types of networks and devices. Routers are used to provide Internet access, VPNs, telephony services, and TV services. Dense wavelength-division multiplexing (DWDM) is an exception that is often used in addition to routers to increase the amount of throughput that is available via a single strand of optical fiber.

Cisco IP NGN Architecture

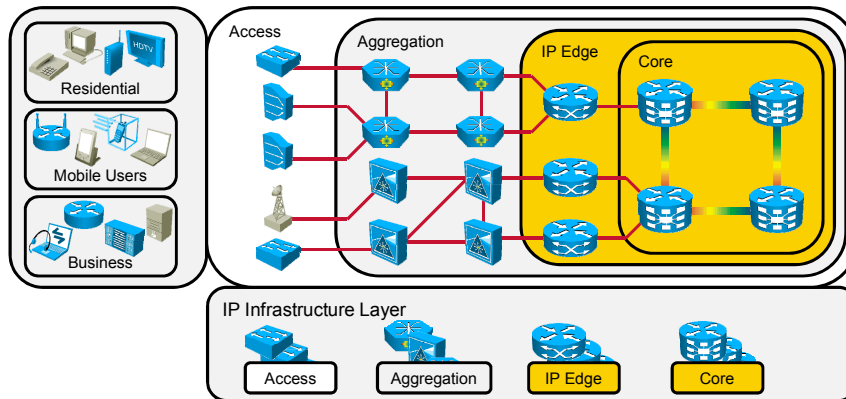
This topic shows the Cisco IP NGN Architecture.



In earlier days, service providers were specialized for different types of services, such as telephony, data transport, and Internet service. The popularity of the Internet, through telecommunications convergence, has evolved into the usage of the Internet for all types of services. Development of interactive mobile applications, increasing video and broadcasting traffic, and the adoption of IPv6 have pushed service providers to adopt new architecture to support new services on the reliable IP infrastructure with a good level of performance and quality.

Cisco IP Next-Generation Network (NGN) is the next-generation service provider architecture for providing voice, video, mobile, and cloud or managed services to users. The general idea of Cisco IP NGN is to provide all-IP transport for all services and applications, regardless of access type. IP infrastructure, service, and application layers are separated in NGN networks, thus enabling addition of new services and applications without any changes in the transport network.

Cisco IP NGN Infrastructure Layer



- The IP infrastructure layer provides connectivity between customer and service provider.
- MPLS is used in the core and edge network.
- P routers are in the core, and PE routers are in the edge segment.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1.7

The IP infrastructure layer is responsible for providing reliable infrastructure for running upper layer services. It includes these things:

- Core network
- IP edge network
- Aggregation network
- Access network

It provides the reliable, high-speed, and scalable foundation of the network. End users are connected to a service provider through a customer premises equipment (CPE) device, using any possible technology. Access and aggregation network devices are responsible for enabling connectivity between customer equipment and service provider edge equipment. A core network is used for fast switching packets between edge devices.

MPLS is a technology that is primarily used in the service provider core and the IP edge portion of the IP infrastructure layer. In service provider networks, the result of using MPLS is that only the routers on the edge of the MPLS domain perform routing lookup; all other routers forward packets based on labels. What really makes MPLS useful in service provider (and large enterprise) networks is that it enhances Border Gateway Protocol (BGP) routing and provides different services and applications, such as Layer 2 and 3 VPNs, QoS, and traffic engineering (TE).

These are service provider transport technologies in the core portion of the Cisco IP NGN model:

- SONET/SDH
- DWDM and ROADM
- IP over DWDM
- 10/40/100 Gigabit Ethernet

SONET/SDH

This topic describes SONET/SDH.

SONET/SDH

- SONET (Synchronous Transport Signal)—STS-<n>
- SDH (Synchronous Transport Module)—STM-<n>

Bit Rate (Mb/s)	SONET			SDH			Abbrev.
	Signal	Channels		Signal	Channels		Speed
		DS1	DS3		E1	E4	(Gb/s)
51.84	STS-1	28	1	STM-0	21	0	
155.52	STS-3	84	3	STM-1	63	1	
622.08	STS-12	336	12	STM-4	252	4	
2488.32	STS-48	1344	48	STM-16	1008	16	2.5
9953.28	STS-192	5376	192	STM-64	4032	64	10.0
39813.10	STS-768	21504	768	STM-256	16128	256	40.0

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—1-8

SONET/SDH was initially designed to carry 64K Pulse Code Modulation (PCM) voice channels that are commonly used in telecommunication. The basic underlying technology that is used in the SONET/SDH system is synchronous Time Division Multiplexing (TDM).

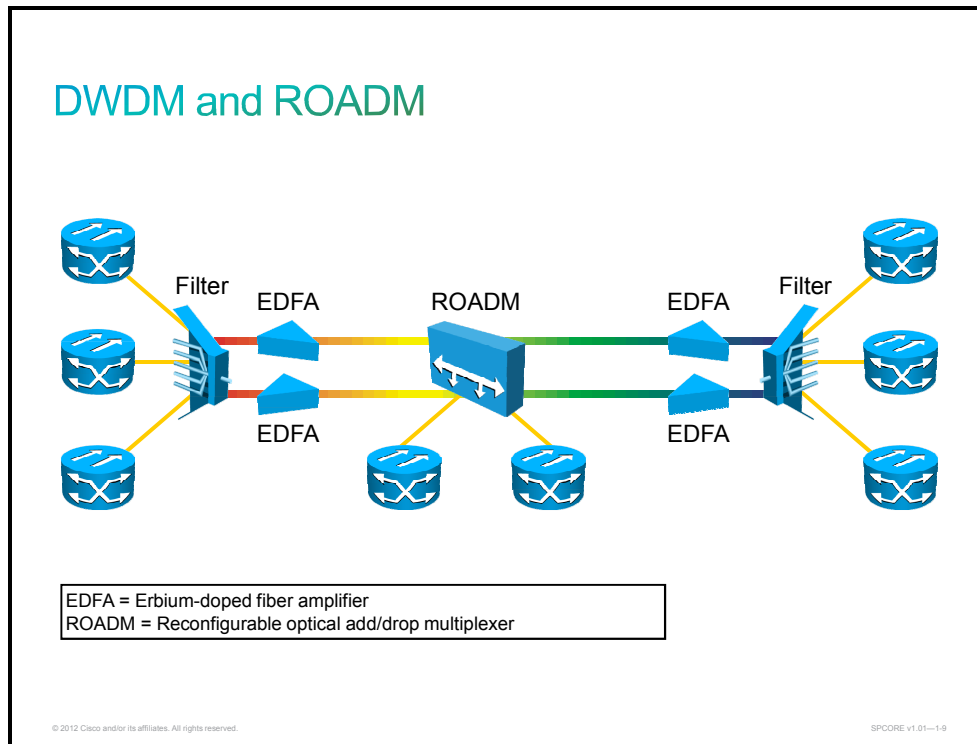
The major difference between SONET and SDH is the terminology that is used to describe them. For example, a SONET OC-3 signal is called an SDH STM-1 signal by the ITU-T.

The SONET/SDH standard specifies standards for communication over fiber optics as well as electrical carriers for lower-speed signaling rates (up to 155 Mb/s). The standard describes the frame format that should be used to carry the different types of payload signals as well as the control signaling that is needed to keep a SONET/SDH connection operational.

The SONET standard is mainly used in the United States, while the SDH standard is mainly European. In the United States, the SDH standard is used for international connections.

DWDM and ROADM

This topic describes DWDM and ROADM.



Wavelength division multiplexing (WDM) is a technology that multiplexes a number of optical carrier signals into a single optical fiber by using different wavelengths of laser light. Dense WDM, or DWDM, refers to optical signals that are multiplexed within the 1550 nm band.

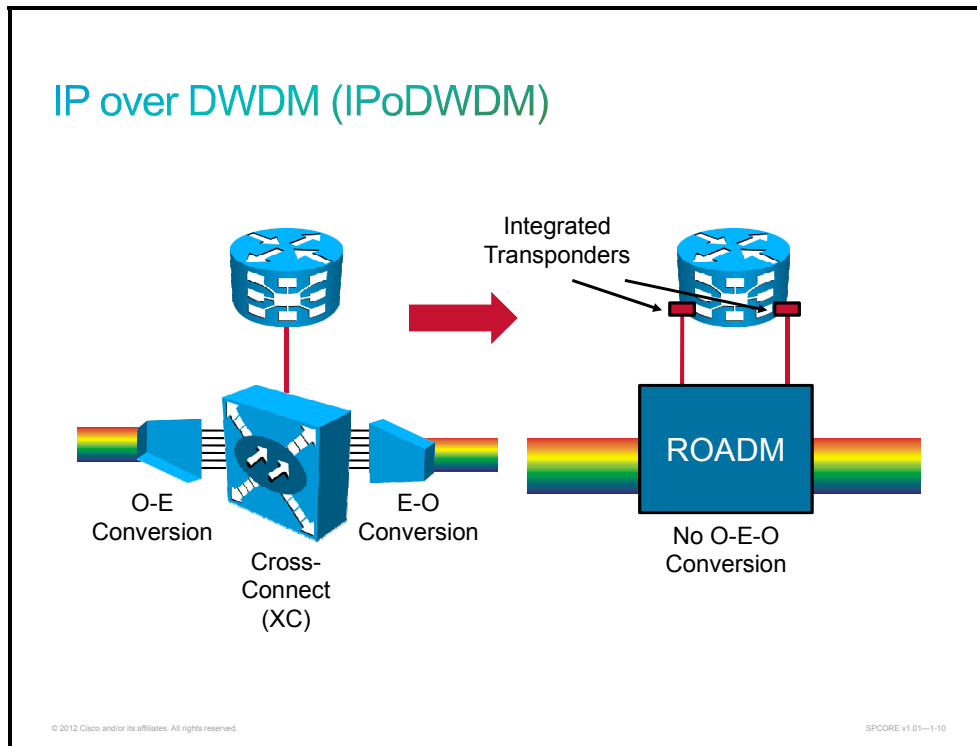
Intermediate optical amplification sites in DWDM systems may allow for the dropping and adding of certain wavelength channels. In earlier systems, adding or dropping wavelengths required manually inserting or replacing wavelength-selective cards. This was costly and in some systems required that all active traffic be removed from the DWDM system, because inserting or removing the wavelength-specific cards interrupts the multi-wavelength optical signal.

A reconfigurable optical add-drop multiplexer (ROADM) is a form of optical add-drop multiplexer (OADM) that adds the ability to switch traffic remotely, from a WDM system at the wavelength layer. This capability allows individual or multiple wavelengths that are carrying data channels to be added or dropped from a transport fiber without the need to convert the signals on all of the WDM channels to electronic signals, and back again to optical signals.

With a ROADM, network operators can reconfigure the multiplexer remotely by sending soft commands. The architecture of the ROADM is such that dropping or adding wavelengths does not interrupt the “pass-through” channels. Numerous technological approaches are utilized for various commercial ROADMs, the trade-off being between cost, optical power, and flexibility.

IP over DWDM (IPoDWDM)

This topic describes IPoDWDM.



Service providers continue to look for the best economics for increasing network capacity to accommodate the continued growth in IP traffic that is driven by data, voice, and primarily video traffic. The reasons for integrating IP and DWDM are simply to deliver a significant reduction in capital expenditures and improve the operational efficiency of the network.

IP over DWDM (IPoDWDM) is a technology that integrates DWDM on routers. Routers must support the ITU-T G.709 standard so that they can monitor the optical path. Element integration refers to the capability to take multiple, separate elements that operate in the network and collapse them into a single device without losing any of the desired functions for continued operation.

The integration of core routers with the optical transport platform eliminates the need for optical-electrical-optical (OEO) modules (transponders) in the transport platform.

10/40/100 Gigabit Ethernet Standards

This topic describes the 10/40/100 Gigabit Ethernet Standards.

10/40/100 Gigabit Ethernet Standards

- Support full-duplex operation only
- Preserve the 802.3 / Ethernet frame format, utilizing the 802.3 MAC
- Preserve minimum and maximum frame size of current 802.3 standard
- Support a BER better than or equal to 10⁻¹² at the MAC/PLS service interface
- Provide appropriate support for OTN
- Support a MAC data rate of 40/100 Gb/s
- Provide physical layer specifications that support 40/100 Gb/s operation over various pluggable modules

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-11

IEEE 802.3ba is an IEEE standard of the 802.3 family of data link layer standards for Ethernet LAN and WAN applications, whose objective is to support speeds faster than 10 Gb/s. The standard supports 40 Gb/s and 100 Gb/s transfer rates. The decision to include both speeds comes from the demand to support the 40 Gb/s rate for local server applications and the 100 Gb/s rate for internet backbones.

The 40/100 Gigabit Ethernet standards include a number of different Ethernet physical layer (PHY) specifications, so a networking device may support different pluggable modules.

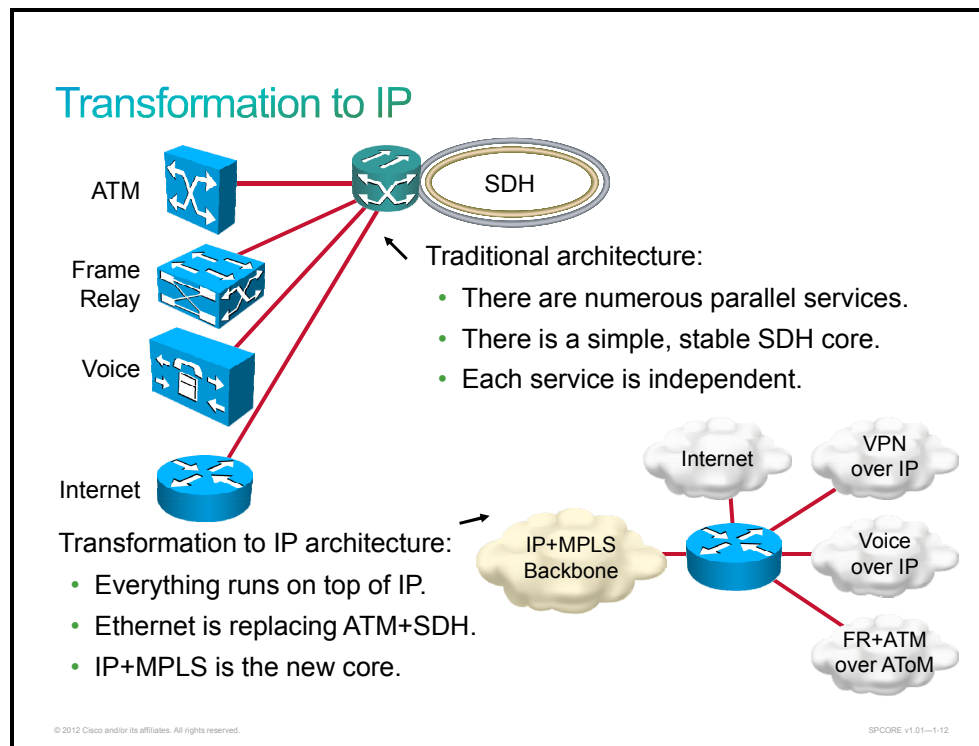
The main objectives are these:

- Support full-duplex operation only
- Preserve the 802.3 / Ethernet frame format utilizing the 802.3 MAC
- Preserve the minimum and maximum frame size of the current 802.3 standard
- Support a bit error rate (BER) better than or equal to 10⁻¹² at the MAC-physical layer signaling sublayer (PLS) service interface
- Provide appropriate support for optical transport network (OTN)
- Support a MAC data rate of 40 Gb/s
- Provide physical layer specifications that support 40 Gb/s operation over:
 - At least 10 km on single-mode fibre (SMF)
 - At least 100 m on OM3 multimode fibre (MMF)
 - At least 7 m over a copper cable assembly
 - At least 1 m over a backplane

- Support a MAC data rate of 100 Gb/s
- Provide physical layer specifications that support 100 Gb/s operation over:
 - At least 40 km on SMF
 - At least 10 km on SMF
 - At least 100 m on OM3 MMF
 - At least 7 m over a copper cable assembly

Transformation to IP

This topic describes traditional Service Providers Transformation to IP.



Traditional service provider architecture was based on numerous parallel services with a simple, stable SDH core, where each service was independent.

Modern service providers usually transform to IP protocol architecture, where everything runs on top of IP protocol. In this scenario, Ethernet replaces ATM or SDH; IP, in combination with MPLS, is used in the core network.

Among the factors that drive service providers to transition to IP could be changed usage patterns:

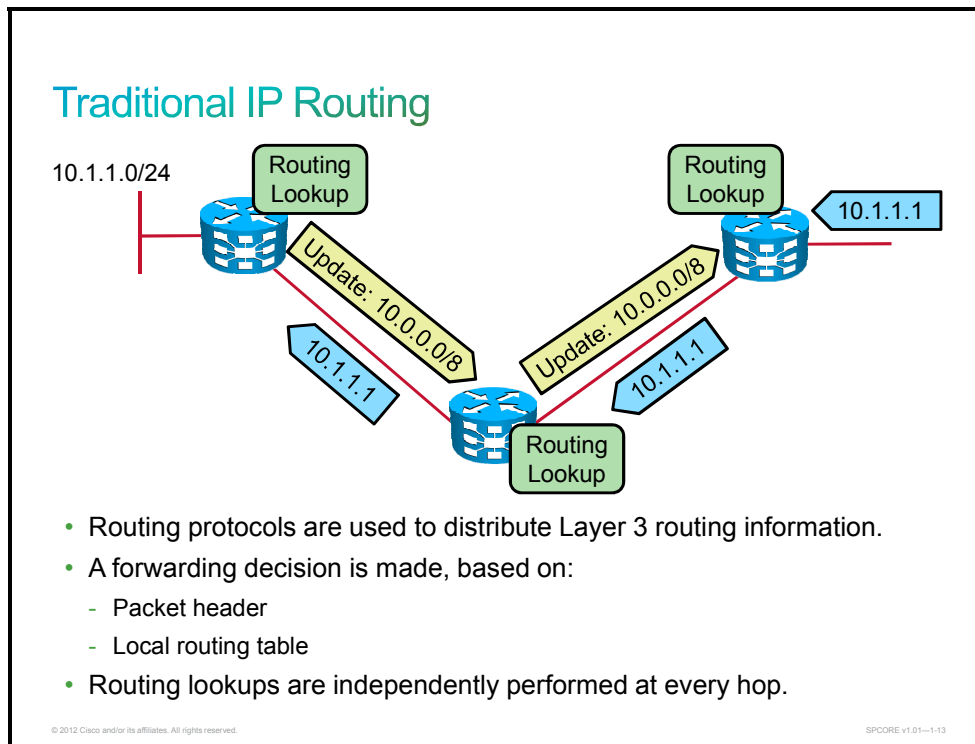
- Traditional traffic is slowly growing.
- IP traffic is exploding.
- Everything-over-IP makes business sense.
- Customers accept just-good-enough solutions.

Among the factors that drive service providers to transition to IP could be technology changes:

- Packet switching is cheaper than circuit switching.
- MPLS provides reasonable virtual circuit capabilities.
- MPLS recovery times are comparable to SDH.
- Ethernet is cheaper than ATM or SDH.
- Rigid QoS is not needed on very high-speed links.

Traditional IP Routing

This topic describes traditional IP routing where packet forwarding discussions are based on the IP address.



Before basic MPLS functionality is explained, these three foundations of traditional IP routing need to be highlighted:

- Routing protocols are used on all devices to distribute routing information.
- Each router analyzes the Layer 3 header of each packet, compared to the local routing table, and makes a decision about where to forward the packet. Regardless of the routing protocol, routers forward packets contingent on a destination address-based routing lookup.

Note The exception to this rule is policy-based routing (PBR), where routers will bypass the destination-based routing lookup.

- The routing lookup is performed independently on every router in the network.

MPLS Introduction

This topic describes a MPLS at a high level.

MPLS

- Provides an intermediate encapsulation between an OSI Layer 3 IP header and an arbitrary OSI Layer 2 header
- Can also encapsulate a non-IP payload
- Bases forwarding on a label, regardless of the payload
- Result:
 - Different protocols can be used to determine the path.
 - Different payloads can be used to provide different services.
 - Any traditional telco service or a functional equivalent can be implemented in an MPLS-enabled environment.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-14

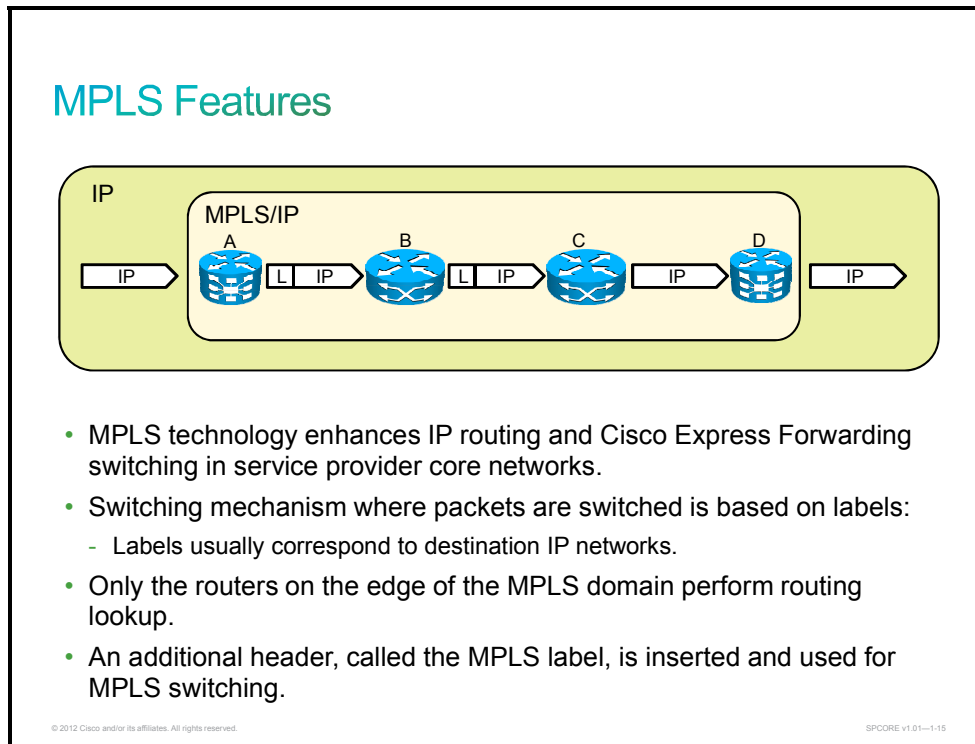
MPLS provides an intermediate encapsulation between an Open Systems Interconnection (OSI) Layer 3 IP header and an arbitrary OSI Layer 2 header. This function enables the forwarding of packets through label-switched paths (LSPs) that can be created using various methods and protocols, depending on the required results. Additionally, the payload can be any Layer 3 or even Layer 2 protocol.

MPLS enables service providers to provide services that are the same or functionally equivalent to service that was available with traditional telcos:

- Internet access can be provided.
- ATM or Frame Relay VPNs can be replaced by Layer 3 MPLS VPNs, or if they are required, they can be retained using Layer 2 MPLS VPNs.
- SONET/SDH can be implemented using DWDM, or the same type of quality of service (QoS) characteristics can be implemented using Layer 2 MPLS VPNs in combination with QoS implementation.

MPLS Features

This topic describes MPLS forwarding based on the MPLS label.



MPLS is a technology that is primarily used in service provider core networks. MPLS improves classic IP routing using Cisco Express Forwarding by introducing an additional header into a packet. This additional header is called the MPLS label. MPLS switches packets that are based on labels lookup instead of IP address lookup. Labels usually correspond to destination IP networks; each destination has a corresponding label on each MPLS-enabled router.

In service provider networks, the result of using MPLS is that only the routers on the edge of MPLS domain perform routing lookup; all other routers forward packets that are based on labels.

MPLS is a packet-forwarding technology that uses appended labels to make forwarding decisions for packets.

- Within the MPLS network, the Layer 3 header analysis is done just once (when the packet enters the MPLS domain). Labels are appended to the packet, and then the packet is forwarded into the MPLS domain.
- Simple label inspection that is integrated with Cisco Express Forwarding switching drives subsequent packet forwarding.

Note The Cisco Express Forwarding switching mechanism will be covered later in this course.

MPLS Benefits

This topic describes the benefits of MPLS.

MPLS Benefits

- Decreases forwarding overhead on core routers
- Can support forwarding of non-IP protocols
- Enhances BGP routing
- Supports multiple applications:
 - Unicast and multicast IP routing
 - VPN
 - Traffic engineering (TE)
 - QoS
 - AToM

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-18

In modern routers, MPLS label switching is not much faster than IP routing, but MPLS is not used just because of its switching performance.

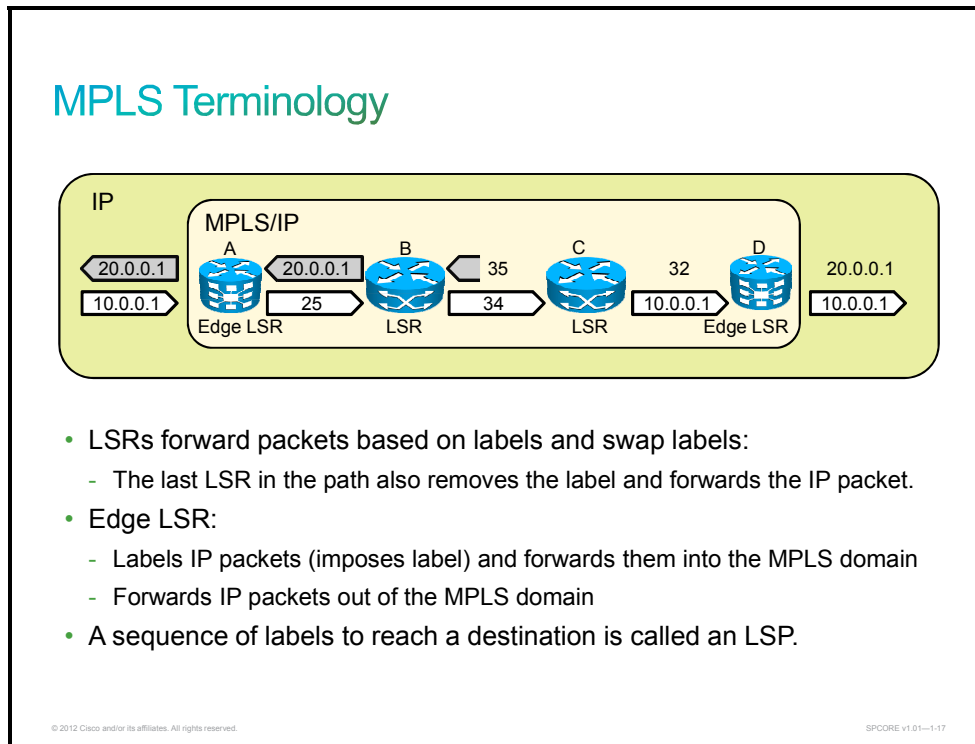
There are several other benefits to MPLS:

- MPLS decreases the forwarding overhead on the core routers.
- MPLS supports multiple useful applications:
 - Unicast and multicast IP routing
 - VPN
 - TE
 - QoS
 - Any Transport over MPLS (AToM)
- MPLS supports the forwarding of non-IP protocols, because MPLS technologies are applicable to any network layer protocol.

MPLS is very useful in service provider (and large enterprise) networks because it enhances BGP routing and provides different services and applications, such as Layer 2 and 3 VPNs, QoS, and TE.

MPLS Terminology

This topic describes the LSR, Edge LSR and LSP terminologies.



In an MPLS domain, there are two types of routers:

- **Label-switched router (LSR):** A device that forwards packets that are primarily based on labels
- **Edge LSR:** A device that primarily labels packets or forwards IP packets out of an MPLS domain

LSRs and edge LSRs are usually capable of doing both label switching and IP routing. Their names are based on their positions in an MPLS domain. Routers that have all interfaces enabled for MPLS are called LSRs because they mostly forward labeled packets (except for the penultimate LSR). Routers that have some interfaces that are not enabled for MPLS are usually at the edge of an MPLS domain. Ingress edge LSR forwards packets that are based on IP destination addresses, and labels them if the outgoing interface is enabled for MPLS. Egress LSR forwards IP packets that are based on routing lookup outside the MPLS domain.

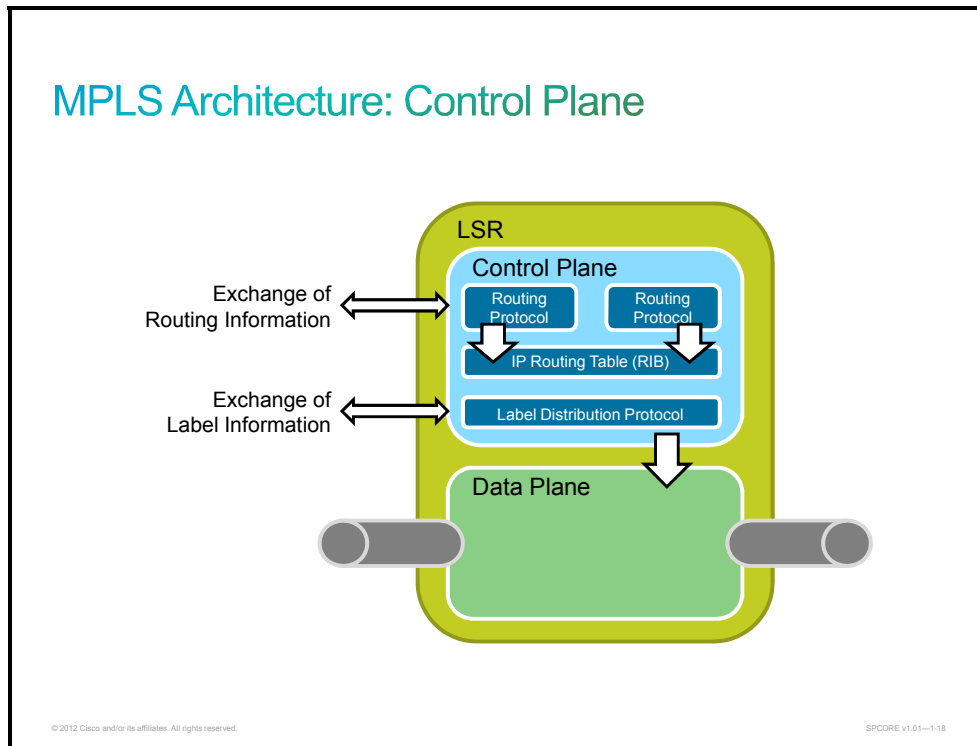
A sequence of labels that is used to reach a destination is called a label-switched path (LSP). LSPs are unidirectional; that means that the return traffic uses a different LSP. A penultimate LSR router in an LSP path removes a label and forwards the IP packet to the egress edge LSR router, which routes the IP packet, based on routing lookup. The removing of a label on the penultimate LSR is called penultimate hop popping (PHP).

For example, an edge LSR receives a packet for destination 10.0.0.1, imposes label 25, and forwards the frame to the LSR in the MPLS backbone. The first LSR swaps label 25 for label 34, and forwards the frame. The second (penultimate) LSR removes the label and forwards the IP packet to the edge LSR. The edge LSR forwards the packet, based on IP destination address 10.0.0.1.

Note	PHP is implemented because of the increased performance on the egress edge LSR. Without PHP, the edge LSR would receive a labeled packet, where two lookups would be needed. The first one would be based on labels, and the result would be to remove the label. The second lookup would be to route the IP packet based on destination IP address and routing table.

MPLS Architecture: Control Plane

This topic describes the MPLS Control Plane.



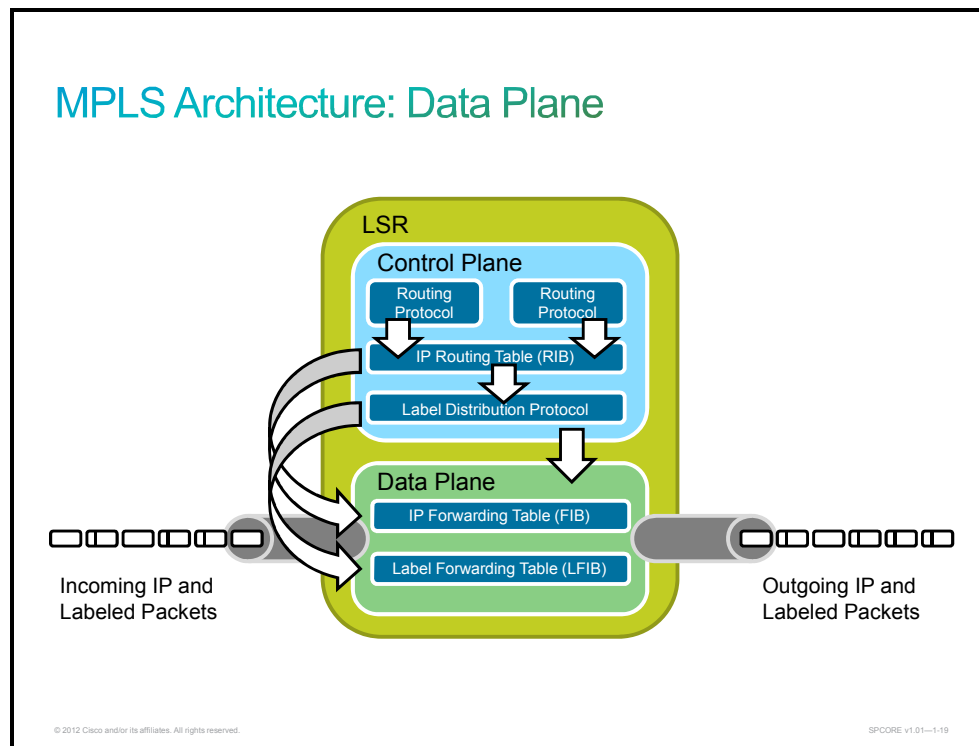
The control plane builds a routing table (routing information base [RIB]) that is based on the routing protocol. Various routing protocols, such as Open Shortest Path First (OSPF), Interior Gateway Routing Protocol (IGRP), Enhanced Interior Gateway Routing Protocol (EIGRP), Intermediate System-to-Intermediate System (IS-IS), Routing Information Protocol (RIP), and BGP can be used in the control plane for managing Layer 3 routing.

The control plane uses a label exchange protocol to create and maintain labels internally, and to exchange these labels with other MPLS-enabled devices. The label exchange protocol binds labels to networks that are learned via a routing protocol. Label exchange protocols include MPLS Label Distribution Protocol (LDP), the older Cisco Tag Distribution Protocol (TDP), and BGP (used by MPLS VPN). Resource Reservation Protocol (RSVP) is used by MPLS TE to accomplish label exchange.

The control plane also builds two forwarding tables, a forwarding information base (FIB) from the information in the RIB, and a label forwarding information base (LFIB) table, based on the label exchange protocol and the RIB. The LFIB table includes label values and associations with the outgoing interface for every network prefix.

MPLS Architecture: Data Plane

This topic describes the MPLS Data Plane.

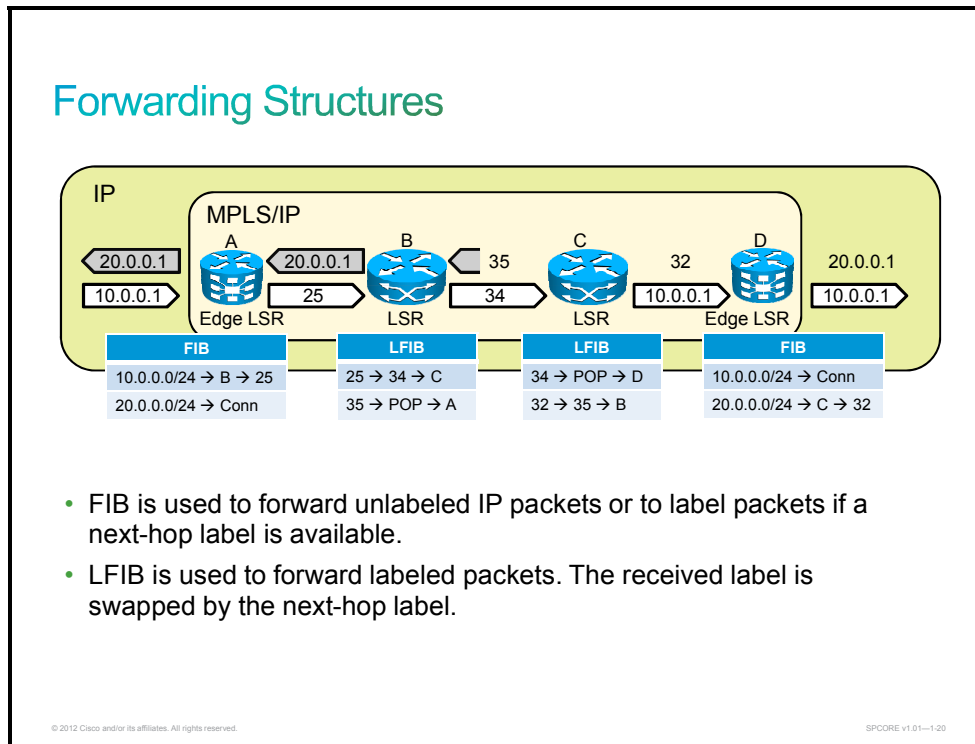


The data plane takes care of forwarding, based on either destination addresses or labels; the data plane is also known as the forwarding plane.

The data plane is a simple forwarding engine that is independent of the type of routing protocol or label exchange protocol being used. The data plane forwards packets to the appropriate interface, based on the information in the LFIB or the FIB tables.

Forwarding Structures

This topic describes the FIB and LFIB.



The data plane on a router is responsible for forwarding packets, based on decisions done by routing protocols (which run in the router control plane). The data plane on an MPLS-enabled router consists of two forwarding structures:

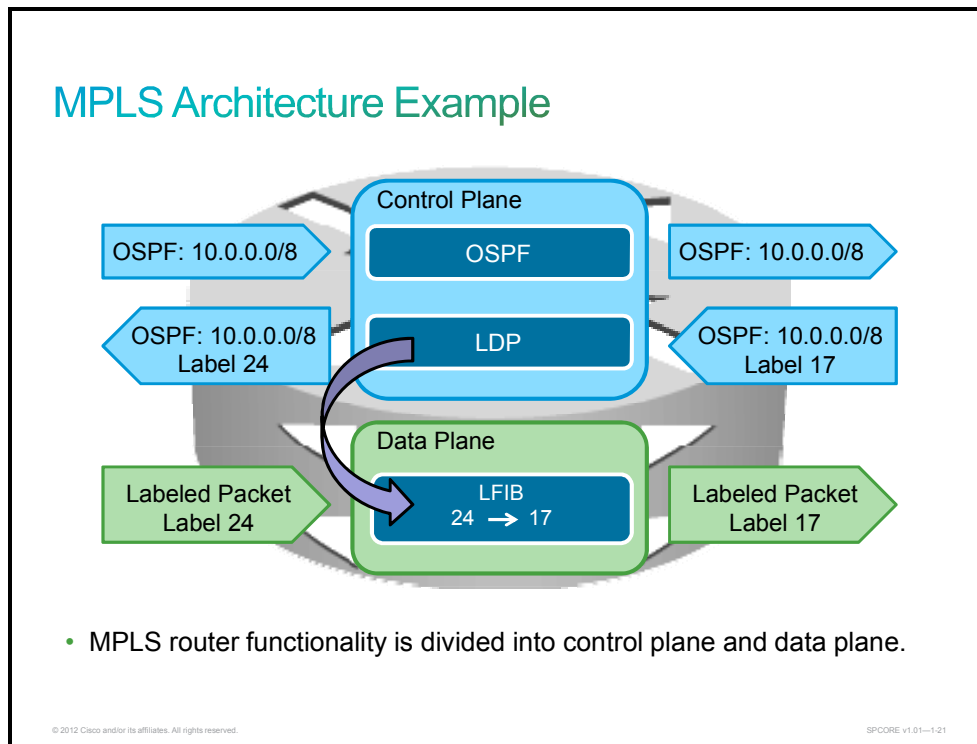
- **Forwarding information base (FIB):** When a router is enabled for Cisco Express Forwarding, the FIB is used to forward IP packets, based on decisions made by routing protocols. The FIB is populated from a routing table and includes destination networks, next hops, outgoing interfaces, and pointers to Layer 2 addresses. The FIB on an MPLS-enabled router also contains an outgoing label, if an outgoing interface is enabled for MPLS. FIB lookup is done when an IP packet is received. Based on the result, the router can send out an IP packet or a label can be imposed.
- **Label forwarding information base (LFIB):** The LFIB is used when a labeled packet is received. The LFIB contains the incoming and outgoing label, outgoing interface, and next-hop router information. When an LFIB lookup is done, the result can be to swap a label and send a labeled packet or to remove a label and send an IP packet.

These combinations of forwarding packets are possible:

- A received IP packet (FIB) is forwarded, based on the IP destination address, and is sent as an IP packet.
- A received IP packet (FIB) is forwarded, based on the IP destination address, and is sent as a labeled packet.
- A received labeled packet (LFIB) is forwarded, based on the label; the label is changed (swapped) and the labeled packet is sent.
- A received labeled packet (LFIB) is forwarded, based on the label; the label is removed and the IP packet is sent.

MPLS Architecture Example

This topic shows an example of the protocols used in the MPLS Control Plane and the LFIB in the Data Plane.



In the example LSR architecture, the control plane uses these protocols:

- A routing protocol (OSPF), which receives and forwards information about IP network 10.0.0.0/8
- A label exchange protocol (LDP), which receives label 24 to be used for packets with destination address 10.0.0.0/8

(A local label 17 is generated and is sent to upstream neighbors so that these neighbors can label packets with the appropriate label.)

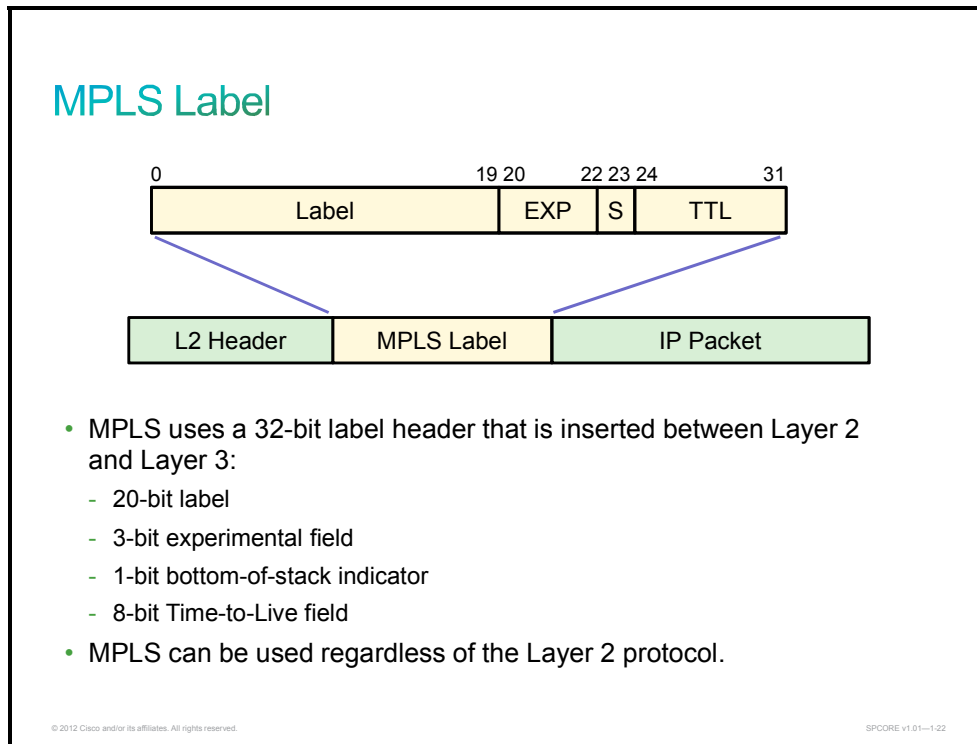
The data plane uses an LFIB to forward packets based on labels:

- The LFIB receives an entry from LDP, where label 24 is mapped to label 17. When the data plane receives a packet labeled with a 24, it replaces label 24 with label 17 and forwards the packet through the appropriate interfaces.

Note In the example, both packet flow and routing and label updates are from left to right.

MPLS Labels

This topic describes MPLS label.



The figure presents an MPLS label that is used for MPLS switching. This label is inserted between the Layer 2 and Layer 3 header and can be used regardless of which Layer 2 protocol is used.

The label is 32 bits long and consists of the following fields:

Field	Description
20-bit label	The actual label used for switching. Values 0 to 15 are reserved.
3-bit experimental (EXP) field	Undefined in the RFC. Used by Cisco to define a class of service (CoS) (IP precedence).
Bottom-of-stack bit	MPLS allows multiple labels to be inserted. The bottom-of-stack bit determines if this label is the last label in the packet. If this bit is set (1), it indicates that this is the last label.
8-bit Time to Live (TTL) field	Has the same purpose as the TTL field in the IP header.

The edge LSR then does these tasks:

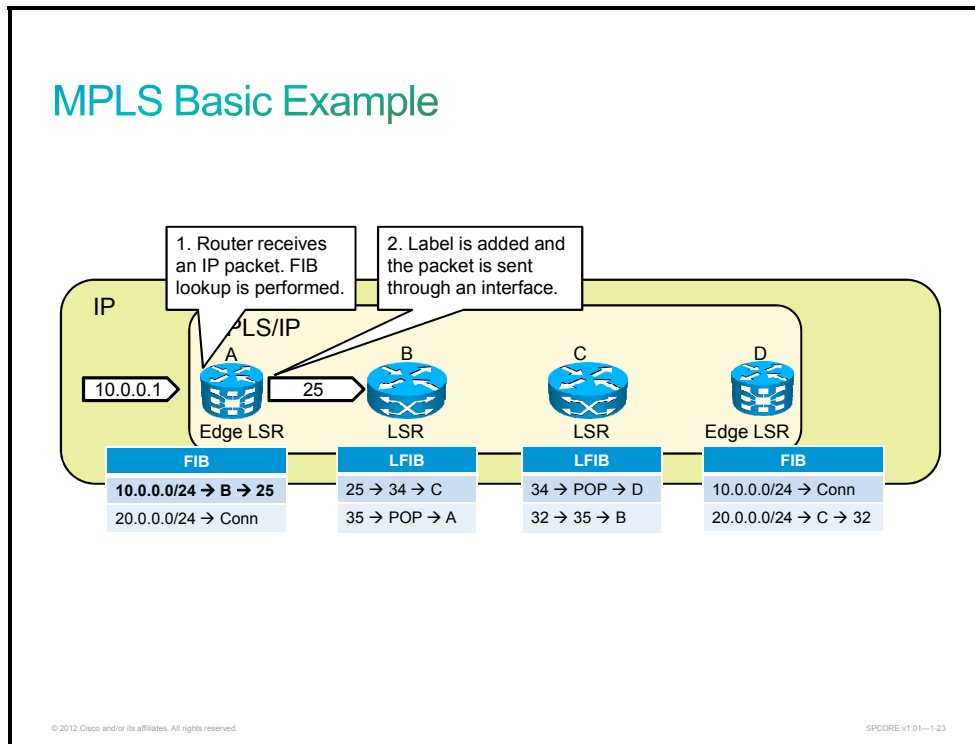
- The router performs routing lookup to determine the outgoing interface.
- The router inserts a label between the Layer 2 frame header and the Layer 3 packet header, if the outgoing interface is enabled for MPLS and if a next-hop label for the destination exists. This inserted label is also called the shim header.
- The router then changes the Layer 2 protocol identifier (PID) or EtherType value in the Layer 2 frame header to indicate that this is a labeled packet. For example, EtherType 0x8847 means a MPLS unicast packet.
- The router sends the labeled packet.

Note	Other routers in the MPLS core simply forward packets based on the received label.

MPLS is designed for use on virtually any media and Layer 2 encapsulation.

MPLS Packet Flow Basic Example

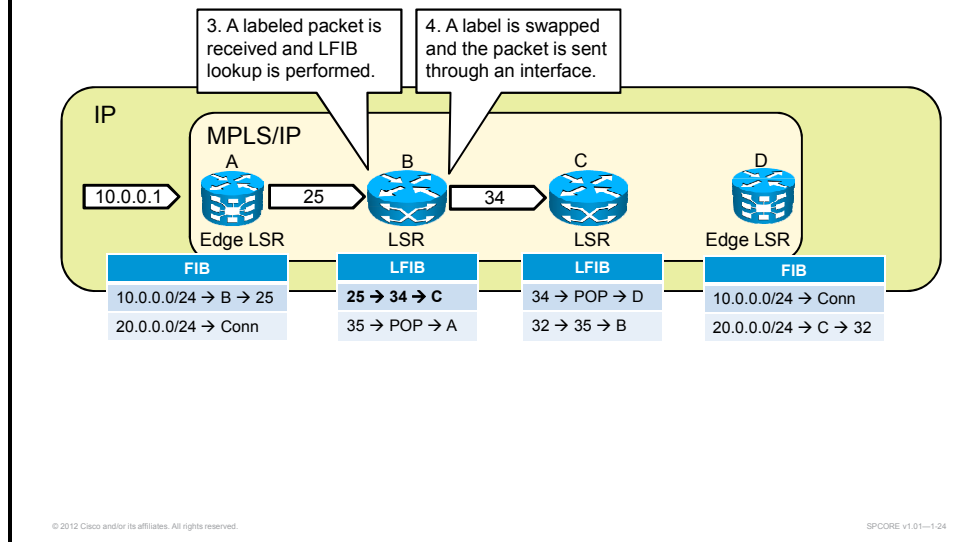
This topic explains how an IP packet is forwarded using MPLS label switching.



The figure shows an example of the way in which a packet traverses an MPLS-enabled network. Router A receives an IP packet destined for 10.0.0.1.

- Step 1** Router A performs a FIB lookup. The FIB for that destination states that the packet should be labeled using label 25 and sent to router B.
- Step 2** Router A adds a label 25 and the packet is sent out the interface that connects to router B.

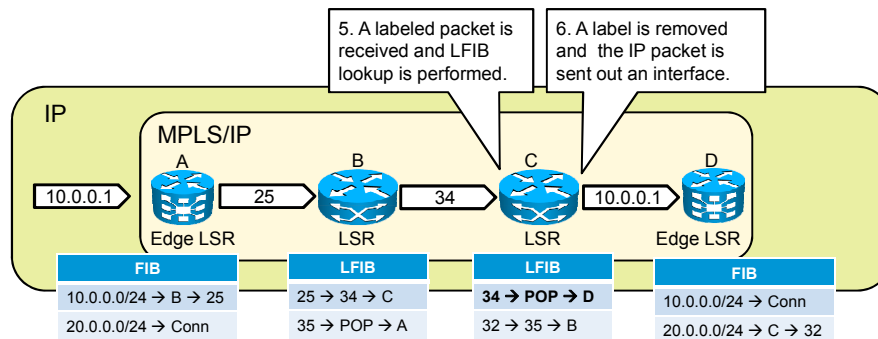
MPLS Basic Example (Cont.)



Step 3 Router B receives an IP packet that is labeled with label 25. Router B performs an LFIB lookup, which states that label 25 should be swapped with label 34.

Step 4 The label is swapped and the packet is sent to router C.

MPLS Basic Example (Cont.)



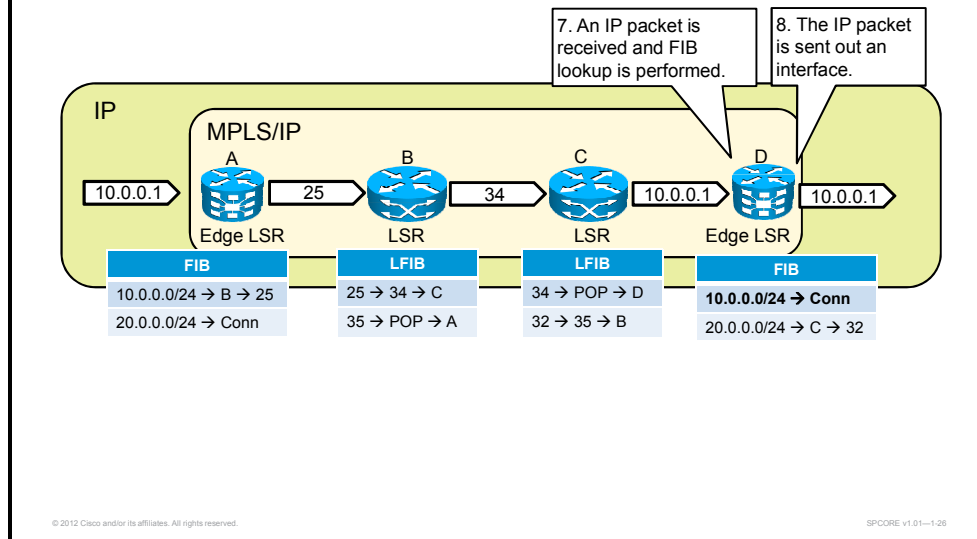
© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.25

- Step 5** Router C receives an IP packet that is labeled with label 34. Router C performs an LFIB lookup, which states that label 34 should be removed (penultimate hop popping), and the unlabeled IP packet should be sent out the interface that connects to router D. “POP” is often used as a label value that indicates that a label should be removed.
- Step 6** The label is removed and the unlabeled IP packet is sent out the interface that connects to router D.

Note A router will actually display a value of IMP-NULL (implicit null) instead of POP. An implicit null label means that the label should be removed. An IMP-NULL label uses the value 3 from a reserved range of labels.

MPLS Basic Example (Cont.)



Step 7 Finally, router D receives an IP packet. Router D performs a FIB lookup, which states that the destination network is directly connected.

Step 8 The IP packet is sent out the directly connected interface.

MPLS Label Stack

This topic describes the MPLS label stack.

MPLS Label Stack

- Usually only one label is assigned to a packet, but multiple labels in a **label stack** are supported.
- These scenarios may produce more than one label:
 - MPLS VPNs (two labels): The top label points to the egress router and the second label identifies the VPN.
 - MPLS TE (two or more labels): The top label points to the endpoint of the traffic engineering tunnel and the second label points to the destination.
 - MPLS VPNs combined with MPLS TE (three or more labels)

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-27

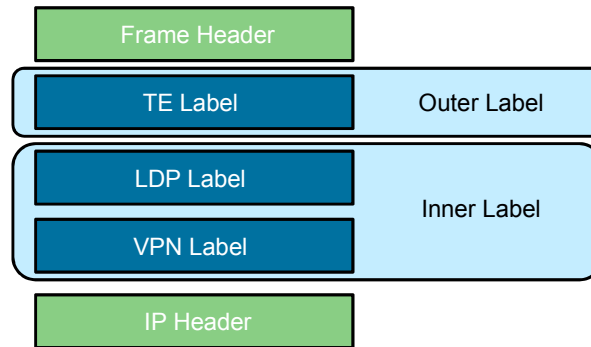
Simple MPLS uses just one label in each packet. However, MPLS does allow multiple labels in a label stack to be inserted in a packet.

These applications may add labels to packets:

- **MPLS VPNs:** With MPLS VPNs, Multiprotocol Border Gateway Protocol (MP-BGP) is used to propagate a second label that is used in addition to the one propagated by LDP or TDP.
- **Cisco MPLS Traffic Engineering (MPLS TE):** MPLS TE uses RSVP to establish LSP tunnels. RSVP also propagates labels that are used in addition to the one propagated by LDP or TDP.

A combination of these mechanisms and other advanced features might result in three or more labels being inserted into one packet.

MPLS Label Stack (Example)



- The outer label is used for switching the packet in the MPLS network (points to the TE destination).
- Inner labels are used to separate packets at egress points (point to an egress router and identify a VPN).

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1-28

The figure shows an example of a label stack where both MPLS TE and MPLS VPN are enabled.

The outer label is used to switch the MPLS packet across the network. In this case, the outer layer is a TE label pointing to the endpoint of a TE tunnel.

The inner labels are ignored by the intermediary routers. In this case, the inner labels are used to point to the egress router and to identify the VPN for the packet.

MPLS Applications

This topic describes MPLS applications in a service provider environment.

MPLS Applications

MPLS is already used in many different applications:

- Unicast IP routing
- Multicast IP routing
- MPLS TE
- QoS
- MPLS VPNs:
 - Layer 2 MPLS VPNs
 - Layer 3 MPLS VPNs

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.20

MPLS is a technology that is used for the delivery of IP Services. MPLS can be used in different applications, as outlined here:

- Unicast IP routing is the most common application for MPLS.
- Multicast IP routing is treated separately because of different forwarding requirements.
- MPLS TE is an add-on to MPLS that provides better and more intelligent link usages.
- Differentiated QoS can also be provided with MPLS.
- MPLS VPNs are implemented, using labels to allow overlapping address space between VPNs.
- AToM is a solution for transporting Layer 2 packets over an IP or MPLS backbone.

MPLS support for a label stack allows implementation of enhanced applications, such as VPNs, TE, and enhanced QoS.

MPLS Unicast IP Routing

This topic describes MPLS support for Unicast IP routing.

MPLS Unicast IP Routing

- Basic MPLS service supports unicast IP routing.
- MPLS unicast IP routing provides enhancement over traditional IP routing.
 - It has the ability to use labels for packet forwarding:
 - The FEC corresponds to a destination address stored in the IP routing table.
 - Label-based forwarding provides greater efficiency.
 - Labels support connection-oriented services.
 - It has the capability to carry a stack of labels assigned to a packet:
 - Label stacks allow implementation of enhanced applications.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-30

Basic MPLS supports unicast IP routing.

There are two significant enhancements that MPLS unicast IP routing provides over traditional IP routing:

- The ability to use labels for packet forwarding
- The capability to carry a stack of labels assigned to a packet

Using labels for packet forwarding increases efficiency in network core devices because the label swapping operation is less CPU-intensive than a routing lookup. MPLS can also provide connection-oriented services to IP traffic due to forwarding equivalence class (FEC)-based forwarding.

Note The MPLS unicast IP traffic FEC corresponds to a destination network stored in the IP routing table.

MPLS Multicast IP Routing

This topic describes MPLS support for Multicast IP routing.

MPLS Multicast Routing

- MPLS can also support multicast IP routing:
 - A dedicated protocol is not needed to support multicast traffic across an MPLS domain.
 - **Cisco Protocol Independent Multicast Version 2** with extensions for MPLS is used to propagate routing information and labels.
 - The FEC is equal to a destination multicast address that is stored in the multicast routing table.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-31

Multicast IP routing can also use MPLS. Cisco Protocol Independent Multicast (PIM) version 2 with extensions for MPLS is used to propagate routing information and labels.

The FEC is equal to a destination multicast address.

MPLS VPNs

This topic describes MPLS support for VPNs.

MPLS VPNs

- MPLS VPNs are highly scalable and support IP services:
 - Multicast
 - QoS
 - Telephony support within a VPN
 - Centralized services, including content and web hosting to a VPN
- Networks are learned via an IGP from a customer or via BGP from other MPLS backbone routers.
- Labels are propagated via MP-BGP. Two labels are used:
 - The top label points to the egress router.
 - The second label identifies the outgoing interface on the egress router or a routing table where a routing lookup is performed.
 - FEC is equivalent to a VPN site descriptor or VPN routing table.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-32

MPLS enables highly scalable VPN services to be supported. For each MPLS VPN user, the network appears to function as a private IP backbone, over which the user can reach other sites within the VPN organization, but not the sites of any other VPN organization. MPLS VPNs are a common application for service providers. Building VPNs in Layer 3 allows delivery of targeted services to a group of users represented by a VPN.

MPLS VPNs are seen as private intranets, and they support IP services such as those listed here:

- Multicast
- QoS
- Telephony support within a VPN
- Centralized services including content and web hosting to a VPN

Customer networks are learned via an Interior Gateway Protocol (IGP), OSPF, EBGp, EIGRP, or Routing Information Protocol version 2 (RIPv2), via static routing from a customer, or via BGP from other MPLS backbone routers.

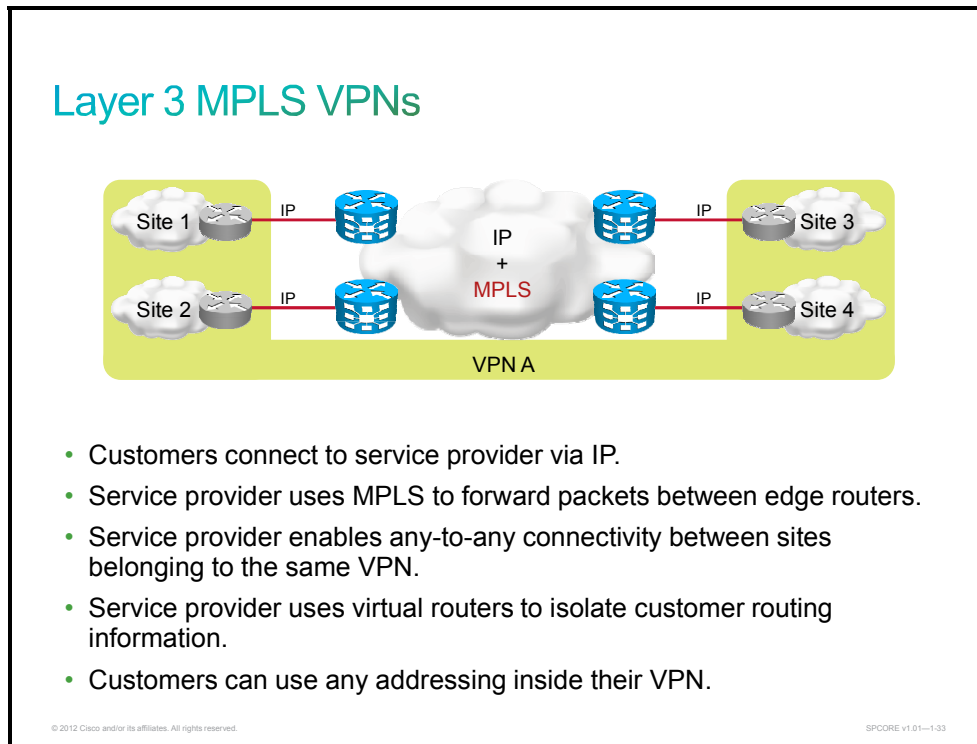
MPLS VPNs use two labels:

- The top label points to the egress router.
- The second label identifies the outgoing interface on the egress router or a routing table where a routing lookup is performed.

LDP is needed in the top label to link edge LSRs with a single LSP tunnel. MP-BGP is used in the second label to propagate VPN routing information and labels across the MPLS domain. The MPLS VPN FEC is equivalent to a VPN site descriptor or a VPN routing table.

Layer 3 MPLS VPNs

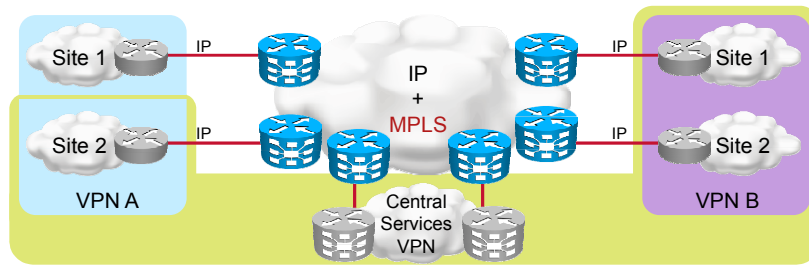
This topic describes MPLS support for Layer 3 VPNs.



The main characteristic of Layer 3 MPLS VPNs is that customers connect to a service provider via IP. They need to establish IP routing (static or dynamic) in order to exchange routing information between customer sites belonging to the same VPN. As different customers may use the same private IP address ranges, the service provider cannot perform normal IP forwarding. MPLS must be used instead, to ensure isolation in the data plane between packets belonging to different customers, yet potentially having the same IP addresses. Virtual routers (virtual routing and forwarding [VRF] instances) are used on service provider routers to isolate customer routing information. MPLS seamlessly provides any-to-any connectivity between sites belonging to the same VPN.

The most basic VPN is a so-called simple VPN or an intranet. This type of VPN is a collection of sites that are given full connectivity within the VPN while isolating the VPN from any other component in the network (other VPNs, Internet).

Layer 3 MPLS VPNs (Cont.)



- Service provider enables any-to-any connectivity between VPN sites.
- All or selected sites have access to the central VPN.
- Customers can use any addressing inside their VPNs.
- Customers must use nonoverlapping addresses to access the central VPN.

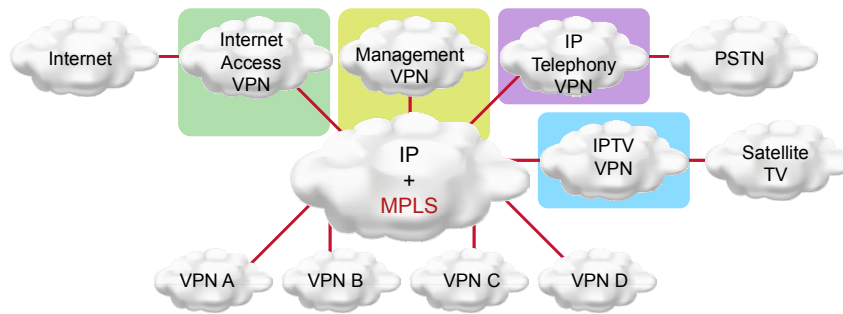
© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1-34

The figure illustrates overlapping Layer 3 MPLS VPNs where multiple customer VPNs are provided access to a central service VPN. Both VPN A (only site 2) and VPN B (both sites) in the example are able to communicate with the central services VPN. VPN A and VPN B are still isolated from each other.

The only requirement in the case of overlapping VPNs is that the VPNs use unique addressing, at least when accessing the resources available in other VPNs.

Layer 3 MPLS VPN (Cont.)



- Usage scenarios:
 - Internet access
 - Centralized management of managed customer devices
 - IP telephony
 - IPTV

© 2012 Cisco and/or its affiliates. All rights reserved.

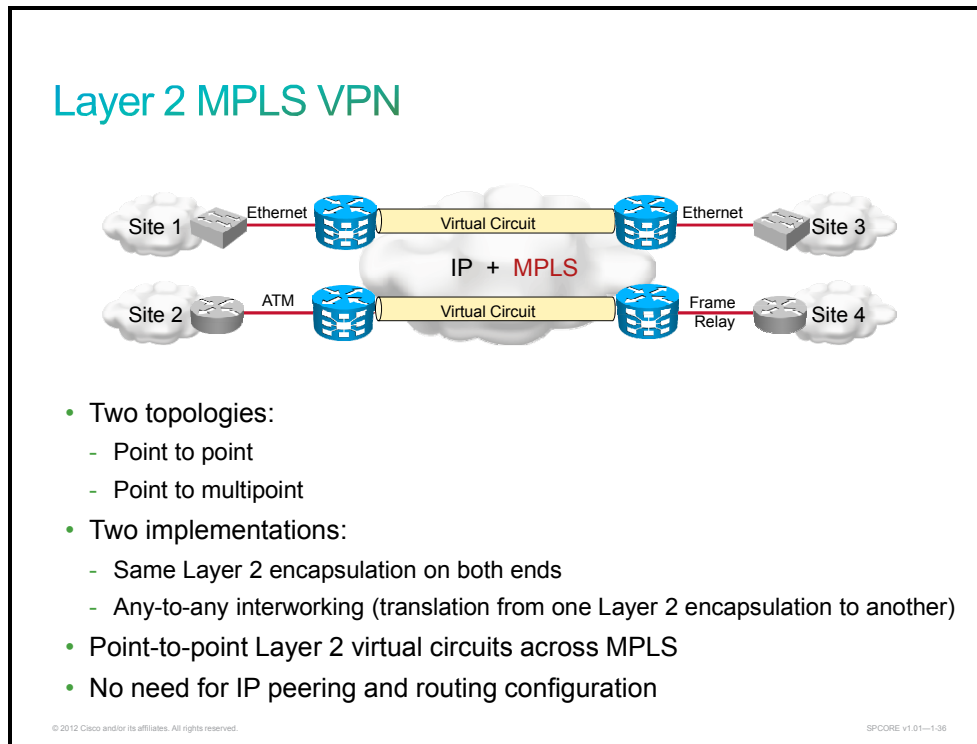
SPCORE v1.01--1.35

The figure illustrates a few examples of central services VPNs:

- Internet access can be provided through a dedicated Layer 3 MPLS VPN. Service providers can also offer wholesale services to other service providers, where they can choose their upstream Internet service provider.
- A management VPN is often used to manage the network infrastructure and services. This management VPN can also be used to manage customer routers inside Layer 3 MPLS VPNs in case the customer devices are owned and managed by the service provider.
- IP telephony can be isolated and provided to customers through a dedicated VPN.
- Even IPTV can now be isolated and provided through a dedicated VPN.

Layer 2 MPLS VPNs

This topic describes MPLS support for Layer 2 VPNs.



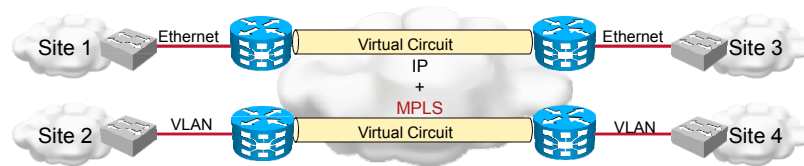
Layer 2 MPLS VPNs enable service providers to offer point-to-point or multipoint Layer 2 connections between distant customer sites. The top figure illustrates a point-to-point Ethernet connection between a pair of customer LAN switches across a virtual circuit that is implemented using MPLS. The other example illustrates interworking where one customer site uses ATM and the other site uses Frame Relay. The MPLS network translates between the two technologies similarly to what most ATM switches were able to do.

The main advantage of Layer 2 MPLS VPNs is that they do not require any IP signaling between the customer and the provider.

Ethernet over MPLS (EoMPLS) can be implemented in two ways:

- Point-to-point Ethernet over MPLS, where all Ethernet traffic is exchanged over a single virtual circuit (LSP)
- Virtual Private LAN Services (VPLS), where multiple sites can be interconnected over a full mesh of virtual circuits

Layer 2 MPLS VPN (Cont.)



Point-to-point Ethernet over MPLS

- Two modes of operation:
 - **Port mode**: Entire Ethernet frames are encapsulated into an MPLS LSP.
 - **VLAN mode**: Selected VLANs are extracted and encapsulated into dedicated MPLS LSPs.

© 2012 Cisco and/or its affiliates. All rights reserved.

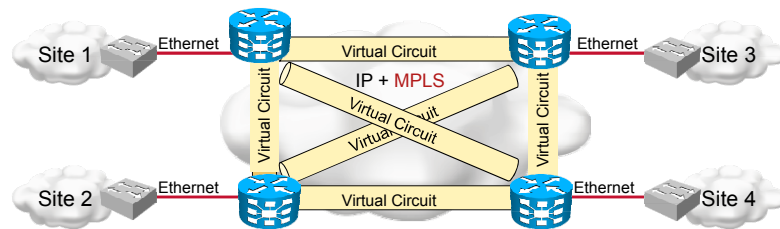
SPCORE v1.01--1.37

The most common application of Layer 2 MPLS VPNs is to provide point-to-point Ethernet connectivity between customer sites.

Ethernet over MPLS (EoMPLS) can be implemented in two ways:

- **Port mode**: Entire Ethernet frames are encapsulated into an MPLS LSP. This option allows one physical interface to be routed to a single distant remote site, but it can use IEEE 802.1Q VLANs end to end.
- **VLAN mode**: Selected VLANs are extracted and encapsulated into dedicated MPLS LSPs. This option allows a central customer site to use a single physical link with multiple VLANs that are then routed to multiple individual remote sites in different locations.

Layer 2 MPLS VPN (Cont.)



VPLS

- Multipoint Ethernet over MPLS
- MPLS network is like a virtual switch.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1-38

VPLS enables service providers with MPLS networks to offer geographically dispersed Ethernet Multipoint Service (EMS), or Ethernet Private LAN Service, as defined by the Metropolitan Ethernet Forum (MEF).

The figure illustrates VPLS implementation between four customer LAN switches in different locations. A full mesh of LSPs ensures optimal forwarding for learned MAC addresses between any pair of sites in the same virtual switch.

Note Refer to the *Implementing Cisco Service Provider Next-Generation Edge Network Services* course for detailed coverage on Layer 3 and Layer 2 MPLS VPN implementations.

MPLS Traffic Engineering

This topic describes MPLS support for Traffic Engineering.

MPLS Traffic Engineering

- MPLS TE supports constraint-based routing.
- MPLS TE enables the network administrator to:
 - Control traffic flow in the network
 - Reduce congestion in the network
 - Make best use of network resources
- MPLS TE requires OSPF or ISIS with extensions to hold the entire network topology in their databases.
- OSPF and IS-IS should also have some additional information about network resources and constraints.
- RSVP is used to establish TE tunnels and to propagate labels.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-39

Another application of MPLS is TE. MPLS TE enables an MPLS backbone to replicate and expand upon the TE capabilities of Layer 2 ATM and Frame Relay networks. MPLS TE supports constraint-based routing, in which the path for a traffic flow is the shortest path that meets the resource requirements (constraints) of the traffic flow. Factors such as bandwidth requirements, media requirements, and the priority of one traffic flow versus another can be taken into account. TE capabilities enable the administrator of a network to accomplish these goals:

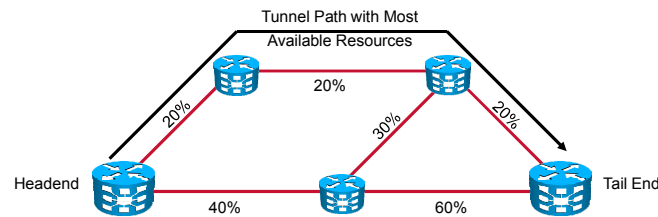
- Control traffic flow in the network
- Reduce congestion in the network
- Make best use of network resources

MPLS TE has these special requirements:

- Every LSR must see the entire topology of the network (only OSPF and IS-IS hold the entire topology).
- Every LSR needs additional information about links in the network. This information includes available resources and constraints. OSPF and IS-IS have extensions to propagate this additional information.
- RSVP is used to establish TE tunnels and to propagate the labels.

Every edge LSR must be able to create an LSP tunnel on demand. RSVP is used to create an LSP tunnel and to propagate labels for TE tunnels.

MPLS Traffic Engineering (Cont.)



- Redundant networks may experience unequal load in their network.
- It is difficult to optimize resource utilization using routing protocols with default destination-based routing.
- MPLS TE tunnels are used to enable traffic flow across any path.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1-40

The primary reason for using MPLS TE, as the name suggests, is to engineer traffic paths. Redundant networks may experience unequal loads in their networks, due to the calculated best paths that are typically determined based on IGP metrics. It is difficult to optimize resource utilization (link utilization) using routing protocols with default destination-based routing. The figure illustrates a scenario where the least-cost path would flow through the most congested link, thus making it even more congested, and possibly resulting in drops and increased delays. MPLS TE can be used to divert some traffic to “less optimal” paths; this capability will result in better utilization of resources (more network throughput) and lower delays.

MPLS QoS

This topic describes MPLS support for QoS.

MPLS QoS

- MPLS QoS provides differentiated types of service across an MPLS network.
- MPLS QoS offers capability:
 - Packet classification
 - Congestion avoidance
 - Congestion management
- MPLS QoS is an extension to unicast IP routing that provides differentiated services.
- Extensions to LDP are used to propagate different labels for different classes.
- The FEC is a combination of a destination network and a class of service.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.41

MPLS QoS enables network administrators to provide differentiated types of service across an MPLS network. MPLS QoS offers packet classification, congestion avoidance, and congestion management.

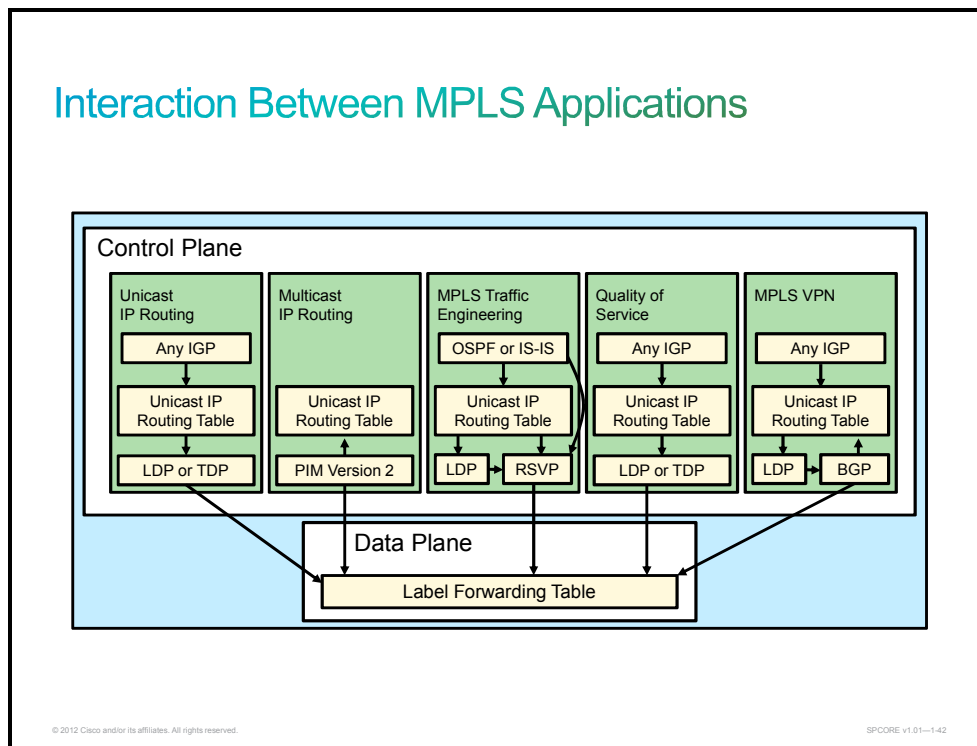
Note MPLS QoS functions map nearly one-for-one to IP QoS functions on all interface types.

Differentiated QoS is achieved by using MPLS experimental bits or by creating separate LSP tunnels for different classes. Extensions to LDP are used to create multiple LSP tunnels for the same destination (one for each class).

The FEC for MPLS QoS is equal to a combination of a destination network and a class of service (CoS).

Interaction between MPLS Applications

This topic describes the interaction Between MPLS Applications.



The figure shows the overall architecture when multiple applications are used.

Regardless of the application, the functionality is always split into the control plane and the data (forwarding) plane, as discussed here:

- The applications may use different routing protocols and different label exchange protocols in the control plane.
- The applications all use a common label-switching data (forwarding) plane.
- Edge LSR Layer 3 data planes may differ to support label imposition and disposition.
- Typically, a label is assigned to an FEC.

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- Traditional ISP provided Internet access, while traditional Telco provided VPN services.
- Modern service providers provide Internet access, VPN, telephony and QoS using IP and MPLS.
- Cisco IP NGN is the next-generation service provider architecture for providing voice, video, mobile, and cloud or managed services to users.
- The SONET standard is mainly used in the United States, while the SDH standard is mainly European.
- Wavelength division multiplexing (WDM) is a technology that multiplexes a number of optical carrier signals into a single optical fiber by using different wavelengths of laser light.
- IP over DWDM (IPoDWDM) is a technology that integrates DWDM on routers.
- IEEE 802.3ba is part of the 802.3 family of data link layer standards for Ethernet LAN and WAN applications, whose objective is to support speeds faster than 10 Gb/s.
- Traditional service provider architecture was based on numerous parallel services with a simple, stable SDH core.
- Traditional IP routing forwards packets based on the destination IP address.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—1-43

Summary (Cont.)

- MPLS enables the forwarding of packets through label-switched paths that can be created using various methods and protocols, depending on the required results.
- MPLS switches packets based on label lookup instead of IP address lookup. Labels usually correspond to destination IP networks.
- MPLS is very useful in service provider (and large enterprise) networks because it enhances BGP routing and provides different services and applications.
- In an MPLS domain, there are two types of routers: LSRs and edge LSRs.
- The control plane builds a routing table that is based on the routing protocol.
- The data plane takes care of forwarding, based on either destination addresses or labels.
- The data plane on an MPLS-enabled router consists of two forwarding structures: FIB and LFIB.
- Control plane on an MPLS-enabled router usually uses link-state routing protocol to exchange IP prefixes and LDP to exchange MPLS labels.
- An MPLS label is a 4-byte identifier that is used for making forwarding decisions.
- If an MPLS-enabled router receives a labeled packet, the router performs LFIB lookup.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—1-44

Summary (Cont.)

- MPLS allows multiple labels in a label stack to be inserted in a IP packet.
- MPLS is used in many applications: IP routing, MPLS VPNs, MPLS TE, and QoS.
- One of the significant enhancements of unicast MPLS routing over IP routing is the capability to carry a stack of labels assigned to a packet.
- Multicast IP routing can also use MPLS. Cisco Protocol Independent Multicast (PIM) version 2 with extensions for MPLS is used to propagate routing information and labels.
- MPLS supports highly scalable VPN services.
- The main characteristic of Layer 3 MPLS VPNs is that a customer transparently connects his networks through a service provider network via IP.
- Layer 2 MPLS VPNs enable service providers to offer point-to-point or multipoint Layer 2 connections between distant customer sites.
- MPLS TE supports constraint-based routing, in which the path for a traffic flow is the shortest path that meets the resource requirements of the traffic flow.
- MPLS QoS provides differentiated types of service across an MPLS network.
- MPLS application may use different protocols at control and data plane.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-45

Label Distribution Protocol

Overview

This lesson takes a detailed look at the Label Distribution Protocol (LDP) neighbor discovery process via hello messages and by the type of information that is exchanged. It also describes the events that occur during the negotiation phase of LDP session establishment, as well as the nonadjacent neighbor discovery process, providing a further understanding of the Multiprotocol Label Switching (MPLS) technology. This lesson discusses how label allocation and distribution function in an MPLS network, the use of penultimate hop popping (PHP), and how the MPLS data structures are built. These topics are essential for understanding the fundamentals of the way that information gets distributed and placed into the appropriate tables, for both labeled and unlabeled packet usage.

This lesson presents LDP convergence issues, describes how routing protocols and MPLS convergence interact, and concludes with an explanation of the switching mechanisms on various Cisco platforms.

Objectives

Upon completing this lesson, you will be able to describe the LDP process and operation in a service provider network. You will be able to meet these objectives:

- Describe LDP is the protocol used to exchange the MPLS labels
- Describe how LDP neighbor adjacency is established
- Describe the LDP Link Hello Message
- Describe the LDP Session Negotiation
- Describe the use of the LDP Targeted Hello Message to form LDP neighbor adjacency between non directly connected LSRs
- Describe LDP Session Protection using a backup targeted hello
- Describe LDP Graceful Restart and NonStop Routing (NSR)
- Describe how the forwarding structures used by MPLS are populated
- Explain the LSP
- Explain the MPLS Label Allocation and Distribution process
- Show how IP packets are propagated across an MPLS domain

- Define the steady state condition when all the labels are exchanged by LDP and the LIBs, LFIBs and FIBs are completely populated
- Explain Label Advertisement Control and Label Acceptance Control
- Explain the how IP Aggregation in the core can break an LSP into two segments
- Describe loop detection using the MPLS TTL field
- Describe the disabling of TTL propagation to hide the core routers in the MPLS domain
- Show a steady state condition in the MPLS domain
- Show how a link failure is managed in an MPLS domain
- Show how a link recovery is managed in an MPLS domain
- Describe the three IP switching mechanisms (Process Switching, Fast Switching and Cisco Express Forwarding)
- Explain the sequence of events that occurs when process switching and fast switching are used for destinations that are learned through BGP
- Explain the sequence of events that occurs when CEF switching is used for destinations that are learned through BGP
- Describe CEF on Cisco IOS XE and Cisco IOS XR platforms
- Describe the show commands used to monitor CEF operations

Label Distribution Protocol (LDP)

This topic describes LDP as the protocol used to exchange the MPLS labels.

Label Distribution Protocol

- MPLS introduces a label field that is used for forwarding decisions.
- Although labels are locally significant, they must be advertised to directly reachable peers.
 - Option 1 is to include this parameter in existing IP routing protocols.
 - Option 2 is to create a new protocol to exchange labels.
- The second option has been used, because there are too many existing IP routing protocols that would have to be modified to carry labels.
- The new protocol is called Label Distribution Protocol (LDP).

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1.4

One application of MPLS is unicast IP routing. A label is assigned to destination IP networks and is later used to label packets sent toward those destinations.

Note In MPLS terminology, a forwarding equivalence class (FEC) in MPLS unicast IP routing equals an IP destination network.

Standard or vendor-specific routing protocols are used to advertise IP routing information. MPLS adds a new piece of information that must be exchanged between adjacent routers.

Here are the two possible approaches to propagating this additional information (labels) between adjacent routers:

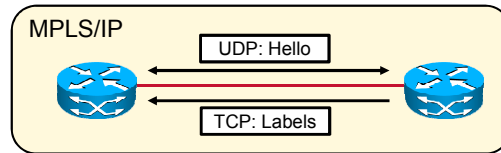
- Extend the functionality of existing routing protocols
- Create a new protocol dedicated to exchanging labels

The first approach requires much more time and effort because of the large number of different routing protocols: Open Shortest Path First (OSPF), Intermediate System-to-Intermediate System (IS-IS), Enhanced Interior Gateway Routing Protocol (EIGRP), Interior Gateway Routing Protocol (IGRP), Routing Information Protocol (RIP), and so on. The first approach also causes interoperability problems between routers that support this new functionality and those that do not. Therefore, the IETF selected the second approach and defined Label Distribution Protocol (LDP) in RFC 3036.

LDP Neighbor Adjacency Establishment

This topic describes how LDP neighbor adjacency is established.

LDP Neighbor Session Establishment



- LDP establishes a session in two steps:
 - Hello messages are periodically sent on all MPLS-enabled interfaces.
 - MPLS-enabled routers respond to received hello messages by attempting to establish a session with the source of the hello messages.
- An LDP link hello message is a UDP packet that is sent to the “all routers on this subnet” multicast address (224.0.0.2 or FF02::2).
- TCP is used to establish the session.
- Both TCP and UDP use well-known LDP port number 646.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.6

LDP is a standard protocol used to exchange labels between adjacent routers.

Before labels can be exchanged, MPLS-enabled routers must first establish adjacencies. This is done in two steps:

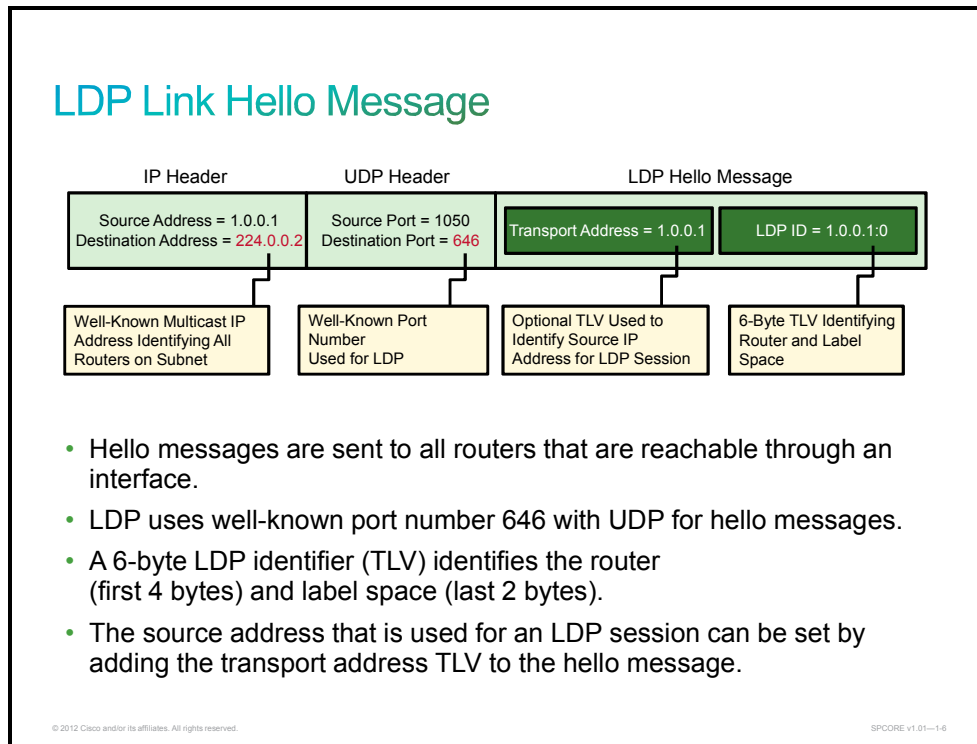
- **LDP discovery:** MPLS routers first discover neighbors using hello messages that are sent to all the routers on the subnet as User Datagram Protocol (UDP) packets with a multicast destination address of 224.0.0.2 (FF02::2 for IPv6) and destination port number of 646.
- **LDP adjacency:** A neighboring MPLS router, that received hello packets, will respond by opening a TCP session with the same destination port number 646, and the two routers begin to establish an LDP session through unicast TCP.

LDP periodically sends hello messages (every 5 seconds). If the label switch router (LSR) is adjacent or one hop from its neighbor, the LSR sends out LDP link hello messages to all the routers on the subnet as UDP packets with a multicast destination address of 224.0.0.2 or FF02::2 (“all routers on a subnet”) and a destination port number of 646.

A neighboring LSR enabled for LDP will respond by opening a TCP session with the same destination port number 646, and the two routers begin to establish an LDP session through unicast TCP.

LDP Link Hello Message

This topic describes the LDP Link Hello Message.



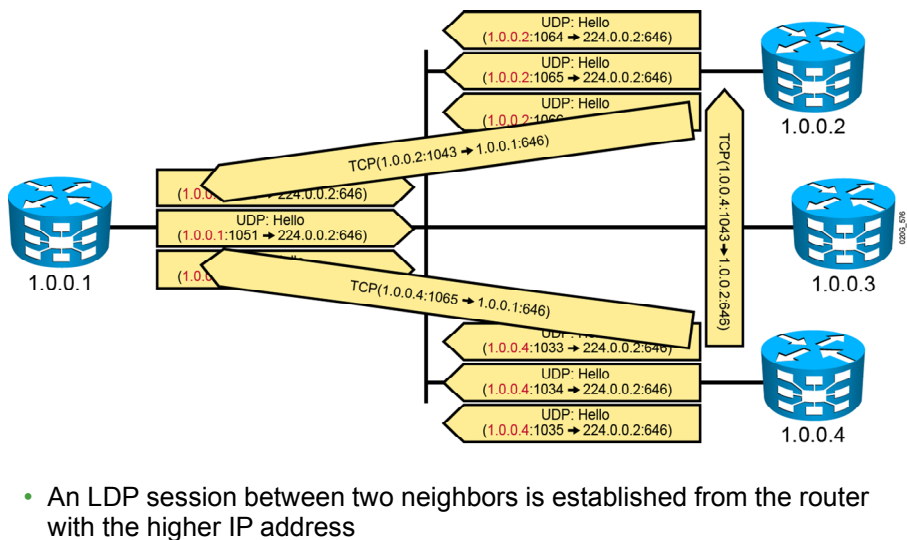
These are the contents of an LDP link hello message:

- Destination IP address (224.0.0.2 for IPv4 or FF02::2 for IPv6), which reaches all routers on the subnetwork
- Destination port, which equals the LDP well-known port number 646
- The actual hello message, which may optionally contain a transport address type, length, value (TLV) to instruct the peer to open the TCP session to the transport address instead of the source address found in the IP header. The LDP identifier (LDP ID) is used to uniquely identify the neighbor and the label space.

Note Label space defines the way MPLS assigns labels to destinations. Label space can be either per-platform or per-interface.

On Cisco routers, for all interface types, except the label-controlled ATM interfaces (running cell-mode MPLS-over-ATM), per-platform label space will be used where all the interfaces of the router share the same set of labels. For per-platform label space, the last two bytes of the LDP ID are always both 0. Multiple LDP sessions can be established between a pair of LSRs if they use multiple label spaces. For example, label-controlled cell-mode ATM interfaces use virtual path identifiers/virtual circuit identifiers (VPIs/VCI) for labels. Depending on its configuration, 0, 1, or more interface-specific label spaces can be used.

LDP Neighbor Discovery



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1.7

In the figure, three out of four routers periodically send out LDP hello messages (the fourth router is not MPLS-enabled).

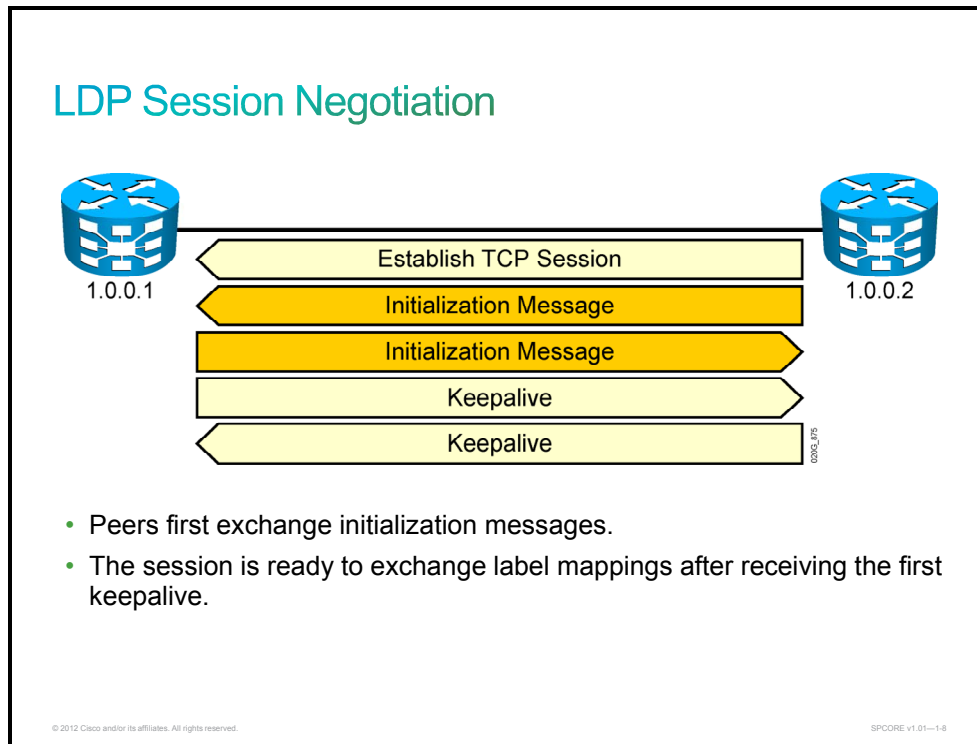
Routers that have the higher LDP router ID must initiate the TCP session. For instance, the router with the LDP router ID 1.0.0.2 initiates a TCP session to the router with LDP router ID 1.0.0.1.

If the LDP router ID is not manually configured, the highest IP address of all loopback interfaces on a router is used as the LDP router ID. If no loopback interfaces are configured on the router, the highest IP address of a configured interface that was operational at LDP startup is used as the LDP router ID.

On Cisco IOS XR Software, if the LDP router ID is not configured, the router can also default to the use of the global router ID as the LDP router ID. After the TCP session is established, routers will keep sending LDP hello messages to potentially discover new peers or to identify failures.

LDP Session Negotiation

This topic describes the LDP Session Negotiation.



LDP session negotiation is a three-step process:

- Step 1** Establish the TCP session.
- Step 2** Exchange initialization messages that contain information such as the label distribution method, the session keepalive time, the fault-tolerant (FT) TLV and so on. The LDP neighbor responds with an initialization message if the parameters are acceptable. If parameters are not acceptable, the LDP neighbor sends an error notification message.
- Step 3** Exchange initial keepalive messages.

Note LDP keepalives are sent every 60 seconds.

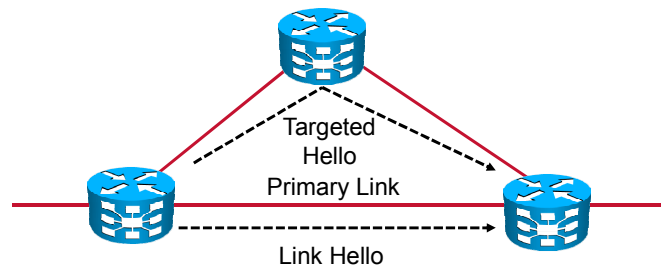
After these steps, the two peers start exchanging labels for networks that they have in their main routing tables.

LDP Discovery of Nonadjacent Neighbors

This topic describes the use of the LDP Targeted Hello Message to form LDP neighbor adjacency between non-directly connected LSRs.

LDP Discovery of Nonadjacent Neighbors

- LDP neighbor discovery of nonadjacent neighbors differs from normal discovery only in the addressing of hello packets:
 - Hello packets use unicast IP addresses instead of multicast addresses.
- When a neighbor is discovered, the mechanism to establish a session is the same



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-9

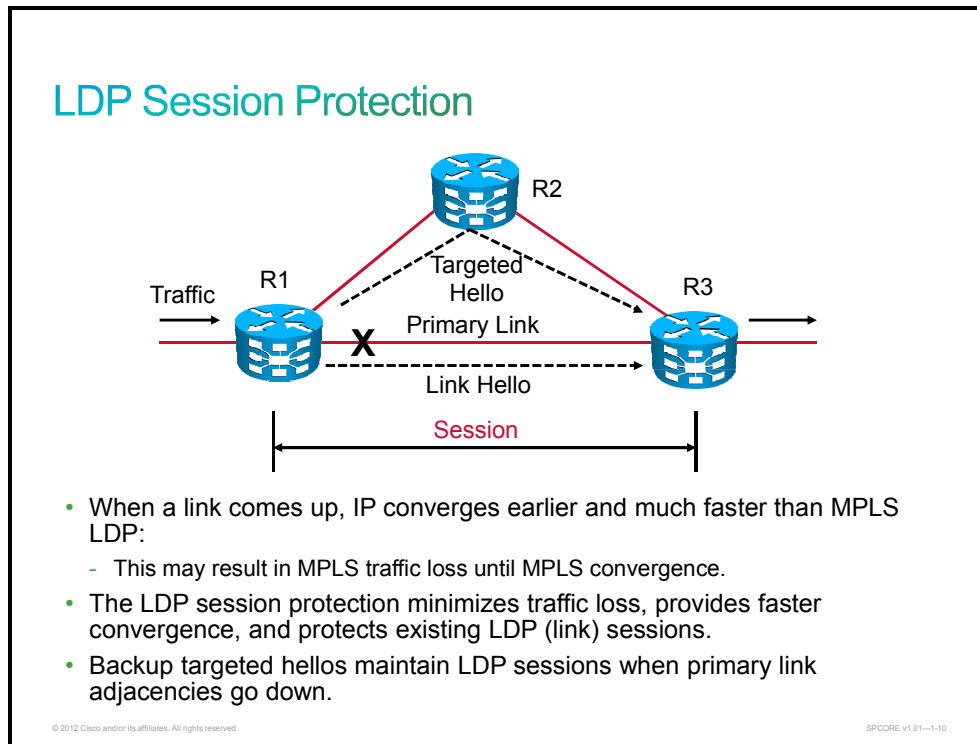
If the LSR is more than one hop from its neighbor, it is not directly connected or adjacent to its neighbor. The LSR can be configured to send a directed hello message as a unicast UDP packet that is specifically addressed to the nonadjacent neighbor LSR. The directed hello message is called an LDP targeted hello.

The rest of the session negotiation is the same as for adjacent routers. The LSR that is not directly connected will respond to the hello message by opening a unicast TCP session with the same destination port number 646, and the two routers begin to establish an LDP session.

For example, when you use the MPLS traffic engineering tunnel interface, a label distribution session is established between the tunnel headend and the tunnel tail end routers. To establish this not-directly-connected MPLS LDP session, the transmission of targeted LDP Hello messages is used.

LDP Session Protection

This topic describes LDP Session Protection using a backup targeted hello.



Another example of using targeted LDP hello messages is between directly connected MPLS label switch routers, when MPLS label forwarding convergence time is an issue.

For example, when a link comes up, IP converges earlier and much faster than MPLS LDP and may result in MPLS traffic loss until MPLS convergence. If a link flaps, the LDP session will also flap due to loss of link discovery. LDP session protection minimizes traffic loss, provides faster convergence, and protects existing LDP (link) sessions by using a “parallel” source of targeted discovery hello. An LDP session is kept alive and neighbor label bindings are maintained when links are down. Upon reestablishment of primary link adjacencies, MPLS convergence is expedited because LDP does not need to relearn the neighbor label bindings.

LDP session protection lets you configure LDP to automatically protect sessions with all or a given set of peers (as specified by the peer-acl). When it is configured, LDP initiates backup targeted hellos automatically for neighbors for which primary link adjacencies already exist. These backup targeted hellos maintain LDP sessions when primary link adjacencies go down.

The Session Protection as shown in the figure illustrates LDP session protection between the R1 and R3 LDP neighbors. The primary link adjacency between R1 and R3 is a directly connected link, and the backup targeted adjacency is maintained between R1 and R3 through R2. If the direct link fails, the direct LDP link adjacency is destroyed, but the LDP session is kept functional using targeted hello adjacency (through R2). When the direct link comes back up, there is no change in the LDP session state and LDP can converge quickly and begin forwarding MPLS traffic.

LDP Graceful Restart and NonStop Routing (NSR)

This topic describes LDP Graceful Restart and NonStop Routing (NSR).

LDP Graceful Restart and Nonstop Routing

- LDP graceful restart provides a control plane mechanism to ensure high availability and allows detection and recovery from failure conditions while preserving NSF services.
- Graceful restart recovers from control plane failures without impacting forwarding.
- Without LDP graceful restart, when an established session fails, the corresponding forwarding states are cleaned immediately from the restarting and peer nodes:
 - In this example, LDP forwarding restarts from the beginning, causing a potential loss of data and connectivity.
- LDP NSR functionality makes failures, such as RP failover, invisible to routing peers with minimal to no disruption of convergence performance.
- LDP NSR does not require protocol extensions and does not force software upgrades on other routers in the network.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.11

LDP graceful restart provides a control plane mechanism to ensure high availability and allows detection and recovery from failure conditions while preserving Nonstop Forwarding (NSF) services. Graceful restart is a way to recover from signaling and control plane failures without impacting forwarding.

Without LDP graceful restart, when an established LDP session fails, the corresponding forwarding states are cleaned immediately from the restarting and peer nodes. The LDP forwarding restarts from the beginning, causing a potential loss of data and connectivity.

The LDP graceful restart capability is negotiated between two peers during session initialization time, in the FT session type length value (TLV). In this TLV, each peer advertises the following information to its peers:

- An LSR indicates that it is capable of supporting LDP graceful restart by including the FT session TLV in the LDP initialization message and setting the L (Learn from Network) flag to 1.
- **Reconnect time:** Advertises the maximum time that the other peer will wait for this LSR to reconnect after control channel failure.
- **Recovery time:** Advertises the maximum time that the other peer will retain its MPLS forwarding state that it preserved across the restart. The recovery time should be long enough to allow the neighboring LSRs to resynchronize their MPLS forwarding state in a graceful manner. This time is used only during session reestablishment after earlier session failure.

Once the graceful restart session parameters are conveyed and the session is functioning, graceful restart procedures are activated.

LDP nonstop routing (NSR) functionality makes failures, such as route processor (RP) or distributed route processor (DRP) failover, invisible to routing peers with minimal to no disruption of convergence performance.

Unlike graceful restart functionality, LDP NSR does not require protocol extensions and does not force software upgrades on other routers in the network, nor does LDP NSR require peer routers to support NSR.

MPLS Forwarding Structures

This topic describes how the forwarding structures used by MPLS are populated.

Forwarding Structures

- Forwarding structures that are used by MPLS need to be populated.
- The FIB is populated two ways:
 - A routing table, which is populated by a routing protocol
 - An MPLS label that is added to the FIB by LDP
- The LFIB is populated by LDP.
- LDP is responsible for the advertisement and redistribution of MPLS labels between MPLS routers.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-12

Forwarding structures that are used by MPLS need to be populated with labels. Label distribution protocol (LDP), which runs in the router control plane, is responsible for label allocation, distribution, and storage.

The forwarding information base (FIB) table, which consists of destination networks, next hops, outgoing interfaces, and pointers to Layer 2 addresses, is populated by using information from the routing table and from the Address Resolution Protocol (ARP) cache. The routing table is in turn populated by a routing protocol. Additionally, the MPLS label is added to destination networks if an outgoing interface is enabled for MPLS and a label has been received from the next hop router. LDP is responsible for adding a label to the FIB table entries.

The label forwarding information base (LFIB) table contains incoming (locally assigned) and outgoing (received from next hop) labels. LDP is responsible for exchanging labels and storing them into the LFIB.

Label Switched Path (LSP)

This topic explains the LSP.

Label-Switched Path

- An **LSP** is a sequence of LSRs that forwards labeled packets of a certain forwarding equivalence class.
 - MPLS unicast IP forwarding builds LSPs based on the output of IP routing protocols.
 - LDP advertises labels only for individual segments in the LSP.
- LSPs are **unidirectional**.
 - Return traffic uses a different LSP (usually the reverse path because most routing protocols provide symmetrical routing).
- An LSP can take a different path from the one chosen by an IP routing protocol (MPLS TE).

© 2012 Cisco and/or its affiliates. All rights reserved.

SPICORE v1.01--1-13

A label-switched path (LSP) is a sequence of LSRs that forwards labeled packets for a particular Forwarding Equivalence Class (FEC). Each LSR swaps the top label in a packet traversing the LSP. An LSP is similar to Frame Relay or ATM virtual circuits.

In MPLS unicast IP forwarding, the FECs are determined by destination networks that are found in the main routing table. Therefore, an LSP is created for each entry that is found in the main routing table. Border Gateway Protocol (BGP) entries are the only exceptions for that rule.

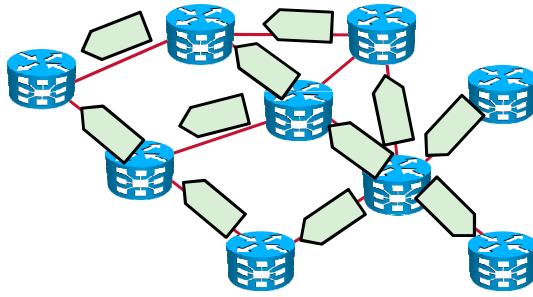
In an ISP environment where a routing table may contain more than 100,000 routes, to minimize the number of labels that are needed in such networks, an exception was made for BGP-derived routing information. All BGP entries in the main routing table use the same label that is used to reach the BGP next-hop router (the PE router). Only the PE routers are required to run BGP. All the core (P) routers run an IGP to learn about the BGP next-hop addresses. The core routers run LDP to learn about labels for reaching the BGP next-hop addresses. This results in one single label being used for all networks that are learned from the BGP neighbor.

An Interior Gateway Protocol (IGP) is used to populate the routing tables in all routers in an MPLS domain. LDP is used to propagate labels for these networks and build LSPs.

LSPs are unidirectional. Each LSP is created over the shortest path, selected by the IGP, toward the destination network. Packets in the opposite direction use a different LSP. The return LSP is usually over the same LSRs, except that packets form the LSP are in the opposite order.

Cisco MPLS Traffic Engineering (MPLS TE) can be used to change the default IGP shortest path selection.

LSP Building



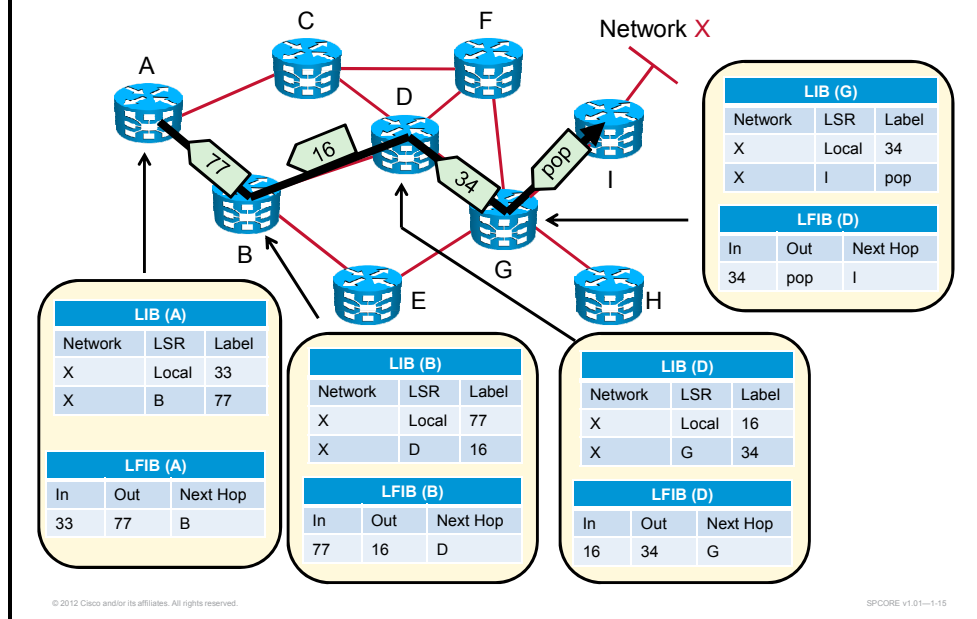
- The IP routing protocol determines the path.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.14

The figure illustrates how an IGP, such as OSPF, IS-IS, or EIGRP, propagates routing information to all routers in an MPLS domain. Each router determines its own shortest path. LDP, which propagates labels for those networks, adds labels to the FIB and LFIB tables.

LSP Building (Cont.)



The figure shows the contents of LFIB and LIB tables. MPLS uses a liberal label retention mode, which means that each LSR keeps all labels received from LDP peers, even if they are not the downstream peers (the next-hop) for reaching network X. With a liberal retention mode, an LSR can almost immediately start forwarding labeled packets after IGP convergence, but the numbers of labels maintained for a particular destination will be larger and thus will consume more memory.

LIB and LFIB tables are shown on the routers for label switched path A-B-D-G-I. On each router for this path, only local labels, and labels received from adjacent routers forming this path, are shown in LIB tables.

Note Notice that router G receives a pop label from final destination router I. The pop action results in the removal of the label rather than swapping labels. This allows the regular IP packet to be forwarded out on the router I interface that is directly connected to network X when the packet leaves the MPLS domain.

Label Allocation and Distribution

This topic explains the MPLS Label Allocation and Distribution process.

Label Allocation and Distribution

Label allocation and distribution in a MPLS network follows these steps:

1. IP routing protocols build the IP routing table.
2. Each LSR assigns a label to every destination in the IP routing table independently.
3. LSRs announce their assigned labels to all other LSRs.
4. Every LSR builds its LIB, LFIB, and FIB data structures based on received labels.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-18

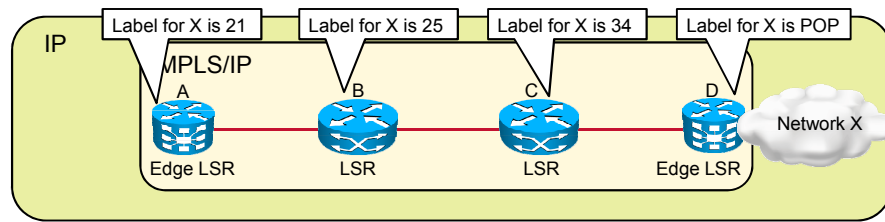
Unicast IP routing and MPLS functionality can be divided into these steps:

- Routing information is exchanged using standard or vendor-specific IP routing protocols (OSPF, IS-IS, EIGRP, and so on).
- Local labels are generated. (One locally unique label is assigned to each IP destination found in the main routing table and stored in the LIB table.)
- Local labels are propagated to adjacent routers, where these labels might be used as next-hop labels (stored in the FIB and LFIB tables to enable label switching).

These data structures contain label information:

- LSRs store labels and related information inside a data structure called a label information base (LIB). The FIB and LFIB contain labels only for the currently used best LSP segment, while the LIB contains all labels known to the LSR, whether the label is currently used for forwarding or not. The LIB in the control plane is the database that is used by LDP; an IP prefix is assigned a locally significant label, which is mapped to a next-hop label that has been learned from a downstream neighbor.
- The LFIB, in the data plane, is the database used to forward labeled packets that are received by the router. Local labels, previously advertised to upstream neighbors, are mapped to next-hop labels, previously received from downstream neighbors.
- The FIB, in the data plane, is the database used to forward unlabeled IP packets that are received by the router. A forwarded packet is labeled if a next-hop label is available for a specific destination IP network. Otherwise, a forwarded packet is not labeled.

Label Allocation



- Each router generates a label for each network in a routing table:
 - Labels have local significance.
 - Label allocation is asynchronous.
- For path discovery and loop avoidance, LDP relies on routing protocols.
- Networks originating on the outside of the MPLS domain are not assigned any label on the edge LSR; instead, the POP label is advertised.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1-17

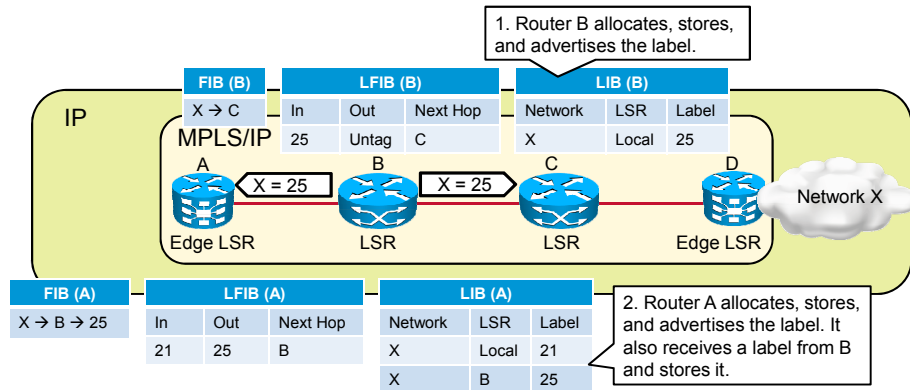
First, each MPLS-enabled router must locally allocate a label for each network that is known to a router. Labels are locally significant (a label for the same network has a different value on different routers), and allocation of labels is asynchronous (routers assign labels, independent of each other).

The LDP is not responsible for finding a shortest, loop-free, path to destinations. Instead, the LDP relies on routing protocols to find the best path to destinations. If, however, a loop does occur, a Time to Live (TTL) field in the MPLS label prevents the packet from looping indefinitely.

On the edge LSR, networks originating on the outside of the MPLS domain are not assigned a label. Instead, the POP (implicit null) label is advertised, which instructs the penultimate router to remove the label.

In the example, all routers except router D assign a label for network X. Router D assigns an implicit null label for that network.

Label Advertisement



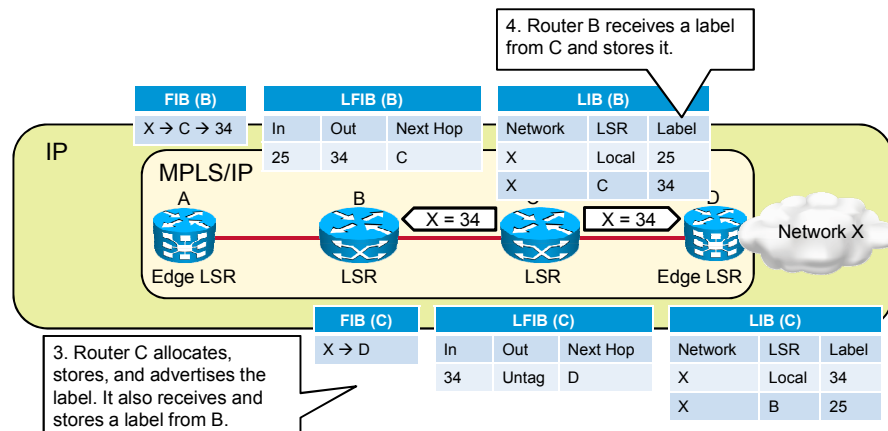
- A router that receives a label from a next hop also stores the label in the FIB.

After a label has been assigned locally, each router must advertise a label to neighbors. The figure shows how a label is assigned and advertised to neighbors on router B.

- Step 1** Router B allocates label 25 for network X. The allocated label is first stored in the label information base (LIB), which stores local labels and labels received from neighbors as well. The label is also stored in a LFIB table as an incoming label. The outgoing label has not yet been set, because router B has not received a label for network X from the next hop router (router C) yet. The allocated label is also advertised to the neighbors (routers A and C), regardless of whether a neighbor actually is a next hop for a destination or not.
- Step 2** Router A allocates his own label for network X (21 in the example). This label is again stored in the LIB and in the LFIB as an incoming label. Router A also receives a label 25 from router B and stores the label in the LIB. Because label 25 has been received from a next hop for destination X, router A also stores label 25 in the LFIB as an outgoing label. Router A also sets the label 25 for destination X in the FIB table, because the label has been received from the next hop.

If a packet for network X was received by router A (not shown in the figure), a FIB lookup would be done. The packet would be labeled using label 25 and sent to router B. Router B would perform an LFIB lookup, which would state that the label should be removed, because the outgoing label had not yet been received from the next hop router (router C).

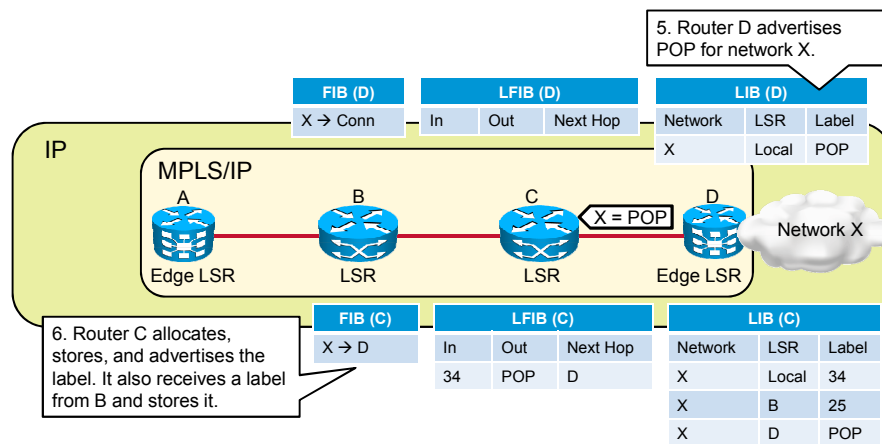
Label Advertisement (Cont.)



- A router stores a label from a neighbor, even if the neighbor is not a next hop for a destination.

- Step 3** Router C allocates label 34 for network X. The allocated label is first stored in the LIB. The label is also stored in the LFIB table as an incoming label. The outgoing label has not yet been set, because router C has not received a label for network X from the next hop router (router D) yet. The allocated label is also advertised to the neighbors (routers B and D), regardless of whether a neighbor actually is a next hop for a destination or not. Router C also received a label 25 from router B and stores it in its LIB, even though router B is not a next hop for destination X.
- Step 4** Router B receives a label 34 from router C and stores the label in the LIB. Because label 34 has been received from a next hop for destination X, router B also stores label 34 in an LFIB as an outgoing label. Router B also sets the label 34 for destination X in the FIB table, because the label has been received from the next hop.

Label Advertisement (Cont.)



- Networks originating on the outside of the MPLS domain are not assigned any label on the edge LSR; instead, the POP label is advertised.

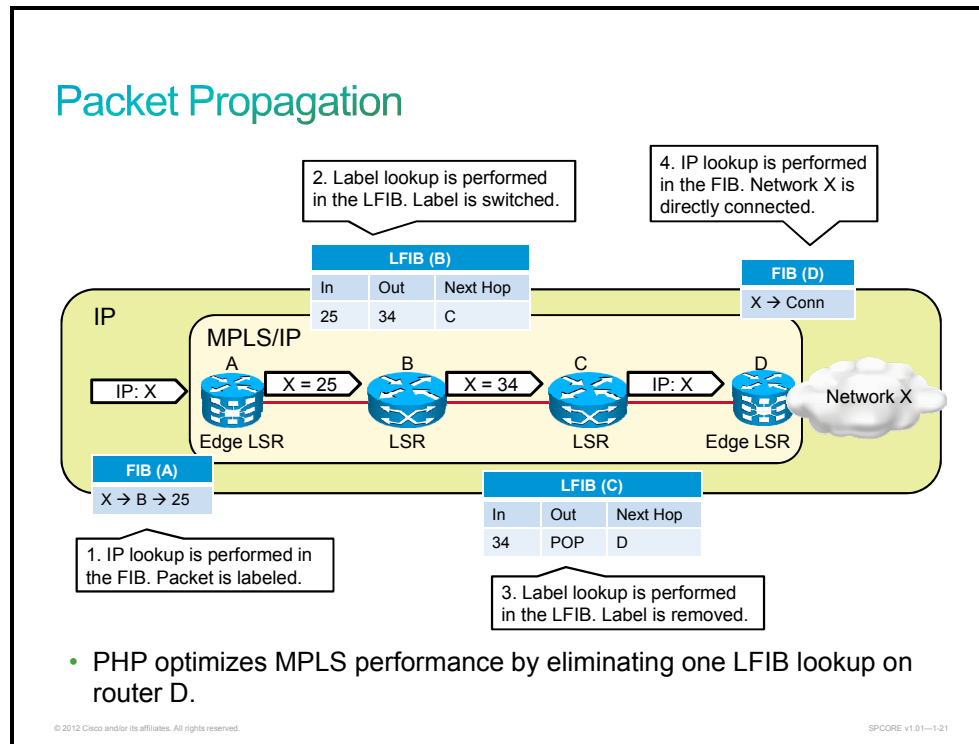
© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1.20

- Step 5** Router D allocates the implicit null label for network X. The allocated label is first stored in the LIB. The implicit null label is also advertised to router C. The implicit null label indicates to the upstream router that it should perform a label removal (pop) operation.
- Step 6** Router C receives the implicit null label from router D and stores the label in the LIB. Because the label has been received from a next hop for destination X, router C also stores the label in LFIB as an outgoing label. Because the implicit null label indicates that label should be removed, router C does set the label in the FIB table. The LSP for a network X is now established.

Packet Propagation across an MPLS Domain

This topic shows how IP packets are propagated across an MPLS domain.



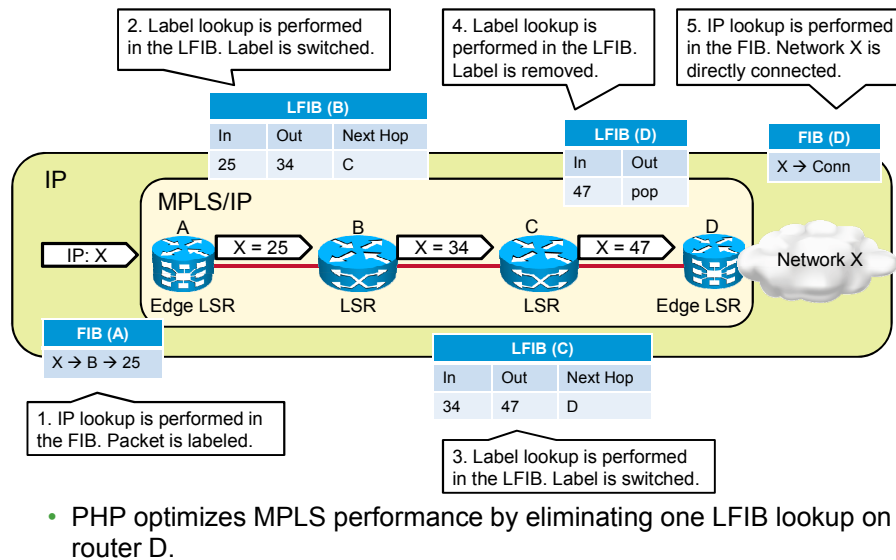
The figure illustrates how IP packets are propagated across an MPLS domain. The steps are as follows:

- Step 1** Router A labels a packet destined for network X by using the next-hop label 25 (Cisco Express Forwarding, switching by using the FIB table).
- Step 2** Router B swaps label 25 with label 34 and forwards the packet to router C (label switching by using the LFIB table).
- Step 3** Router C removes (pops) the label and forwards the packet to router D (label switching by using the LFIB table).
- Step 4** Router D performs IP lookup in the FIB table. Network X is directly connected.

The figure assumes that the implicit null label, which corresponds to the pop action in the LFIB, has been propagated from the egress router (router D) to router C. The term “pop” means to remove the top label in the MPLS label stack instead of swapping it with the next-hop label. The last router before the egress router therefore removes the top label. The process is called penultimate hop popping (PHP), which is enabled by default on all MPLS-enabled routers.

PHP optimizes MPLS performance by eliminating one LFIB lookup at the egress router, as only FIB lookup is needed on router D, the egress router.

Packet Propagation Without PHP



The figure illustrates how IP packets would be propagated across an MPLS domain *without* the PHP process. The steps are as follows:

- Step 1** Router A labels a packet destined for network X by using the next-hop label 25 (Cisco Express Forwarding switching by using the FIB table).
 - Step 2** Router B swaps label 25 with label 34 and forwards the packet to router C (label switching by using the LFIB table).
 - Step 3** As PHP is not enabled, router D does not advertise the implicit null (pop) label to the router C. Therefore, router C swaps label 34 with label 47 and forwards the packet to router D (label switching by using the LFIB table).
 - Step 4** Router D removes the label (label switching by using the LFIB table).
 - Step 5** Router D performs IP lookup in the FIB table. Network X is directly connected.
- As you see, one more LFIB lookup is needed on router D if PHP is not enabled.

MPLS Steady State Condition

This topic defines the steady state condition when all the labels are exchanged by LDP and the LIBs, LFIBs and FIBs are completely populated.

Steady State

- Steady state occurs after all the labels are exchanged and LIB, LFIB, and FIB structures are completely populated.
- It takes longer for LDP to exchange labels than it takes for a routing protocol to converge.
- There is no network downtime before the LDP fully exchanges labels.
- In the meantime, packets can be routed using FIB, if labels are not yet available.
- After the steady state is reached, all packets are label-switched, except on the ingress and egress routers.

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01--1-23

MPLS is fully functional when the routing protocol and LDP have populated all the tables:

- Main IP routing table
- LIB table
- FIB table
- LFIB table

Such a state is called the steady state. After the steady state, all packets are label-switched, except on the ingress and egress routers (edge LSR).

Although it takes longer for LDP to exchange labels (compared with a routing protocol), a router can use the FIB table in the meantime. Therefore, there is no routing downtime while LDP exchanges labels between adjacent LSRs.

MPLS Label Control Methods

This topic explains Label Advertisement Control and Label Acceptance Control.

Label Control

Label Advertisement Control:

- For scalability and security reasons
- LDP configured to perform outbound filtering for local label advertisement, for one or more prefixes to one more peers
- Also referred to as LDP outbound

Label Acceptance Control:

- For security reasons, or to conserve memory
- A label binding acceptance configured for a set of prefixes from a given peer
- Also referred to as LDP inbound label filtering

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01--1.24

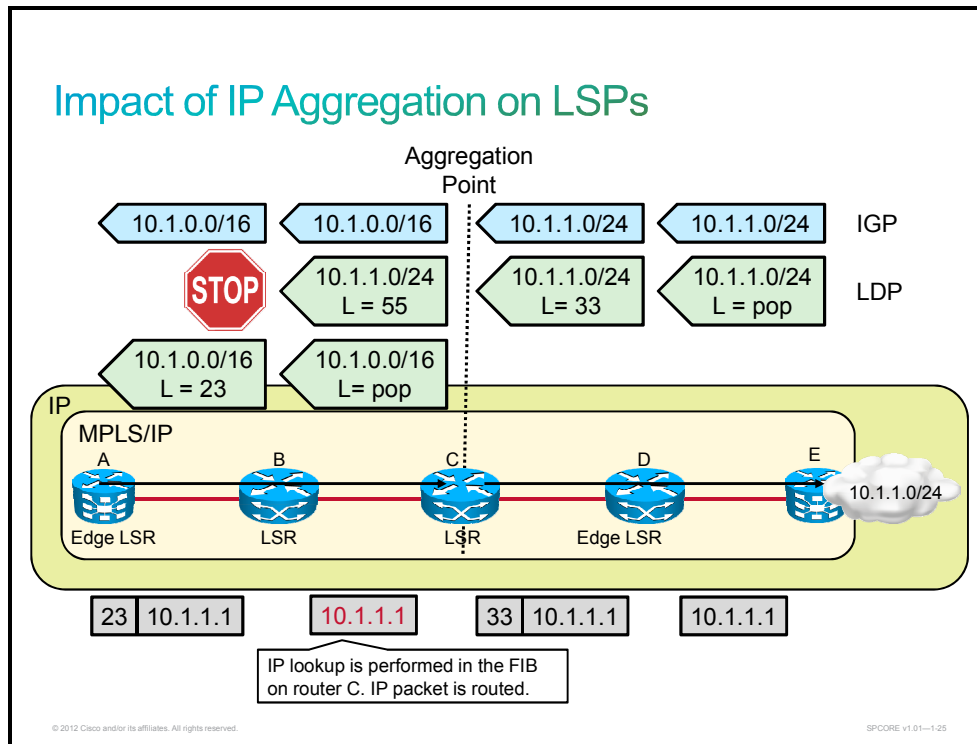
By default, LDP advertises labels for all the prefixes to all its neighbors. When this is not desirable (for scalability and security reasons), you can configure LDP to perform outbound filtering for local label advertisement for one or more prefixes to one or more LDP peers. This feature is known as LDP outbound label filtering, or local label advertisement control.

By default, LDP accepts labels (as remote bindings) for all prefixes from all LDP peers. LDP operates in liberal label retention mode, which instructs LDP to keep remote label bindings from all LDP peers for a given prefix, even if the LDP peer is not the next-hop router. For security reasons, or to conserve memory, you can override this behavior by configuring label binding acceptance for a set of prefixes from a given LDP peer.

The ability to filter remote bindings for a defined set of prefixes is also referred to as LDP inbound label filtering.

Impact of IP Aggregation on LSPs

This topic explains the how IP Aggregation in the core can break an LSP into two segments.



The figure illustrates a potential problem in an MPLS domain.

An IGP propagates the routing information for network 10.1.1.0/24 from router E to other routers in the network. Router C uses a route summarization mechanism to stop the proliferation of all subnetworks of network 10.1.0.0/16. Only the summary network 10.1.0.0/16 is sent to routers B and A.

LDP propagates labels concurrently with the IGP. The LSR that is the endpoint of an LSP always propagates the “pop” label.

Router C has both networks in the routing table:

- 10.1.1.0/24 (the original network)
- 10.1.0.0/16 (the summary route)

Router C, therefore, sends a label, 55 in the example, for network 10.1.1.0/24 to router B. Router C also sends an implicit null (pop) label for the new summary network 10.1.0.0/16 that originates on router C. Router B, however, can use the implicit null (pop) label only for the summary network 10.1.0.0/16 because it has no routing information about the more specific network 10.1.1.0/24; this information was suppressed on router C.

The route summarization results in two LSPs for destination network 10.1.1.0/24. The first LSP ends on router C, where a routing lookup is required to assign the packet to the second LSP.

Impact of IP Aggregation on LSPs (Cont.)

- IP aggregation breaks an LSP into two segments.
- Aggregation should not be used where end-to-end LSPs are required; these are some examples:
 - MPLS VPNs
 - MPLS TEs
 - MPLS-enabled ATM network
 - Transit BGP where the core routers are not running BGP

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.28

Aggregation should also not be used where an end-to-end LSP is required. Here are some typical examples of networks that require end-to-end LSPs:

- An MPLS VPN backbone
- A network that uses MPLS TE
- An MPLS-enabled ATM network
- A transit BGP autonomous system (AS) where the core routers are not running BGP

Loop Detection using the MPLS TTL field

This topic describes loop detection using the MPLS TTL field.

Loop Detection

- LDP relies on loop detection mechanisms that are built into the IGPs that are used to determine the path.
- If, however, a loop is generated (that is, misconfiguration with static routes), the TTL field in the label header is used to prevent the indefinite looping of packets.
- TTL functionality in the label header is equivalent to TTL in the IP headers.
- TTL is usually copied from the IP headers to the label headers (**TTL propagation**).

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1-27

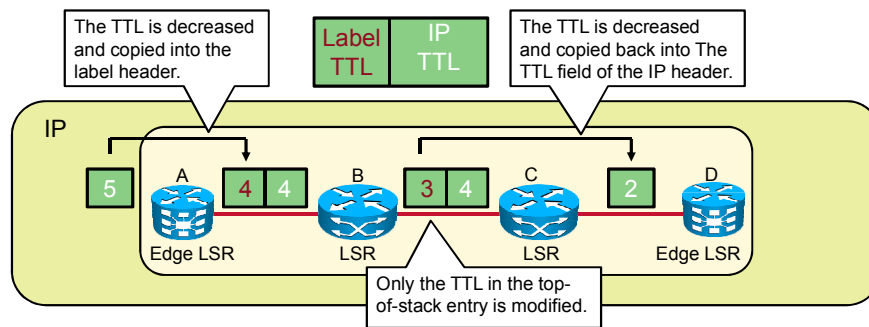
Loop detection in an MPLS-enabled network relies on more than one mechanism.

Most routing loops are prevented by the IGP that is used in the network. MPLS for unicast IP forwarding simply uses the shortest paths as determined by the IGP. These paths are typically loop-free.

If, however, a routing loop does occur (for example, because of misconfigured static routes), MPLS labels also contain a Time-to-Live (TTL) field that prevents packets from looping indefinitely.

The TTL functionality in MPLS is equivalent to that of traditional IP forwarding. Furthermore, when an IP packet is labeled, the TTL value from the IP header is copied into the TTL field in the label. This process is called TTL propagation.

Normal TTL Operation



- Cisco routers have TTL propagation enabled by default.
- On ingress, the TTL is copied from the IP header to the label header.
- On egress, the TTL is copied from the label header to the IP header.
- Labeled packets are dropped when the TTL is decreased to 0.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1.28

The figure illustrates how the TTL value 5 in the IP header is decreased and copied into the TTL field of the label when a packet enters an MPLS domain.

All other LSRs decrease the TTL field only in the label. The original TTL field is not changed until the last label is removed when the label TTL is copied back into the IP TTL.

TTL propagation provides a transparent extension of IP TTL functionality into an MPLS-enabled network.

The packet looping between these two routers is eventually dropped because the value of its TTL field reaches 0.

Disabling TTL Propagation

This topic describes the disabling of TTL propagation to hide the core routers in the MPLS domain.

Disabling TTL Propagation

- TTL propagation can be disabled.
- The IP TTL value **is not** copied into the TTL field of the label, and the label TTL **is not** copied back into the IP TTL.
- Instead, the value 255 is assigned to the label header TTL field on the ingress LSR.
- Disabling TTL propagation **hides core routers** in the MPLS domain.
- Traceroute across an MPLS domain does not show any core routers.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.29

TTL propagation can be disabled to hide the core routers from the end users. Disabling TTL propagation causes routers to set the value 255 into the TTL field of the label when an IP packet is labeled. The network is still protected against indefinite loops, but it is unlikely that the core routers will ever have to send an Internet Control Message Protocol (ICMP) reply to user-originated traceroute packets.

With TTL propagation disabled, the MPLS TTL is calculated independent of the IP TTL, and the IP TTL remains constant for the length of the LSP. Because the MPLS TTL is unlikely to drop from 255 to 0, none of the LSP router hops will trigger an ICMP TTL exceeded message, and consequently these router hops will not be recorded in the traceroute output.

Impact of Disabling TTL Propagation

- Traceroute across an MPLS domain does not show core routers.
- TTL propagation must be disabled on all label switch routers.
- Mixed configurations (some LSRs with TTL propagation enabled and some LSRs with TTL propagation disabled) could result in faulty traceroute output.
- TTL propagation can be enabled for forwarded traffic only.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-30

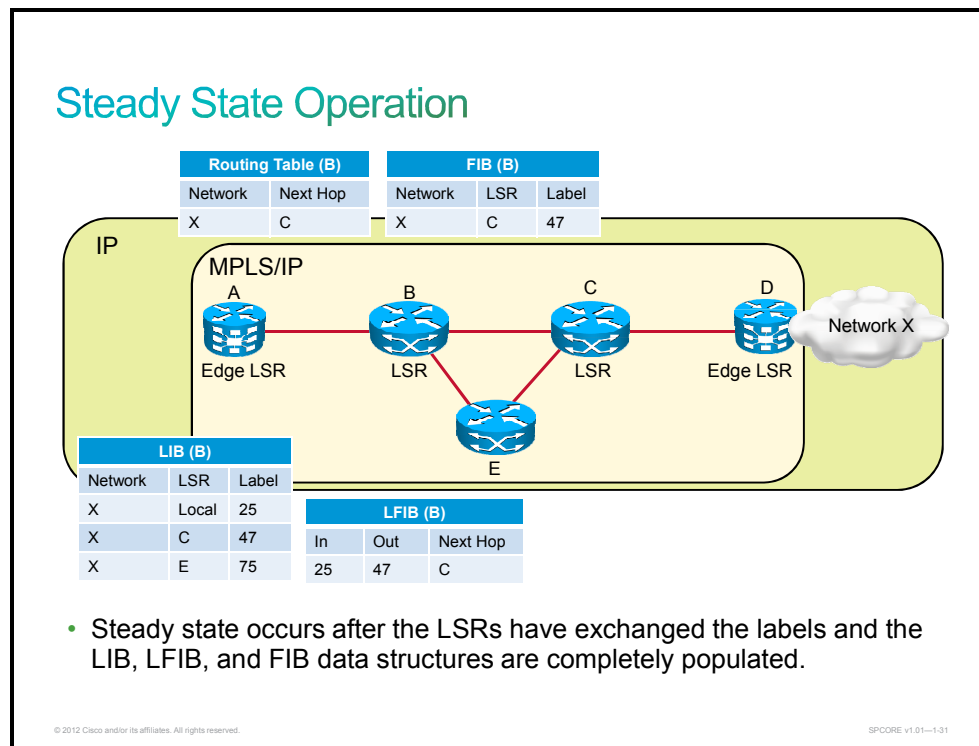
Cisco routers have TTL propagation enabled by default.

If TTL propagation is disabled, it must be disabled on all routers in an MPLS domain to prevent unexpected behavior.

TTL can be optionally disabled for forwarded traffic only, which allows administrators to use traceroute from the routers to troubleshoot problems in the network.

Steady State Condition

This topic shows a steady state condition in the MPLS domain.



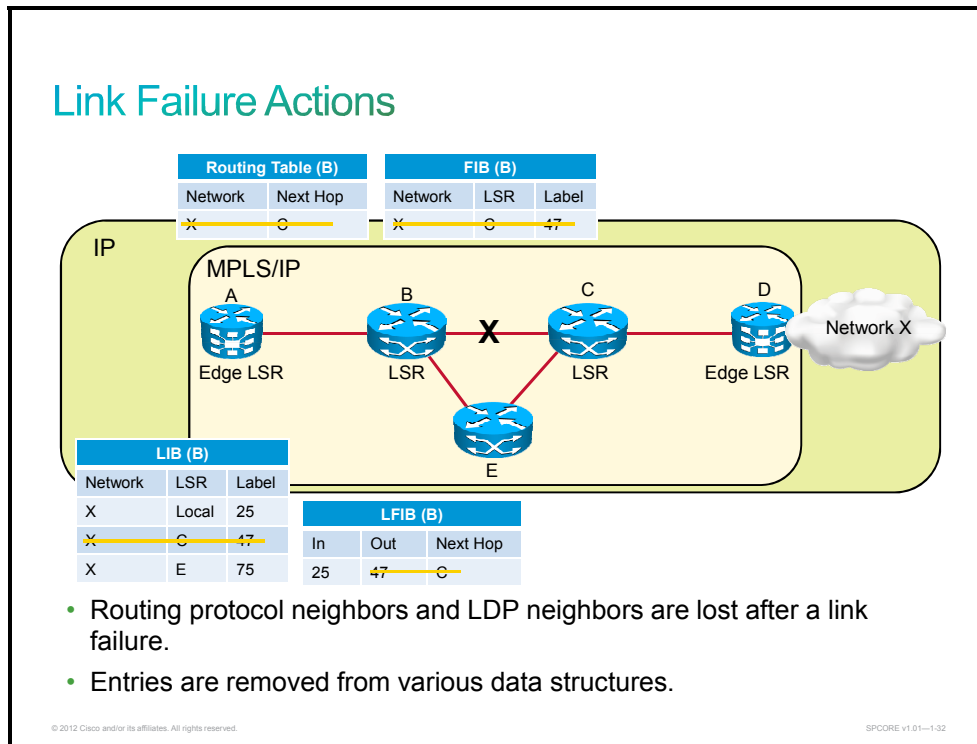
MPLS is fully functional when the IGP and LDP have populated all the tables:

- Main IP routing (routing information base [RIB]) table
- LIB table
- FIB table
- LFIB table

Although it takes longer for LDP to exchange labels (compared with an IGP), a network can use the FIB table in the meantime; therefore, there is no routing downtime while LDP exchanges labels between adjacent LSRs.

Link Failure MPLS Convergence Process

The topic shows how a link failure is managed in an MPLS domain.



The figure illustrates how a link failure is managed in an MPLS domain:

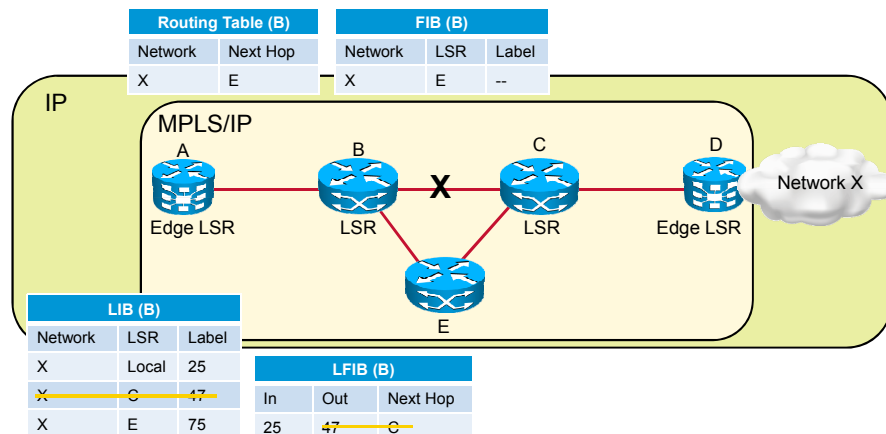
- The overall convergence fully depends on the convergence of the IGP used in the MPLS domain.
- The link between router B and router C goes down.
- Entries regarding router C are removed from the LIB, LFIB, FIB, and RIB (routing table).
- When router B determines that router E should be used to reach network X, the label learned from router E can be used to label-switch the packets.

LDP stores all labels in the LIB table, even if the labels are not used, because the IGP has decided to use another path.

This label storage is shown in the figure, where two next-hop labels were available in the LIB table on router B. This is the label status of router B just before MPLS label convergence:

- Label 47 was learned from router C and is currently unavailable; therefore, because of the failure, label 47 must be removed from the LIB table.
- Label 75 was learned from router E, and can now be used at the moment that the IGP decides that router E is the next hop for network X.

Routing Protocol Convergence

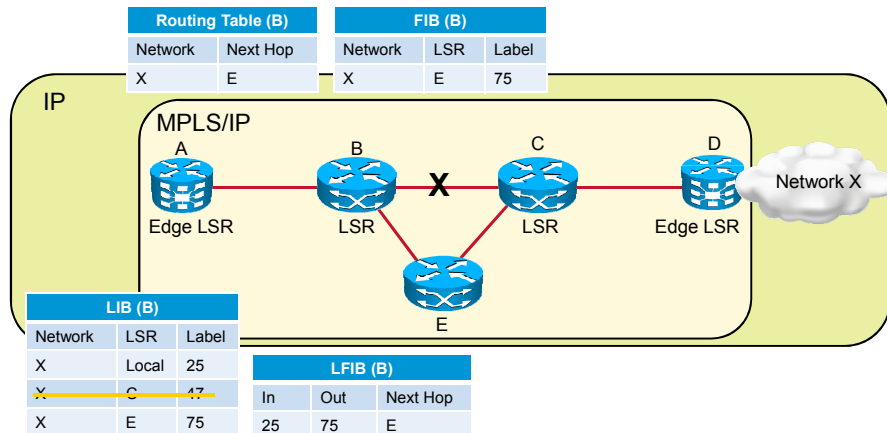


- Routing protocols rebuild the IP routing table and the IP forwarding table.

The figure illustrates how two entries are removed, one from the LIB table and one from the LFIB table, when the link between routers B and C fails. This can be described as follows:

- When the IGP determined that the next hop was no longer reachable, Router B removed the entry from the FIB table.
- Router B removed the entry from the LIB table and the LFIB table because LDP has determined that router C is no longer reachable.

MPLS Convergence



- The LFIB and labeling information in the FIB are rebuilt immediately after the routing protocol convergence, based on labels stored in the LIB.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.34

After the IGP determines that there is another path available, a new entry is created in the FIB table. This new entry points toward router E, and there is already a label available for network X via router E in the LIB table. This information is then used in the FIB table and in the LFIB table to reroute the LSP tunnel via router E.

MPLS Convergence After a Link Failure

- MPLS convergence in frame-mode MPLS does not affect the overall convergence time.
- MPLS convergence occurs immediately after the routing protocol convergence, based on the labels that are already stored in the LIB.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-35

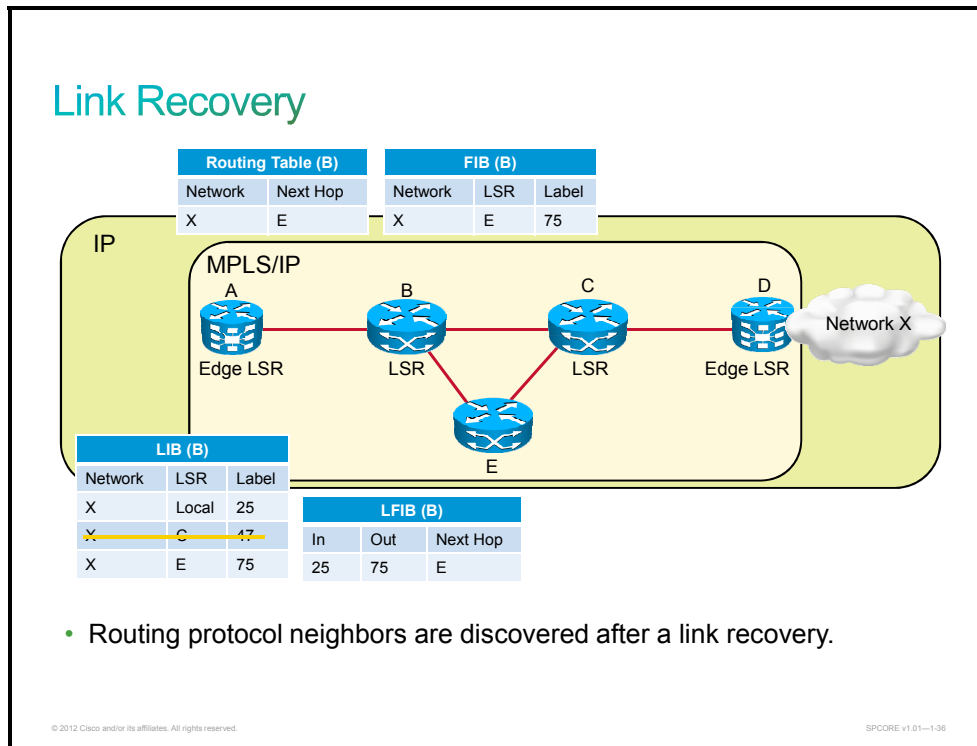
The overall convergence in an MPLS network is not affected by LDP convergence when there is a link failure. Frame-mode MPLS uses liberal label retention mode, which enables routers to store all received labels, even if the labels are not being used. These labels can be used, after the network convergence, to enable immediate establishment of an alternative LSP tunnel.

MPLS uses a 32-bit label field that is inserted between Layer 2 and Layer 3 headers (frame-mode MPLS). In frame-mode MPLS, routers that are running MPLS exchange labeled IP packets as well as unlabeled IP packets (PHP) with one another, in an MPLS domain.

MPLS over ATM uses the ATM header as the label (cell-mode MPLS). In cell-mode MPLS, the LSRs in the core of the MPLS network are ATM switches that forward data based on the ATM header. Cell-mode MPLS operations will not be covered in this course.

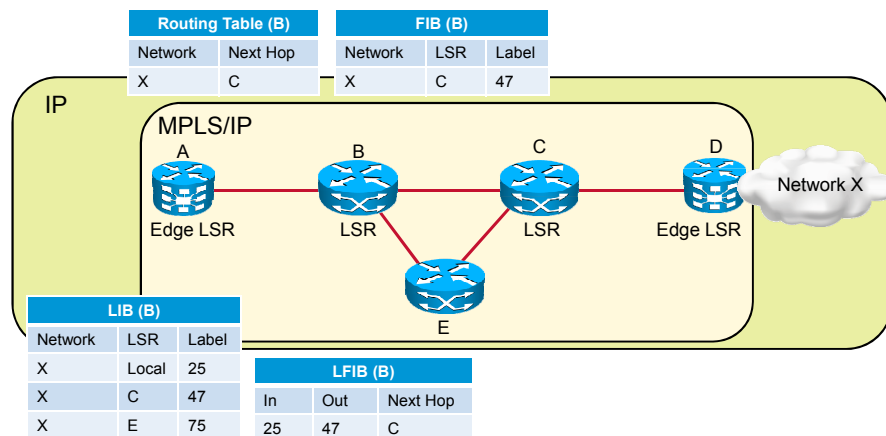
Link Recovery MPLS Convergence Process

The topic shows how a link recovery is managed in an MPLS domain.



The figure illustrates the state of the B tables of the router at the time the link between routers B and C becomes available again, but before the network reconverge.

Link Recovery (Cont.)



- IP routing protocols rebuild the IP routing table.
- The FIB and the LFIB are also rebuilt, but the label information might be lacking.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1-37

The IGP determines that the link between routers B and C is available again, and changes the next-hop address for network X to point to router C. However, when router B also tries to set the next-hop label for network X, it has to wait for the LDP session between routers B and C to be reestablished.

A pop action is used in the LFIB table on router B while the LDP establishes the session between routers B and C. This process adds to the overall convergence time in an MPLS domain. The downtime for network X is not influenced by LDP convergence, because normal IP forwarding is used until the new next-hop label is available.

As shown in the figure, after the LDP session between routers B and C is reestablished, router B will update its tables with an outgoing label of 47, with router C as the next-hop for reaching network X.

Link Recovery: MPLS Convergence

- Routing protocol convergence optimizes the forwarding path after a link recovery.
- The LIB might not contain the label from the new next hop by the time the IGP convergence is complete.
- End-to-end MPLS connectivity might be intermittently broken after link recovery.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-38

Link recovery requires that an LDP session be established (or reestablished), which adds to the convergence time of LDP. Networks may be temporarily unreachable because of the convergence limitations of routing protocols. Cisco MPLS TE can be used to prevent long downtime when a link fails or when it is recovering.

IP Switching Mechanisms

This topic describes the three IP switching mechanisms (Process Switching, Fast Switching and Cisco Express Forwarding).

Evolution of Switching Mechanisms

The Cisco IOS platform supports three IP switching mechanisms:

- Routing table-driven switching—process switching
 - Full lookup for every packet
- Cache-driven switching—fast switching
 - Most recent destinations entered in the cache
 - First packet always process-switched
- Topology-driven switching
 - Cisco Express Forwarding (prebuilt FIB table)

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.39

The first and the oldest switching mechanism that is available in Cisco routers is process switching. Because process switching must find a destination in the routing table (possibly a recursive lookup) and construct a new Layer 2 frame header for every packet, it is very slow and is normally not used.

To overcome the slow performance of process switching, Cisco IOS platforms support several switching mechanisms that use a cache to store the most recently used destinations. The cache uses a faster searching mechanism, and it stores the entire Layer 2 frame header to improve the encapsulation performance. The first packet whose destination is not found in the fast-switching cache is process-switched, and an entry is created in the cache. The subsequent packets are switched in the interrupt code using the cache to improve performance.

The latest and preferred Cisco IOS platform-switching mechanism is Cisco Express Forwarding, which incorporates the best of the previous switching mechanisms. Cisco Express Forwarding supports per-packet load balancing (previously supported only by process switching), per-source or per-destination load balancing, fast destination lookup, and many other features not supported by other switching mechanisms.

The Cisco Express Forwarding cache, or FIB table, is essentially a replacement for the standard routing table.

Standard IP Switching Example

This topic explains the sequence of events that occurs when process switching and fast switching are used for destinations that are learned through BGP.

Cisco Express Forwarding

Cisco IOS XE Software:

- Label switching requires that Cisco Express Forwarding be enabled on the router.
- Cisco Express Forwarding requires a software image that includes Cisco Express Forwarding and IP routing enabled on the device.
- Cisco Express Forwarding is enabled by default on the Cisco ASR 1000 Series Aggregation Services Routers.

```
Router# show ip cef
Prefix      Next Hop      Interface
[...]
10.2.61.8/24  192.168.100.1  FastEthernet1/0/0
              192.168.101.1  FastEthernet2/1/0
```

- Reveals if Cisco Express Forwarding is enabled by default on your platform.
- If not, enable it with **ip cef** command.

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01--140

There is a specific sequence of events that occurs when process switching and fast switching are used for destinations that are learned through BGP.

The figure illustrates this process. Here is a description of the sequence of events:

- When a BGP update is received and processed, an entry is created in the routing table.
- When the first packet arrives for this destination, the router tries to find the destination in the fast-switching cache. Because the destination is not in the fast-switching cache, process switching has to switch the packet when the process is run. The process performs a recursive lookup to find the outgoing interface. The process switching may possibly trigger an Address Resolution Protocol (ARP) request or may find the Layer 2 address in the ARP cache. Finally, it creates an entry in the fast-switching cache.
- All subsequent packets for the same destination are fast-switched, as follows:
 - The switching occurs in the interrupt code (the packet is processed immediately).
 - Fast destination lookup is performed (no recursion).
 - The encapsulation uses a pregenerated Layer 2 header that contains the destination and Layer 2 source (MAC) address. (No ARP request or ARP cache lookup is necessary.)

Whenever a router receives a packet that should be fast-switched, but the destination is not in the switching cache, the packet is process-switched. A full routing table lookup is performed, and an entry in the fast-switching cache is created to ensure that the subsequent packets for the same destination prefix will be fast-switched.

CEF Switching Example

This topic explains the sequence of events that occurs when CEF switching is used for destinations that are learned through BGP.

Cisco Express Forwarding

Cisco IOS XR Software:

- Label switching on a Cisco router requires that Cisco Express Forwarding be enabled.
- Cisco Express Forwarding is mandatory for Cisco IOS XR software, and it does not need to be enabled explicitly.
- Cisco Express Forwarding offers these benefits:
 - Improved performance
 - Scalability
 - Resilience

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01--1.41

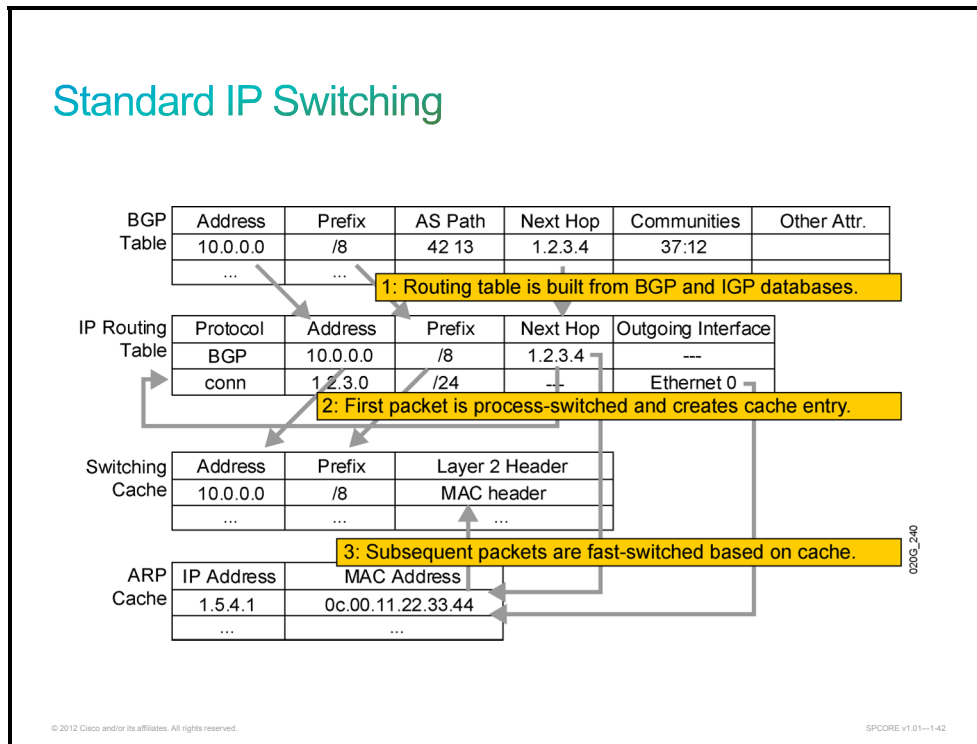
Cisco Express Forwarding uses a different architecture from process switching or any other cache-based switching mechanism. Cisco Express Forwarding uses a complete IP switching table, the FIB table, which holds the same information as the IP routing table. The generation of entries in the FIB table is not packet-triggered but change-triggered. When something changes in the IP routing table, the change is also reflected in the FIB table.

Because the FIB table contains the complete IP switching table, the router can make definitive decisions based on the information in it. Whenever a router receives a packet that should be switched with Cisco Express Forwarding, but the destination is not in the FIB, the packet is dropped.

The FIB table is also different from other fast-switching caches in that it does not contain information about the outgoing interface and the corresponding Layer 2 header. That information is stored in a separate table, the adjacency table. The adjacency table is similar to a copy of the ARP cache, but instead of holding only the destination MAC address, it holds the Layer 2 header.

CEF in IOS XE and IOS XR

This topic describes CEF on Cisco IOS XE and Cisco IOS XR platforms.



Cisco Express Forwarding is an advanced Layer 3 IP switching technology. It optimizes network performance and scalability for all kinds of networks: those that carry small amounts of traffic and those that carry large amounts of traffic in complex patterns, such as the Internet, and networks characterized by intensive web-based applications or interactive sessions.

Cisco Express Forwarding requires a software image that includes Cisco Express Forwarding and IP routing enabled on the device.

Cisco Express Forwarding is enabled by default on the Cisco ASR 1000 Series Aggregation Services Routers.

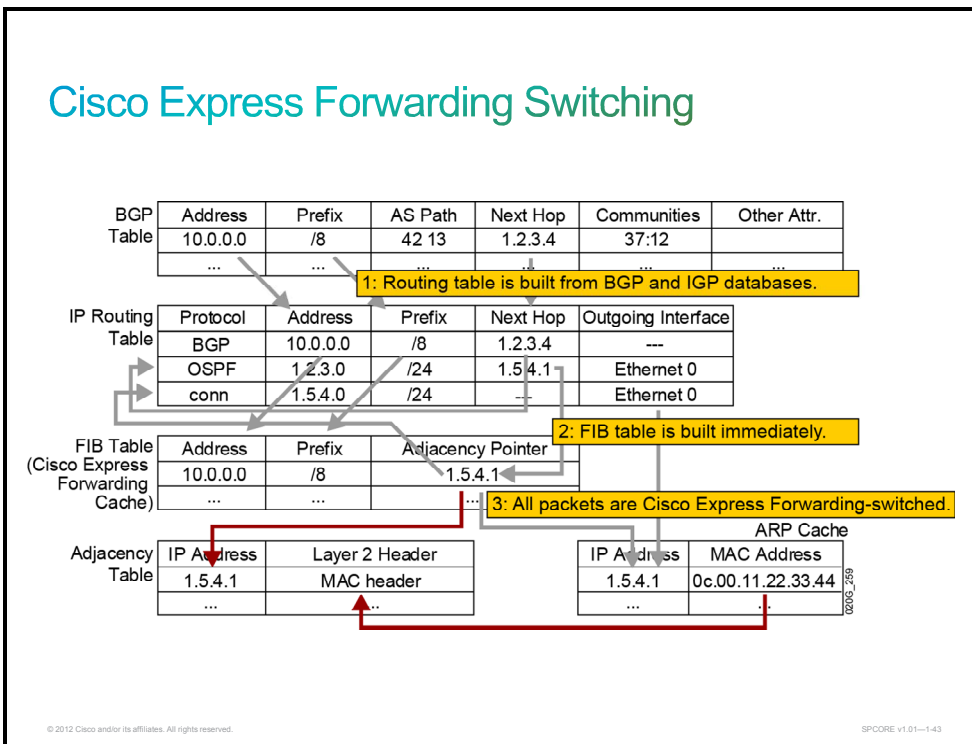
To find out if Cisco Express Forwarding is enabled by default on your platform, enter the **show ip cef** command. If Cisco Express Forwarding is enabled, you receive output that looks like this:

```
Router# show ip cef
Prefix                Next Hop                Interface
[...]
10.2.61.8/24          192.168.100.1           FastEthernet1/0/0
                      192.168.101.1           FastEthernet2/1/0
[...]
```

If Cisco Express Forwarding is not enabled on your platform, the output for the **show ip cef** command looks like this:

```
Router# show ip cef
%CEF not running
```

If Cisco Express Forwarding is not enabled on your platform, use the **ip cef** command to enable Cisco Express Forwarding or the **ip cef distributed** command to enable distributed Cisco Express Forwarding.



Cisco Express Forwarding offers the following benefits:

- **Improved performance:** Cisco Express Forwarding is less CPU-intensive than fast-switching route caching. More CPU processing power can be dedicated to Layer 3 services, such as quality of service (QoS) and encryption.
- **Scalability:** Cisco Express Forwarding offers full switching capacity at each modular services card (MSC) on the Cisco CRS routers.
- **Resilience:** Cisco Express Forwarding offers an unprecedented level of switching consistency and stability in large dynamic networks. In dynamic networks, fast-switched cache entries are frequently invalidated due to routing changes. These changes can cause traffic to be process-switched, using the routing table, rather than fast-switched using the route cache. Because the FIB lookup table contains all known routes that exist in the routing table, it eliminates route cache maintenance and the fast-switch or process-switch forwarding scenario. Cisco Express Forwarding can switch traffic more efficiently than typical demand caching schemes.

The prerequisite required to implement MPLS forwarding is installed composite mini-image and the MPLS package, or a full composite image. Label switching on a Cisco router requires that Cisco Express Forwarding be enabled. Cisco Express Forwarding is mandatory for Cisco IOS XR Software, and it does not need to be enabled explicitly.

Cisco IOS XR Software Cisco Express Forwarding always operates in Cisco Express Forwarding mode with two distinct components: a FIB database and an adjacency table, a protocol-independent adjacency information base (AIB).

Cisco Express Forwarding is a primary IP packet-forwarding database for Cisco IOS XR Software. Cisco Express Forwarding is responsible for these functions:

- Software switching path
- Maintaining forwarding table and adjacency tables (which are maintained by the AIB) for software and hardware forwarding engines

These Cisco Express Forwarding tables are maintained in Cisco IOS XR Software:

- IPv4 Cisco Express Forwarding database
- IPv6 Cisco Express Forwarding database
- MPLS label forwarding database (LFD)
- Multicast forwarding table (MFD)

Monitoring IPv4 Cisco Express Forwarding

This topic describes the show commands used to monitor CEF operations.

Monitoring IPv4 Cisco Express Forwarding

```
RP/0/RSP0/CPU0:PE1# show cef ipv4 192.168.178.0/24 detail
Mon Oct 24 07:05:52.465 UTC
192.168.178.0/24, version 1, attached, connected, internal 0xc0000c1 (ptr 0xad95
8254) [1], 0x0 (0xacf50cb0), 0x0 (0x0)
Updated Oct 20 21:36:09.871
remote adjacency to GigabitEthernet0/0/0/1
Prefix Len 24, traffic index 0, precedence routine (0)
gateway array (0xaccf8560) reference count 1, flags 0x0, source rib (4),
[2 type 3 flags 0x10101 (0xace24758) ext 0x0 (0x0)]
LW-LDI[type=3, refc=1, ptr=0xacf50cb0, sh-ldi=0xace24758]
via GigabitEthernet0/0/0/1, 4 dependencies, weight 0, class 0 [flags 0x8]
path-idx 0
remote adjacency

Load distribution: 0 (refcount 2)

Hash OK Interface Address
0 Y GigabitEthernet0/0/0/1 remote
```

To display the IPv4 Cisco Express Forwarding table, use the **show cef ipv4** command in EXEC mode on Cisco IOS XR Software.

```
RP/0/RSP0/CPU0:PE7# show cef ipv4
Mon Oct 24 07:08:01.177 UTC
```

Prefix	Next Hop	Interface
0.0.0.0/0	drop	default handler
0.0.0.0/32	broadcast	
10.7.1.1/32	receive	Loopback0
10.8.1.0/24	attached	GigabitEthernet0/0/0/1
192.168.178.0/24	attached	GigabitEthernet0/0/0/1
192.168.178.0/32	broadcast	GigabitEthernet0/0/0/1
192.168.178.70/32	receive	GigabitEthernet0/0/0/1
192.168.178.255/32	broadcast	GigabitEthernet0/0/0/1
224.0.0.0/4	point2point	
224.0.0.0/24	receive	
255.255.255.255/32	broadcast	

To display unresolved entries in the FIB table or to display a summary of the FIB, use this form of the **show cef ipv4** EXEC command: **show cef ipv4 [unresolved | summary]**.

To display specific entries in the FIB table based on IP address information, use this form of the **show cef ipv4** command in EXEC mode: **show cef ipv4 [network [mask [longer-prefix]]] [detail]**.

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- MPLS uses LDP to exchange labels.
- UDP multicast is used to discover adjacent LDP neighbors, while TCP is used to establish a session.
- LDP link hello message contains destination IP address, destination port and the actual hello message.
- LDP session negotiation is a three-step process.
- An MPLS-enabled router can be configured to send a directed LDP hello message as a unicast UDP packet that is specifically addressed to the nonadjacent router.
- LDP session protection lets you configure LDP to automatically protect sessions with all or a given set of peers.
- LDP graceful restart provides a control plane mechanism to ensure high availability and allows detection and recovery from failure condition.
- MPLS uses two forwarding structures which have to be populated using routing protocol and LDP.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.48

Summary (Cont.)

- A label-switched path (LSP) is a sequence of LSRs that forwards labeled packets for a particular Forwarding Equivalence Class (FEC).
- Labels are generated locally and then advertised to adjacent routers.
- PHP optimizes MPLS performance (one less LFIB lookup).
- MPLS is fully functional when the routing protocol and LDP have populated all the tables. Such a state is called the steady state.
- You can configure LDP to perform outbound filtering for local label advertisement for one or more prefixes to one or more LDP peers.
- Route summarization in an MPLS-enabled networks breaks LSP into two paths.
- The TTL functionality in MPLS is equivalent to that of traditional IP forwarding.
- TTL propagation can be disabled to hide the core routers from the end users.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.48

Summary (Cont.)

- Although it takes longer for LDP to exchange labels (compared with an IGP), a network can use the FIB table in the meantime.
- Link recovery requires that an LDP session be established (or reestablished), which adds to the convergence time of LDP.
- The Cisco IOS platform supports three IP switching mechanisms: process switching, fast switching, and CEF.
- In standard IP switching, the first packet that arrives is process switched and all subsequent packets are fast switched.
- In CEF switching, the FIB table is built in advance before a packet for a destination is received.
- On IOS and IOS XE, CEF is required for MPLS.
- CEF is enabled on Cisco IOS XR and cannot be disabled.
- To monitor CEF, you can use various **show** commands.

Implementing MPLS in the Service Provider Core

Overview

This lesson describes how to configure Multiprotocol Label Switching (MPLS) on various Cisco platforms. The essential configuration tasks and commands, including their correct syntax, are discussed. Also addressed are advanced configurations such as the label-switching maximum transmission unit (MTU), IP Time-to-Live (TTL) propagation, Label Distribution Protocol (LDP) session protection, LDP graceful restart, LDP interior gateway protocol (IGP) synchronization, LDP autoconfiguration, and conditional label distribution. The lesson also describes the procedures for monitoring MPLS by using syntax and parameter descriptions, interfaces, neighbor nodes, label information base (LIB), and label forwarding information base (LFIB) tables. Also outlined are the usage guidelines for the commands.

The lesson concludes with a look at some of the common issues that arise in MPLS networks. For each issue discussed, there is a recommended troubleshooting procedure to resolve the issue.

Objectives

Upon completing this lesson, you will be able to configure, monitor, and troubleshoot MPLS on Cisco IOS Software, Cisco IOS XE Software, and Cisco IOS XR Software platforms. You will be able to meet these objectives:

- Describe MPLS configuration difference in Cisco IOS XR vs Cisco IOS/IOS XE
- Describe mandatory and optional MPLS configuration tasks
- Explain a basic MPLS configuration
- Describe the MTU requirements on a label switching router interface
- Explain the configuration used to increase the MPLS MTU size on a label switching router interface
- Explain IP TTL Propagation
- Explain the configuration used to disable IP TTL Propagation
- Explain LDP Session Protection Configuration
- Explain LDP Graceful Restart and NSR Configuration
- Explain LDP IGP Synchronization Configuration

- Explain how to enable LDP Autoconfiguration
- Explain Label Advertisement Control Configuration
- Describe the show commands used to monitor MPLS operations
- Describe the MPLS and LDP debug commands
- Describe the Classic Ping and Traceroute operations
- Describe the MPLS Ping and Traceroute operations
- Describe how to troubleshoot common MPLS issues

MPLS Configuraton on Cisco IOS XR vs Cisco IOS/IOS XE

This topic describes MPLS configuration difference in Cisco IOS XR vs Cisco IOS/IOS XE.

MPLS Configuration

- Cisco **IOS XR Software**:
 - MPLS forwarding is enabled when you enable LDP on an interface under MPLS LDP configuration mode.
 - Cisco Express Forwarding is mandatory for Cisco IOS XR Software, and it does not need to be enabled explicitly.
- Cisco **IOS and IOS XE Software**:
 - MPLS forwarding is enabled when you enable MPLS on an interface under interface configuration mode.
 - Cisco Express Forwarding is enabled by default on most Cisco IOS and Cisco IOS XE platforms, including the Cisco ASR 1000 Series Aggregation Services Routers.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-4

Basic configuration of MPLS is simple. On Cisco IOS XR platforms, MPLS is enabled by enabling LDP on each interface under MPLS LDP configuration mode. On Cisco IOS and IOS XE platforms, MPLS is enabled on each interface under interface configuration mode.

MPLS Configuration Tasks

This topic describes mandatory and optional MPLS configuration tasks.

MPLS Configuration Tasks

Mandatory:

- Enable LDP on an interface under MPLS LDP configuration mode (Cisco IOS XR Software).
- Enable MPLS on an interface under interface configuration mode (Cisco IOS and Cisco IOS XE Software).

Optional:

- Configure the MPLS Router ID.
- Configure MTU size for labeled packets.
- Configure IP TTL propagation.
- Configure conditional label advertising.
- Configure access lists to prevent customers from running LDP with PE routers.

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01--1.6

To enable MPLS on a router that is running Cisco IOS XR Software, enable LDP on an interface under MPLS LDP configuration mode. To enable MPLS on a router that is running Cisco IOS Software or Cisco IOS XE Software, enable MPLS on an interface under interface configuration mode.

Optionally, the maximum size of labeled packets may be changed. A stable LDP router ID is required at either end of the link to ensure that the link discovery (and session setup) is successful. If you do not manually assign the LDP router ID on the Cisco IOS XR routers, the Cisco IOS XR routers will default to use the global router ID as the LDP router ID. Global router ID configuration is only available on Cisco IOS XR (not available on Cisco IOS and IOS XE Software).

You can override the global **router-id** command in Cisco IOS XR by further configuring a **router-id** command within a given protocol. However, configuring different router IDs per protocol makes router management more complicated.

By default, the TTL field is copied from the IP header and placed in the MPLS label TTL field when a packet enters an MPLS network. To prevent core routers from responding with (Internet Control Message Protocol [ICMP]) TTL exceeded messages, disable TTL propagation. If TTL propagation is disabled, the value in the TTL field of the MPLS label is set to 255.

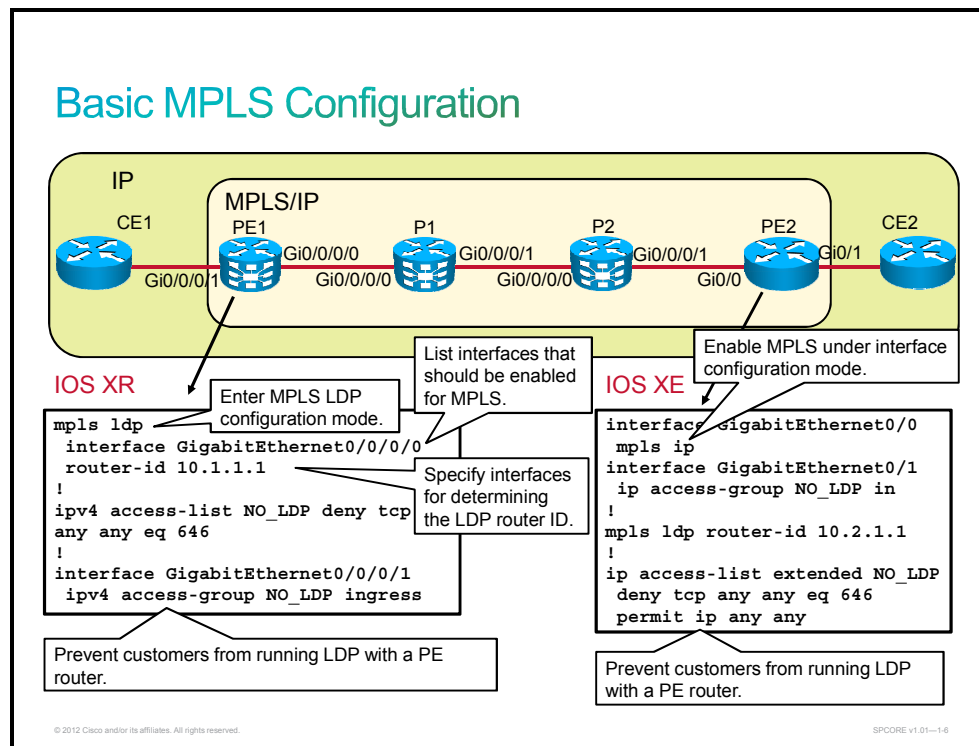
Note Ensure that TTL propagation is either enabled in all routers or disabled in all routers. If TTL is enabled in some routers and disabled in others, the result may be that a packet that is leaving the MPLS domain will have a larger TTL value than when it entered.

By default, a router will generate and propagate labels for all networks that it has in the routing table. If label switching is required for only a limited number of networks (for example, only for router loopback addresses), configure conditional label advertising.

To prevent customers from running LDP with PE routers, configure access lists that will block an LDP well-known TCP port.

Basic MPLS Configuration

This topic explains a basic MPLS configuration.



To enable MPLS on the Cisco IOS XR router, first enter MPLS LDP configuration mode using the **mpls ldp** command. Then specify the interfaces that should be enabled for MPLS by using the **interface** command. In the example, the MPLS for router PE1 is enabled on the GigabitEthernet0/0/0/0 interface. The configuration includes an access control list (ACL) that denies any attempt to establish an LDP session from an interface that is not enabled for MPLS. In the example shown in the figure, router PE1 has the NO_LDP access list applied to interface GigabitEthernet0/0/0/1, which is not enabled for MPLS.

Note Enable MPLS on all core interfaces in your network. On routers P1 and P2, both interfaces GigabitEthernet0/0/0/0 and GigabitEthernet0/0/0/1 should be enabled for MPLS.

To enable MPLS on Cisco IOS and IOS XE routers, first enter interface configuration mode for a desired interface. Then enable MPLS, using the **mpls ip** command. In the example, MPLS is enabled for router PE2 on the GigabitEthernet0/0 interface. The configuration includes an ACL that denies any attempt to establish an LDP session from an interface that is not enabled for MPLS. In the example, router PE2 has the NO_LDP access list applied to the interface GigabitEthernet0/1, which is not enabled for MPLS.

A stable router ID is required at either end of the link to ensure the link discovery (and session setup) is successful. In the example, routers PE1 and PE2 have the LDP router ID set to the IP address of interface loopback 0.

MTU Requirements

This topic describes the MTU requirements on a label switching router interface.

Label Switching MTU

- Label switching increases the maximum MTU requirements on an interface because of the additional label header.
- Interface MTU is automatically increased on WAN interfaces; IP MTU is automatically decreased on LAN interfaces.
- Label-switching MTU can be increased on LAN interfaces (resulting in jumbo frames) to prevent IP fragmentation.
- Jumbo frames must be enabled on the switch.
- Jumbo frames are not supported by all LAN switches.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—1.7

There are three different MTU values:

- The interface MTU is used to determine the maximum size of any packets that can be sent on the interface. In Cisco IOS XR Software, this is the L2 MTU. In IOS and IOS XE Software, this is the Layer 3 payload size.
- The IP MTU is used to determine whether a non-labeled IP packet that is forwarded through an interface has to be fragmented (the IP MTU has no impact on labeled IP packets).
- The MPLS MTU determines the maximum size of a labeled IP packet (MPLS shim header + IP payload size). If the overall length of the labeled packet (including the shim header) is greater than the MPLS MTU, the packet is fragmented. The default MPLS MTU is the MTU that is configured for the interface.

Label switching increases the maximum MTU requirements on an interface, because of the additional label header. The interface MTU is automatically increased on WAN interfaces while the IP MTU is automatically decreased on LAN interfaces.

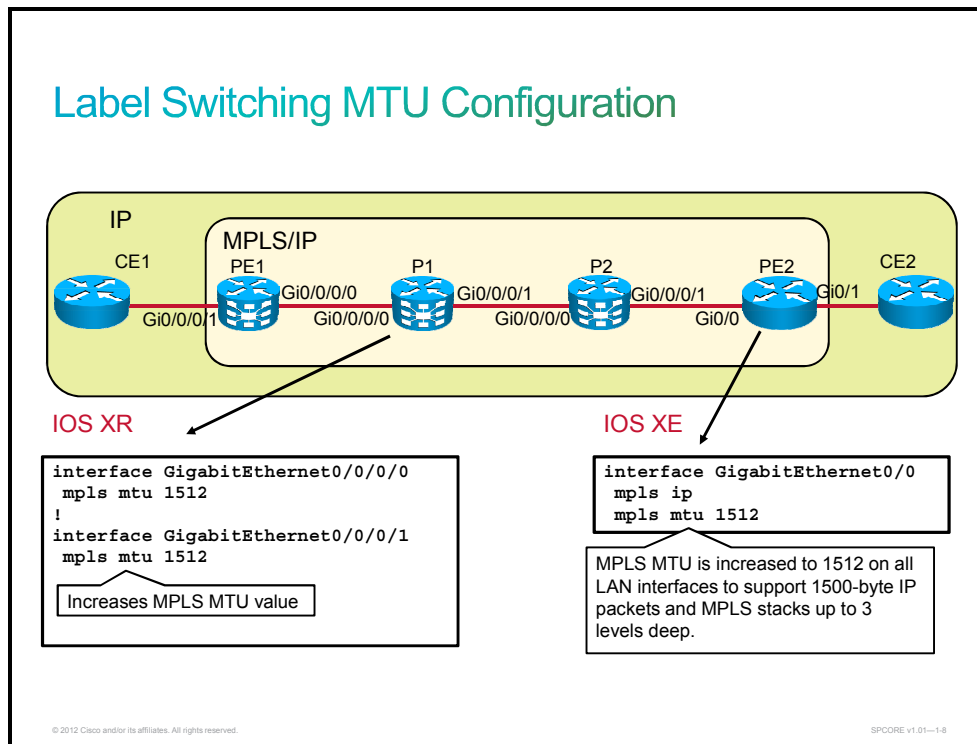
One way of preventing labeled packets from exceeding the maximum size (and being fragmented as a result) is to increase the MTU size of labeled packets for all segments in the label-switched path (LSP) tunnel. This problem will typically occur on LAN switches, where it is more likely that a device does not support oversized packets (also called jumbo frames, giants, or baby giants). Some devices support jumbo frames, and some need to be configured to support them.

The interface MTU size (and therefore also the MPLS MTU) is increased automatically on WAN interfaces but the MPLS MTU needs to be increased manually on LAN interfaces.

The MPLS MTU size has to be increased on all LSRs that are attached to a LAN segment. Additionally, the LAN switches that are used to implement switched LAN segments need to be configured to support jumbo frames.

MPLS MTU Configuration

This topic explains the configuration used to increase the MPLS MTU size on a label switching router interface.



The figure shows a label switching MTU configuration on LAN interfaces for routers P1 and PE2. MPLS MTU is increased to 1512 on the Ethernet interfaces of router P1 to support 1500-byte IP packets and MPLS stacks up to 3 levels deep (3 times 4-byte label).

To configure the maximum packet size or MTU size on an MPLS interface (for Cisco IOS XR Software and Cisco IOS and IOS XE Software), use the **mpls mtu** command in interface configuration mode. To disable this feature, use the **no** form of this command.

IP TTL Propagation

This topic explains IP TTL Propagation.

IP TTL Propagation

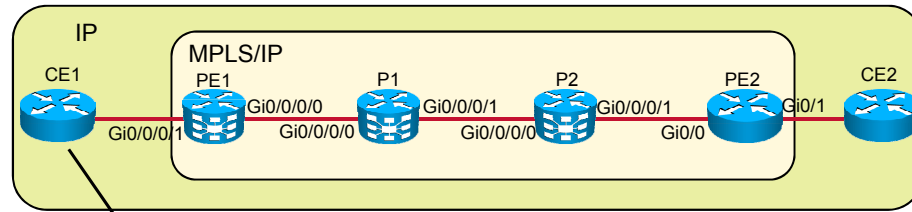
- By default, IP TTL is copied into the MPLS label at label imposition, and the MPLS label TTL is copied (back) into the IP TTL at label removal.
- IP TTL and label TTL propagation can be disabled.
 - TTL value of 255 is inserted into the label header.
- The TTL propagation must be disabled on ingress and egress edge LSRs.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-9

Remember that by default, IP TTL is copied into the MPLS label at label imposition, and the MPLS label TTL is copied (back) into the IP TTL at label removal. IP TTL and label TTL propagation can be disabled if it is desired to hide the core routers from the traceroute output; a TTL value of 255 is inserted in the label header. The TTL propagation must be disabled, at least on ingress and egress edge LSRs, but it is advisable that all routers have TTL propagation enabled, or all disabled.

IP TTL Propagation



```
CE1# traceroute CE2

Type escape sequence to abort.
Tracing the route to CE2

 1 PE1 4 msec 0 msec 0 msec
 2 P1  0 msec 4 msec 0 msec
 3 P2  0 msec 4 msec 0 msec
 4 PE2 0 msec 0 msec 0 msec
 5 CE2 4 msec *  0 msec
```

The **traceroute** command, executed on a customer router, displays all routers in the path.

The figure illustrates typical traceroute behavior in an MPLS network. Because the label header of a labeled packet carries the TTL value from the original IP packet, the routers in the path can drop packets when the TTL is exceeded. Traceroute will therefore show all the routers in the path. This is the default behavior.

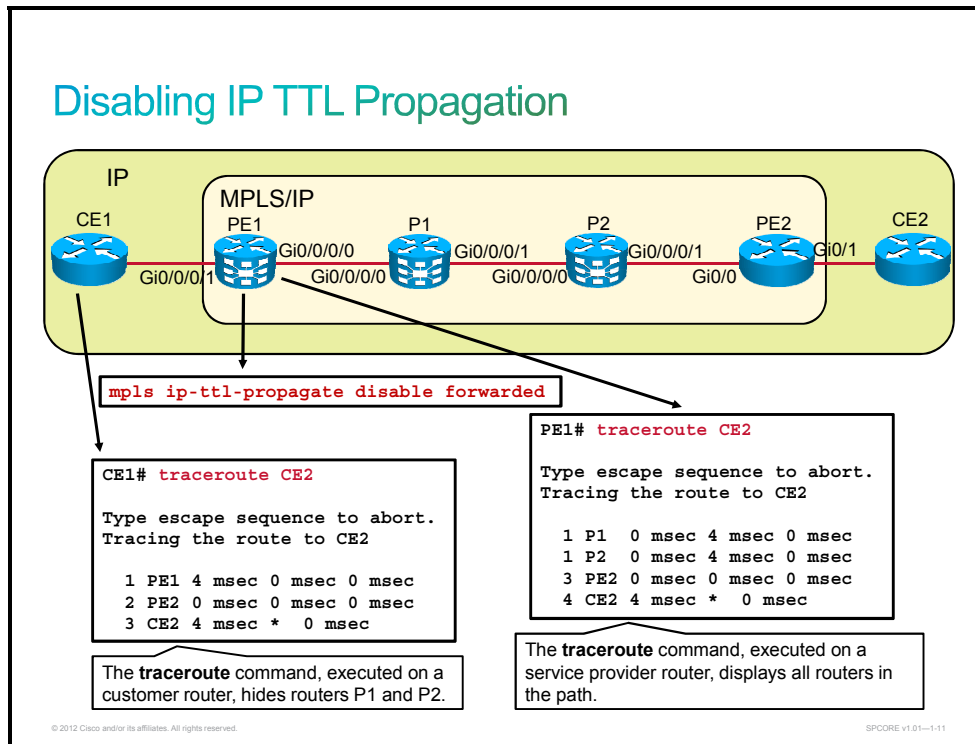
In the example, router CE1 is executing a **traceroute** command that results in this behavior. The steps for this process are as follows:

- Step 1** The first packet is an IP packet with TTL = 1. Router PE1 decreases the TTL and drops the packet because it reaches 0. An ICMP TTL exceeded message is sent to the source.
- Step 2** The second packet sent is an IP packet with TTL = 2. Router PE1 decreases the TTL, labels the packet (the TTL from the IP header is copied into the MPLS label TTL field), and forwards the packet to router P1.
- Step 3** Router P1 decreases the MPLS TTL value, drops the packet, and sends an ICMP TTL exceeded message to the source.
- Step 4** Processing for the third packet is similar, with TTL = 3. Router P2 sends an ICMP TTL exceeded message to the source.
- Step 5** The fourth packet (TTL = 4) experiences processing that is similar to the previous packets, except that router PE2 is dropping the packet based on the TTL in the IP header. Router P2, because of penultimate hop popping (PHP), previously removed the label, and the TTL was copied back to the IP header.

The fifth packet (TTL = 5) reaches the final destination, where the TTL of the IP packet is examined.

Disabling IP TTL Propagation

This topic explains the configuration used to disable IP TTL Propagation.



If TTL propagation is disabled, the TTL value is not copied into the label header. Instead, the label TTL field is set to 255. The probable result is that the TTL field in the label header will not decrease to 0 for any router inside the MPLS domain (unless there is a forwarding loop inside the MPLS network).

If the **traceroute** command is used, ICMP replies are received only from those routers that see the real TTL that is stored in the IP header.

Typically, a service provider likes to hide the backbone network from outside users, but allow inside traceroute to work for easier troubleshooting of the network.

This goal can be achieved by disabling TTL propagation for forwarded packets only, as described here:

- If a packet originates in the router, the real TTL value is copied into the label TTL.
- If the packet is received through an interface, the TTL field in a label is assigned a value of 255.

The result is that someone using traceroute on a provider router will see all of the backbone routers. Customers will see only edge routers.

Use the **mpls ip-ttl-propagate** (Cisco IOS XR Software) or **mpls ip propagate-ttl** (Cisco IOS XE Software) global configuration command to control generation of the TTL field in the label when the label is first added to the IP packet. By default, this command is enabled, which means that the TTL field is copied from the IP header and inserted into the MPLS label. This aspect allows a **traceroute** command to show all of the hops in the network.

To use a fixed TTL value (255) for the first label of the IP packet, use the **no** form of the **mpls ip propagate-ttl** command on Cisco IOS XE Software. To use a fixed TTL value (255) for the first label of the IP packet, use the **mpls ip-ttl-propagate disable** command on Cisco IOS XR Software. This action hides the structure of the MPLS network from a **traceroute** command.

Specify the types of packets to be hidden by using the **forwarded** and **local** arguments. Specifying the **forwarded** parameter allows the structure of the MPLS network to be hidden from customers, but not from the provider. Selective IP TTL propagation hides the provider network from the customer but still allows troubleshooting.

Cisco IOS/IOS-XE:

```
PE2(config)# no mpls ip propagate-ttl ?
    forwarded  Propagate IP TTL for forwarded traffic
    local      Propagate IP TTL for locally originated traffic
    <cr>
```

Cisco IOS-XR configuration:

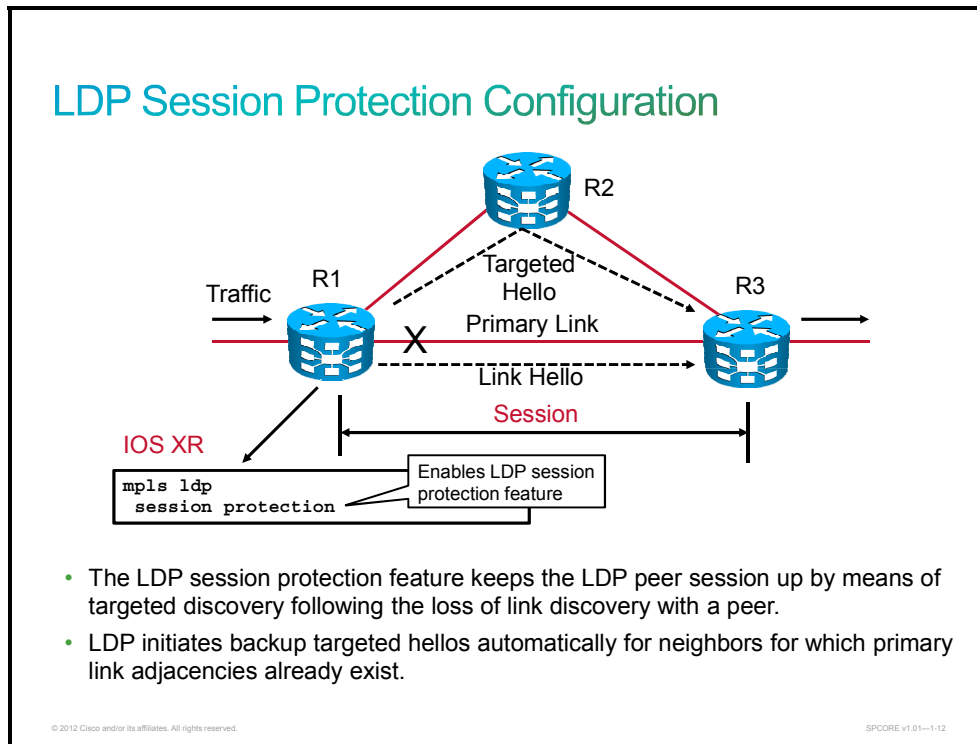
```
RP/0/RSP0/CPU0:PE1(config)# mpls ip-ttl-propagate disable
    forwarded  Disable IP TTL propagation for only forwarded MPLS
packets
    local      Disable IP TTL propagation for only locally generated
MPLS packets
    <cr>
```

mpls ip propagate-ttl (mpls ip-ttl-propagate) Syntax Description

Parameter	Description
forwarded	(Optional) Hides the structure of the MPLS network from a traceroute command only for forwarded packets; prevents the traceroute command from showing the hops for forwarded packets
local	(Optional) Hides the structure of the MPLS network from a traceroute command only for local packets; prevents the traceroute command from showing the hops only for local packets

LDP Session Protection Configuration

This topic explains LDP Session Protection Configuration.



LDP session protection lets you configure LDP to automatically protect sessions with all or a given set of peers (as specified by the peer ACL). When it is configured, LDP initiates backup targeted hellos automatically for neighbors for which primary link adjacencies already exist. These backup targeted hellos maintain LDP sessions when primary link adjacencies go down.

To enable the LDP session protection feature for keeping the LDP peer session up by means of targeted discovery following the loss of link discovery with a peer, use the **session protection** command in MPLS LDP configuration mode in Cisco IOS XR Software. To return to the default behavior, use the **no** form of this command:

session protection [*duration seconds* | *infinite*] [*for peer-acl*]

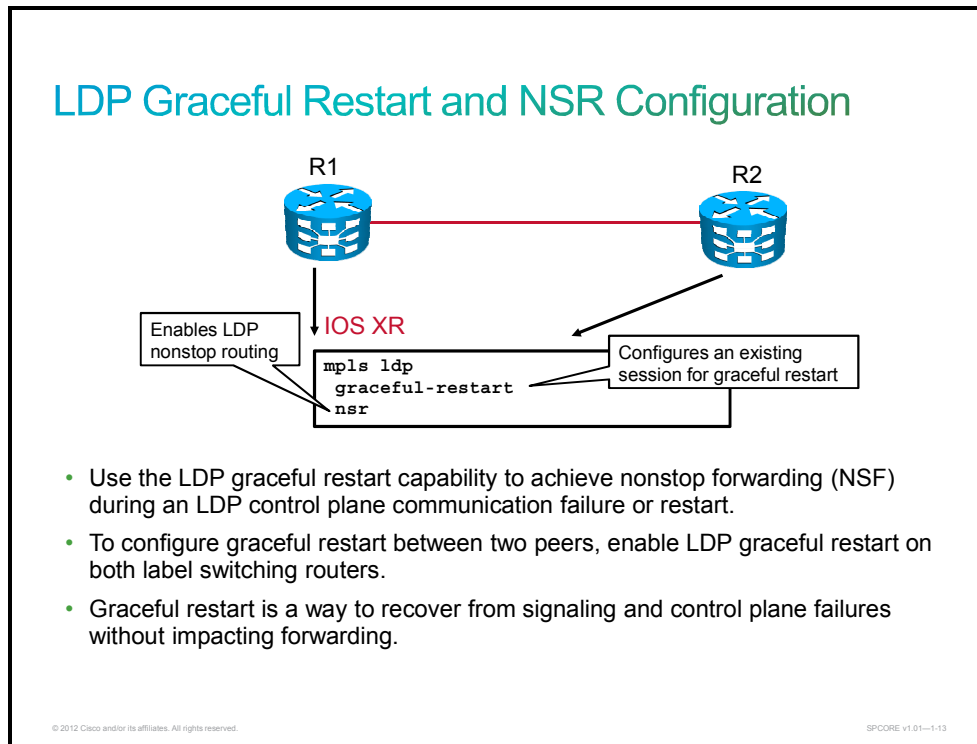
no session protection

By default, session protection is disabled. When it is enabled without peer ACL and duration, session protection is provided for all LDP peers, and continues for 24 hours after a link discovery loss. This LDP session protection feature allows you to enable the automatic setup of targeted hello adjacencies with all or a set of peers, and specify the duration for which session needs to be maintained using targeted hellos after loss of link discovery. LDP supports only IPv4 standard access lists.

On Cisco IOS and IOS XE Software, the similar command to enable LDP session protection is the **mpls ldp session protection** global configuration command.

LDP Graceful Restart and NSR Configuration

This topic explains LDP Graceful Restart and NSR Configuration.



LDP graceful restart provides a control plane mechanism to ensure high availability, and allows detection and recovery from failure conditions while preserving nonstop forwarding (NSF) services. Graceful restart is a way to recover from signaling and control plane failures without impacting forwarding.

Use the LDP graceful restart capability to achieve nonstop forwarding during an LDP control plane communication failure or restart. To configure graceful restart between two peers, enable LDP graceful restart on both label switching routers (LSRs).

To configure graceful restart, use the **graceful-restart** command in MPLS LDP configuration mode. To return to the default behavior, use the **no** form of this command.

graceful-restart [**reconnect-timeout** *seconds* | **forwarding-state-holdtime** *seconds*]

no graceful-restart [**reconnect-timeout** | **forwarding-state-holdtime**]

graceful-restart Syntax Description

Parameter	Description
forwarding-state-holdtime <i>seconds</i>	(Optional) Time that the local forwarding state is preserved (without being reclaimed) after the local LDP control plane restarts. The range is 60 to 600 seconds. The default is 180 seconds.
reconnect-timeout <i>seconds</i>	(Optional) Time that the local LDP sends to its graceful "restartable" peer, indicating how long its neighbor should wait for reconnection if there is an LDP session failure. The range is 60 to 300 seconds. The default is 120 seconds.

LDP NSR functionality makes failures, such as route processor (RP) or distributed route processor (DRP) failover, invisible to routing peers with minimal to no disruption of convergence performance. To enable LDP NSR on Cisco IOS XR Software use the **nsr** command in **mpls ldp** configuration mode.

When you enable MPLS LDP graceful restart on a router that peers with an MPLS LDP SSO/NSF enabled router, the SSO/NSF enabled router can maintain its forwarding state when the LDP session between them is interrupted. While the SSO/NSF enabled router recovers, the peer router forwards packets using stale information. This enables the SSO/NSF enabled router to become operational more quickly.

When an LDP graceful restart session is established and there is control plane failure, the peer LSR starts graceful restart procedures, initially keeps the forwarding state information pertaining to the restarting peer, and marks this state as stale. If the restarting peer does not reconnect within the reconnect timeout, the stale forwarding state is removed. If the restarting peer reconnects within the reconnect time period, it is provided recovery time to resynchronize with its peer. After this time, any unsynchronized state is removed.

The value of the forwarding state hold time keeps the forwarding plane state associated with the LDP control-plane in case of a control-plane restart or failure. If the control plane fails, the forwarding plane retains the LDP forwarding state for twice the forwarding state hold time. The value of the forwarding state hold time is also used to start the local LDP forwarding state hold timer after the LDP control plane restarts. When the LDP graceful restart sessions are renegotiated with its peers, the restarting LSR sends the remaining value of this timer as the recovery time to its peers. Upon local LDP restart with graceful restart enabled, LDP does not replay forwarding updates to MPLS forwarding until the forwarding state hold timer expires.

To display the status of the LDP graceful restart, use the **show mpls ldp graceful-restart** command in EXEC mode. You can also check to see if the router is configured for graceful restart with the **show mpls ldp neighbor brief** command in EXEC mode.

```
RP/0/RP0/CPU0:router# show mpls ldp neighbor brief
```

```
v
Peer                GR Up Time                Discovery Address
-----
3.3.3.3:0           Y  00:01:04                3            8
2.2.2.2:0           N  00:01:02                2            5
```

```
RP/0/RP0/CPU0:router# show mpls ldp graceful-restart
```

```
Forwarding State Hold timer : Not Running
```

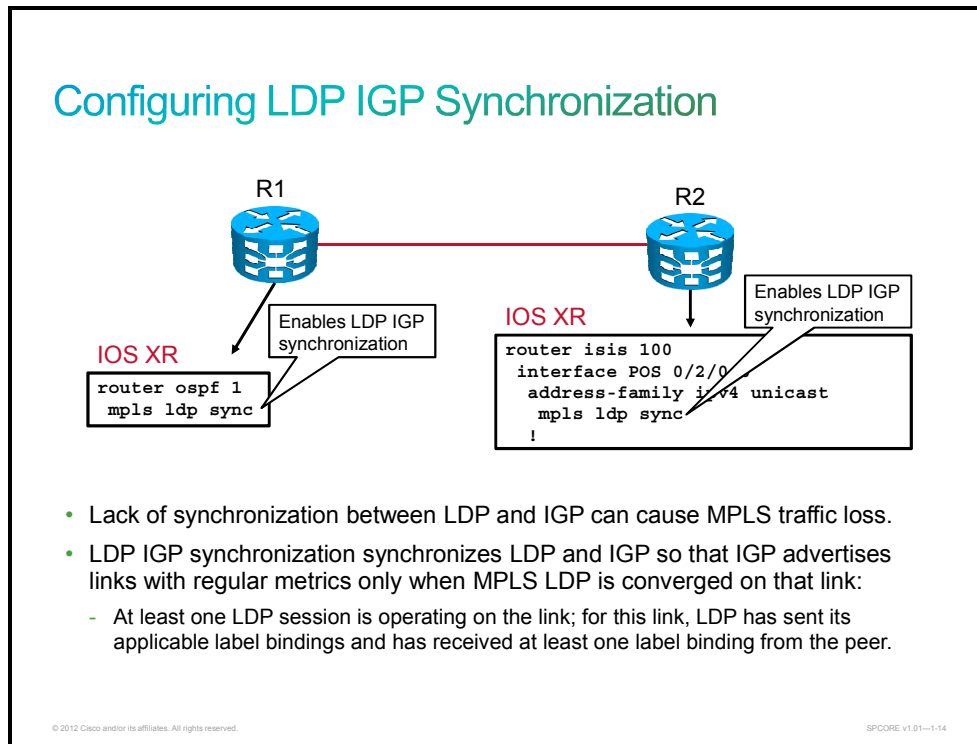
```
GR Neighbors                : 1
```

```
Neighbor ID      Up  Connect Count  Liveness Timer  Recovery Timer
-----
3.3.3.3          Y      1              -                -
```

On Cisco IOS and IOS XE Software, a similar command that will enable LDP graceful restart is the **mpls ldp graceful-restart** global configuration command.

LDP IGP Synchronization Configuration

This topic explains LDP IGP Synchronization Configuration.



Lack of synchronization between LDP and IGP can cause MPLS traffic loss. Upon link up, for example, IGP can advertise and use a link before LDP convergence has occurred; or, a link may continue to be used in the IGP after an LDP session goes down.

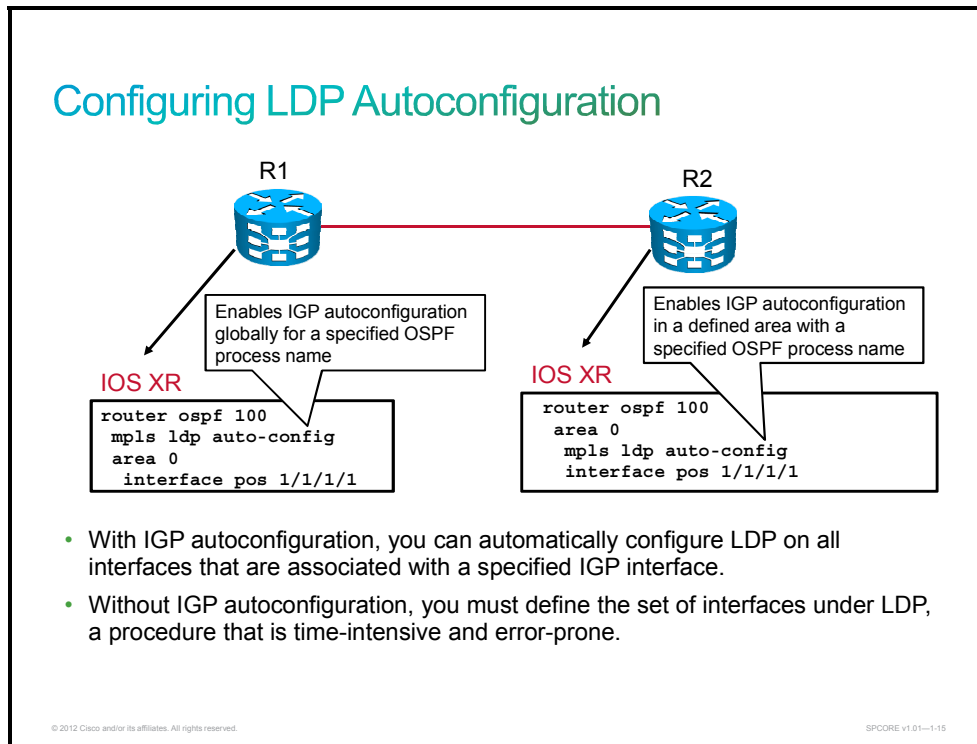
LDP IGP synchronization synchronizes LDP and IGP so that IGP advertises links with regular metrics only when MPLS LDP is converged on that link. LDP considers a link converged when at least one LDP session is operating on the link for which LDP has sent its applicable label bindings and has received at least one label binding from the peer. LDP communicates this information to IGP upon link up or session down events and IGP acts accordingly, depending on the synchronization state.

Normally, when LDP IGP synchronization is configured, LDP notifies IGP as soon as LDP is converged. When the delay timer is configured, this notification is delayed. Under certain circumstances, it might be required to delay declaration of resynchronization to a configurable interval. LDP provides a configuration option to delay synchronization for up to 60 seconds.

The LDP IGP synchronization feature is only supported for OSPF or IS-IS. To enable Label LDP IGP synchronization on Cisco IOS, IOS XE, and IOS XR Software, use the **`mpls ldp sync`** command in the appropriate mode (OSPF or IS-IS configuration mode). To disable LDP IGP synchronization, use the **`no`** form of this command.

LDP Autoconfiguration

This topic explains how to enable LDP Autoconfiguration.



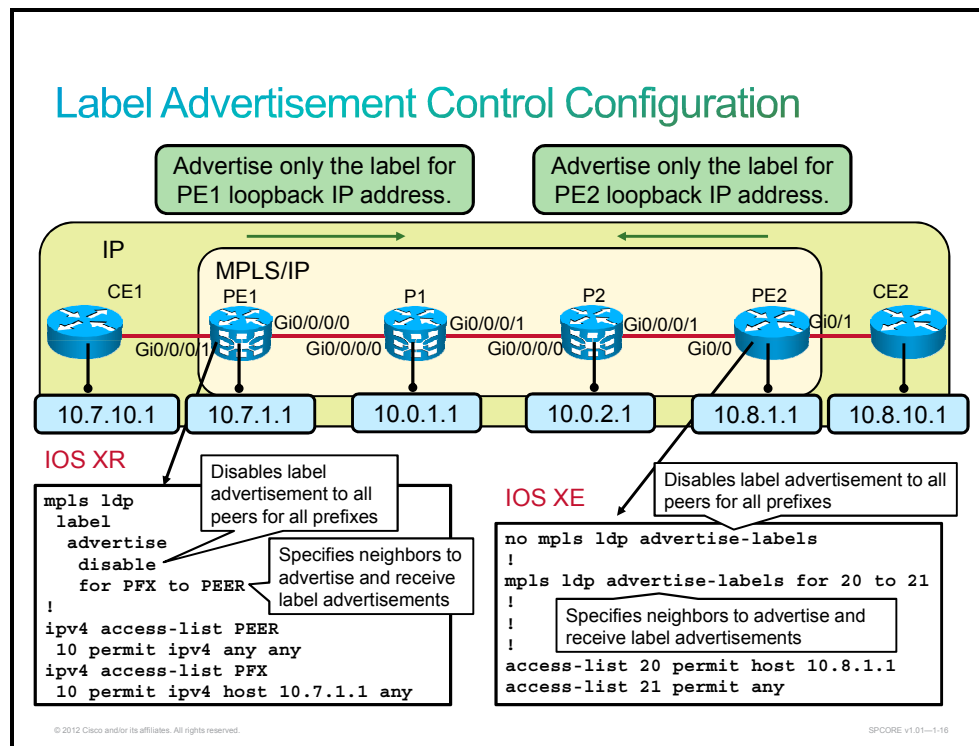
To enable LDP on many interfaces, IGP autoconfiguration allows you to automatically configure LDP on all interfaces that are associated with a specified OSPF or IS-IS interface. However, there must be one IGP (OSPF or IS-IS) that is set up to enable LDP autoconfiguration.

Without IGP autoconfiguration, you must define the set of interfaces under LDP, a procedure that is time-intensive and error-prone.

Similarly, on Cisco IOS and IOS XE Software, the `mpls ldp autoconfig` router OSPF or IS-IS configuration command is used to enable LDP autoconfiguration.

Label Advertisement Control Configuration

This topic explains Label Advertisement Control Configuration.



By default, LDP advertises labels for all the prefixes to all its neighbors. When this is not desirable (for scalability and security reasons), you can configure LDP to perform outbound filtering for local label advertisement, for one or more prefixes, to one or more peers. This feature is known as LDP outbound label filtering, or local label advertisement control.

The example describes where conditional label advertising can be used. The existing network still performs normal IP routing, but the MPLS LSP tunnel between the loopback interfaces of the LSR routers is needed to enable MPLS VPN functionality. Using one contiguous block of IP addresses for loopbacks on the provider edge (PE), routers can simplify the configuration of conditional advertising.

In the figure, the PE1 router (running Cisco IOS XR Software) should advertise only the label of the loopback prefix for PE1 (10.7.1.1/32), and not the loopback prefix of CE1 (10.7.10.1/32). In the same manner, the PE2 router (running Cisco IOS XE Software) should advertise only the label for the loopback prefix of PE2 (10.8.1.1/32), and not the loopback prefix of CE1 (10.8.10.1/32).

To control the advertisement of local labels on Cisco IOS XR Software, use the **label advertise** command in MPLS LDP configuration mode. To return to the default behavior, use the **no** form of this command:

label advertise {**disable** | **for prefix-acl** [**to peer-acl**] | **interface interface**}

no label advertise {**disable** | **for prefix-acl** [**to peer-acl**] | **interface interface**}

Example

```
mpls ldp
label advertise disable
label advertise for PFX to PEER
```

Syntax Description

Parameter	Description
for <i>prefix-access-list</i>	(Optional) This parameter specifies which destinations should have their labels advertised.
to <i>peer-access-list</i>	(Optional) This parameter specifies which LSR neighbors should receive label advertisements. An LSR is identified by its router ID, which consists of the first 4 bytes of its 6-byte LDP identifier.
interface <i>interface</i>	(Optional) This parameter specifies an interface for label allocation and advertisement of its interface IP address.

This command is used to control which labels are advertised to which LDP neighbors. On Cisco IOS and IOS XE Software, use the **mpls ldp advertise-labels** command in global configuration mode. To prevent the distribution of locally assigned labels, use the **no** form of this command, as shown:

- **mpls ldp advertise-labels** [**for** *prefix-access-list* [**to** *peer-access-list*]]
- **no mpls ldp advertise-labels** [**for** *prefix-access-list* [**to** *peer-access-list*]]

The configuration in the figure for router PE1 disables label advertisement to all peers for all prefixes, except for prefix 10.7.1.1/32. The configuration in the figure for router PE2 disables label advertisement to all peers for all prefixes, except for prefix 10.8.1.1/32.

Label Advertisement Control Verification

LIB table of P1 **before** label advertisement control is configured:

```
RP/0/RSP0/CPU0:P1# show mpls ldp bindings
10.7.10.1/32, rev 85
  Local binding: label: 16021
  Remote bindings: (2 peers)
    Peer          Label
    -----
    10.0.2.1:0    16022
    10.7.1.1:0    16025
```

The label for the loopback prefix of CE1 is received on P1 from PE1.

LIB table of P1 **after** label advertisement control is configured:

```
RP/0/RSP0/CPU0:P1# show mpls ldp bindings
10.7.10.1/32, rev 85
  Local binding: label: 16021
  Remote bindings: (1 peer)
    Peer          Label
    -----
    10.0.2.1:0    16022
```

The label for the loopback prefix of CE1 is not received on P1 from PE1.

To verify content of the LIB table, use the **show mpls ldp bindings**. The output displays local labels for each destination network, as well the labels that are received from all LDP neighbors.

This output was taken from P1 before label advertisement control was configured. In the example, the local label for network 10.7.10.1/32 is 16021, the label received from 10.0.2.1 neighbor (P2) is 16022, and the label received from 10.7.1.1 neighbor (PE1) is 16025. Note that P1 received the label for network 10.7.10.1/32 from neighbor PE1, as highlighted in this output:

```
RP/0/RSP0/CPU0:P1# show mpls ldp bindings
10.7.1.1/32, rev 61
  Local binding: label: 16013
  Remote bindings: (3 peers)
    Peer          Label
    -----
    10.0.2.1:0    16013
    10.7.1.1:0    IMP-NULL

10.7.10.1/32, rev 85
  Local binding: label: 16021
  Remote bindings: (3 peers)
    Peer          Label
    -----
    10.0.2.1:0    16022
    10.7.1.1:0    16025
```

This output was taken from P1 after label advertisement control was configured. In the example, the local label for network 10.7.10.1/32 is 16021, and the label received from 10.0.2.1 neighbor (P2) is 16022. Note that P1 did not receive the label for network 10.7.10.1/32 from neighbor PE1.

```
RP/0/RSP0/CPU0:P1# show mpls ldp bindings
```

```
10.7.1.1/32, rev 61
```

```
Local binding: label: 16013
```

```
Remote bindings: (3 peers)
```

Peer	Label
-----	-----
10.0.2.1:0	16013
10.7.1.1:0	IMP-NULL

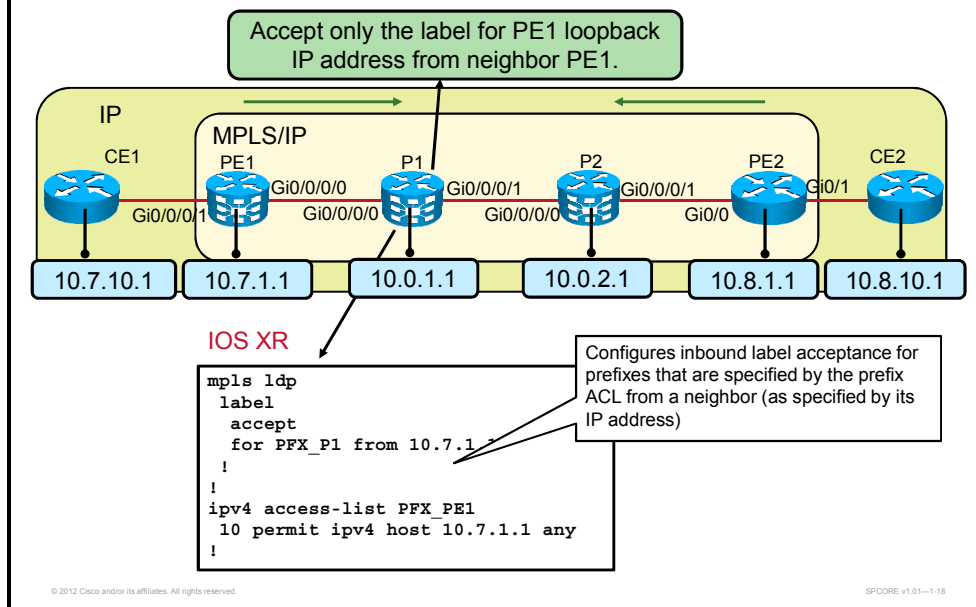
```
10.7.10.1/32, rev 85
```

```
Local binding: label: 16021
```

```
Remote bindings: (2 peers)
```

Peer	Label
-----	-----
10.0.2.1:0	16022

Label Acceptance Control Configuration



By default, LDP accepts labels (as remote bindings) for all prefixes from all peers. LDP operates in liberal label retention mode, which instructs LDP to keep remote bindings from all peers for a given prefix. For security reasons, or to conserve memory, you can override this behavior by configuring label binding acceptance for set of prefixes from a given peer.

The ability to filter remote bindings for a defined set of prefixes is also referred to as LDP inbound label filtering.

To control the receipt of labels (remote bindings) on Cisco IOS XR Software for a set of prefixes from a peer, use the **label accept** command in MPLS LDP configuration mode. To return to the default behavior, use the **no** form of this command.

label accept for *prefix-acl* from *A.B.C.D*

no label accept for *prefix-acl* from *A.B.C.D*

Example

```
mpls ldp
```

label accept for PFX_PE1 from 10.7.1.1 Syntax Description

Parameter	Description
for <i>prefix-acl</i>	Accepts and retains remote bindings for prefixes that are permitted by the prefix access list <i>prefix-acl</i> .
from <i>A.B.C.D</i>	Displays Peer IP address

The configuration in the figure for router P1 accepts only the label for the PE1 loopback IP address from neighbor PE1.

Monitor MPLS

This topic describes the **show** commands used to monitor MPLS operations.

MPLS Monitoring Commands

```
show mpls ldp parameters
```

- Displays LDP parameters on the local router

```
show mpls interfaces
```

- Displays MPLS status on individual interfaces

```
show mpls ldp discovery
```

- Displays all discovered LDP neighbors

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01--1-19

To display available LDP parameters, use the **show mpls ldp parameters** command in privileged EXEC mode.

To display information about one or more interfaces that have the MPLS feature enabled, use the **show mpls interfaces** [*interface*] [**detail**] command in EXEC mode.

To display the status of the LDP discovery process (Hello protocol), use these commands in privileged EXEC mode:

- **show mpls ldp discovery** [*vrf vpn-name*]
- **show mpls ldp discovery** [**all**]

The **show mpls ldp discovery** command displays all MPLS-enabled interfaces and the neighbors that are present on the interfaces.

Display MPLS LDP Parameters

```
RP/0/RSP0/CPU0:PE1# show mpls ldp parameters

LDP Parameters:
  Role: Active
  Protocol Version: 1
  Router ID: 10.7.1.1
  Null Label: Implicit
  Session:
    Hold time: 180 sec
    Keepalive interval: 60 sec
    Backoff: Initial:15 sec, Maximum:120 sec
    Global MD5 password: Disabled
  Discovery:
    Link Hellos:      Holdtime:15 sec, Interval:5 sec
    Targeted Hellos:  Holdtime:90 sec, Interval:10 sec
  Graceful Restart:
    Disabled
  NSR: Disabled, Not Sync-ed
  Timeouts:
    Local binding: 300 sec
    Forwarding state in LSD: 15 sec
  Max:
    1050 interfaces (800 attached, 250 TE tunnel), 1000 peers
  OOR state
  Memory: Normal
```

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1-20

To display the current LDP parameters, use the **show mpls ldp parameters** command in EXEC mode.

show mpls ldp parameters Field Description

Field	Description
Protocol version	Identifies the version of LDP that is running on the platform
Router ID	Currently used router ID
Null label	Identifies LDP use of implicit or explicit null labels for prefixes where it has to use a null label
Session hold time	The time that an LDP session is to be maintained with an LDP peer, without receiving LDP traffic or an LDP keepalive message from the peer
Session keepalive interval	The interval of time between consecutive transmissions of LDP keepalive messages to an LDP peer
Session backoff	Initial maximum backoff time for sessions
Discovery link hellos	The time to remember that a neighbor platform wants an LDP session without receiving an LDP hello message from the neighbor (hold time), and the time interval between the transmission of consecutive LDP hello messages to neighbors (interval)
Discovery targeted hellos	Indicates the time: ▪ To remember that a neighbor platform wants an LDP session when the neighbor platform is not directly connected to the router or the neighbor platform has not sent an LDP hello message. This intervening interval is known as <i>hold time</i>. ▪ Interval between the transmission of consecutive hello messages to a neighbor that is not directly connected to the router, and if targeted hellos are being accepted, displaying peer-acl (if any)
Graceful restart	Graceful restart status

Field	Description
NSR	Nonstop routing status
Timeouts	Various timeouts (of interest) that LDP is using. One timeout is <i>binding no route</i> , which indicates how long LDP will wait for an invalid route before deleting it. It also shows Restart recovery time for LSD and LDP.
OOR state	Out of resource memory state: Normal, Major, or Critical.

Display MPLS Configured Interfaces

```
RP/0/RSP0/CPU0:PE1# show mpls interfaces
Tue Oct 18 12:35:17.016 UTC
Interface                LDP      Tunnel  Enabled
-----
GigabitEthernet0/0/0/0   Yes      No      Yes
GigabitEthernet0/0/0/2   Yes      No      Yes

RP/0/RSP0/CPU0:PE1# show mpls interfaces detail
Tue Oct 18 12:36:06.585 UTC
Interface GigabitEthernet0/0/0/0:
    LDP labelling enabled
    LSP labelling not enabled
    MPLS enabled
Interface GigabitEthernet0/0/0/2:
    LDP labelling enabled
    LSP labelling not enabled
    MPLS enabled
```

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-21

To display information about LDP-enabled interfaces, use the **show mpls ldp interfaces** command in EXEC mode. To display additional information, use the **show mpls ldp interfaces detail** command.

Display Status of LDP Discovery Process

```
RP/0/RSP0/CPU0:PE7# show mpls ldp discovery
Tue Oct 18 12:36:43.084 UTC

Local LDP Identifier: 10.7.1.1:0
Discovery Sources:
  Interfaces:
    GigabitEthernet0/0/0/0 : xmit
    GigabitEthernet0/0/0/2 : xmit/recv
    LDP Id: 10.0.1.1:0, Transport address: 10.0.1.1
    Hold time: 10 sec (local:15 sec, peer:10 sec)
```

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.22

To display the status of the LDP discovery process, use the **show mpls ldp discovery** command in EXEC mode. The **show mpls ldp discovery** command shows both link discovery and targeted discovery. When no interface filter is specified, this command generates a list of interfaces that are running the LDP discovery process. This command also displays neighbor discovery information for the default routing domain.

show mpls ldp discovery Field Description

Field	Description
Local LDP identifier	LDP identifier for the local router. An LDP identifier is a 6-byte construct displayed in the form of IP address:number. By convention, the first 4 bytes of the LDP identifier constitute the router ID; integers, starting with 0, constitute the final two bytes of the IP address:number construct.
Interfaces	The interfaces that are engaging in LDP discovery activity, as described here: ▪ The xmit field indicates that the interface is transmitting LDP discovery hello packets. ▪ The recv field indicates that the interface is receiving LDP discovery hello packets. The LDP identifiers indicate LDP (or TDP) neighbors discovered on the interface.
Transport address	Address that is associated with this LDP peer (advertised in hello messages).
LDP ID	LDP identifier of the LDP peer.
Hold time	State of the forwarding hold timer and its current value.

LDP Monitoring Commands

```
show mpls ldp neighbor
```

- Displays individual LDP neighbors

```
show mpls ldp neighbor detail
```

- Displays more details about LDP neighbors

```
show mpls ldp bindings
```

- Displays LIB table

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-1-23

To display the status of LDP sessions, use the **show mpls ldp neighbor** command. To display the contents of the LIB, use the **show mpls ldp bindings** command.

Display the Status of the LDP Session

```
RP/0/RSP0/CPU0:PE1# show mpls ldp neighbor
Tue Oct 18 12:37:15.213 UTC

Peer LDP Identifier: 10.0.1.1:0
TCP connection: 10.0.1.1:646 - 10.7.1.1:52952
Graceful Restart: No
Session Holdtime: 180 sec
State: Oper; Msgs sent/rcvd: 382/421; Downstream-Unsolicited
Up time: 05:32:14
LDP Discovery Sources:
  GigabitEthernet0/0/0/2
Addresses bound to this peer:
  10.0.1.1          10.10.10.18      192.168.1.1      192.168.2.1
  192.168.11.1     192.168.21.1     192.168.31.1     192.168.51.1
  192.168.61.1     192.168.71.1
```

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.24

To display the status of LDP sessions, use the **show mpls ldp neighbor** command:

show mpls ldp neighbor [*A.B.C.D* | *type interface-path-id* | **gr** | **non-gr** | **sp** | | **standby** | **brief** | **detail**]

The status of the LDP session is indicated by “State: Oper” (operational).

The **show mpls ldp neighbor** command provides information about all LDP neighbors in the entire routing domain; conversely, the **show** output is filtered to display:

- LDP neighbors with specific IP addresses
- LDP neighbors on a specific interface
- LDP neighbors that can be gracefully restarted
- LDP neighbors that cannot be gracefully restarted
- LDP neighbors enabled with session protection

show mpls ldp neighbor Field Description

Field	Description
Peer LDP identifier	This field is the LDP identifier of the neighbor (peer) for this session
Graceful restart	This field is the graceful restart status (Y or N)
TCP connection	This field displays the TCP connection used to support the LDP session, shown in the format that follows: ■ neighbor IP address.peer port ■ local IP address.local port
State	This field displays the state of the LDP session. Generally, this is “Oper” (operational), but “transient” is another possible state.

Field	Description
Msgs sent/rcvd	This field displays number of LDP messages sent to and received from the session peer. The count includes the transmission and receipt of periodic keepalive messages, which are required for maintenance of the LDP session.
Up time	This field displays the length of time that the LDP session has existed.
LDP discovery sources	This field displays the source (or sources) of LDP discovery activity that led to the establishment of this LDP session.
Addresses bound to this peer	This field displays the known interface addresses of the LDP session peer. These are addresses that might appear as next-hop addresses in the local routing table. They are used to maintain the LFIB.

Display the Detailed Status of the LDP Session

```
RP/0/RSP0/CPU0:PE1# show mpls ldp neighbor detail
Tue Oct 18 12:39:44.893 UTC

Peer LDP Identifier: 10.0.1.1:0
TCP connection: 10.0.1.1:646 - 10.7.1.1:52952
Graceful Restart: No
Session Holdtime: 180 sec
State: Oper; Msgs sent/rcvd: 385/424; Downstream-Unsolicited
Up time: 05:34:44
LDP Discovery Sources:
  GigabitEthernet0/0/0/2
Addresses bound to this peer:
  10.0.1.1          10.10.10.18      192.168.1.1      192.168.2.1
  192.168.11.1     192.168.21.1     192.168.31.1     192.168.51.1
  192.168.61.1     192.168.71.1
Peer holdtime: 180 sec; KA interval: 60 sec; Peer state: Estab
NSR: Disabled
Capabilities:
  Sent:
    0x50b (Typed Wildcard FEC)
  Received:
    0x50b (Typed Wildcard FEC)
```

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-25

To display the detailed status of LDP sessions, use the **show mpls ldp neighbor detail** command.

Display the Contents of the LIB Table

```
RP/0/RSP0/CPU0:P1# show mpls ldp bindings
Tue Oct 18 06:32:04.302 UTC

10.0.0.0/8, rev 67
  Local binding: label: 16019
  Remote bindings: (1 peers)
    Peer          Label
    -----
    10.0.2.1:0    16019

10.7.10.1/32, rev 85
  Local binding: label: 16021
  Remote bindings: (3 peers)
    Peer          Label
    -----
    10.0.2.1:0    16022
    10.3.1.1:0    16025
```

To verify content of the LIB table, use the **show mpls ldp bindings** command. The output displays local labels for each destination network, as well as the labels that have been received from all LDP neighbors.

To display the contents of the label information base (LIB), use the **show mpls ldp bindings** command in EXEC command:

show mpls ldp bindings [*prefix {mask | length}*] [**advertisement-acls**] [**detail**] [**local**] [**local-label label**] [**to label**] [**neighbor address**] [**remote-label label**] [**to label**] [**summary**]

You can choose to view the entire database or a subset of entries according to the following criteria:

- Prefix
- Input or output label values or ranges
- Neighbor advertising the label

In the example, the local label for network 10.7.10.1/32 is 16021, the label for that network received from 10.0.2.1 neighbor is 16022 and the label received from 10.3.1.1 neighbor is 16025.

Note The **show mpls ldp bindings summary** command displays summarized information from the LIB and is used when you are testing scalability or when it is deployed in a large-scale network.

show mpls ldp bindings Field Descriptions

Field	Description
a.b.c.d/n	This field is the IP prefix and mask for a particular destination (network/mask).
Rev	This field is the revision number (rev) that is used internally to manage label distribution for this destination.
Local binding	A locally assigned label for a prefix
Remote bindings	Outgoing labels for this destination that are learned from other LSRs. Each item in this list identifies the LSR from which the outgoing label was learned and reflects the label that is associated with that LSR. Each LSR in the transmission path is identified by its LDP identifier.
(Rewrite)	Binding has been written into MPLS forwarding and is in use.
(No route)	Route is not valid. LDP times it out before the local binding is deleted.

Monitor Label Switching

```
show mpls forwarding
```

- Displays contents of LFIB table

```
show cef
```

- Displays contents of FIB table

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01--1.27

To display the contents of the MPLS LFIB, use the **show mpls forwarding** command in EXEC mode.

To display the contents of the FIB Cisco Express Forwarding table, use the **show cef** command in EXEC mode.

Display the Contents of the LFIB Table

```
RP/0/RSP0/CPU0:PE1# show mpls forwarding
Wed Oct 19 11:00:43.683 UTC
```

Local Label	Outgoing Label	Prefix or ID	Outgoing Interface	Next Hop	Bytes Switched
16000	Pop	10.0.1.1/32	Gi0/0/0/2	192.168.71.1	0
16001	16000	10.0.2.1/32	Gi0/0/0/2	192.168.71.1	31354
16002	16010	10.5.1.1/32	Gi0/0/0/2	192.168.71.1	0
16003	16011	10.6.1.1/32	Gi0/0/0/2	192.168.71.1	0
16021	16009	192.168.42.0/24	Gi0/0/0/2	192.168.71.1	0
16023	16018	10.4.1.1/32	Gi0/0/0/2	192.168.71.1	0
16024	16004	192.168.108.0/24	Gi0/0/0/2	192.168.71.1	0
16025	Unlabelled	10.7.10.1/32	Gi0/0/0/0	192.168.107.71	945410
16026	16023	10.8.1.1/32	Gi0/0/0/2	192.168.71.1	0
16027	16024	10.8.10.1/32	Gi0/0/0/2	192.168.71.1	0

To display the contents of the MPLS LFIB, use the **show mpls forwarding** command in EXEC mode:

show mpls forwarding [**detail** | {**label** *label number*} | **interface** *interface-path-id* | **labels** *value* | **location** | **prefix** [*network/mask* | *length*] | **private** | **summary** | **tunnels** *tunnel-id*]

The output displays the incoming and outgoing label for each destination, together with the outgoing interface and next hop. In the example, the incoming (Local) label for 192.168.42.0/24 network is 16021 (allocated by this router), and the outgoing label is 16009 (as advertised by the next hop). In the example, you can also see network 10.0.1.1/32 that has a POP label set as the outgoing label. This means that the router learned from a neighbor that a label should be removed from a labeled packet. There is also network 10.7.10.1/32 that does not have an outgoing label set (Unlabeled); this means that the label has not yet been received from a neighbor, or that the network is outside the MPLS domain and the router is an edge LSR.

show mpls forwarding Field Descriptions

Field	Description
Local label	This field is the IP prefix and mask for a particular destination (network/mask).
Outgoing label	This field is the label that was assigned by the next hop or downstream peer. Some of the entries that display in this column are these: - Unlabeled : No label for the destination from the next hop, or label switching is not enabled on the outgoing interface - Pop Label: Next hop advertised an implicit-null label for the destination
Prefix or tunnel ID	This field is the address or tunnel where packets with this label are going.
Outgoing interface	This field is the interface through which packets with this label are sent.
Next hop	This field is the IP address of the neighbor that assigned the outgoing label.
Bytes switched	This field is the number of bytes switched with this incoming label.

Display the Contents of the FIB Table

```
RP/0/RSP0/CPU0:PE1# show cef
Wed Oct 19 11:34:31.879 UTC

Prefix                Next Hop                Interface
<...output omitted...>
10.0.1.1/32           192.168.71.1           GigabitEthernet0/0/0/2
10.0.2.1/32           192.168.71.1           GigabitEthernet0/0/0/2
10.5.1.1/32           192.168.71.1           GigabitEthernet0/0/0/2
10.6.1.1/32           192.168.71.1           GigabitEthernet0/0/0/2
10.7.10.1/32          192.168.107.71         GigabitEthernet0/0/0/0
10.8.1.1/32           192.168.71.1           GigabitEthernet0/0/0/2
10.8.10.1/32          192.168.71.1           GigabitEthernet0/0/0/2
192.168.42.0/24       192.168.71.1           GigabitEthernet0/0/0/2
192.168.51.0/24       192.168.71.1           GigabitEthernet0/0/0/2
<...output omitted...>
```

- Use command **show cef 192.168.42.0** to show details for specific prefix

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1.29

To display information about packets forwarded by FIB Cisco Express Forwarding, use the **show cef** command in EXEC mode:

show cef [*prefix* [*mask*]] [**hardware** {**egress** | **ingress**}] [**detail**] [**location** {*node-id* | **all**}]

In the figure, the next hop for network 192.168.42.0/24 is 192.168.71.1 and the outgoing interface is GigabitEthernet0/0/0/2.

To verify the FIB Cisco Express Forwarding table, use the **show cef** command, followed by a desired prefix:

```
RP/0/RSP0/CPU0:PE7# show cef 192.168.42.0
Wed Oct 19 12:08:03.213 UTC
192.168.42.0/24, version 0, internal 0x4004001 (ptr 0xad958e70) [1],
0x0 (0xacf50a94), 0x450 (0xadffc4b0)
Updated Oct 19 05:51:07.981
remote adjacency to GigabitEthernet0/0/0/2
Prefix Len 24, traffic index 0, precedence routine (0)
  via 192.168.71.1, GigabitEthernet0/0/0/2, 4 dependencies, weight 0,
class 0 [flags 0x0]
  path-idx 0
  next hop 192.168.71.1
  remote adjacency
    local label 16021      labels imposed {16009}
```

Debugging MPLS and LDP

This topic describes the MPLS and LDP debug commands.

Debugging MPLS and LDP

```
debug mpls ldp ...
```

- Debugs LDP adjacencies, session establishment, and label bindings exchange

```
debug mpls packet [interface]
```

- Debugs labeled packets switched by the router

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01-1-30

A large number of **debug** commands are associated with MPLS. The **debug mpls ldp** commands debug various aspects of the LDP protocol, from label distribution to exchange of the application layer data between adjacent LDP-speaking routers.

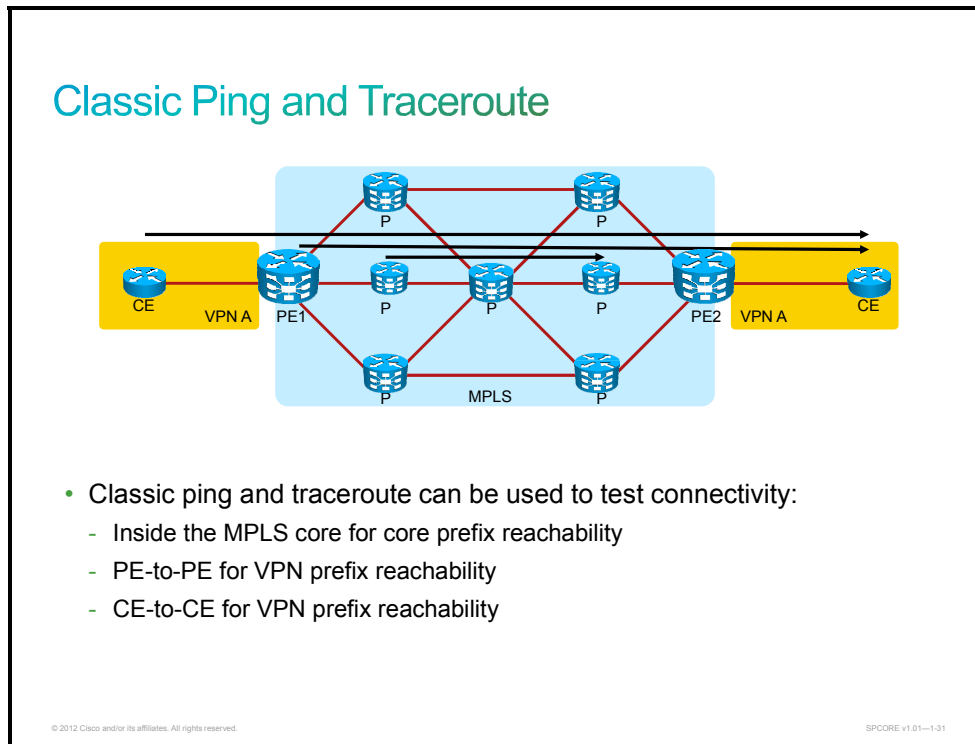
Note Use **debug** commands with caution. Enabling debugging can disrupt the operation of the router under high load conditions. Before you start a **debug** command, always consider the output that the command may generate and the amount of time this may take. You should also look at your CPU load before debugging by using the **show processes cpu** command. Verify that you have ample CPU capacity available before beginning the debugging process.

The **debug mpls packet** command displays all labeled packets that are switched by the router (through the specified interface).

Caution Use the **debug mpls packet** command with care, because it generates output for every packet that is processed. Furthermore, enabling the **debug mpls packet** command causes fast and distributed label switching to be disabled for the selected interfaces. To avoid adversely affecting other system activity, use this command only when traffic on the network is at a minimum.

Classic Ping and Traceroute

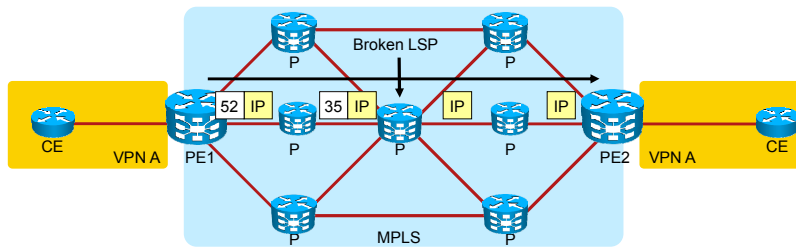
This topic describes the Classic Ping and Traceroute operations.



Standard ping and traceroute tools can be used in MPLS environments to test reachability in three different scenarios:

- Test the reachability of prefixes that are reachable through the global routing table via IP forwarding or label switching. The tools can be used on PE and P routers.
- Test the reachability of Layer 3 MPLS VPN prefixes that are reachable through a virtual routing and forwarding (VRF) routing table via label switching. The tools can be used on PE routers configured with the required VRF.
- Customers can use the tools to test Layer 3 MPLS VPN connectivity end to end.

Classic Ping and Traceroute (Cont.)



- Broken LSPs may not always be revealed:
 - Broken LSPs revert back to IP forwarding.
 - Ping and traceroute succeed.
- Cisco IOS Software does encode MPLS information in ICMP replies.
- Even multiple paths can sometimes be detected.

```
RP/0/RSP0/CPU0:PE1# traceroute 172.16.1.14
Type escape sequence to abort.
Tracing the route to 172.16.1.14
 0 192.168.1.46 [MPLS: Label 34 Exp 0] 8 msec
 1 192.168.1.42 [MPLS: Label 38 Exp 0] 12 msec
 2 192.168.1.46 [MPLS: Label 34 Exp 0] 24 msec
 3 192.168.1.14 [MPLS: Label 37 Exp 0] 48 msec
 4 192.168.1.18 [MPLS: Label 33 Exp 0] 8 msec
 5 192.168.1.14 [MPLS: Label 37 Exp 0] 8 msec
 6 192.168.1.38 52 msec
 7 192.168.1.34 8 msec *
```

The figure illustrates a redundant network where the standard traceroute tool was used to determine the path from one PE to another PE. Cisco routers in the path will encode some MPLS information into ICMP replies to be displayed by the router that is initiating route tracing.

The sample output shows how labels are displayed and how even multiple paths can be detected, although not always reliably (equal paths on subsequent hops will typically not be displayed by classic traceroute).

MPLS Ping and Traceroute

This topic describes the MPLS Ping and Traceroute operations.

MPLS Ping and Traceroute

- Designed for monitoring and troubleshooting MPLS LSPs
- Encapsulates UDP requests directly into selected LSP
- More choices in generating requests:
 - Exp field, TTL, reply mode, output interface, and so on
 - Explicit null label usage
 - Not subject to TTL propagation disabling
- More information in replies:
 - Labels, interfaces, many other LSP diagnostic details
- Can be used to monitor:
 - LDP LSPs
 - MPLS TE tunnel LSPs
 - Layer-2 MPLS VPN LSPs

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-33

Special MPLS ping and MPLS traceroute were designed for monitoring and troubleshooting MPLS LSPs. These features provide a means to check connectivity and isolate a failure point, thus providing the MPLS Operation, Administration, and Maintenance (OAM) solution.

Normal ICMP ping and traceroute are used to help diagnose the root cause when a forwarding failure occurs. However, they may not detect LSP failures because an ICMP packet can be forwarded via IP to the destination when an LSP breakage occurs, whereas MPLS LSP ping and traceroute can be used to identify LSP breakages.

A forwarding equivalence class (FEC) must be selected to choose the associated LSP. FECs can be any of these:

- IP prefix in the FIB table with label next-hop information
- Layer 2 MPLS VPN virtual circuit (pseudowire)
- MPLS traffic engineering tunnel LSP

MPLS ping and traceroute will use UDP packets with loopback destination addresses to encode requests and label them with the selected FEC label.

Enable MPLS OAM by using the **mpls oam** command on all routers in the MPLS network.

MPLS Ping

- UDP request generated for selected LSP
- Uses two UDP (port 3503) messages
 - MPLS echo request
 - MPLS echo reply
- Labeled packet with IP (UDP) payload
 - Source address: Routable address sender
 - Destination address: Random from 127/8
 - Destination port: 3503
 - TTL: 255

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-34

MPLS ping uses UDP on port 3503 to encode two types of messages:

- **MPLS echo request:** The MPLS echo request message includes the information about the tested FEC (prefix) which is encoded as one of the type length values (TLVs). Additional TLVs can be used to request additional information in replies. The downstream mapping TLV can be used to request information about additional information such as downstream router and interface, MTU, and multipath information from the router where the request is processed.
- **MPLS echo reply:** The MPLS echo reply uses the same packet format as the request, except that it may include additional TLVs to encode the information. The basic reply is, however, encoded in the reply code field.

The MPLS echo request will use the outgoing interface IP address as the source and a loopback IP address, which is configurable, as the destination (127.0.0.1). The TTL in MPLS ping is set to 255. Using the 127/8 address in the IP header destination address field will cause the packet not to be forwarded by any routers using the IP header, if the LSP is broken somewhere inside the MPLS domain.

The initiating router can also request a reply mode, which can be one of the following:

- The default reply mode uses IPv4 and MPLS to return the reply.
- Router alert mode that forces every router in the return path to perform process switching of the return packet, which in turn forces the use of the IP forwarding table (avoids any confusion if the return LSP is broken). This functionality is achieved by adding label 1 onto the label stack for the reply packet.
- It is also possible to request that no reply be sent.

MPLS Ping Example 1

```
RP/0/RSP0/CPU0:PE1# ping mpls ipv4 172.16.1.14 255.255.255.255
Sending 5, 100-byte MPLS Echos to 172.16.1.14/32,
    timeout is 2 seconds, send interval is 0 msec:

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
'L' - labeled output interface, 'B' - unlabeled output interface,
'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
'M' - malformed request, 'm' - unsupported tlvs, 'N' - no label entry,
'P' - no rx intf label prot, 'p' - premature termination of LSP,
'R' - transit router, 'I' - unknown upstream index,
'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.
!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 16/17/20 ms
```

- IPv4 FEC from the global IPv4 routing table

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--138

The sample MPLS ping illustrates the command syntax. The main difference from the standard ping is the need to exactly specify the FEC (a prefix in the FIB table) from which the router will learn the next-hop label, output interface, and Layer 2 forwarding information.

A successful reply is also represented by exclamation marks. A number of other results are possible, depending on the return code that can map to any of the characters described in the legend portion of the MPLS ping output.

ping mpls {ipv4 addr/mask} [destination {start address} {end address} {address increment}] | [dsmap] | [exp exp bits in MPLS header] | [force-explicit-null] | [interval send interval between requests in msec] | [output interface echo request output interface] [pad pad TLV pattern] | [repeat repeat count] | [reply dscp differentiated services codepoint value] | [reply mode ipv4 | router-alert | no-reply] | [reply pad-tlv] | [revision echo packet tlv versioning] | [{size packet size} | [source source specified as an IP address] | {sweep {min value} {max value} {increment}}] | [timeout timeout in seconds] | [ttl time to live] | [verbose]

MPLS Ping Example 2

```
RP/0/RSP0/CPU0:PE1# ping mpls ipv4 172.16.1.14 255.255.255.255 ttl 1 dsmmap repeat 1
Sending 1, 100-byte MPLS Echos to 172.16.1.14/32,
    timeout is 2 seconds, send interval is 0 msec:

Codes: '.' - success, 'Q' - request not sent, '.' - timeout,
'L' - labeled output interface, 'B' - unlabeled output interface,
'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
'M' - malformed request, 'm' - unsupported tlvs, 'N' - no label entry,
'P' - no rx intf label prot, 'p' - premature termination of LSP,
'R' - transit router, 'I' - unknown upstream index,
'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.
L
Echo Reply received from 192.168.1.2
  DSMAP 0, DS Router Addr 127.0.0.1, DS Intf Addr 0
    Depth Limit 0, MRU 1500 [Labels: 33 Exp: 0]
    Multipath Addresses:

Success rate is 0 percent (0/1)
```

- The downstream map (**dsmmap**) option can be used to retrieve the details for a given hop.
- MPLS traceroute can be used instead to display detailed information for all hops.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-36

The example illustrates how you can request the downstream map (**dsmmap**) information and select from which hop. The reply contains the downstream information, including MTU.

The **dsmmap** optional parameter interrogates a transit router for downstream map information.

MPLS Traceroute Example

```
RP/0/RSP0/CPU0:PE1# traceroute mpls ipv4 172.16.1.14 255.255.255.255
Tracing MPLS Label Switched Path to 172.16.1.14/32, timeout is 2 seconds

Codes: '.' - success, 'Q' - request not sent, '.' - timeout,
'L' - labeled output interface, 'B' - unlabeled output interface,
'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
'M' - malformed request, 'm' - unsupported tlvs, 'N' - no label entry,
'P' - no rx intf label prot, 'p' - premature termination of LSP,
'R' - transit router, 'I' - unknown upstream index,
'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.
 0 192.168.1.1 127.0.0.1 MRU 1500 [Labels: 33 Exp: 0]
I 1 192.168.1.2 127.0.0.1 MRU 1500 [Labels: 33 Exp: 0] 8 ms, ret code 6
I 2 192.168.1.14 127.0.0.1 MRU 1504 [Labels: implicit-null Exp: 0] 12 ms, ret code 6
I 3 192.168.1.34 12 ms, ret code 3
```

- Labels and MTU can be determined using MPLS traceroute.
- Detailed error information is retrieved upon failure somewhere in the path.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--137

The sample MPLS traceroute shows that the downstream mapping information is reported from each hop where the TTL expired. The output shows the entire path with labels and maximum receive unit (MRU), as well as the round-trip time and IP addresses on routers in the path. MRU includes the maximum size of the IP packet including the label stack that could be forwarded out of the particular interface.

To learn the routes that packets follow when traveling to their destinations, use the **traceroute mpls** command in EXEC mode.

traceroute mpls *{ {ipv4 addr/mask} | {traffic-eng tunnel tunnel intf num} }* **[destination** *{start address} {end address} {address increment} }* **| [exp** *exp bits in MPLS header* **]** **| [flags** *fec* **]** **| [force-explicit-null]** **| [output interface** *echo request output interface* **]** **| [reply dscp** *DSCP bits in reply IP header* **]** **| [reply mode** *[ipv4 | router-alert | no-reply]* **]** **| [revision** *echo packet tlv versioning* **]** **| [source** *source specified as an IP address* **]** **| [timeout** *timeout in seconds* **]** **| [ttl** *time to live* **]** **| [verbose]**

Troubleshoot MPLS

This topic describes how to troubleshoot common MPLS issues.

Symptoms of Common MPLS Issues

- The LDP session does not start.
- Labels are not allocated.
- Labels are not distributed.
- Packets are not labeled, although the labels have been distributed.
- MPLS intermittently breaks after an interface failure.
- Large packets are not propagated across the network.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-38

Here are the common issues that can be encountered while you are troubleshooting a frame mode MPLS network:

- The LDP session does not start.
- The LDP session starts, but the labels are not allocated or distributed.
- Labels are allocated and distributed, but the forwarded packets are not labeled.
- MPLS stops working intermittently after an interface failure, even on interfaces totally unrelated to the failed interface.
- Large IP packets are not propagated across the MPLS backbone, even though the packets were successfully propagated across the pure IP backbone.

LDP Session Startup Issues

- Symptom:
 - LDP neighbors are not discovered.
 - The **show mpls ldp discovery** command does not display the expected LDP neighbors.
- Diagnosis:
 - MPLS is not enabled on the adjacent router.
- Verification:
 - Verify with the **show mpls interface** command on the adjacent router.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-39

Symptom: If MPLS is enabled on an interface, but no neighbors are discovered, it is likely that MPLS is not enabled on the neighbor.

The router is sending discovery messages, but the neighbor is not replying because it does not have LDP enabled.

Solution: Enable MPLS on the neighboring router.

LDP Session Startup Issues (Cont.)

- Symptom:
 - LDP neighbors are discovered; the LDP session is not established.
 - The **show mpls ldp neighbor** command does not display a neighbor in operational state.
- Diagnosis:
 - The connectivity between loopback interfaces is broken; the LDP session is usually established between loopback interfaces of adjacent LSRs.
- Verification:
 - Verify connectivity with the extended **ping** command.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-40

Symptom: LDP neighbors are exchanging hello packets, but the LDP session is never established.

Solution: Check the reachability of the loopback interfaces, because they are typically used to establish the LDP session. Make sure that the loopback addresses are exchanged via the IGP that is used in the network.

Label Distribution Issues

- Symptom:
 - Labels are allocated, but not distributed.
 - Using the **show mpls ldp bindings** command on the adjacent LSR does not display labels from this LSR.
- Diagnosis:
 - There are problems with conditional label distribution.
- Verification:
 - Debug label distribution with the **debug mpls ldp advertisements** command.
 - Examine the neighbor LDP router IP address with the **show mpls ldp discovery** command.
 - Verify that the neighbor LDP router IP address is matched by the access list specified in the **mpls ldp label advertise** command.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—1-41

Symptom: Labels are generated for local routes on one LSR but are not received on neighboring LSRs.

Solution: Check whether conditional label advertising is enabled and verify both access lists that are used with the command.

Intermittent MPLS Failures After Interface Failure

- Symptom:
 - The overall MPLS connectivity in a router intermittently breaks after an interface failure.
- Diagnosis:
 - The IP address of a physical interface is used for the LDP identifier. Configure a loopback interface on the router.
- Verification:
 - Verify the local LDP identifier with the **show mpls ldp neighbors** command.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--1-42

Symptom: MPLS connectivity is established, labels are exchanged, and packets are labeled and forwarded. However, an interface failure can sporadically stop an MPLS operation on unrelated interfaces in the same router.

Details: LDP sessions are established between IP addresses that correspond to the LDP router ID. If the LDP router ID has not been manually configured, the LDP router ID is assigned using the algorithm that is also used to assign an OSPF or a BGP router ID.

This algorithm selects the highest IP address of an active interface if there are no loopback interfaces configured on the router. If that interface fails, the LDP router ID is lost and the TCP session that is carrying the LDP data is torn down, resulting in loss of all neighbor-assigned label information.

The symptom can be easily verified with the **show mpls ldp neighbors** command, which displays the local and remote LDP router ID. Verify that both of these IP addresses are associated with a loopback interface.

Solution: Manually configure the LDP router ID, referencing a loopback interface that is reachable by the IGP.

Packet Propagation Issues

- Symptom:
 - Large packets are not propagated across the network.
 - Use of the extended ping command with varying packet sizes fails for packet sizes almost to 1500 packets.
 - In some cases, MPLS might work, but MPLS VPN will fail.
- Diagnosis:
 - There are label MTU issues or switches that do not support jumbo frames in the forwarding path.
- Verification:
 - Issue the traceroute command through the forwarding path; identify all LAN segments in the path.
 - Verify the label MTU setting on routers attached to LAN segments.
 - Check for low-end switches in the transit path.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--143

Symptom: Packets are labeled and sent, but they are not received on the neighboring router. A LAN switch between the adjacent MPLS-enabled routers may drop the packets if it does not support jumbo frames. In some cases, MPLS might work, but MPLS VPN will fail.

Solution: Change the MPLS MTU size, taking into account the maximum number of labels that may appear in a packet.

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- On Cisco IOS XR platforms, MPLS is enabled by enabling LDP on each interface. On Cisco IOS and IOS XE platforms, MPLS is enabled on each interface.
- It is recommended to manually set the router ID.
- To enable MPLS on the Cisco IOS XR router, first enter MPLS LDP configuration mode and then list the interfaces.
- Label switching increases the maximum MTU requirements on an interface, because of the additional label header.
- To configure the maximum packet size or MTU size on an MPLS interface, use the **mpls mtu** command in interface configuration mode.
- By default, IP TTL is copied into the MPLS label at label imposition, and the MPLS label TTL is copied into the IP TTL at label removal.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—1-44

Summary (Cont.)

- If TTL propagation is disabled, the TTL value is not copied into the label header. Instead, the label TTL field is set to 255.
- When LDP session protection is configured, LDP initiates backup targeted hellos automatically for neighbors for which primary link adjacencies already exist.
- Graceful restart is a way to recover from signaling and control plane failures without impacting forwarding.
- LDP IGP synchronization synchronizes LDP and IGP so that IGP advertises links with regular metrics only when MPLS LDP is converged on that link.
- To enable LDP on many interfaces, IGP autoconfiguration allows you to automatically configure LDP on all interfaces that are associated with a specified OSPF or IS-IS interface.
- LDP outbound label filtering performs outbound filtering for local label advertisement, for one or more prefixes, to one or more peers.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—1-45

Summary (Cont.)

- You can use various **show** commands to monitor MPLS.
- When debugging MPLS and LDP in production environments, use the **debug** commands with extreme cautions.
- Standard ping and traceroute tools can be used in MPLS environments to test reachability.
- Special MPLS ping and MPLS traceroute were designed for monitoring and troubleshooting MPLS LSPs.
- If an LDP session does not come up, verify if MPLS is enabled on the neighboring router.

Module Summary

This topic summarizes the key points that were discussed in this module.

Module Summary

- MPLS features, concepts, and terminology, and MPLS label format were discussed. LSR architecture and operations were also explained in this module.
- The assignment and distribution of labels in an MPLS network, including neighbor discovery and session establishment procedures, were discussed. Label distribution, control, and retention modes were described.
- The details of implementing MPLS on Cisco IOS, IOS XE, and IOS XR platforms were explained, and detailed configuration, monitoring, and debugging guidelines for a typical service provider network were discussed.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—1-1

This module explained the features of Multiprotocol Label Switching (MPLS) compared with those of traditional hop-by-hop IP routing. MPLS concepts and terminology, along with MPLS label format and label switch router (LSR) architecture and operations, were explained in this module. The module also described the assignment and distribution of labels in an MPLS network, including neighbor discovery and session establishment procedures. Label distribution, control, and retention modes were also covered.

The module also explained the process for implementing MPLS on Cisco IOS, IOS XE, and IOS XR platforms, giving detailed configuration, monitoring, and debugging guidelines for a typical service provider network.

Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

- Q1) What are three foundations of traditional IP routing? (Choose three.) (Source: Introducing MPLS)
- A) Routing protocols are used on all devices to distribute routing information.
 - B) Regardless of protocol, routers always forward packets based on only the IP destination address (except for using PBR).
 - C) Routing lookups are performed on every router.
 - D) Routing is performed by assigning a label to an IP destination.
 - E) Routing lookup is performed only at the first router on the path.
- Q2) Which three statements about MPLS are true? (Choose three.) (Source: Introducing MPLS)
- A) MPLS uses labels to forward packets.
 - B) MPLS works only in IP networks.
 - C) MPLS labels can correspond to a Layer 3 destination address, QoS, source address, or Layer 2 circuit.
 - D) MPLS does not require a routing table lookup on core routers.
 - E) MPLS performs routing lookup on every router.
- Q3) The MPLS label field consists of how many bits? (Source: Introducing MPLS)
- A) 64 bits
 - B) 32 bits
 - C) 16 bits
 - D) 8 bits
- Q4) Which two statements about LSRs are true? (Choose two.) (Source: Introducing MPLS)
- A) An edge LSR is a device that inserts labels on packets or removes labels and forwards packets based on labels.
 - B) An LSR is a device that primarily labels packets or removes labels.
 - C) An LSR is a device that forwards packets based on labels.
 - D) An end LSR is a device that primarily inserts labels on packets or removes labels.
- Q5) Which two statements about RSVP are true? (Choose two.) (Source: Introducing MPLS)
- A) RSVP is used to create an LSP tunnel.
 - B) RSVP propagates labels for TE tunnels.
 - C) RSVP assigns labels for TE tunnels.
 - D) RSVP is not used to create an LSP tunnel.
- Q6) In MPLS VPN networks, which statement is true? (Source: Introducing MPLS)
- A) Labels are propagated via LDP or TDP.
 - B) Next-hop addresses, instead of labels, are used in an MPLS VPN network.
 - C) Labels are propagated via MP-BGP.
 - D) Two labels are used; the top label identifies the VPN, and the bottom label identifies the egress router.

- Q7) Which two statements about interactions between MPLS applications are true? (Choose two.) (Source: Introducing MPLS)
- A) The forwarding plane is the same for all applications.
 - B) Differences exist in the forwarding plane depending on the MPLS application.
 - C) The control plane is the same for all applications.
 - D) Differences exist in the control plane depending on the MPLS application.
- Q8) What does per-platform label space require? (Source: Label Distribution Protocol)
- A) It requires only one LDP session.
 - B) It requires one session per interface.
 - C) It requires multiple sessions for parallel links.
 - D) "Per-platform" is not a proper term in MPLS terminology.
- Q9) LDP uses which two well-known port numbers? (Choose two.) (Source: Label Distribution Protocol)
- A) 464
 - B) 646
 - C) 711
 - D) 171
- Q10) Which three pieces of information are contained in the LFIB? (Choose three.) (Source: Label Distribution Protocol)
- A) local generated label
 - B) outgoing label
 - C) local address
 - D) next-hop address
 - E) destination IP network
- Q11) When an IP packet is to be label-switched as it traverses an MPLS network, which table is used to perform the label switching? (Source: Label Distribution Protocol)
- A) LIB
 - B) FIB
 - C) FLIB
 - D) LFIB
- Q12) Which statement is correct? (Source: Label Distribution Protocol)
- A) An IP forwarding table resides on the data plane; LDP runs on the control plane; and an IP routing table resides on the data plane.
 - B) An IP forwarding table resides on the data plane; LDP runs on the control plane; and an IP routing table resides on the control plane.
 - C) An IP forwarding table resides on the control plane; LDP runs on the control plane; and an IP routing table resides on the data plane.
 - D) An IP forwarding table resides on the control plane; LDP runs on the control plane; and an IP routing table resides on the control plane.

- Q13) Which two tables contain label information? (Choose two.) (Source: Label Distribution Protocol)
- A) LIB
 - B) main IP routing table
 - C) FLIB
 - D) LFIB
- Q14) Which two statements about LSPs are correct? (Choose two.) (Source: Label Distribution Protocol)
- A) LSPs are bidirectional.
 - B) LSPs are unidirectional.
 - C) LDP advertises labels for the entire LSP.
 - D) LDP advertises labels only for individual segments in the LSP.
- Q15) Which statement about TTL propagation being disabled is correct? (Source: Label Distribution Protocol)
- A) The label TTL is copied back into the IP TTL.
 - B) The IP TTL is copied back into the TTL of the label.
 - C) The IP TL is not copied back into the TTL of the label.
 - D) TTL label propagation can not be disabled.
- Q16) Upon a link failure, which three tables are updated to reflect the failed link? (Choose three.) (Source: Label Distribution Protocol)
- A) LIB
 - B) LFIB
 - C) FIB
 - D) FLIB
 - E) BIL
- Q17) Which statement correctly describes how a link failure is handled in an MPLS network? (Source: Label Distribution Protocol)
- A) Overall convergence depends on LDP.
 - B) Overall convergence depends on the IGP that is used.
 - C) Upon a link failure, only LDP convergence is affected.
 - D) Upon a link failure, only the IGP convergence is affected.
- Q18) Upon a link recovery, which three tables are updated to reflect the failed link? (Choose three.) (Source: Label Distribution Protocol)
- A) LFIB
 - B) FLIB
 - C) FIB
 - D) LIB
 - E) BIL
- Q19) Which statement correctly describes convergence in MPLS network after a link failure has occurred and been restored? (Source: Label Distribution Protocol)
- A) MPLS convergence occurs after IGP convergence.
 - B) MPLS convergence occurs before IGP convergence peer to peer.
 - C) If a failure occurs with the IGP, MPLS convergence is not affected.
 - D) If a failure occurs with the IGP, MPLS will not be able to converge after the IGP failure has been corrected unless the MPLS process is bounced.

- Q20) What is another name for topology-driven switching? (Source: Label Distribution Protocol)
- A) Cisco Express Forwarding
 - B) fast switching
 - C) cache switching
 - D) process switching
- Q21) If IP TTL propagation is not allowed, what is the value that is placed in the MPLS header? (Source: Implementing MPLS in the Service Provider Core)
- A) 0
 - B) 1
 - C) 254
 - D) 255
- Q22) What needs to be configured to specify which neighbors would selectively receive label advertisements? (Source: Implementing MPLS in the Service Provider Core)
- A) Controlled label distribution needs to be configured.
 - B) Conditional label distribution needs to be configured.
 - C) Unsolicited label distribution needs to be configured.
 - D) No configuration is necessary; all neighbors will receive all labels.
- Q23) Which command is used to display the contents of the LIB table? (Source: Implementing MPLS in the Service Provider Core)
- A) **show mpls ldp labels**
 - B) **show mpls ldp bindings**
 - C) **show mpls ldp neighbors**
 - D) **show mpls forwarding-table**
- Q24) Which is the correct command to enable MPLS in Cisco IOS Software? (Source: Implementing MPLS in the Service Provider Core)
- A) Router(config)#**ip mpls**
 - B) Router(config-if)#**ip mpls**
 - C) Router(config)#**mpls ip**
 - D) Router(config-if)#**mpls ip**

Module Self-Check Answer Key

Q1)	A, B, C
Q2)	A, C, D
Q3)	B
Q4)	A, C
Q5)	A, B
Q6)	C
Q7)	A, D
Q8)	A
Q9)	B, F
Q10)	A, B, D
Q11)	D
Q12)	B
Q13)	A, D
Q14)	B, D
Q15)	C
Q16)	A, B, C
Q17)	B
Q18)	A, C, D
Q19)	A
Q20)	A
Q21)	D
Q22)	B
Q23)	B
Q24)	D

MPLS Traffic Engineering

Overview

This module on Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) technology discusses the requirement for TE in modern service provider networks that must attain optimal resource utilization. The traffic-engineered tunnels provide a means of mapping traffic streams onto available networking resources in a way that prevents the overuse of subsets of networking resources while other subsets are underused. All the concepts and mechanics that support TE are presented, including tunnel path discovery with link-state protocols and tunnel path signaling with Resource Reservation Protocol (RSVP). Some of the advanced features of TE, such as autobandwidth and guaranteed bandwidth, are introduced as well.

This module discusses the requirement for implementing MPLS TE, and includes the configuration of routers to enable basic traffic tunnels, assignment of traffic to a tunnel, control of path selection, and performance of tunnel protection and tunnel maintenance. Configurations are shown for various Cisco platforms.

Module Objectives

Upon completing this module, you will be able to discuss the requirement for traffic engineering in modern service provider networks that must attain optimal resource utilization. This ability includes being able to meet these objectives:

- Describe the concepts that allow service providers to map traffic through specific routes to optimize network resources, especially bandwidth
- Describe the details of link attribute propagation with an IGP and constraint-based path computation
- Describe MPLS TE commands for the implementation of MPLS traffic tunnels
- Describe the MPLS TE commands for link and node protection

Introducing MPLS Traffic Engineering Components

Overview

This lesson explains the components of Multiprotocol Label Switching (MPLS) traffic engineering (TE), such as traffic tunnels (along with associated characteristics and attributes), tunnel path discovery based on link-state protocols, and tunnel setup signaling with Resource Reservation Protocol (RSVP).

Objectives

Upon completing this lesson, you will be able to describe the concepts that allow service providers to map traffic through specific routes to optimize network resources, especially the bandwidth. You will be able to meet these objectives:

- Describe basic Traffic Engineering concepts
- Describe Traffic Engineering with a Layer 2 Overlay Model
- Describe a Layer 3 routing model without Traffic Engineering
- Describe Traffic Engineering with a layer 3 routing model
- Describe Traffic Engineering with the MPLS TE Model
- Describe basic concept of MPLS TE traffic tunnels
- Describe MPLS TE traffic tunnels attributes
- Describe the link resource attributes
- Describe Constraint-Based Path Computation
- Describe the MPLS TE process
- Describe the Role of RSVP in Path Setup Procedures
- Describe the Path Setup Procedures using RSVP
- Describe the three methods to forward traffic to a tunnel
- Describe the Autoroute feature

Traffic Engineering Concepts

This topic describes basic Traffic Engineering concepts.

What Is Traffic Engineering?

- Traffic engineering is manipulating your traffic to fit your network.
- Network engineering is building your network to carry your predicted traffic.
- TE is a process of measures, models, and controls of traffic to achieve various goals.
- TE for data networks provides an integrated approach to managing traffic at Layer 3.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.3

TE can be contrasted with network engineering:

- Traffic engineering is manipulating your traffic to fit your network.
- Network engineering is building your network to carry your predicted traffic.

TE has been widely used in voice telephony. TE means that the traffic is measured and analyzed. A statistical model is then applied to the traffic pattern to make a prognosis and estimates.

If the anticipated traffic pattern does not match the network resources well, the network administrator remodels the traffic pattern. Such decisions can be made to achieve a more optimal use of the resources or to reduce costs by selecting a cheaper transit carrier.

In the data communications world, traffic engineering provides an integrated approach to engineering traffic at Layer 3 in the Open Systems Interconnection (OSI) model. The integrated approach means that routers are configured to divert traffic from destination-based forwarding and move the traffic load from congested parts of the network to uncongested parts.

Traditionally, this diversion has been done using overlay networks where routers use carefully engineered ATM permanent virtual circuits (PVCs) or Frame Relay PVCs to distribute the traffic load on Layer 2.

Traffic Engineering Motivations

- Reduce the overall **cost** of operations by more efficient use of bandwidth resources .
- Prevent a situation where some parts of a network are **overutilized** (congested), while other parts remain underutilized.
- Implement traffic **protection against failures**.
- Enhance **SLA** in combination with QoS.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—2-4

Cost reduction is the main motivation for TE. WAN connections are an expensive item in the service provider budget. A cost savings, which results from a more efficient use of resources, will help to reduce the overall cost of operations. Additionally, more efficient use of bandwidth resources means that a service provider can avoid a situation where some parts of a network are congested, while other parts are underutilized.

Because TE can be used to control traffic flows, it can also be used to provide protection against link or node failures by providing backup tunnels.

Finally, when combined with quality of service (QoS) functionality, TE can provide enhanced service level agreements (SLAs).

Business Drivers for Traffic Engineering

- Routers forward traffic along the least-cost route that is discovered by routing protocols.
- Network bandwidth may not be efficiently utilized:
 - The least-cost route may not be the only possible route.
 - The least-cost route may not have enough resources to carry all the traffic.
 - Alternate paths may be underutilized.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.6

In a Layer 3 routing network, packets are forwarded hop by hop. In each hop, the destination address of the packet is used to make a routing table lookup. The routing tables are created by an interior gateway protocol (IGP), which finds the least-cost route, according to its metric, to each destination in the network.

In many networks, this method works well. But in some networks, the destination-based forwarding results in the overutilization of some links, while others are underutilized. This imbalance will happen when there are several possible routes to reach a certain destination. The IGP selects one of them as the best, and uses only that route. In the extreme case, the best path may have to carry so large a volume of traffic that packets are dropped, while the next-best path is almost idle.

One solution to the problem would be to adjust the link bandwidths to more appropriate values. The network administrator could reduce the bandwidth on the underutilized link and increase the bandwidth on the overutilized one. However, making this adjustment is not always possible.

The alternate path is a backup path. In a primary link failure, the backup must be able to forward at least the major part of the traffic volume that is normally forwarded by the primary path. Therefore, it may not be possible to reduce the bandwidth on the backup path. And without a cost savings, the budget may not allow an increase to the primary link bandwidth.

To provide better network performance within the budget, network administrators move a portion of the traffic volume from the overutilized link to the underutilized link. During normal operations, this move results in fewer packet drops and quicker throughput. If there is a failure to any of the links, all traffic is forwarded over the remaining link, which then, of course, becomes overutilized.

Moving portions of the traffic volume cannot be achieved by traditional hop-by-hop routing using an IGP for path determination.

Business Drivers for Traffic Engineering (Cont.)

- Lack of resources results in congestion in two ways:
 - When network resources themselves are **insufficient to accommodate the offered load**
 - When traffic streams are **inefficiently mapped** onto available resources
- Some resources are overutilized while others remain underutilized.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2-6

Network congestion, caused by too much traffic and too few network resources, cannot be solved by moving portions of the traffic between different links. Moving the traffic will help only in the case where some resources are overutilized and others are underutilized. The traffic streams in normal Layer 3 routing are inefficiently mapped onto the available resources.

Good mapping of the traffic streams onto the resources constitutes better use of the money invested in the network.

Cost savings that results in a more efficient use of bandwidth resources helps to reduce the overall cost of operations. These reductions, in turn, help service providers and organizations gain an advantage over their competitors. This advantage becomes more important as the service provider market becomes even more competitive.

A more efficient use of bandwidth resources means that a provider could avoid a situation where some parts of the network are congested while other parts are underutilized.

Congestion Avoidance and Traffic Engineering

Network congestion can be addressed in two ways:

- Expansion of capacity or classical congestion control techniques (queuing, rate limiting, and so on)
- Traffic engineering, if the problems result from inefficient resource allocation

The focus of TE is not on congestion that is created as a result of a short-term burst, but on congestion problems that are prolonged.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—2.7

TE does not solve temporary network congestion that is caused by traffic bursts. This type of problem is better managed by an expansion of capacity or by classic techniques such as various queuing algorithms, rate limiting, and intelligent packet dropping. TE does not solve problems when the network resources themselves are insufficient to accommodate the required load.

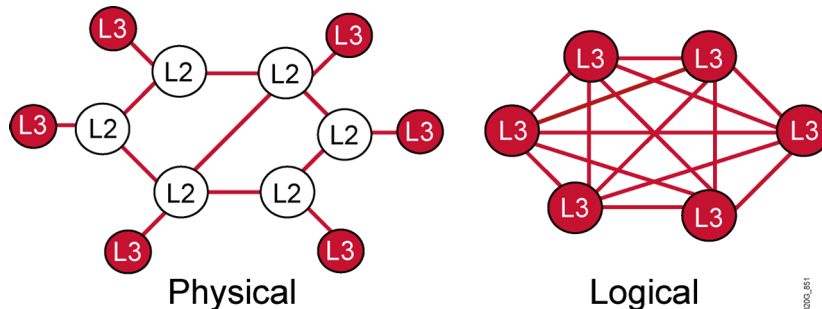
TE is used when the problems result from inefficient mapping of traffic streams onto the network resources. In such networks, one part of the network suffers from prolonged congestion, possibly continuously, while other parts of the network have spare capacity.

Traffic Engineering with a Layer 2 Overlay Model

This topic describes Traffic Engineering with a Layer 2 Overlay Model

Traffic Engineering with a Layer 2 Overlay Model

- The use of the explicit Layer 2 transit layer allows very exact control of the way that traffic uses the available bandwidth.
- PVCs or SVCs carry traffic across Layer 2.
- Layer 3 at the edge sees a complete mesh.



In the Layer 2 overlay model, the routers (Layer 3 devices) are overlaid on the Layer 2 topology. The routers are not aware of the physical structure and the bandwidth that is available on the links. The IGP views the Layer 2 PVCs or switched virtual circuits (SVCs) as point-to-point links and makes its forwarding decisions accordingly.

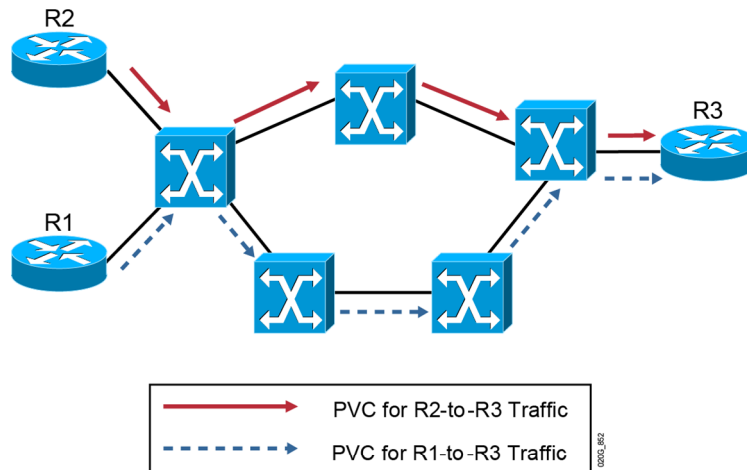
All traffic engineering is done at Layer 2. PVCs are carefully engineered across the network, normally using an offline management system. SVCs are automatically established by using signaling, and their way across the Layer 2 network is controlled by integrated path determination, such as the Private Network-to-Network Interface (PNNI) protocol.

In the Layer 2 overlay model, PVCs or SVCs carry the traffic across the network. With a Frame Relay network, PVC setup is usually made using a management tool. This tool helps the network administrator calculate the optimum path across the Layer 2 network, with respect to available bandwidth and other constraints that may be applied on individual links.

ATM may use the same type of tools as Frame Relay for PVC establishment, or may use the SVC approach, where routers use a signaling protocol to dynamically establish an SVC.

If the Layer 2 network provides a full mesh between all routers, the Layer 3 IGP sees all the other routers as directly connected and is likely to use the direct logical link whenever it forwards a packet to another router. The full mesh gives Layer 2 full control of the traffic load distribution. Manual engineering of PVCs and the configuration of PNNI parameters are the tools that allow very exact control of the way traffic uses the available bandwidth.

Traffic Engineering with a Layer 2 Overlay Model: Example



Traffic engineering in Layer 2, using the overlay model, allows detailed decisions about which link should be used to carry various traffic patterns.

In this example, traffic from R2 to R3 uses the top PVC (solid arrows), which takes the shortest path using the upper transit switch.

However, traffic from R1 to R3 uses the bottom PVC (dashed arrows), which does not take the shortest path. TE on Layer 2 has been applied to let the second PVC use links that would otherwise have been underutilized. This approach avoids overutilization of the upper path.

Traffic Engineering with a Layer 2 Overlay Model (Cont.)

Drawbacks of the Layer 2 overlay solution

- Extra network devices
- More complex network management:
 - Two-level network without integrated network management
 - Additional training, technical support, field engineering
- IGP routing scalability issue for meshes
- Additional bandwidth overhead (“cell tax”)
- No differential service (class of service)

© 2012 Cisco and/or its affiliates. All rights reserved.

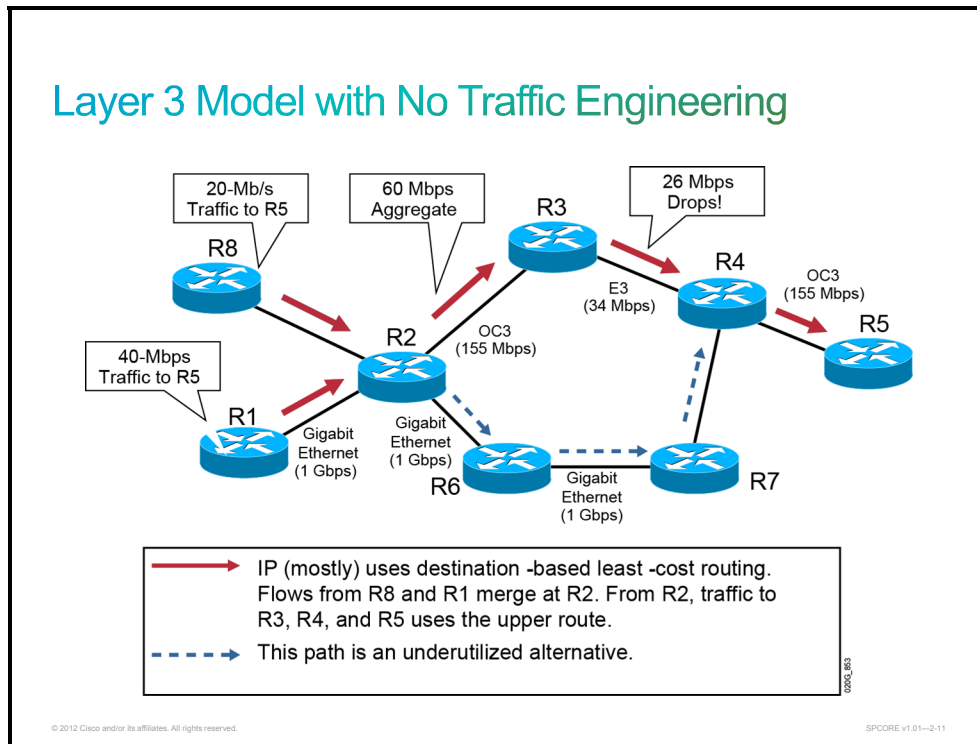
SPCORE v1.01-2-10

Using the Layer 2 overlay model has several drawbacks:

- The routers are not physically connected to other routers. The Layer 2 network introduces the need for an additional device, the ATM or Frame Relay switch.
- Two networks must be managed. The Layer 2 network requires its own management tools, which manage several other tasks, and support TE as well. At the same time, the router network (Layer 3), with its IGP and tuning parameters, must be managed. Both of these management tasks require trained staff for technical support and in the field.
- The Layer 3 network must be highly meshed to take advantage of the benefits that are provided by the Layer 2 network. The highly meshed network may cause scalability problems for the IGP due to the large number of IGP neighbors.
- Overlay networks always require an extra layer of encapsulation. A Frame Relay header must be added to the IP packets, or, when ATM is used, the IP packet must be segmented into cells, each of which must have its own header. The extra layer of encapsulation causes bandwidth overhead.
- The Layer 2 devices do not have any Layer 3 knowledge. After the router has transmitted the IP packet across the physical link to the first switch, all the IP information is unknown to the Layer 2 devices (ATM/Frame Relay switches). When congestion does occur in the Layer 2 network, the switches have no ability to selectively discard IP packets or to requeue them. Thus, no IP differentiated services can be used within the Layer 2 switch network.

Layer 3 routing model without Traffic Engineering

This topic describes a Layer 3 routing model without Traffic Engineering



If the same network topology is created using routers (Layer 3 devices), TE must be performed differently. In the example here, if no traffic engineering is applied to the network, traffic from both R8 and R1 toward R5 will use the least-cost path (the upper path, which has one less hop). This flow may result in the overutilization of the path R2, R3, R4, R5, while the lower path R2, R6, R7, R4, R5 (with the one extra hop) will be underutilized.

Traffic Engineering with a layer 3 routing model

This topic describes Traffic Engineering with a layer 3 routing model

Traffic Engineering with a Layer 3 Model

The “destination-based” forwarding paradigm is clearly inadequate:

- Path computation that is based only on an IGP metric is not enough.
- Support for “explicit” routing (source routing) is not available.
- The supported alternatives are static routes and policy routing.
- It does provide controlled backup and recovery.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2-12

The destination-based forwarding paradigm that is currently used in Layer 3 networks cannot resolve the problem of overutilization of one path while an alternate path is underutilized.

The IGP uses its metric to compute a single best way to reach each destination. There are problems with Layer 3 TE:

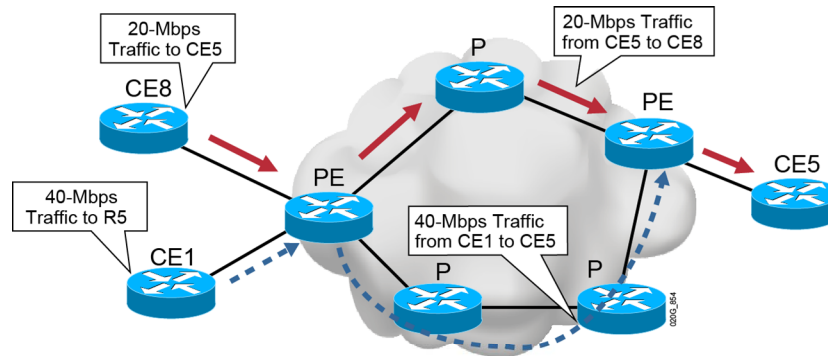
- IP source routing could be used to override the IGP-created routing table in each of the intermediate routers. However, in a service provider network, source routing is most often prohibited. The source routing would also require the host to create the IP packets to request source routing. The conclusion is that source routing is not an available tool for TE.
- Static routing, which overrides the IGP, can be used to direct some traffic to take a different path than that of other traffic. However, static routing does not discriminate among various traffic flows based on the source. Static routing also restricts how redundancy in the network can be used, and it is not a scalable solution.
- Policy-based routing (PBR) is able to discriminate among packet flows, based on the source, but it suffers from low scalability and the same static routing restrictions on using redundancy.

Traffic Engineering with the MPLS TE Model

This topic describes Traffic Engineering with the MPLS TE Model

Traffic Engineering with the MPLS TE Model

- A tunnel is assigned labels that represent the path (LSP) through the system.
- Forwarding within the MPLS network is based on the labels (no Layer 3 lookup).



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2-13

In the MPLS TE implementation, routers use MPLS label switching with TE.

The aim is to control the paths along which data flows, rather than relying simply on destination-based routing. MPLS TE uses tunnels to control the data flow path. An MPLS TE tunnel is simply a collection of data flows that share some common attribute. This attribute might be all traffic sharing the same entry point to the network and the same exit point.

A TE tunnel maps onto an MPLS label-switched path (LSP). After the data flows and the TE tunnels are defined, MPLS technology is used to forward traffic across the network. Data is assigned an MPLS TE, which defines the route for traffic to take through the network. The packets that are forwarded under MPLS TE have a stack of two labels that are imposed by the ingress router. The topmost label identifies a specific LSP or TE tunnel to use to reach another router at the other end of the tunnel. The second label indicates what the router at the far end of the tunnel should do with the packet.

Traffic Engineering with the MPLS TE Model (Cont.)

- The MPLS TE LSPs are created by RSVP.
- The actual path can be specified:
 - Explicitly, as defined by the system administrator
 - Dynamically, as defined using the underlying IGP protocol

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2-14

For MPLS TE, manual assignment and configuration of the labels can be used to create LSPs to tunnel the packets across the network on the desired path. However, to increase scalability, the Resource Reservation Protocol (RSVP) is used to automate the procedure.

By selecting the appropriate LSP, a network administrator can direct traffic via explicitly indicated routers. The explicit path across identified routers provides benefits that are similar to those of the overlay model, without introducing a Layer 2 network. This approach also eliminates the risk of running into IGP scalability problems due to the many neighbors that exist in a full mesh of routers.

MPLS TE provides mechanisms equivalent to those described previously in this lesson, along with the Layer 2 overlay network. For circuit-style forwarding, instead of using ATM or Frame Relay virtual circuits, the MPLS TE tunnel is used. For signaling, RSVP is used with various extensions to set up the MPLS TE tunnels.

For constraint-based routing (CBR) that is used in MPLS TE, either Intermediate System-to-Intermediate System (IS-IS) or Open Shortest Path First (OSPF) with extensions is used to carry resource information, such as available bandwidth on the link. Both link-state protocols use new attributes to describe the nature of each link with respect to the constraints. A link that does not have the required resource is not included in the MPLS TE tunnel.

To actually direct the traffic onto the MPLS TE tunnels, network administrators need extensions to IS-IS and OSPF. Directing the traffic into tunnels results in the addition of entries in the Forwarding Information Base (FIB). The IP packets are directed into the MPLS TE tunnel by imposing the correct label stack.

MPLS TE Traffic Tunnels

This topic describes basic concept of MPLS TE traffic tunnels.

Traffic Tunnels: Concepts

The concept of MPLS TE traffic tunnels was introduced to overcome the limitations of hop-by-hop IP routing:

- A tunnel is an aggregation of traffic flows that are placed inside a common MPLS label-switched path.
- Flows are then forwarded along a common path within a network.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2-16

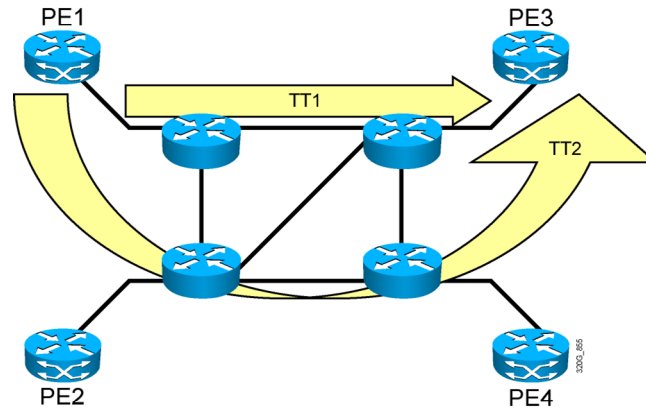
The aim of TE is to control the paths along which data flows, rather than relying simply on traditional destination-based routing. To fulfill this aim, the concept of a “traffic tunnel” has been introduced.

A traffic tunnel is simply a collection of data flows that share some common attribute:

- Most simply, this attribute might be the sharing of the same entry point to the network and the same exit point. In practice, in an ISP network, there is usually a definable data flow from the points of presence (POPs), where the customers attach to the ISP network. There are also the Internet exchange points (IXPs), where data typically leaves the ISP network to traverse the Internet.
- In a more complex situation, this attribute could be augmented by defining separate tunnels for different classes of service. For example, in an ISP model, leased-line corporate customers could be given a preferential throughput over dial-up home users. This preference might be greater guaranteed bandwidth or lower latency and higher precedence. Even though the traffic enters and leaves the ISP network at the same points, different characteristics could be assigned to these types of users by defining separate traffic tunnels for their data.

Traffic Tunnels: Concepts (Cont.)

- The unidirectional **single class of service** model encapsulates all of the traffic between an ingress and an egress router.
- The **different classes of service** model assigns traffic into separate tunnels with different characteristics.



Defining traffic trunks (tunnels) requires an understanding of the traffic flows in the network. By understanding the ingress and corresponding egress points, a picture of the traffic flows in the network can be produced.

In the example, there are two traffic tunnels (TT1 and TT2) that are defined for data from PE1 to PE3. These tunnels are unidirectional; they identify the traffic flows *from* PE1.

Note In practice, there are probably similar tunnels operating in the opposite direction, *to* PE1 *from* PE3.

There may also be tunnels that are defined from all the other routers to each other. Defining tunnels from every router in the network to every router might sound like an administrative nightmare. However, this is not usually the case for the following reasons:

- The routers that are identified as the tunnel headends are usually on the edge of the network. The traffic tunnels link these routers across the core of the network.
- In most networks it is relatively easy to identify the traffic flows, and they rarely form a complete “any-to-any” mesh.
- For example, in ISP networks, the traffic tunnels generally form a number of star formations, with their centers at the IXPs and the points at the POPs. Traffic in an ISP network generally flows from the customers that are connected at the POPs to the rest of the Internet (reached via the IXPs). A star-like formation can also exist in many networks centering on the data center. This tendency is found in both ISP networks (providing web-hosting services) and enterprise networks.

After the data flows, and therefore the traffic tunnels are defined, the technology that they use to forward the data across the network is MPLS. Data that enters a traffic tunnel is assigned an MPLS label-switched path (LSP). The LSP defines the route that is taken through the network.

Traffic Tunnels: Characteristics

- A traffic tunnel is distinct from the MPLS LSP through which it traverses:
 - More than one TE tunnel can be defined between two points:
 - Each tunnel may pick the same or different paths through the network.
 - Each tunnel will use different MPLS labels.
 - A traffic tunnel can be moved from one path onto another, based on resources in the network.
- A traffic tunnel is configured by defining its required attributes and characteristics.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2-17

In two important ways, traffic tunnels are distinct from the MPLS LSPs that they use:

- There is a one-to-one mapping of traffic tunnels onto MPLS LSPs. Two tunnels may be defined between two points and may happen to pick the same path through the network. However, they will use different MPLS labels.
- Traffic tunnels are not necessarily bound to a particular path through the network. As resources change in the core, or perhaps as links fail, the traffic tunnel may reroute, picking up a new MPLS LSP as it does.

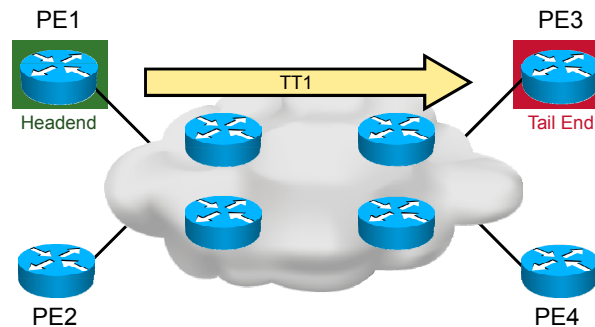
Configuring the traffic tunnels includes defining the characteristics and attributes that it requires. In fact, defining the characteristics and attributes of traffic tunnels is probably the most important aspect of TE. Without a specification of the requirements of the data in a traffic tunnel, the data might as well be left to route as it did previously, based only on destination information over the least-cost path.

Traffic Tunnels Attributes

This topic describes MPLS TE traffic tunnels attributes

Traffic Tunnels: Attributes

- Attributes are explicitly assigned through administrative action.
- A traffic tunnel has several characteristics:
 - Its ingress (headend) and egress (tail end) label switch routers
 - The forwarding equivalence class that is mapped onto it
 - A set of attributes that determine its characteristics



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2-18

A traffic tunnel is a set of data flows sharing some common feature, attribute, or requirement. If there is no characteristic in the data flow that is in common with some other flow, there is nothing to define that data as part of a flow or group of flows.

Therefore, the traffic tunnel must include attributes that define the commonality between the data flows making up the tunnel. The attributes that characterize a traffic tunnel include the following:

- **Ingress and egress points:** These points are, fundamentally, the routers at the ends of the tunnel. They are the most basic level of commonality of data flows, given that the flows in a tunnel all start in the same place and end in the same place.
- **Complex characteristics of the data flows:** Examples are bandwidth and latency and precedence requirements.
- **Class of data:** This attribute defines what data is part of this tunnel and what is not. This definition includes such characteristics as traffic flow, class of service, and application class.

The network administrator defines the attributes of a traffic tunnel when the tunnel itself is defined. However, some of these attributes are, in part, influenced by the underlying network and protocols.

Note MPLS TE setup is a control plane function.

Traffic Tunnels: Attributes (Cont.)

The administrator enters the relevant information (attributes) at the headend of the traffic tunnel:

- **Traffic parameter:** Resources required for tunnel (for example, required bandwidth)
- **Generic path selection and management:** Path can be administratively specified or computed by the IGP
- **Resource class affinity:** Can include or exclude certain links for certain traffic tunnels
- **Adaptability:** Traffic tunnel to be reoptimized?
- **Priority and preemption:** Importance of a traffic tunnel and possibility for a preemption of another tunnel
- **Resilience:** Desired behavior under fault conditions

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2-19

The general tunnel characteristics must be configured by the network administrator to create the tunnel. This configuration includes some or all of these attributes:

- **Traffic parameters:** Traffic parameters are the resources that are required by the tunnel, such as the minimum required bandwidth.
- **Generic path selection and management:** This category refers to the path selection criteria. The actual path that is chosen through the network could be statically configured by the administrator or could be assigned dynamically by the network, based on information from the IGP, which is IS-IS or OSPF.
- **Resource class affinity:** This category refers to restricting the choice of paths by allowing the dynamic path to choose only certain links in the network.

Note This restriction can also be accomplished by using the IP address exclusion feature.

- **Adaptability:** Adaptability is the ability of the path to reroute on failure or to optimize on recovery or discovery of a better path.
- **Priority and preemption:** Traffic tunnels can be assigned a priority (0 to 7) that signifies their “importance.” When you are setting up a new tunnel or rerouting, a higher-priority tunnel can tear down (preempt) a lower-priority tunnel; in addition, a new tunnel of lower priority may fail to set up because some tunnels of a higher priority already occupy the required bandwidth of the lower-priority tunnel.
- **Resilience:** Resilience refers to how a traffic tunnel responds to a failure in the network. Does it attempt to reroute around failures, or not?

Link Resource Attributes

This topic describes the link resource attributes

Network Links and Link Attributes

Resource attributes (link availability) are configured locally on the router interfaces:

- **Maximum bandwidth**
 - The amount of bandwidth available
- **Link affinity string**
 - Allows the operator to administratively include or exclude links in path calculations
- **Constraint-based specific metric**
 - The TE default metric

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.20

For the tunnel to dynamically discover its path through the network, the headend router must be provided with information on which to base this calculation. Specifically, it needs to be provided with this information:

- **Maximum bandwidth:** The maximum bandwidth is the amount of bandwidth that is available on each link in the network. Because there are priority levels for traffic tunnels, the availability information must be sent for each priority level, for each link. Including priority levels means that the path decision mechanism is given the opportunity to choose a link with some bandwidth already allocated to a lower-priority tunnel, forcing that lower-priority tunnel to be “bounced” off the link.
- **Link resource class:** For administrative reasons, the network administrator may decide that some tunnels are not permitted to use certain links. To accomplish this goal, for each link, a link resource class must be defined and advertised. The definition of the tunnel may include a reference to particular “affinity bits.” The tunnel affinity bits are matched against the link resource class to determine whether a link may be used as part of the LSP.
- **Constraint-based specific metric:** Each link has a cost or metric for calculating routes in the normal operation of the IGP. It may be that, when calculating the LSP for traffic tunnels, the link should use a different metric. Thus, a constraint-based specific metric may be specified.

Constraint-Based Path Computation

This topic describes Constraint-Based Path Computation

Constraint-Based Path Computation

- Constraint-based routing is demand-driven.
- Resource-reservation-aware routing paradigm:
 - Based on criteria including, but not limited to, network topology
 - Calculated at the edge of a network:
 - Modified Dijkstra algorithm at tunnel headend (CSPF [constraint-based SPF] or PCALC [path calculation])
 - Output is a sequence of IP interface addresses (next-hop routers) between tunnel endpoints

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.21

In traditional networks, the IGP calculates paths through the network, based on the network topology alone. Routing is destination-based, and all traffic to a given destination from a given source uses the same path through the network. That path is based simply on what the IGP regards as the “least cost” between the two points (source and destination).

MPLS TE employs CBR in which the path for a traffic flow is the shortest path that meets the resource requirements (constraints) of the traffic flow.

Constrained Shortest Path First (CSPF) or path calculation (PCALC) is an extension of shortest path first (SPF) algorithms. The path that is computed by using CSPF or PCALC is the shortest path fulfilling a set of constraints.

CBR behaves in these ways:

- It augments the use of link cost by also considering other factors, such as bandwidth availability or link attributes, when choosing the path to a destination.
- It tends to be carried out at the edge of the network, discovering a path across the core to some destination elsewhere at the other edge of the network. Typically, this discovery uses the CSPF calculation (a version of SPF that is used by IS-IS and OSPF, that considers other factors in addition to cost, such as bandwidth availability).
- It produces a sequence of IP addresses that correspond to the routers that are used as the path to the destination; these addresses are the next-hop addresses for each stage of the path.

A consequence of CBR is that, from one source to one destination, many different paths can be used through the network, depending on the requirements of those data flows.

Constraint-Based Path Computation (Cont.)

- Constraint-based routing takes into account these three elements :
 - Policy constraints associated with the tunnel and physical links
 - Physical resource availability
 - Network topology state
- Two types of tunnels can be established across those links with matching attributes:
 - Dynamic—Using the least-cost path computed by OSPF or IS-IS
 - Explicit—Using a path that is defined with Cisco IOS configuration commands

© 2012 Cisco and/or its affiliates. All rights reserved.

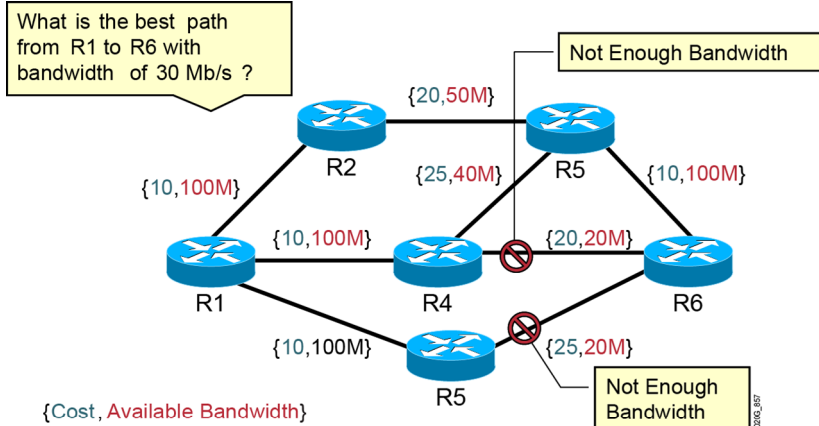
SPCORE v1.01-2.22

When choosing paths through the network, the CBR system takes into account these factors:

- The topology of the network, including information about the state of the links (the same information that is used by normal hop-by-hop routing)
- The resources that are available in the network, such as the bandwidth not already allocated on each link and at each of the eight priority levels (priority 0 to 7)
- The requirements that are placed on the constraint-based calculation that is defining the policy or the characteristics of this traffic tunnel

Of course, CBR is a dynamic process, which responds to a request to create a path and calculates (or recalculates) the path, based on the status of the network at that time. The network administrator can explicitly define the traffic tunnel and can also mix static and dynamic computation.

Constraint-Based Path Computation (Cont.)



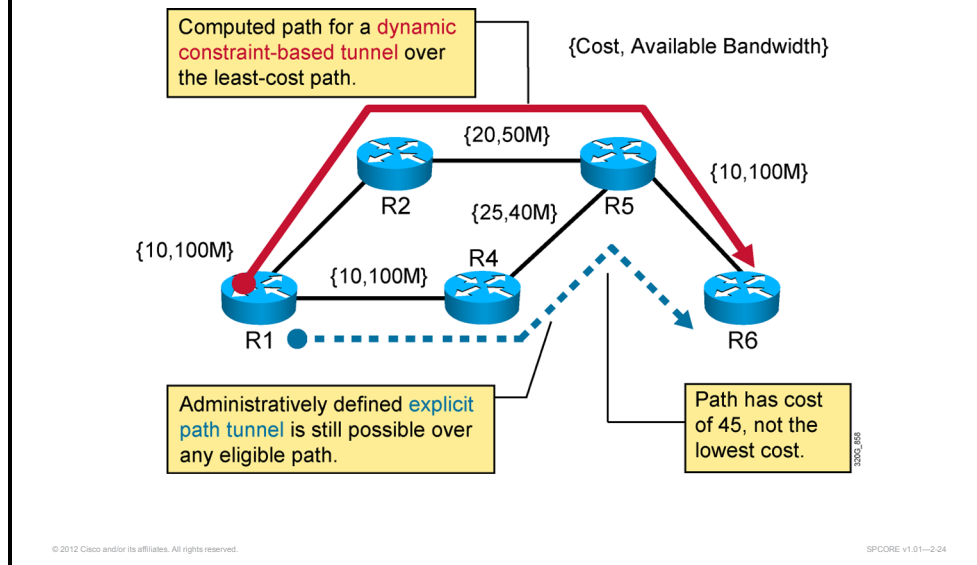
An example network is shown in the figure. Each link specifies a link cost for metric calculation and a bandwidth available for reservation; for example, a metric of 10 and an available bandwidth of 100 Mb/s is shown for the link between R1 and R2. Other than these criteria, no links are subject to any policy restriction that would disallow their use for creating traffic tunnels.

The requirement is to create a tunnel from R1 to R6 with a bandwidth of 30 Mb/s.

Based simply on the link costs, the least-cost path from R1 to R6 is R1-R4-R6 with a cost of 30. However, the link from R4 to R6 has only 20 Mb/s of bandwidth available for reservation and therefore cannot fulfill the requirements of the tunnel.

Similarly, the link R3-R6 has only 20 Mb/s available as well, so no paths can be allocated via R3.

Constraint-Based Path Computation (Cont.)



The diagram now shows only those links that can satisfy the requirement for 30 Mb/s of available bandwidth.

Over this topology, two tunnel paths are shown:

- The top (solid arrow) path shows the result of a dynamic constraint-based path calculation. The calculation ignores any links that do not satisfy the bandwidth requirement (those shown in the previous figure but not shown here, such as the connections to R3) and then executes a CSPF calculation on what remains. This calculation has yielded the path R1-R2-R5-R6 with a path cost of 40.
- The network administrator has statically defined the bottom (dashed arrow) path (R1-R4-R5-R6). Had the administrator attempted to define a path that did not have the required free bandwidth, tunnel establishment would have failed. This tunnel does indeed fulfill the minimum bandwidth requirement. However, adding the link costs yields a total of 45, which is not the lowest cost possible.

MPLS TE Process

This topic describes the MPLS TE process.

Traffic Engineering Processes

- Information distribution
- Path selection and calculation
- Path setup
- Tunnel admission control
- Forwarding of traffic on to tunnel
- Path maintenance

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.25

There are six TE processes to understand:

- **Information distribution:** Because the resource attributes are configured locally for each link, they must be distributed to the headend routers of traffic tunnels. These resource attributes are flooded throughout the network using extensions to link-state intradomain routing protocols, either IS-IS or OSPF. The flooding takes place under these conditions:
 - Link-state changes occur.
 - The resource class of a link changes (this could happen when a network administrator reconfigures the resource class of a link).
 - The amount of available bandwidth crosses one of the preconfigured thresholds.The frequency of flooding is bounded by the OSPF and IS-IS timers. There are up thresholds and down thresholds. The up thresholds are used when a new trunk is admitted. The down thresholds are used when an existing trunk goes away.
- **Path selection:** Path selection for a traffic tunnel takes place at the headend routers of the traffic tunnels. Using extended IS-IS or OSPF, the edge routers have knowledge of both network topology and link resources. For each traffic tunnel, the tail-end router starts from the destination of the traffic tunnel and attempts to find the shortest path toward the headend router (using the CSPF algorithm). The CSPF calculation does not consider the links that are explicitly excluded by the resource class affinities of the traffic tunnel or the links that have insufficient bandwidth. The output of the path selection process is an explicit route consisting of a sequence of label switching routers. This path is used as the input to the path setup procedure.

- **Path setup:** Path setup is initiated by the headend routers. RSVP is the protocol that establishes the forwarding state along the path that is computed in the path selection process. The headend router sends an RSVP PATH message for each traffic tunnel it originates.
- **Tunnel admission control:** Tunnel admission control manages the situation when a router along a computed path has insufficient bandwidth to honor the resource that is requested in the RSVP PATH message.
- **Forwarding of traffic to a tunnel:** Traffic can be forwarded to a tunnel by several means, including these:
 - Static routing
 - Policy routing from the global routing table
 - Autoroute
- **Path maintenance:** Path maintenance refers to two operations: path reoptimization and restoration.

Role of RSVP in Path Setup Procedures

This topic describes the Role of RSVP in Path Setup Procedures

Role of RSVP in Path Setup Procedures

- When the path has been determined, a signaling protocol is needed:
 - To establish and maintain label-switched paths (LSPs) for traffic tunnels
 - For creating and maintaining resource reservation states across a network (bandwidth allocation)
- Resource Reservation Protocol (RSVP) was adopted by the MPLS workgroup of the IETF.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.28

The result of the constraint-based calculation is a list of routers that form the path to the destination. The path is a list of IP addresses that identify each next hop along the path.

However, this list of routers is known only to the router at the headend of the tunnel, the one that is attempting to build the tunnel. Somehow, this now-explicit path must be communicated to the intermediate routers. It is not up to the intermediate routers to make their own CSPF calculations; they merely abide by the path that is provided to them by the headend router.

Therefore, some signaling protocol is required to confirm the path, to check and apply the bandwidth reservations, and finally to apply the MPLS labels to form the MPLS LSP through the routers. The MPLS working group of the IETF has adopted RSVP to confirm and reserve the path and apply the labels that identify the tunnel. Label Distribution Protocol (LDP) is used to distribute the labels for the underlying MPLS network.

Note RSVP is needed for both explicit and dynamic path setup.

Path Setup and Admission Control with RSVP

This topic describes the Path Setup Procedures using RSVP

Path Setup with RSVP

- When the path has been calculated, it must be signaled across the network.
 - Reserve any bandwidth to avoid “double booking” from other TE reservations.
 - Priority can be used to preempt low priority existing tunnels.
- RSVP is used to set up TE LSP.
 - PATH message (from head to tail) carries LABEL_REQUEST.
 - RESV message (from tail to head) carries LABEL.
- When the RESV message reaches the headend, the tunnel interface is up.
- RSVP messages exist for LSP teardown and error signaling.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.27

To signal the calculated path across the network, an RSVP PATH message is sent to the tail-end router by the headend router for each traffic tunnel the headend originates.

Note This process occurs in the MPLS control plane.

The RSVP PATH message carries the explicit route (the output of the path selection process) computed for this traffic tunnel, consisting of a sequence of label switching routers. The RSVP PATH message always follows this explicit route. Each intermediate router along the path performs trunk admission control after receiving the RSVP PATH message. When the router at the end of the path (tail-end router) receives the RSVP PATH message, it sends an RSVP RESV message in the reverse direction toward the headend of the traffic tunnel. As the RSVP RESV message flows toward the headend router, each intermediate node reserves bandwidth and allocates labels for the traffic tunnel. When the RSVP RESV message reaches the headend router, the LSP for the traffic tunnel is established.

RSVP messages also provide support for LSP teardown and error signaling.

RSVP and Tunnel Admission Control

- On receipt of PATH message:
 - Router checks whether there is bandwidth available to honor the reservation.
 - If bandwidth is available, then RSVP is accepted.
- On receipt of a RESV message:
 - Router actually reserves the bandwidth for the TE LSP.
 - If preemption is required, lower priority LSPs are torn down.
- OSPF or IS-IS updates are triggered.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.28

Trunk admission control is used to confirm that each device along the computed path has sufficient provisioned bandwidth to support the resource requested in the RSVP PATH message. When a router receives an RSVP PATH message, it checks whether there is enough bandwidth to honor the reservation at the setup priority of the traffic tunnel. Priority levels 0 to 7 are supported. If there is enough provisioned bandwidth, the reservation is accepted, otherwise the path setup fails. When the router receives the RSVP RESV message, it reserves bandwidth for the LSP. If preemption is required, the router must tear down existing tunnels with a lower priority. As part of trunk admission control, the router must do local accounting to keep track of resource utilization and trigger IS-IS or OSPF updates when the available resource crosses the configured thresholds.

Forwarding Traffic to a Tunnel

This topic describes the three methods to forward traffic to a tunnel

Forwarding Traffic to a Tunnel

- **IP routing** is separate from LSP routing and does not see internal details of the LSP.
- The traffic must be mapped to the tunnel:
 - **Static routing**: The static route in the IP routing table points to an LSP tunnel interface.
 - **Policy routing**: The next-hop interface is an LSP tunnel.
 - **Autoroute**: SPF enhancement
 - The headend sees the tunnel as a directly connected interface (for modified SPF only).
 - The default cost of a tunnel is equal to the shortest IGP metric, regardless of the path used.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPICORE v1.01-2.29

The traffic tunnel normally does not appear in the IP routing table. The IP routing process does not see the traffic tunnel, so the traffic tunnel is generally not included in any SPF calculations. The IP traffic can be mapped onto a traffic tunnel in four ways:

- Use static routes that point to the tunnel interfaces.
- Use PBR and setting the next hop for the destination to the tunnel interface.
- Use the autoroute feature, an SPF enhancement that includes the tunnel interface in the route calculation as well. The result of the autoroute feature is that the tunnel is seen at the headend (and only there) as a directly connected interface. The metric (cost) of the tunnel is set to the normal IGP metric from the tunnel headend to the tunnel endpoint (over the least-cost path, regardless of whether the tunnel is actually using the least-cost path).

Note With the autoroute feature, the traffic-engineered tunnel appears in the IP routing table as well, but this appearance is restricted to the tunnel headend only.

- Using forwarding adjacency, which allows the tunnel to be announced via OSPF or IS-IS as a point-to-point link to other routers. To be used for data forwarding, such a traffic tunnel has to be set up bidirectionally.

The first two options are not very flexible or scalable. The traffic for each destination that needs to use the tunnel must be manually mapped to the tunnel.

For example, when you are using static routes, the tunnel is used only for the explicit static routes. Any other traffic that is not covered by the explicit static routes, including traffic for the tail-end router (even though the tunnel terminates on that router), will not be able to use the tunnel; instead, it will follow the normal IGP path.

Autoroute

This topic describes the Autoroute feature

IP Forwarding Database Modification with Autoroute

- The autoroute feature enables the headend to see the LSP as a directly connected interface:
 - This feature is used only for the SPF route determination, not for the constraint-based path computation.
 - All traffic that is directed to prefixes topologically behind the tunnel endpoint (tail end) is forwarded onto the tunnel.
- **Autoroute affects the headend only**; other routers on the LSP path do not see the tunnel.
- The tunnel is treated as a directly connected link to the tail end.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.30

To overcome the problems that result from static routing configurations onto MPLS TE tunnels, the autoroute feature was introduced. The autoroute feature enables the headend router to see the MPLS TE tunnel as a directly connected interface. The headend uses the MPLS TE tunnel in its modified SPF computations.

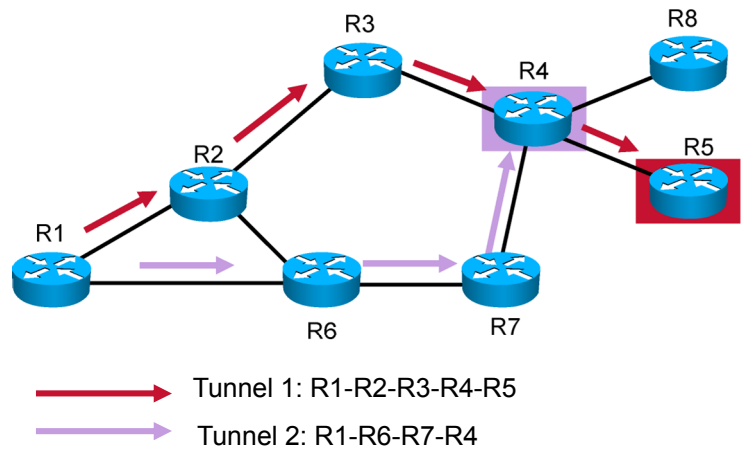
Note	The MPLS TE tunnel is used only for normal IGP route calculation (at the headend only) and is not included in any constraint-based path computation.

When the traffic tunnel is built, there is a directly connected link from headend to tail end.

The autoroute feature enables all the prefixes that are topologically behind the MPLS TE tunnel endpoint (tail end) to be reachable via the tunnel itself. This contrasts with static routing, where only statically configured destinations are reachable via the tunnel.

The autoroute feature affects the headend router only and has no effect on intermediate routers. These routers still use normal IGP routing for all the destinations.

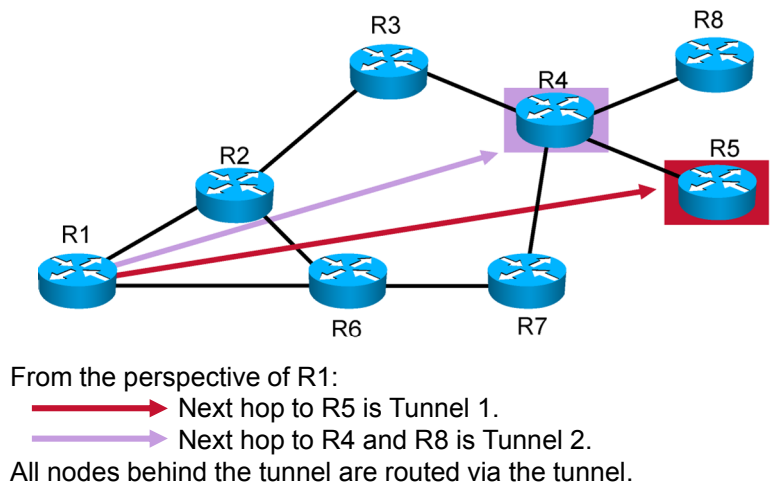
Autoroute Topology (OSPF and IS-IS)



The figure shows an example with two TE tunnels from R1. When the tunnels are up, R4 and R5 appear as directly connected neighbors to R1.

Note The tunnels are seen for routing purposes only by R1, the headend router. Intermediate routers do not see the tunnel, nor do they take it into consideration for route calculations.

Autoroute Topology (OSPF and IS-IS) (Cont.)



From the R1 perspective, next hop to router R5 is interface Tunnel 1, and next hop to router R4 and R8 is Tunnel 2. All nodes behind the tunnel are routed via the tunnel.

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- Traffic engineering is manipulating your traffic to fit your network.
- In TE with a Layer 2 overlay model, PVCs are carefully engineered across the network, normally using an offline management system.
- If TE is not used in a Layer 3 model, some links may be overutilized and other may be underutilized.
- The destination-based forwarding that is currently used in Layer 3 networks cannot resolve the problem of overutilization of one path while an alternate path is underutilized.
- The aim of MPLS TE is to control the path of traffic flow using MPLS labels and LSP.
- A traffic tunnel is a collection of data flows that share some common attribute.
- A traffic tunnel must include attributes that define the commonality between the data flows making up the tunnel.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.33

Summary (Cont.)

- A headend router must be provided with network link attributes in order to calculate a path through a network.
- Constraint-Based Path Computation finds a path for a traffic flow as the shortest path that meets the resource requirements (constraints) of the traffic flow.
- Tunnel admission control manages the situation when a router along a computed path has insufficient bandwidth to honor the resource that is requested in the RSVP PATH message.
- RSVP is used to confirm and reserve the path and apply the labels that identify the tunnel.
- The RSVP PATH message carries the explicit route (the output of the path selection process) computed for this traffic tunnel, consisting of a sequence of label switching routers.
- The IP routing process does not see the traffic tunnel, so the traffic tunnel is generally not included in any SPF calculations.
- The autoroute feature enables the headend router to see the MPLS TE tunnel as a directly connected interface.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.34

MPLS Traffic Engineering Operations

Overview

This lesson describes the details of link attribute propagation with an interior gateway protocol (IGP) and constraint-based path computation. The lesson also describes the details of Multiprotocol Label Switching (MPLS) traffic engineering (TE) tunnels, including path setup procedures and path maintenance. The lesson concludes with an explanation of the methods of assigning traffic into MPLS TE tunnels.

Objectives

Upon completing this lesson, you will be able to describe the concepts that allow service providers to map traffic through specific routes to optimize network resources, especially the bandwidth. You will be able to meet these objectives:

- Describe the attributes needed for performing Constraint-Based Path Computation
- List the four Link Resource Attributes
- Describe the maximum bandwidth and the maximum reservable bandwidth link resource attributes
- Describe the Link Resource Class attribute
- Describe the Constraint-Based Specific Link Metric attribute (Administrative Weight)
- List the Tunnel attributes
- Describe the Traffic Parameter and Path Selection and Management Tunnel Attributes
- Describe the Resource Class Affinity Tunnel Attributes
- Describe the Adaptability, Priority, Preemption Tunnel Attributes
- Describe the Resilience Tunnel Attributes
- Explain the implementation of TE Policies using Affinity Bits
- Explain the Propagating of MPLS TE Link Attributes using a Link-State Routing Protocol
- Explain Constraint-Based Path Computation
- Explain the Path Setup process

- Explain the RSVP functions in the Path Setup process
- Explain the Tunnel and Link Admission Control process
- Explain the Path Rerouting process
- List the three methods to forward traffic to a tunnel
- Explain using static routing to assign traffic to traffic tunnel
- Explain using autoroute to assign traffic to traffic tunnel
- Explain the need to adjust the tunnel default metric using either an absolute or relative value
- Explain adjusting the tunnel metrics using a relative and absolute value
- Describe the Forwarding Adjacency feature

Attributes used by Constraint-Based Path Computation

This topic describes the attributes needed for performing Constraint-Based Path Computation.

Constraint-Based Path Computation

Constraint-based path computation must be provided with several resource attributes before LSP path determination.

- **Link resource attributes** provide information on the resources of each link.
- **Traffic tunnel attributes** characterize the traffic tunnel.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.4

Constraint-based path computation, which takes place at the headend of the traffic-engineered tunnel, must be provided with several resource attributes before the label-switched path (LSP) is actually determined. These attributes include the following:

- Link resource attributes that provide information on the resources of each link
- Traffic tunnel attributes that characterize the traffic tunnel

MPLS TE Link Resource Attributes

This topic lists the four Link Resource Attributes.

MPLS TE Link Resource Attributes

- Maximum bandwidth
- Maximum reservable bandwidth
- Link resource class
- Constraint-based specific link metric

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.6

There are four link resource attributes:

- Maximum bandwidth
- Maximum reservable bandwidth
- Link resource class
- Constraint-based specific link metric

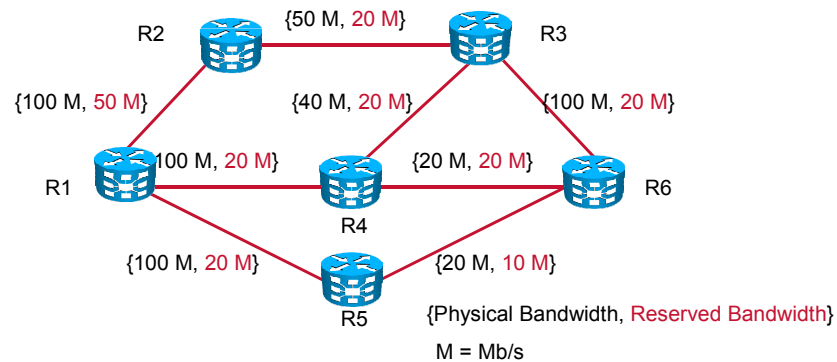
Each of these attributes will be discussed in detail.

MPLS TE Link Resource Attributes: Maximum Bandwidth and Maximum Reservable Bandwidth

This topic describes the maximum bandwidth and the maximum reservable bandwidth link resource attributes

MPLS TE Link Resource Attributes: Maximum Allocation Multiplier

- **Maximum bandwidth:** The maximum bandwidth that can be used on this link in this direction (physical link)
- **Maximum reservable bandwidth:** The maximum amount of bandwidth that can be reserved in this direction on this link



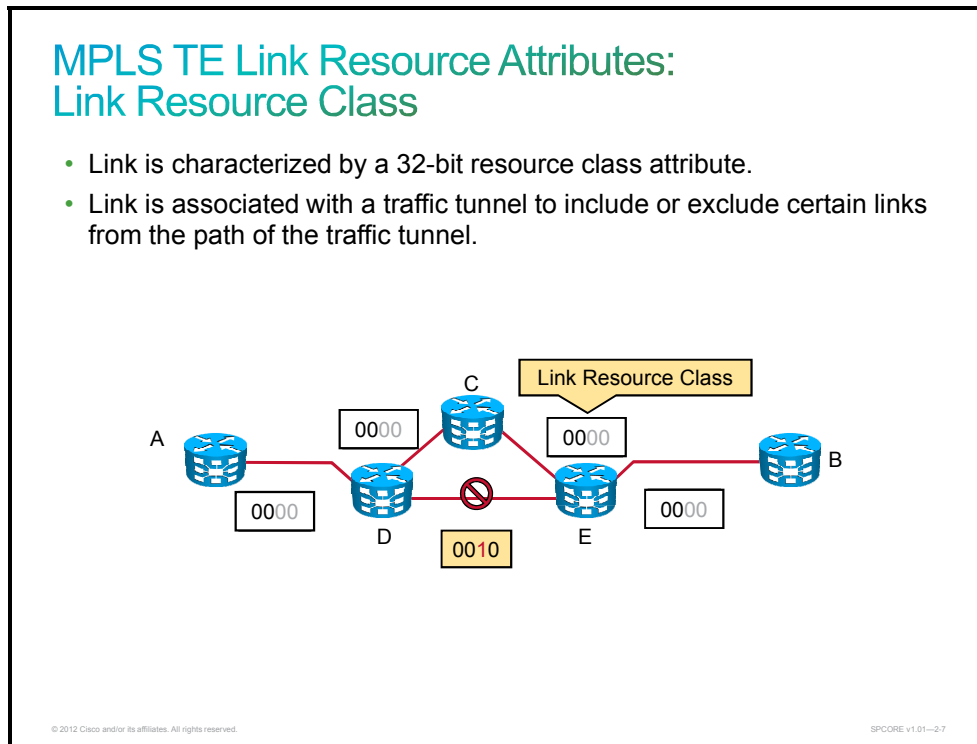
Among the link resource attributes, the most important is the maximum allocation multiplier. This attribute manages the amount of bandwidth that is available on a specified link. “Available” means not yet allocated (as opposed to not presently in use); the attribute is thus a measure of allocation, not utilization. Furthermore, because there are priority levels for traffic tunnels, this availability information needs to be configured for each priority level on the link. The bandwidth at the upper priority level is typically higher than at lower levels (0-7 levels). Because of oversubscriptions, the total amount of bandwidth can exceed the actual bandwidth of the link. There are three components to the link resource attribute:

- **Maximum bandwidth:** This component provides information on the maximum bandwidth that can be used on the link, per direction, given that the traffic tunnels are unidirectional. This parameter is usually set to the configured bandwidth of the link.
- **Maximum reservable bandwidth:** This component provides information on the maximum bandwidth that can be reserved on the link per direction. By default, it is set to 75 percent of the maximum bandwidth.
- **Unreserved bandwidth:** Unreserved bandwidth provides information on the remaining bandwidth that has not yet been reserved.

Note A higher priority can preempt a lower priority, but a lower priority cannot preempt a higher priority.

MPLS TE Link Resource Attributes: Link Resource Class

This topic describes the Link Resource Class attribute.



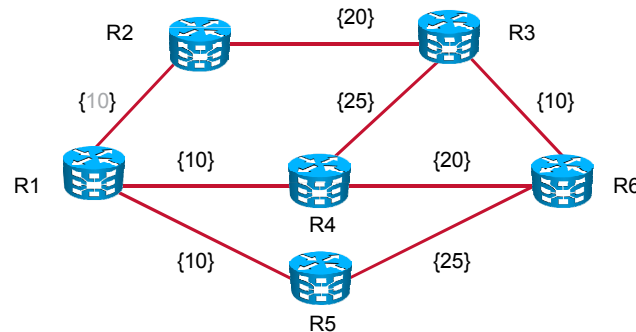
For each link, another link resource attribute, the link resource class, is provided. The link is characterized by a 32-bit link resource class attribute string, which is matched with the traffic tunnel resource class affinity attribute, and allows inclusion or exclusion of the link into the path of the tunnel.

MPLS TE Link Resource Attributes: Constraint-Based Specific Link Metric (Administrative Weight)

This topic describes the Constraint-Based Specific Link Metric attribute (Administrative Weight).

MPLS TE Link Resource Attributes: Constraint-Based Specific Link Metric

- This metric is administratively assigned to present a differently weighted topology to traffic engineering SPF calculations:
 - Administrative weight (TE metric)



Each link has a cost or metric for calculating routes in the normal operation of the IGP. It may be that, when calculating paths for traffic tunnels, the link should use a different metric than the IGP metric. Hence, a constraint-based specific link metric, the administrative weight, may be administratively assigned as well.

MPLS TE Tunnel Attributes

This topic lists the Tunnel attributes.

MPLS TE Tunnel Attributes

- Traffic parameter
- Generic path selection and management
- Tunnel resource class affinity
- Adaptability
- Priority
- Preemption
- Resilience

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.0

Seven tunnel attributes are available to influence path selection:

- Traffic parameter
- Generic path selection and management
- Tunnel resource class affinity
- Adaptability
- Priority
- Preemption
- Resilience

MPLS TE Tunnel Attributes: Traffic Parameter and Path Selection and Management Tunnel Attributes

This topic describes the Traffic Parameter and Path Selection and Management Tunnel Attributes.

MPLS TE Tunnel Attributes

- Traffic parameter:
 - Indicates the resource requirements (for example, bandwidth) of the traffic tunnel
- Generic path selection and management:
 - Specifies how the path for the tunnel is computed:
 - Dynamic LSP: Constraint-based computed paths based on a combination of bandwidth and policies
 - Explicit LSP: Administratively specified off line (typically using CLI)

© 2012 Cisco and/or its affiliates. All rights reserved.

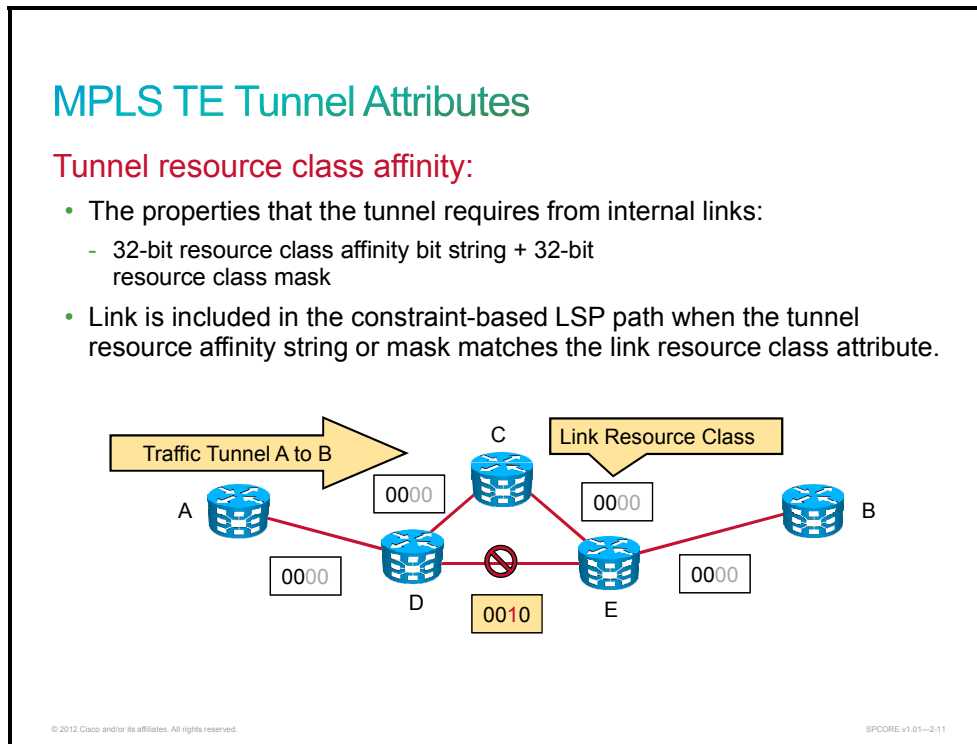
SPCORE v1.01-2-10

Two of the MPLS TE tunnel attributes affect the path setup and maintenance of the traffic tunnel:

- The traffic parameter (bandwidth) attribute specifies (among other traffic characteristics) the amount of bandwidth that is required by the traffic tunnel. The traffic characteristics may include peak rates, average rates, permissible burst size, and so on. From a TE perspective, the traffic parameters are significant because they indicate the resource requirements of the traffic tunnel. These characteristics are useful for resource allocation. A path is not considered for an MPLS TE tunnel if it does not have the bandwidth that is required.
- The path selection and management attribute (path selection policy) specifies the way in which the headend routers should select explicit paths for traffic tunnels. The path can be configured manually or computed dynamically by using the constraint-based path computation; both methods take the resource information and policies into account.

MPLS TE Tunnel Attributes: Tunnel Resource Class Affinity

This topic describes the Resource Class Affinity Tunnel Attributes.



The tunnel resource class affinity attribute allows the network administrator to apply path selection policies by administratively including or excluding network links. Each link may be assigned a resource class attribute. Resource class affinity specifies whether to explicitly include or exclude links with resource classes in the path selection process. The resource class affinity is a 32-bit string that is accompanied by a 32-bit resource class mask. The mask indicates which bits in the resource class need to be inspected. The link is included in the constraint-based LSP when the resource class affinity string or mask matches the link resource class attribute.

MPLS TE Tunnel Attributes: Adaptability, Priority, Preemption

This topic describes the Adaptability, Priority, Preemption Tunnel Attributes.

MPLS TE Tunnel Attributes

- **Adaptability:**
 - If reoptimization is enabled, then a traffic tunnel can be rerouted through different paths by the underlying protocols:
 - Primarily due to changes in resource availability
- **Priority:**
 - Relative importance of traffic tunnels
 - Determines the order in which path selection is done for traffic tunnels at connection establishment and under fault scenarios:
 - Setup priority: Priority for taking a resource
- **Preemption:**
 - Determines whether another traffic tunnel can preempt a specific traffic tunnel:
 - Hold priority: Priority for holding a resource

© 2012 Cisco and/or its affiliates. All rights reserved.

SPICORE v1.01--2.12

The adaptability attribute indicates whether the traffic tunnel should be reoptimized, and consequently rerouted to another path, primarily because of the changes in resource availability.

The priority and preemption tunnel attributes are closely associated and play an important role in competitive situations where traffic tunnels compete for link resources. Two types of priorities are assigned to each traffic tunnel:

- Setup priority (priority) defines the relative importance of traffic tunnels and determines the order in which path selection is done for traffic tunnels at connection establishment and during rerouting because of faulty conditions. Priorities are also important at implementation; they can permit preemption because they can be used to impose a partial order on a set of traffic tunnels according to which preemptive policies can be actualized.
- Holding priority (preemption) defines the preemptive rights of competing tunnels and specifies the priority for holding a resource. This attribute determines whether a traffic tunnel can preempt another traffic tunnel from a given path, and whether another traffic tunnel can preempt a specific traffic tunnel. Preemption can be used to ensure that high-priority traffic tunnels can always be routed through relatively favorable paths within a Differentiated Services (DiffServ) environment. Preemption can also be used to implement various prioritized restoration policies following fault events.

MPLS TE Tunnel Attributes: Resilience

This topic describes the Resilience Tunnel Attributes.

MPLS TE Tunnel Attributes

Resilience:

- Determines the behavior of a traffic tunnel under fault conditions:
 - Do not reroute the traffic tunnel.
 - Reroute through a feasible path with enough resources.
 - Reroute through any available path regardless of resource constraints.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2-13

Two additional tunnel attributes define the behavior of the tunnel in faulty conditions or if the tunnel becomes noncompliant with tunnel attributes (for example, the required bandwidth):

The resilience attribute determines the behavior of the tunnel under faulty conditions; it can specify the following behavior:

- Not to reroute the traffic tunnel at all
- To reroute the tunnel through a path that can provide the required resources
- To reroute the tunnel through any available path, irrespective of available link resources

Implementing TE Policies with Affinity Bits

This topic explains the implementation of TE Policies using Affinity Bits.

Implementing TE Policies with Affinity Bits

- Link is characterized by the link resource class
 - Default value of bits is 0
- Tunnel is characterized by:
 - Tunnel resource class affinity
 - Default value of bits is 0
 - Tunnel resource class affinity mask
 - (0 = do not care, 1 = care)
 - Default value of the tunnel mask is 0x0000FFFF

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.14

The policies during LSP computation can be implemented using the resource class affinity bits of the traffic tunnel and the resource class bits of the links over which the tunnel should pass (following the computed LSP).

Each traffic tunnel is characterized by a 32-bit resource class affinity string, which is accompanied by a respective resource class mask. The zero bits in the mask exclude the respective link resource class bits from being checked.

Each link is characterized by its resource class 32-bit string, which is set to 0 by default. The matching of the tunnel resource class affinity string with the resource class string of the link is performed during the LSP computation.

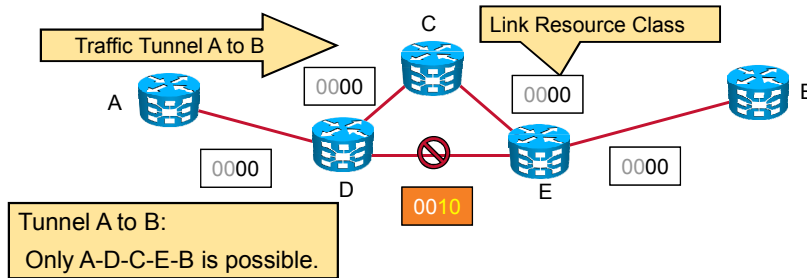
Note You can also exclude links or nodes by using the IP address exclusion feature when you are configuring tunnels.

Implementing TE Policies with Affinity Bits (Cont.)

Using Affinity Bits to Avoid Specific Links

Setting a link bit in the lower half drives all tunnels off the link, except those specially configured.

Tunnel affinity: bits = 0000, mask = 0011



This example shows a sample network with tunnel resource class affinity bits and link resource bits. For simplicity, only the four affinity and resource bits (of the 32-bit string) are shown. The tunnel should be established between routers A (headend) and B (tail end).

With the tunnel resource class affinity bits and the link resource class bits at their default values of 0, the constraint-based path computation would have two possible paths: A-D-E-B or A-D-C-E-B.

Because it is desirable to move all dynamically computed paths away from the link D-E, the link resource class bits were set to a value 0010 and the tunnel mask was set to 0011.

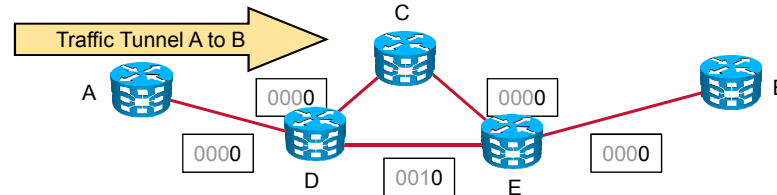
In the example, the tunnel mask requires only the lower two bits to match. The 00 of the traffic affinity does not match the 10 of the link D-E resource class and results in the exclusion of this link as a possible path for the tunnel. The only remaining alternative path is D-C-E, on which the default values of the resource class string (all zeros) match the tunnel affinity bits.

Implementing TE Policies with Affinity Bits (Cont.)

Using the Affinity Bit Mask to Allow all Links

A specific tunnel can then be configured to allow all links by clearing the bit in its affinity attribute mask.

Tunnel affinity: bits = 0000, mask = 0001



Tunnel A to B:

A-D-E-B and A-D-C-E-B are possible.

In this sample network, only the lower bit has been set in the tunnel mask. The tunnel affinity bits remain unchanged, as do the resource class bits on the D-E link.

The matching between the tunnel resource class affinity bits and the link resource class bits is done on the lowest bit only (because the mask setting is 0001). The 0 of the tunnel affinity bit (the lowest bit) matches with the 0 of the link resource class bit (the lowest bit) and therefore the link D-E remains in the possible path computation (along with the D-C-E link).

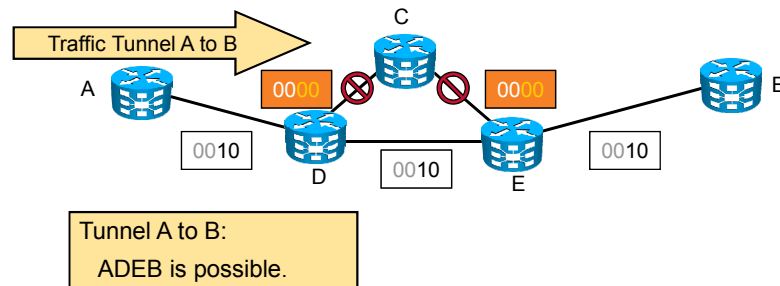
The path that will actually be used depends on other tunnel and link attributes, including the required and available bandwidth.

Implementing TE Policies with Affinity Bits (Cont.)

Using Affinity Bits to Dedicate Links to Specific Purposes

A specific tunnel can be restricted to only some links by turning on the bit in its affinity attribute bits.

Tunnel affinity: bits = 0010, mask = 0011



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2-17

This example deals with setting the tunnel resource class affinity bits and the link resource class bits to force the tunnel to follow a specific path. Links A-D-E-B are all configured with the resource class value 0010.

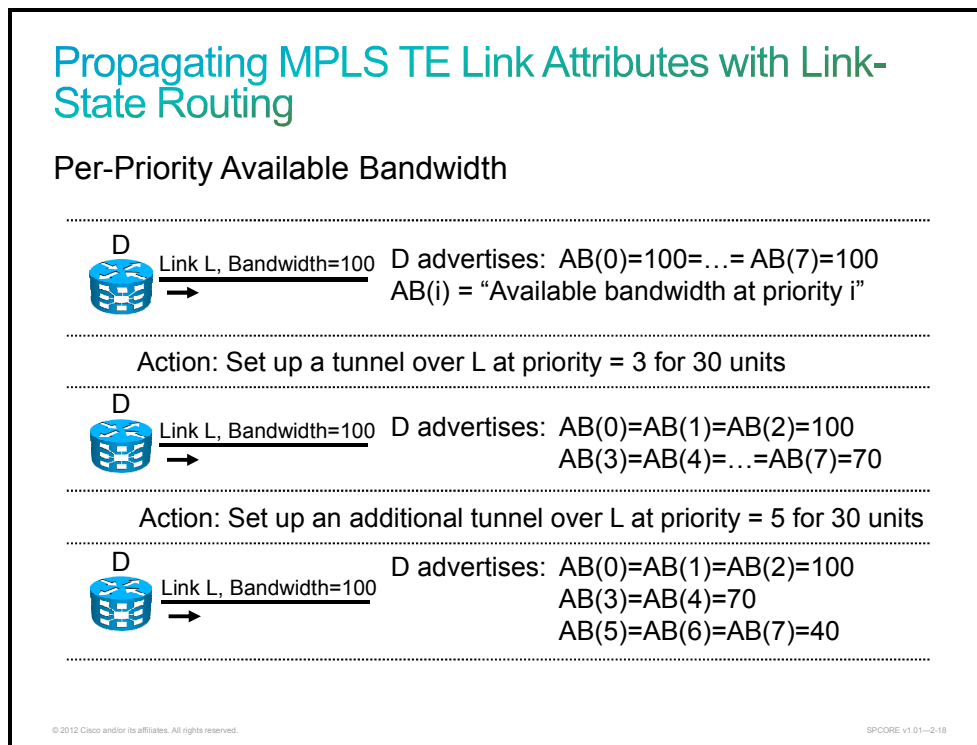
The tunnel resource class affinity bits are set to a value of 0010 and the mask to 0011. Only the lower two bits will be compared in the constraint-based path computation.

The 10 of the tunnel resource class affinity matches the 10 of the link resource class on all links that are configured with that value.

The 10 does not match the 00 that is set on the path D-C-E, and thus only one possible LSP remains (A-D-E-B).

Propagating MPLS TE Link Attributes with Link-State Routing Protocol

This topic explains the Propagating of MPLS TE Link Attributes using a Link-State Routing Protocol.



The link resource attributes must be propagated throughout the network to be available at the headend of the traffic tunnel when the LSP computation takes place.

Because the propagation (flooding) of the attributes can be achieved only by IGPs, OSPF and IS-IS were extended to support the MPLS TE features.

OSPF uses new link-state advertisements (opaque LSAs), and IS-IS uses new type, length, value (TLV) attributes in its link-state packets.

Another important factor in LSP computation is the available bandwidth on the link over which the traffic tunnel will pass. This bandwidth is configured per priority level (8 levels, 0 being the highest, 7 the lowest) and communicated in respective IGP link-state updates, again per priority.

When a certain amount of the bandwidth is reserved at a certain priority level, this amount is subtracted from the available bandwidth at that level and at all levels below. The bandwidth at upper levels remains unchanged.

In the figure, the maximum bandwidth is set to the bandwidth of the link, which is 100 (on a Fast Ethernet link). The system allows the administrator to set the available bandwidth (AB) to a higher value than the interface bandwidth. When the administrator is making a reservation, any bandwidth above the interface bandwidth will be rejected. The available bandwidth is advertised in the link-state packets of router D. The value is 100 at all priority levels before any tunnel is set up.

In the next part of the figure, a tunnel at priority level 3 that requires 30 units of bandwidth is set up across the link L. The available bandwidth at all priority levels above level 3 (0, 1, and 2) remains unchanged at 100. On all other levels, 30 is subtracted from 100, which results in an available bandwidth of 70 at priority level 3 and below (4-7).

Finally, another tunnel is set up at priority level 5 that requires 30 units of bandwidth across the link L. The available bandwidth at all priority levels above level 5 remains unchanged (100 on levels 0 to 2, and 70 on levels 3 and 4). On all other levels, 30 is subtracted from 70, which results in an available bandwidth of 40 at priority level 5 and below (6-7).

Note	All bandwidth reservation is done in the control plane and does not affect the actual traffic rates in the data plane.

Propagating MPLS TE Link Attributes with Link-State Routing (Cont.)

IGP resource flooding takes place in the following situations:

- Link-state changes
- Resource class of a link changes:
 - Manual reconfiguration
 - Amount of available bandwidth crosses one of the preconfigured thresholds
- Periodic (timer-based):
 - A node checks attributes; if they are different, it floods its update status
- On LSP setup failure

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2-19

The flooding of resource attributes by the IGP takes place along with certain conditions and events:

- When the link changes its state (up, down)
- When the resource class of the link changes because of a manual reconfiguration or because a preconfigured threshold is crossed by the available bandwidth
- When a node periodically checks resource attributes, and if the resource attributes were changed, the update is flooded
- When the LSP setup fails

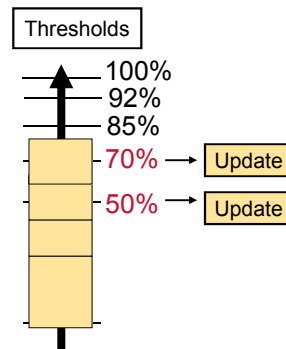
Propagating MPLS TE Link Attributes with Link-State Routing (Cont.)

Significant Change and Preconfigured Thresholds

- For stability reasons, rapid changes should not cause rapid generation of updates:

- Each time a threshold is crossed, an update is sent (different thresholds for up and down).

- It is possible that the headend node thinks it can signal an LSP tunnel via node X, while X does not have the required resources:
 - X refuses the LSP tunnel, and broadcasts an update of its status.



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.20

For stability purposes, significant rapid changes in available link resources should not trigger the updates immediately.

There is a drawback, however, in not propagating the changes immediately. Sometimes the headend sees the link as available for the LSP and includes the link in its path computation, even though the link may be down or does not have the required resource available. When the LSP is actually being established, a node with a link that lacks the required resources cannot establish the path and floods an immediate update to the network.

The thresholds for resources are set both for an up direction (resources exceeding the threshold) and a down direction (resources dropping below the threshold). When the threshold is crossed (in either direction), the node generates an update that carries the new resource information.

The figure shows the threshold values for the up direction (100 percent, 92 percent, 85 percent, 70 percent, and 50 percent) and two updates being sent out. In this example one update is immediately sent when the 50 percent margin is crossed. The second is sent when the 70 percent margin is crossed.

Constraint-Based Path Computation

This topic explains Constraint-Based Path Computation.

Constraint-Based Path Computation

When establishing a tunnel, the edge routers have knowledge of both network topology and link resources within their area:

- Two methods for establishing traffic tunnels:
 - Static
 - Dynamic path setup
- In both cases the result is an explicit route expressed as a sequence of interface IP addresses (for numbered links) or TE router IDs (for unnumbered links) in the path from tunnel endpoints.
- RSVP is used to establish and maintain constraint-based label-switched paths for traffic tunnels along an explicit path.

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01-2.21

The headend of the traffic tunnel has visibility of both the network topology and network resources. This information is flooded throughout the network via a link-state IGP.

The LSP for the traffic tunnel can be statically defined or computed dynamically. The computation takes the available resources and other tunnel and link attributes into account (thus, it represents constraint-based path computation). The result of the constraint-based path computation is a series of IP addresses that represent the hops on the LSP between the headend and tail end of the traffic tunnel.

For LSP signaling and the final establishment of the path, Resource Reservation Protocol (RSVP) is used.

Constraint-Based Path Computation (Cont.)

Dynamic constraint-based path computation is triggered by the headend of the tunnel:

- For a new tunnel
- For an existing tunnel whose current LSP has failed
- For an existing tunnel when you are doing reoptimization

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.22

Constraint-based path computation is always performed at the traffic tunnel headend. The computation is triggered for the following situations:

- A new tunnel
- An existing tunnel whose LSP setup has failed
- The reoptimization of an existing traffic tunnel

The LSP computation is restricted by several factors (constraint-based). The LSP can be computed only if these conditions are met:

- The endpoints of the tunnel are in the same Open Shortest Path First (OSPF) or Intermediate System-to-Intermediate System (IS-IS) area (due to link-state flooding of resources).
- The links that are explicitly excluded via the link resource class bit string, or that cannot provide the required bandwidth, are pruned from the computation.

Constraint-Based Path Computation (Cont.)

Path selection:

- CBR uses its own metric (administrative weight, or TE cost; by default equal to the IGP cost), only during constraint-based computation.
- If there is a tie, select the path with the highest minimum bandwidth.
- If there is still a tie, select the path with the smallest hop count.
- If everything else fails, then pick a path at random.
- LSP path setup: An explicit path is used by RSVP to reserve resources and establish an LSP path.
- Final result: **unidirectional MPLS TE tunnel, seen only at the headend router.**

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.23

Constraint-based path computation selects the path that the traffic tunnel will take, based on the administrative weight (TE cost) of each individual link. This administrative weight is by default equal to the IGP link metric. The value is used only during the constraint-based path computation.

If there are more candidates for the LSP (several paths with the same metric), then the selection criteria is as follows (in sequential order):

- The highest minimum bandwidth on the path takes precedence.
- The smallest hop count takes precedence.

If more than one path still exists after applying both of these criteria, a path is randomly chosen.

When the LSP is computed, RSVP is used to actually reserve the bandwidth, to allocate labels for the path, and finally to establish the path.

The result of a constraint-based path computation is a unidirectional MPLS TE tunnel (traffic tunnel) that is seen only at the tunnel endpoints (headend and tail end).

Constraint-Based Path Computation (Cont.)

- MPLS TE tunnel is not a link for link-state adjacency:
 - Establishment of a tunnel does not trigger any LSA announcements or a new SPF calculation (unless the forwarding adjacency feature is enabled).
 - The tunnel interface is used for MPLS TE tunnel creation and visualization, but the behavior of MPLS TE tunnels is different from other tunnel protocols (for example, GRE).
- Only traffic entering at the headend router will use the tunnel.
- **IP cost:** If autoroute is used, an MPLS TE tunnel in the IP routing table has a cost of the shortest IGP path to the tunnel destination (regardless of the LSP path).

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.24

From the perspective of IGP routing, the traffic tunnel is not seen as an interface at all, and is not included in any IGP route calculations (except for other IP tunnels such as generic routing encapsulation [GRE] tunnels). The traffic-engineered tunnel, when established, does not trigger any link-state update or any SPF calculation.

This behavior can be changed by defining two tunnels in a bidirectional way.

Cisco IOS Software and Cisco IOS XR Software use the tunnel mainly for visualization. The rest of the actions that are associated with the tunnel are done by MPLS forwarding and other MPLS TE-related mechanisms.

The IP traffic that will actually use the traffic-engineered tunnel is forwarded to the tunnel only by the headend of the tunnel. In the rest of the network, the tunnel is not seen at all (no link-state flooding).

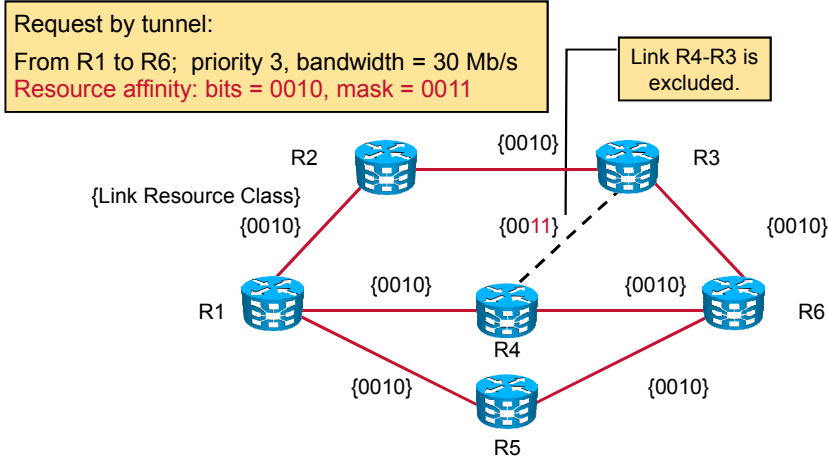
With the autoroute feature, the traffic tunnel has the following characteristics:

- Appears in the routing table
- Has an associated IP metric (cost equal to the best IGP metric to the tunnel endpoint)
- Is also used to forward the traffic for destinations behind the tunnel endpoint

Even with the autoroute feature, the tunnel itself is not used in link-state updates and other networks do not have any knowledge of it.

Constraint-Based Path Computation (Cont.)

Path Selection Considering Policy Constraints



This example of constraint-based path computation and LSP selection requires that the traffic tunnel be established between R1 (headend) and R6 (tail end). The traffic tunnel requirements are as follows:

- The required bandwidth at priority level 3 is 30 Mb/s.
- The resource class affinity bits are set to 0010, and the tunnel mask is 0011. The checking is done only on the lower two bits.

The link R4-R3 should be excluded from the LSP; therefore, its resource class bit string is set to 0011. When the traffic tunnel resource class affinity bits are compared to the link R4-R3 resource class bits, there is no match, and the link is effectively excluded from the LSP computation.

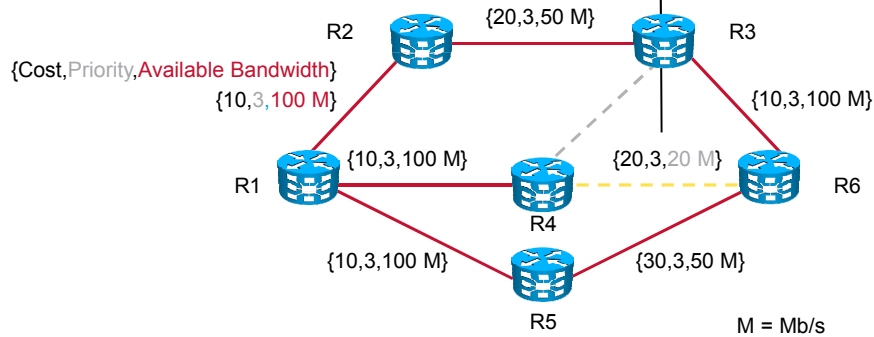
Constraint-Based Path Computation (Cont.)

Path Selection Considering Available Resources

Request by tunnel:

From R1 to R6; priority 3, bandwidth = 30 Mb/s
Resource affinity: bits = 0010, mask = 0011

Not Enough
Bandwidth



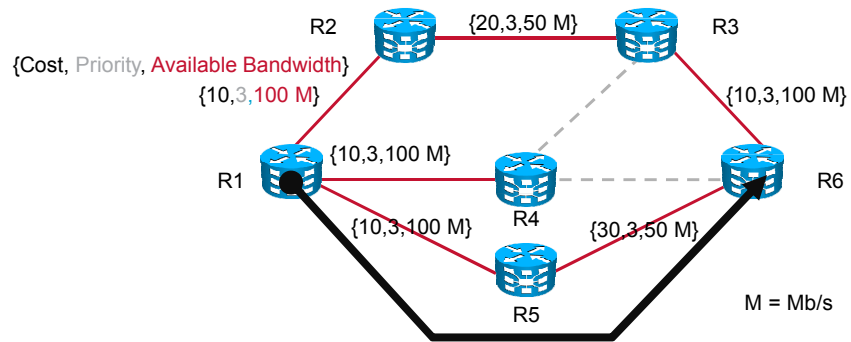
The next parameter that is checked during constraint-based path computation is the TE cost (administrative weight) of each link through which the tunnel will possibly pass. The lowest cost is calculated across the path R1-R4-R6; the overall cost is 30. All other possible paths have a higher overall cost.

When resources are taken into account, constraint-based path computation finds that on the lowest-cost path there is not enough bandwidth to satisfy the traffic tunnel requirements (30 Mb/s required, 20 Mb/s available). As a result, the link R4-R6 is effectively excluded from the LSP computation.

Constraint-Based Path Computation (Cont.)

Selecting the Best Path

The headend router has two possible paths with a total cost of 40: R1-R2-R3-R6 and R1-R5-R6, both offering at least 50 Mb/s (minimum bandwidth). Because of the smaller hop count, R1-R5-R6 is selected.



The resulting LSPs (after exclusion of the links that do not satisfy the traffic tunnel requirements) in the example are: R1-R2-R3-R6 and R1-R5-R6. Both paths have a total cost of 40, and the tie has to be resolved using the tie-break rules.

First, the highest minimum bandwidth on the path is compared. After the comparison, both paths are still candidates because both can provide at least 50 Mb/s of the bandwidth (the minimum bandwidth).

The next rule, the minimum number of hops on the LSP, is applied. Because the lower path (R1-R5-R6) has a lower hop count, this path is finally selected and the constraint-based computation is concluded.

The next step toward final establishment of the LSP for the traffic-engineered tunnel is the signaling of the path via RSVP.

Path Setup

This topic explains the Path Setup process.

Path Setup

- LSP path setup is initiated at the headend of a tunnel.
- The route (list of next-hop routers) is defined:
 - Statically defined
 - Computed by CBR
- The route is used by RSVP :
 - Assign labels
 - Reserve bandwidth on each link
- These tunnel attributes affect path setup:
 - Bandwidth
 - Priority
 - Affinity attributes

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.28

LSP setup is always initiated at the traffic tunnel headend. The explicit route for the traffic tunnel is composed of a list of next-hop routers toward the tunnel endpoint (or tail end). The LSP tunnels can be statically defined or computed with constraint-based routing (CBR) and thus routed away from network failures, congestion, and bottlenecks.

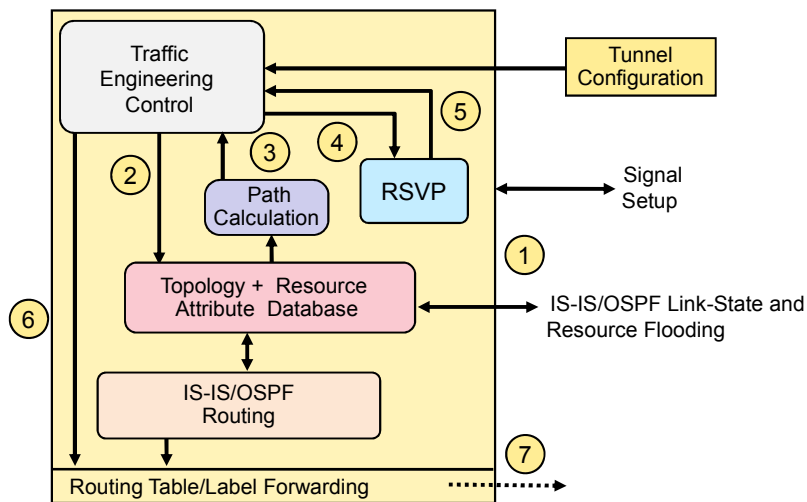
The explicit route is used by RSVP with TE extensions to assign labels and to reserve the bandwidth on each link.

Labels are assigned using the downstream-on-demand allocation mode.

Path setup is affected by the following tunnel attributes:

- Bandwidth
- Priority
- Affinity attributes

Path Setup (Cont.)



The figure shows a conceptual block diagram of the components of CBR and path computation. In the upper left corner is a TE control module, where the control algorithms run. The module looks at the tunnels that have been configured for CBR.

The TE control module periodically checks the CBR topology database (shown in the middle of the block diagram) to calculate the best current path from the current device to the tunnel destination.

After the path is calculated, the module transfers the path to the RSVP module to signal the circuit setup across the network. If the signaling succeeds, the signaling message eventually returns to the device, and RSVP announces back to the TE control module that the tunnel has been established.

Consequently, the TE control module tells the IGP routing module that the tunnel is available for use.

The IGP routing module includes the tunnel information in its routing table calculation and uses it to affect what routes are put into the routing table.

RSVP usage in Path Setup

This topic explains the RSVP functions in the Path Setup process.

RSVP Usage in Path Setup

RSVP makes resource reservations for both unicast and multicast applications:

- RSVP provides support for dynamic membership changes and automatic adaptation to routing changes.
- RSVP sends periodic refresh messages to maintain the state along the reserved path.
- RSVP sessions are used between routers, not between hosts.

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01--2.30

RSVP plays a significant role in path setup for LSP tunnels and supports both unicast and multicast applications.

RSVP dynamically adapts to changes either in membership (for example, multicast groups) or in the routing tables.

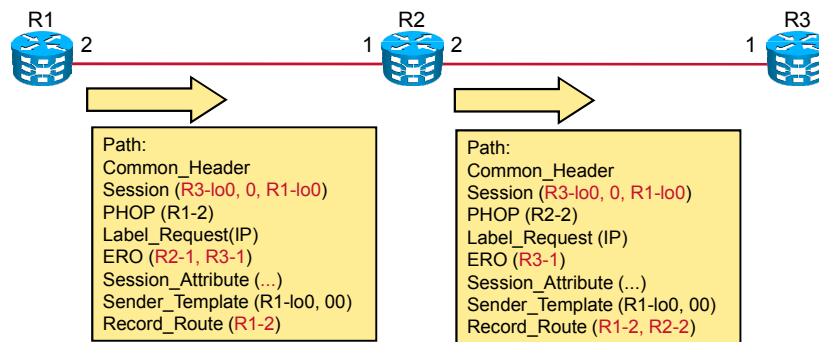
Additionally, RSVP transports traffic parameters and maintains the control and policy over the path. The maintenance is done by periodic refresh messages that are sent along the path to maintain the state.

In the normal usage of RSVP, the sessions are run between hosts. In TE, the RSVP sessions are run between the routers on the tunnel endpoints. The following RSVP message types are used in path setup:

- Path
- Resv
- PathTear
- ResvErr
- PathErr
- ResvConf
- ResvTear

When the RSVP Resv message flows back toward the sender, the intermediate nodes reserve the bandwidth and allocate the label for the tunnel. The labels are placed into the Label object of the Resv message.

Hop-by-Hop Path Setup with RSVP



In the example here, the LSP tunnel path setup is started by the RSVP Path message, which is initiated by the tunnel headend (R1). (Some of the important contents of the message are explained and monitored in the next example.)

The RSVP Path message contains several objects, including the session identification (R3-lo0, 0, R1-lo0 in the example), which uniquely identifies the tunnel. The traffic requirements for the tunnel are carried in the session attribute. The label request that is present in the message is handled by the tail-end router, which allocates the respective label for the LSP.

The explicit route object (ERO) is populated by the list of next hops, which are either manually specified or calculated by CBR (where R2-1 is used to represent the interface labeled “1” at the R2 router in the figure). The previous hop (PHOP) is set to the outgoing interface address of the router. The record route object (RRO) is populated with the same address as well.

Note The sender template is used in assigning unique LSP identifiers (R1-lo0 = loopback interface 0, which identifies the tunnel headend; 00 = the LSP ID). The same tunnel can take two possible LSPs (one primary and another secondary). In such a case the headend must take care of assigning unique IDs to these paths.

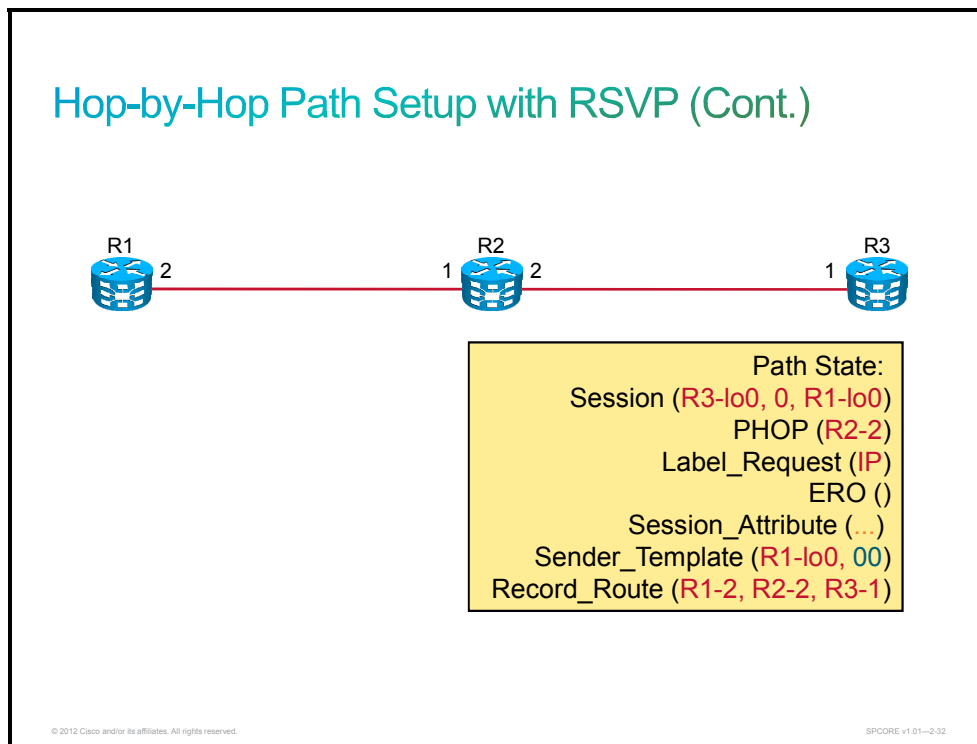
As the next-hop router (R2) receives the RSVP Path message, the router checks the ERO and looks into the L bit (loose) regarding the next-hop information. If this bit is set and the next hop is not on a directly connected network, the node performs a CBR calculation (path calculation, or PCALC) using its TE database and specifies this loose next hop as the destination.

In this way the ERO is augmented by the new results and forms a hop-by-hop path up to the next loose node specification.

Then the intermediate routers along the path (indicated in the ERO) perform the traffic tunnel admission control by inspecting the contents of the session attribute object. If the node cannot meet the requirements, it generates a PathErr message. If the requirements are met, the node is saved in the RRO.

Router R2 places the contents of the ERO into its path state block and removes itself from the ERO (R2 removes the R2-1 entry from the ERO). R2 adjusts the PHOP to the address of its own interface (the 2 interface at R2, R2-2) and adds the address (R2-2) to the RRO. The Path message is then forwarded to the next hop in the ERO.

Note Several other functions are performed at each hop as well, including traffic admission control.



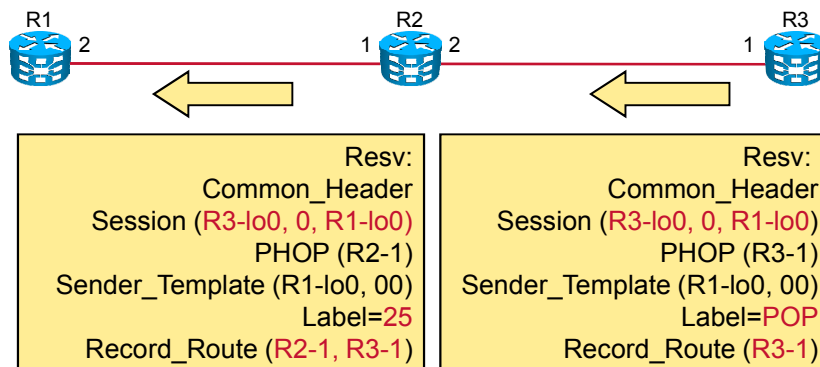
When the RSVP Path message arrives at the tail-end router (the endpoint of the tunnel), the label request message triggers the path label allocation. The label is placed in the corresponding label object of the RSVP Resv message that is generated. The RSVP message is sent back to the headend following the reverse path that is recorded in the ERO, and is stored at each hop in its path state block.

When the RSVP Path message arrives at the tail-end router (R3), the path state block is created and the ERO becomes empty (after removing the address of the router itself from the list), which indicates that it has reached the tail end of the tunnel. The RRO at this moment contains the entire path from the headend router.

The RSVP Resv message must be generated.

The label request object in the RSVP Path message requires the tail-end router to allocate a label for the specified LSP tunnel (session).

Hop-by-Hop Path Setup with RSVP (Cont.)



Because R3 is the tail-end router, it does not allocate a specific label for the LSP tunnel. The implicit null label is used instead (the value “POP” in the label object).

The PHOP in the RSVP Resv message is populated by the interface address of the tail-end router, and this address is copied to the RRO as well.

Note The RRO is reinitiated in the RSVP Resv message.

The Resv message is forwarded to the next-hop address in the path state block of the tail-end router. The next-hop information in the path state block was established when the Path message was traveling in the opposite direction (headend to tail end).

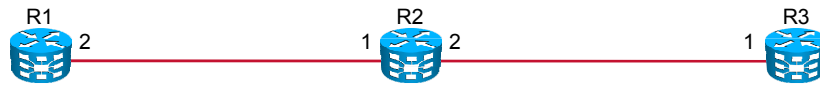
The RSVP Resv message travels back to the headend router. On each hop (in addition to the admission control itself) label handling is performed. As you can see from the RSVP Resv message that is shown in the figure, the following actions were performed at the intermediate hop (R2):

- The interface address of R2 was put into the PHOP field and added to the beginning of the RRO list.
- The incoming label (5) was allocated for the specified LSP.

Note The label switch table is not shown, but contains the information for label switching. In this particular case the label “5” is replaced with an implicit null label (POP).

The Resv message is forwarded toward the next hop that is listed in the path state block of the router. The next-hop information in the path state block was established when the Path message was traveling in the opposite direction (headend to tail end).

Hop-by-Hop Path Setup with RSVP (Cont.)



Resv state:
Session (R3-lo0, 0, R1-lo0)
PHOP (R2-1)
Sender_Template (R1-lo0, 00)
Label=5
Record_Route (R1-2, R2-1, R3-1)

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.34

When the RSVP Resv message arrives at the headend router (R1), the LSP setup is concluded. The label (5) that is allocated by the next-hop router toward the tunnel endpoint (PHOP = R2-1) is stored, and the explicit route that was taken by the tunnel is now present in the RRO. The LSP tunnel is established.

Tunnel and Link Admission Control

This topic explains the Tunnel and Link Admission Control process.

Tunnel and Link Admission Control

Invoked by RSVP Path message:

- First, it determines if resources are available.
- If bandwidth is not available, two things happen:
 - Link-level call admission control (LCAC) says no to RSVP.
 - PathErr message is sent.
- If bandwidth is available, this bandwidth is put aside in a waiting pool.
- Bandwidth is only reserved when a Resv message is received.
 - The Resv message travels from tunnel tail end to tunnel headend.
 - The Resv message triggers IGP information distribution when resource thresholds are crossed.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.35

One of the essential steps that is performed at each hop of the route to the LSP tunnel endpoint (the tunnel) is admission control, which is invoked by the RSVP Path message traveling from the headend to the tail-end router.

Each hop on the way determines if the available resources that are specified in the session attribute object are available.

If there is not enough bandwidth on a specified link through which the traffic tunnel should be established. The link-level call admission control (LCAC) module informs RSVP about the lack of resources, and RSVP generates an RSVP PathErr message with the code “Requested bandwidth unavailable.” Additionally, the flooding of the node resource information (by the respective link-state IGP) can be triggered.

If the requested bandwidth is available, the bandwidth is reserved and is put into a waiting pool for the Resv message to confirm the reservation. Additionally, if the resource threshold is exceeded, the respective IGP triggers the flooding of the resource information.

Tunnel and Link Admission Control (Cont.)

Preemption

- The process of LSP path setup may require the preemption of resources.
- LCAC notifies RSVP of the preemption.
- RSVP sends PathErr or ResvErr or both for the preempted tunnel.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2-38

During admission control, the priorities are checked as well.

If the requested bandwidth is available, but is in use by lower-priority sessions, then the lower-priority sessions (beginning with the lowest priority) may be preempted to free the necessary bandwidth. There are eight levels of priority, 0 being the highest, and 7 being the lowest.

When preemption is supported, each preempted reservation triggers a ResvErr or a PathErr message or both with the code “Policy control failure.”

Path Rerouting

This topic explains the Path Rerouting process.

Path Rerouting

- Path rerouting may result from these events:
 - Reoptimization due to a change in network resources
 - Link failures that affect the LSP path

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.37

Because the tunnel is not linked to the LSP that is carrying it, the actual path can dynamically change without affecting the tunnel.

Path rerouting may result from either of these two circumstances:

- Reoptimization due to a change in network resources
- Link failures that affect the LSP

Path Reoptimization

- **Problem:** Some previously unavailable resources become available, rendering the current path nonoptimal.
- **Solution:** Reoptimization
 - A periodic timer checks for the most optimal path.
 - A better LSP seems to be available:
 - The device attempts to signal the better LSP.
 - If successful, it replaces the old and inferior LSP with the new and better LSP.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.38

The LSP must be rerouted when there are physical (topology) failures or when certain changes in resource usage require it. As resources in another part of the network become available, the traffic tunnels may have to be reoptimized.

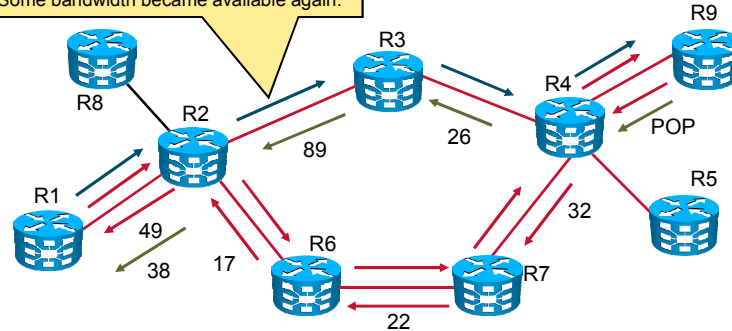
The reoptimization is done on a periodic basis. At certain intervals, a check for the most optimal paths for LSP tunnels is performed and if the current path is not the most optimal, tunnel rerouting is initiated.

The device (headend router) first attempts to signal a better LSP, and only after the new LSP setup has been established successfully, will the traffic be rerouted from the former tunnel to the new one.

Path Reoptimization (Cont.)

Nondisruptive Rerouting: Reoptimization

Some bandwidth became available again.



- Current path (ERO = R1-R2-R6-R7-R4-R9)
- New path (ERO = R1-R2-R3-R4-R9) is shared with current path and reserved for both paths.
- ← Until R9 gets new Path message, current Resv is refreshed. PathTear message can then be sent to remove the old Path (and release resources).

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.39

The figure shows how the nondisruptive rerouting of the traffic tunnel is performed. Initially, the ERO lists the LSP R1-R2-R6-R7-R4-R9, with R1 as the headend and R9 as the tail end of the tunnel.

The changes in available bandwidth on the link R2-R3 dictate that the LSP be reoptimized. The new path R1-R2-R3-R4-R9 is signaled, and parts of the path overlap with the existing path. Still, the current LSP is used.

Note On links that are common to the current and new LSP, resources that are used by the current LSP tunnel should not be released before traffic is moved to the new LSP tunnel, and reservations should not be counted twice (doing so might cause the admission control to reject the new LSP tunnel).

After the new LSP is successfully established, the traffic is rerouted to the new path and the reserved resources of the previous path are released. The release is done by the tail-end router, which initiates an RSVP PathTear message.

The labels that are allocated during the RSVP path setup are shown as well. The tail-end router assigns the implicit null (POP) label.

Path Rerouting: Link Failure

The Goal

- Repair at the headend of the tunnel in the event of failure of an existing LSP:
 - IGP or RSVP alarms the headend.
- New path for LSP is computed, and eventually a new LSP is signaled.
- Tunnel interface goes down if there is no LSP available for 10 sec.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.40

When a link that makes up a certain traffic tunnel fails, the headend of the tunnel detects the failure in one of two ways:

- The IGP (OSPF or IS-IS) sends a new link-state packet with information about changes that have happened.
- RSVP alarms the failure by sending an RSVP PathTear message to the headend.

Link failure detection, without any preconfigured or precomputed path at the headend, results in a new path calculation (using a modified SPF algorithm) and consequently in a new LSP setup.

Note The tunnel interface that is used for the specified traffic tunnel (LSP) goes down if the specified LSP is not available for 10 sec. In the meantime, the traffic that is intended for the tunnel continues using a broken LSP, which results in black hole routing.

Path Rerouting: Link Failure (Cont.)

Link Failure: What Happens

(Example: One link along a dynamic tunnel LSP path goes down.)

- The RSVP PathTear causes the headend to flag the LSP as dead.
- The RSVP session is cleared.
- The PCALC is triggered:
 - No alternative path is found:
 - Headend sets the tunnel down.
 - An alternative path is found:
 - A new LSP is directly signaled.
 - The adjacency table is updated for the tunnel interface.
 - The Cisco Express Forwarding table is updated for all entries that resolve on this tunnel adjacency.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-241

When the router along the dynamic LSP detects a link failure, it sends an RSVP PathTear message to the headend.

This message signals to the headend that the tunnel is down. The headend clears the RSVP session, and a new PCALC is triggered using a modified SPF algorithm.

There are two possible outcomes of the PCALC calculation:

- No new path is found. The headend sets the tunnel interface down.
- An alternative path is found. The new LSP setup is triggered by RSVP signaling, and adjacency tables for the tunnel interface are updated. Also, the CEF table is updated for all the entries that are related to this tunnel adjacency.

Assigning Traffic to Traffic Tunnels

This topic lists the three methods to forward traffic to a tunnel.

Traffic Flow Modifications

- CBR is used to find the path for an LSP tunnel.
- IP routing does not see internal details.
- Tunnels can be used for routing only if they are explicitly specified:
 - The static route in the IP routing table points to a selected LSP tunnel interface.
 - Advertise the tunnel to IP by using autoroute.
 - Policy routing sets the next-hop interface to an LSP tunnel.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.42

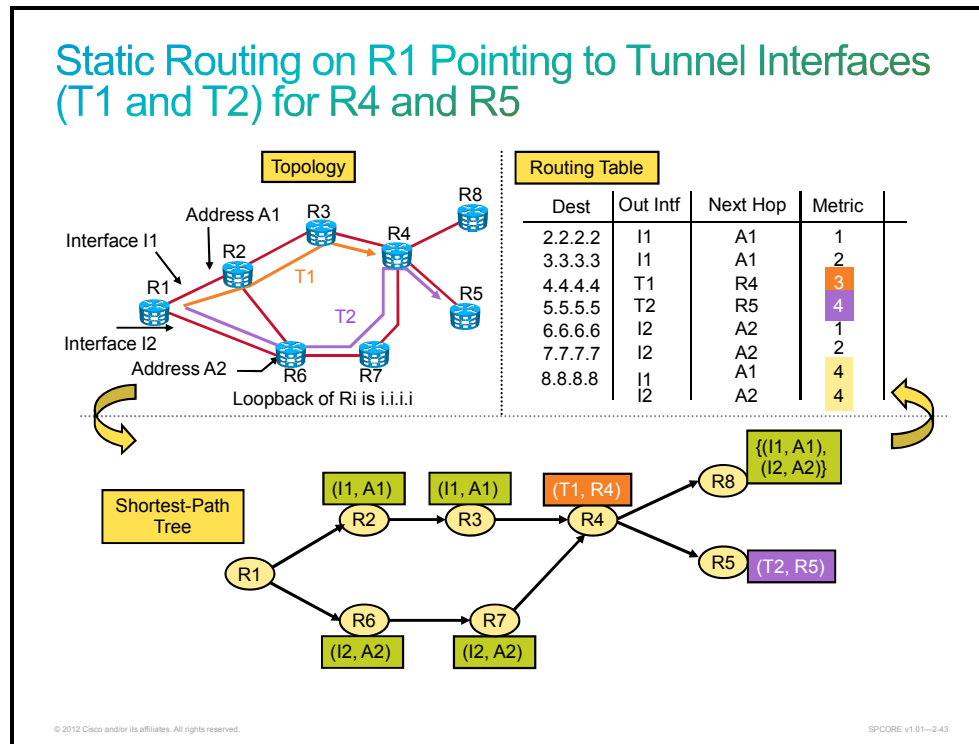
The LSP is computed by CBR, which takes the resource requirements into consideration as well. When the LSP is established for the tunnel, the traffic can flow across it. From the IP perspective, an LSP is a simple tunnel.

These engineered tunnels can be used for IP routing only if the tunnels are explicitly specified for routing in one of two ways:

- Via static routes that point to the tunnel
- Via policy routing that sets a next-hop interface to the tunnel

Using Static Routing to Assign Traffic to Traffic Tunnel

This topic explains using static routing to assign traffic to traffic tunnel.



The example topology here shows two engineered tunnels: T1 (between R1 and R4) and T2 (between R1 and R5). The loopback addresses on each router are in the form i.i.i.i where i is the router number (for example, the R5 loopback address is 5.5.5.5). The metric on each of the interfaces is set to 1.

R1 has two physical interfaces: I1 and I2, and two neighboring routers (next hops) with addresses of A1 and A2, respectively.

The routing table lists all eight loopback routes and associated information. Only the statically configured destinations (R4 and R5) list tunnels as their outgoing interfaces. For all other destinations the normal IGP routing is used and results in physical interfaces (along with next hops) as the outgoing interfaces towards these destinations. The metric to the destination is the normal IGP metric.

Note Even for the destination that is behind each of the tunnel endpoints (R8), the normal IGP routing is performed if there is no static route to the traffic-engineered tunnel.

The SPF algorithm calculates paths to destinations in its usual way; however, a constraint-based computation is performed for the paths and for the tunnels.

Autoroute

This topic explains using autoroute to assign traffic to traffic tunnel.

IP Forwarding Database Modification with Autoroute

- The Autoroute feature enables the headend to see the LSP as a directly connected interface:
 - Only for the SPF route determination, not for the constraint-based path computation.
 - All traffic that is directed to prefixes that are topologically behind the tunnel endpoint (tail end) is forwarded onto the tunnel.
- **Autoroute affects the headend only**; other routers on the LSP path do not see the tunnel.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.44

To overcome the problems that result from static routing configurations onto MPLS TE tunnels, the autoroute feature was introduced.

The autoroute feature enables the headend router to see the MPLS TE tunnel as a directly connected interface and to use it in its modified SPF computations.

The MPLS TE tunnel is used only for normal IGP route calculation (at the headend only) and is not included in any constraint-based path computation.

The autoroute feature enables all the prefixes that are topologically behind the MPLS TE tunnel endpoint (tail end) to be reachable via the tunnel itself (unlike with static routing, where only statically configured destinations are reachable via the tunnel).

The autoroute feature affects the headend router only and has no effect on intermediate routers. These routers still use normal IGP routing for all the destinations.

Autoroute: Path Selection Rules

- The cost of the TE tunnel is equal to the shortest IGP metric to the tunnel endpoint; the metric is tunable.
- The tunnel metric is used in the decision-making process:
 - If the tunnel metric is equal to, or lower than, the native IGP metric, the tunnel replaces the existing next hops; otherwise, the tunnel is not considered for routing.
 - If the tunnel metric is equal to other TE tunnels, the tunnel is added to the existing next hops (parallel paths).
- Tunnels can be load-balanced (Cisco Express Forwarding mechanism); the tunnel bandwidth factor is considered.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-245

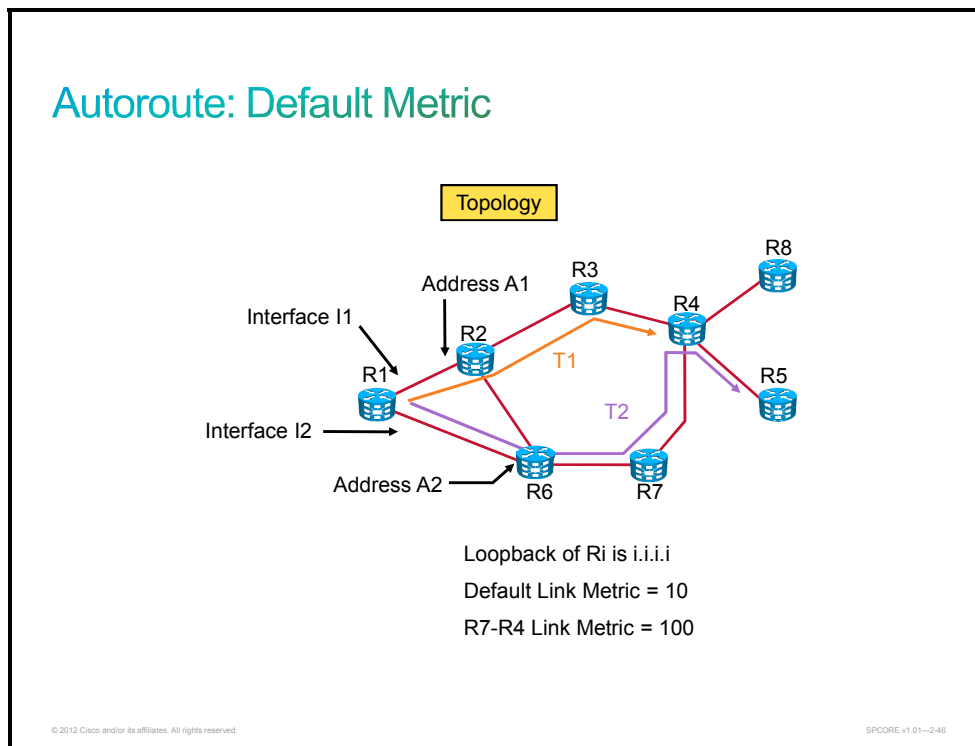
Because the autoroute feature includes the MPLS TE tunnel into the modified SPF path calculation, the metric of the tunnel plays a significant role. The cost of the tunnel is equal to the best IGP metric to the tunnel endpoint regardless of the LSP. The tunnel metric is tunable using either relative or absolute metrics.

During installation of the best paths to the destination, the tunnel metric is compared to other existing tunnel metrics and to all the native IGP path metrics. The lower metric is better and if the MPLS TE tunnel has an equal or lower metric than the native IGP metric, it is installed as a next hop to the respective destinations.

If there are tunnels with equal metrics, they are installed in the routing table and provide for load balancing. The load balancing is done proportionally to the configured bandwidth of the tunnel.

Autoroute: Default Metric

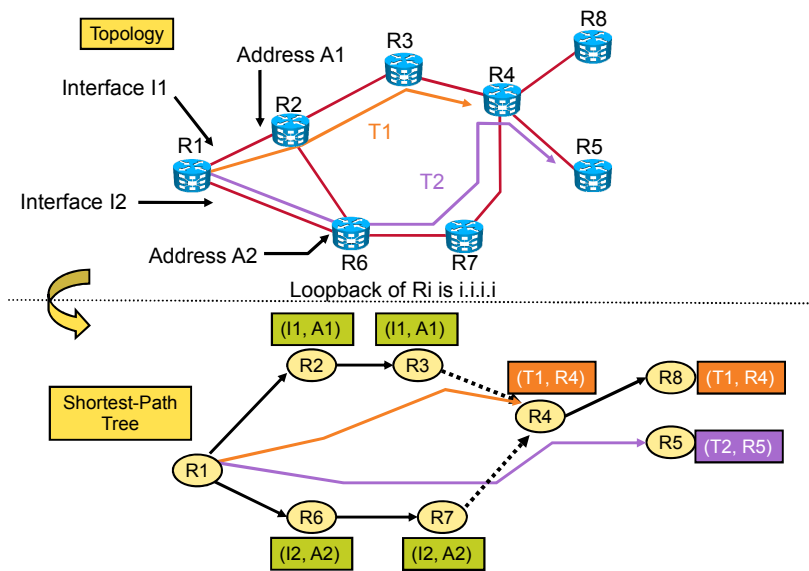
This topic describes default metric for TE tunnel.



The example topology shows two engineered tunnels: T1 (between R1 and R4) and T2 (between R1 and R5). The loopback addresses on each router are in the form i.i.i.i where i is the router number (for example, the R5 loopback address is 5.5.5.5). The metric on each of the interfaces is set to 10.

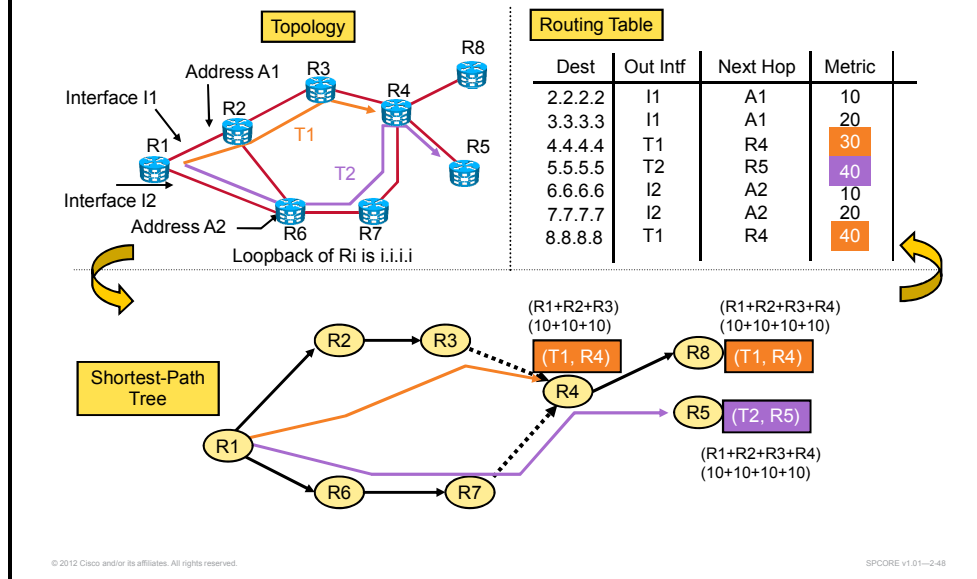
R1 has two physical interfaces, I1 and I2, and two neighboring routers (next hops) with addresses of A1 and A2, respectively.

Autoroute: Default Metric (Cont.)



This example shows the resulting shortest-path tree from router R1. In this situation the tunnels are seen for routing purposes only by the headend. Intermediate routers do not see the tunnel, nor do they take it into consideration for route calculations.

Autoroute: Default Metric (Cont.)



The routing table shows all destinations at the endpoint of the tunnel and behind it (R8) as reachable via the tunnel itself. The metric to the destination is the normal IGP metric.

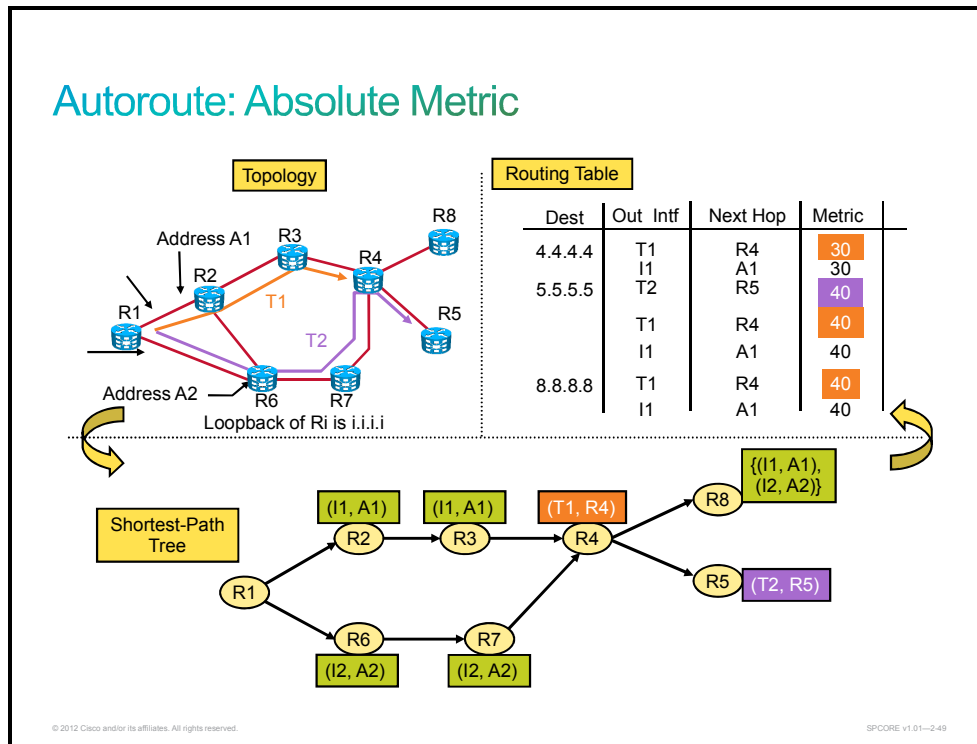
The LSP for T1 follows the path R1-R2-R3-R4 and the tunnel cost is the best IGP metric (30) to the tunnel endpoint. The metric to R8 is 40 (T1 plus one hop).

The LSP for T2 follows the path R1-R6-R7-R4-R5. Although the LSP passes through the R7-R4 link, the overall metric of the tunnel is 40 (the sum of metrics on the best IGP path R1-R2-R3-R4-R5).

In the routing table all the networks that are topologically behind the tunnel endpoint are reachable via the tunnel itself. Because, by default, the MPLS TE tunnel metric is equal to the native IGP metric, the tunnel is installed as a next hop to the respective destinations. This is the effect of the autoroute feature.

Autoroute: Relative and Absolute Metric

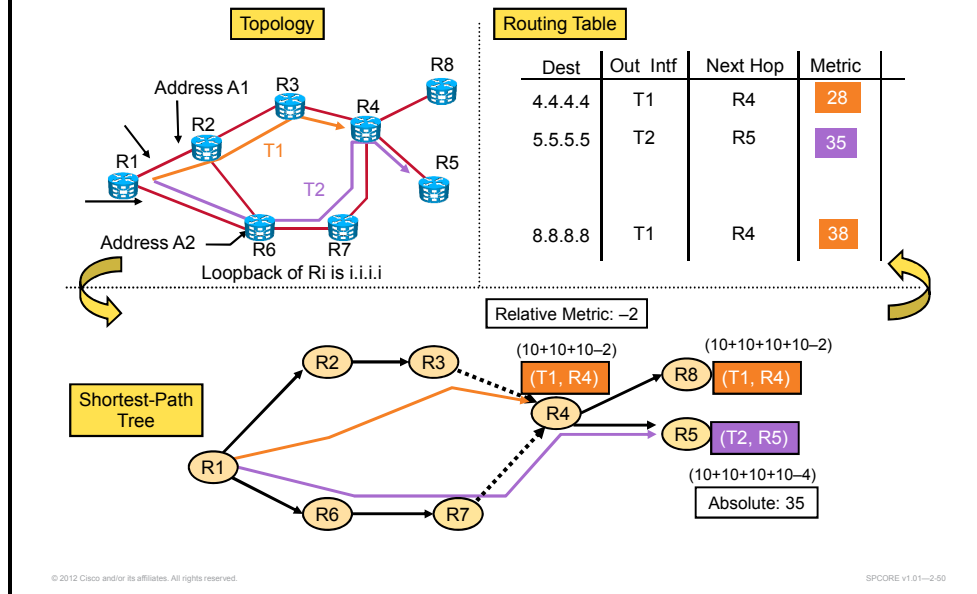
This topic explains adjusting the tunnel metrics using a relative and absolute value.



Depending on the ability of the IGP to support load sharing, the native IP path may also show up in the routing table as a second path option. In this example, there appear to be two paths to R4, while there is only one physical path. For R5 there appear to be three paths, two of which do not follow the desired tunnel path.

The tunnel metrics can be tuned, and either relative or absolute metrics can be used to resolve this issue.

Autoroute: Relative and Absolute Metric



In this example, the relative metric is used to control path selection. T1 is given a value of the native IGP value -2 (28), which makes it preferred to the native IP path. When the tunnel is considered in the IGP calculation, the native IGP metric (30) is greater than the tunnel metric (28) for all the destinations that are topologically behind the tunnel endpoint. As a result, all the destination networks are reachable via the TE tunnel.

T2 could have been given the native IGP value -4 (36), giving it preference to the native IP path and the T1-R4 path. However, in this example, it was given an absolute value of 35.

Forwarding Adjacency

This topic describes the Forwarding Adjacency feature.

Forwarding Adjacency

- Mechanism for:
 - Better intra- and inter-POP load balancing
 - Tunnel sizing independent of inner topology
- Allows the announcement of established tunnel via link-state (LSP) announcements

© 2012 Cisco and/or its affiliates. All rights reserved.

SPICORE v1.01-2.01

The MPLS TE forwarding adjacency feature allows a network administrator to handle a traffic-engineered LSP tunnel as a link in an IGP network, based on the SPF algorithm.

A forwarding adjacency can be created between routers regardless of their location in the network.

Forwarding adjacency is a mechanism to allow the announcement of established tunnels via IGP to all nodes within an area.

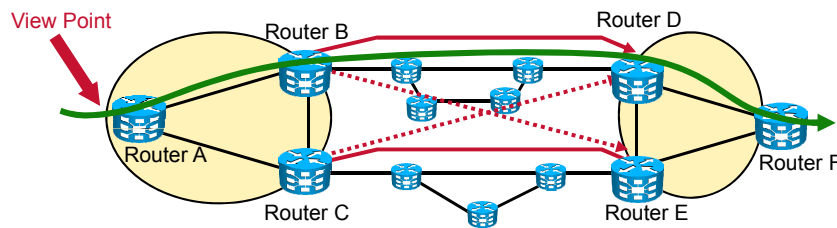
By using forwarding adjacency, you can achieve the following goals:

- Better load balancing when you are creating POP-to-POP tunnels
- Use of tunnels from any upstream node, independent of the inner topology of the network
- Use of tunnels independent of topology changes within the tunneled network area

Forwarding Adjacency (Cont.)

Traffic Flow Without Forwarding Adjacency

- Tunnels are created and announced to IP with autoroute, with equal cost to load-balance.
- All the POP-to-POP traffic exits via the routers on the IGP shortest path:
 - No load balancing
 - All traffic flows on tunnel: A-B-D-F



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.62

Before you consider the real benefits of forwarding agency, it is important to clearly see the limitations of autoroute in certain network topologies.

In this example, established tunnels exist from B to D, from B to E, from C to E, and from C to D; the preferred tunnels are B to D and C to E.

- The path metric from B to D for IS-IS is 30 (assuming default metric).
- The path metric from C to E is 20.

But traffic is entering at router A. Router A has no knowledge about the existence of tunnels between B and D and C and E. It only has its IGP information, indicating that the better path to F leads via routers B and D.

The results are as follows:

- There will be no load balancing.
- All traffic will flow via B and D.

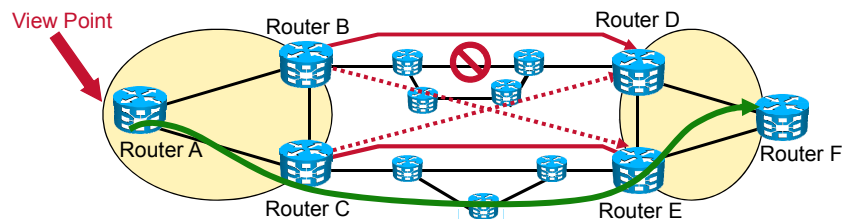
Any change in the core topology will affect the path metric and thus will affect any load balancing for POP-to-POP traffic.

Note You can theoretically prevent this problem by creating tunnels from any router to any other router, but this design does not scale in service provider networks.

Forwarding Adjacency (Cont.)

Traffic Flow Without Forwarding Adjacency

- All the POP-to-POP traffic exits via the routers on the IGP shortest path.
- Change in the core topology does affect the load balancing in the POP:
 - Normal state: All traffic flows A-B-D-F
 - Link failure: All traffic flows A-C-E-F



What happens if a link on the path between B and D gets broken?

Even though rerouting will possibly take place, the IGP metric for the complete path A-B-intermediates-D-F will change.

In this example, the change in the metric will result in a possibly unplanned switchover of the traffic from A to F from the upper path to the lower path.

This switchover may result in a possible congestion on the path from C to E, whereas the protected path from B to D is idled out.

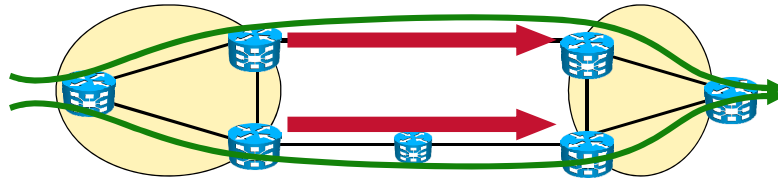
Forwarding Adjacency (Cont.)

POP to POP traffic is better load-balanced:

- In the POP, if the two core routers are used
- In the core, if at least two tunnels are used
- As long as the IGP metric for a path with the forwarding adjacency (for example, 25) is shorter than the forwarding adjacency-free path (for example, 30)

Inner topology does not affect tunnel sizing:

- Change in the core topology does not affect load balancing in the POP.



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.64

By using forwarding adjacency, you can create POP-to-POP tunnels where traffic paths and load balancing can be designed independent of the inner (core) topology of the network and independent of link failures.

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- Dynamic constraint-based path computation is triggered by the headend of the tunnel
- There are four link resource attributes
- The most important link resource attribute is the maximum allocation multiplier. Traffic tunnel attributes affect how the path is set up and maintained
- Link resource class attribute string allows inclusion or exclusion of the link into the path of the tunnel
- Each link has a cost or metric for calculating routes in the normal operation of the IGP
- Seven tunnel attributes are available to influence path selection
- The tunnel path can be configured manually or computed dynamically by using the constraint-based path computation
- The tunnel resource class affinity attribute allows the network administrator to apply path selection policies by administratively including or excluding network links

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.65

Summary (Cont.)

- The adaptability attribute indicates whether the traffic tunnel should be reoptimized, and consequently rerouted to another path
- The resilience attribute determines the behavior of the tunnel under faulty conditions
- The policies during LSP computation can be implemented using the resource class affinity bits of the traffic tunnel and the resource class bits of the links over which the tunnel should pass
- Important factor in LSP computation is the available bandwidth on the link over which the traffic tunnel will pass
- The result of the constraint-based path computation is a series of IP addresses that represent the hops on the LSP between the headend and tail end of the traffic tunnel
- LSP setup is always initiated at the traffic tunnel headend
- Each hop on the way determines if the available resources that are specified in the session attribute object are available.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.66

Summary (Cont.)

- Path rerouting may result from reoptimization due to a change in network resources or link failures that affect the LSP
- TE tunnels can be used for IP routing only if the tunnels are explicitly specified for routing
- Even for the destination that is behind each of the tunnel endpoints, the normal IGP routing is performed if there is no static route to the traffic-engineered tunnel.
- The autoroute feature enables the headend router to see the MPLS TE tunnel as a directly connected interface and to use it in its modified SPF computations.
- When using autoroute feature all the networks in the routing table that are topologically behind the tunnel endpoint are reachable via the tunnel itself.
- The tunnel metrics can be tuned, and either relative or absolute metrics can be used to resolve this issue
- The MPLS TE forwarding adjacency feature allows a network administrator to handle a traffic-engineered LSP tunnel as a link in an IGP network, based on the SPF algorithm

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.67

Implementing MPLS TE

Overview

This lesson describes Multiprotocol Label Switching (MPLS) traffic engineering (TE) commands for the implementation of MPLS traffic tunnels. The configuration commands that are needed to support MPLS TE are explained, and sample setups are presented. This lesson describes advanced MPLS TE commands that are used in path selection in a typical service provider environment. The configuration commands are accompanied by usage guidelines and sample setups.

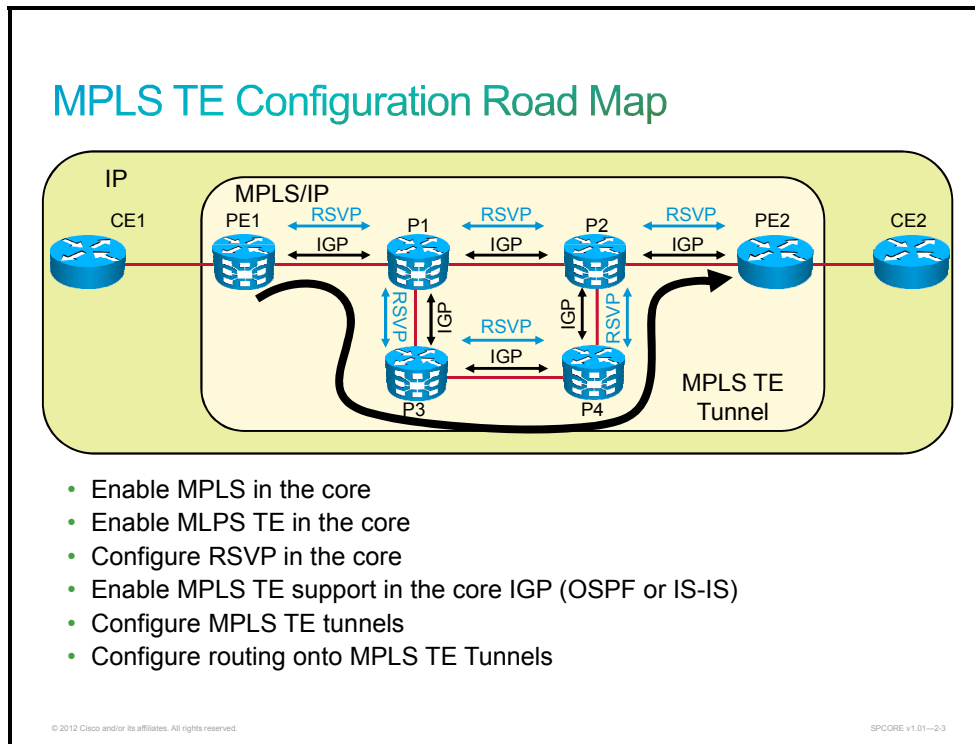
Objectives

Upon completing this lesson, you will be able to describe MPLS TE commands for the implementation of MPLS traffic tunnels. You will be able to meet these objectives:

- List the MPLS TE configuration tasks
- Explain the commands to enable MPLS TE
- Explain the commands to enable RSVP
- Explain the commands to enable OSPF to support MPLS TE
- Explain the commands to enable IS-IS to support MPLS TE
- Explain the commands to enable MPLS TE Tunnels
- Explain the commands to enable Static Routing and Autoroute
- Describe the show commands used to monitor MPLS TE operations
- Explain creating a dynamic MPLS TE tunnel using a case study
- Explain creating an explicit MPLS TE tunnel using a case study
- Explain enabling periodic tunnel optimization using a case study
- Explain enabling Path Selection Restrictions using a case study
- Explain modifying the Administrative Weight using a case study
- Explains enabling Autoroute and Forwarding Adjacency using a case study

MPLS TE Configuration Tasks

This topic lists the MPLS TE configuration tasks.



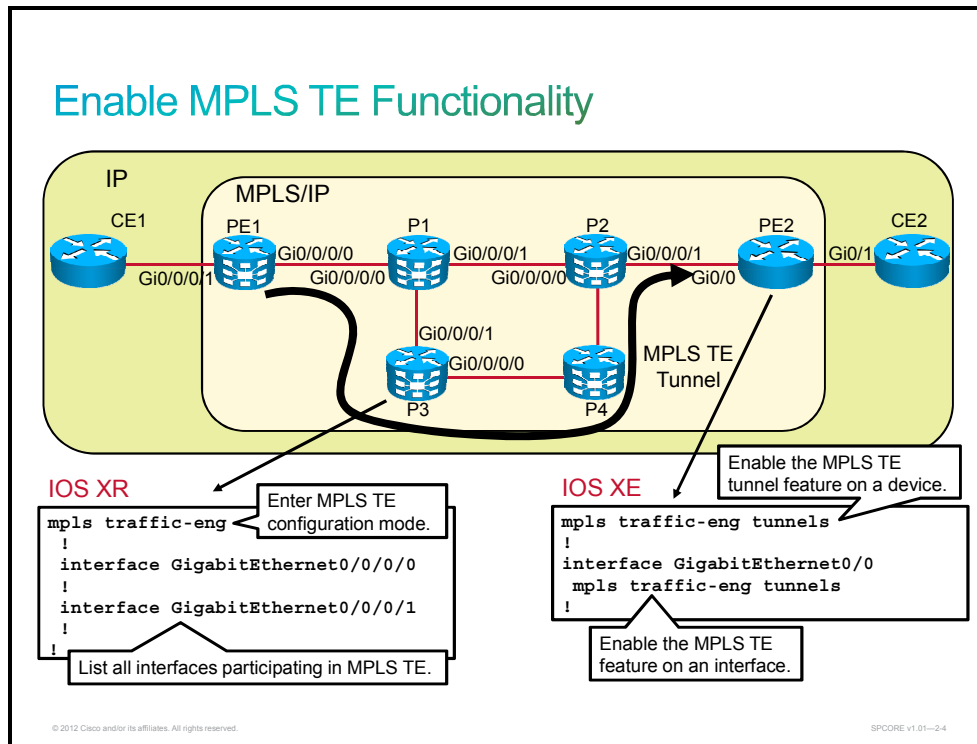
These steps are required to configure MPLS TE tunnels. (To configure MPLS TE tunnels, MPLS should already be enabled in the core network.)

- Enable MPLS TE in the core.
- Configure Resource Reservation Protocol (RSVP) in the core.
- Enable MPLS TE support in the core interior gateway protocol (IGP) with Open Shortest Path First (OSPF) or Intermediate System-to-Intermediate System (IS-IS).
- Configure MPLS TE tunnels.
- Configure routing onto MPLS TE tunnels.

Configuration for each of these required steps for building MPLS TE tunnels between routers will be shown.

MPLS TE Configuration

This topic explains the commands to enable MPLS TE.



To enter MPLS TE configuration mode on Cisco IOS XR Software, use the **mpls traffic-eng** command in global configuration mode. List all interfaces participating in MPLS TE in mpls traffic engineering configuration mode.

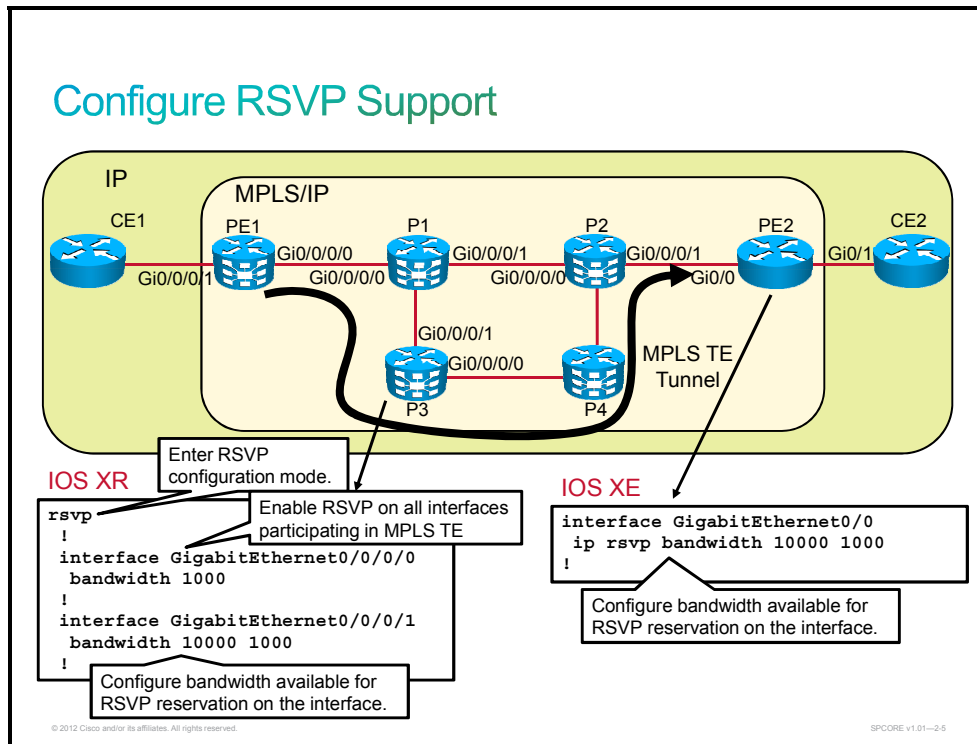
To enable MPLS TE tunnel signaling on a Cisco IOS XE device, use the **mpls traffic-eng tunnels** command in global configuration mode. To disable this feature, use the **no** form of this command.

To enable MPLS TE tunnel signaling on an interface, assuming that it is enabled for the device, use the **mpls traffic-eng tunnels** command in interface configuration mode. An enabled interface has its resource information flooded into the appropriate IGP link-state database, and accepts TE tunnel signaling requests.

Note MPLS TE functionality should be enabled on all routers on the path from headend to tail end of the MPLS TE tunnel.

RSVP Configuration

This topic explains the commands to enable RSVP.



Enable RSVP on all interfaces that are participating in MPLS TE and configure the bandwidth that is available for RSVP reservation.

To enter RSVP configuration mode, use the **rsvp** global configuration command on Cisco IOS XR Software. To enter interface configuration mode for the RSVP protocol, use the **interface** command; use the **bandwidth** command to set the reservable bandwidth, the maximum RSVP bandwidth that is available for a flow and the sub-pool bandwidth on this interface.

bandwidth *total-bandwidth max-flow sub-pool sub-pool-bw*

Syntax Description

Parameter	Description
<i>total-bandwidth</i> (<i>interface-kbps</i>)	(Optional) Maximum amount of bandwidth, in kb/s, that may be allocated by RSVP flows
<i>max-flow</i> (<i>single-flow-kbps</i>)	(Optional) Maximum amount of bandwidth, in kb/s, that may be allocated to a single flow

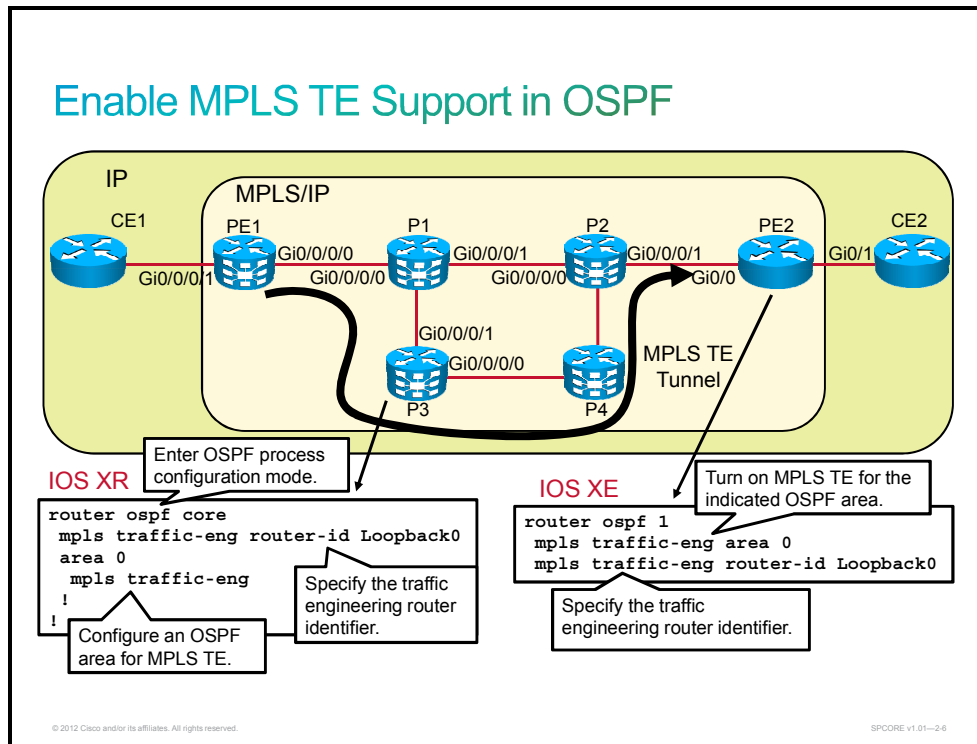
To enable RSVP for IP on an interface, use the **ip rsvp bandwidth** interface configuration command on Cisco IOS XE Software. To disable RSVP, use the **no** form of this command.

ip rsvp bandwidth [*interface-kbps* [*single-flow-kbps*]]

Note RSVP support should be enabled on all routers on the path from headend to tail end of the MPLS TE tunnel.

OSPF Configuration

This topic explains the commands to enable OSPF to support MPLS TE.



One of required steps to configure MPLS TE tunnels is enabling MPLS TE support in the IGP routing protocol (OSPF or IS-IS).

To enable MPLS TE support for OSPF routing protocol on Cisco IOS XR Software, enter OSPF process configuration mode and use the **mpls traffic-eng router-id interface** command to specify that the TE router identifier for the node is the IP address associated with a given OSPF interface.

mpls traffic-eng router-id {router-id | interface-type interface-instance}

A router identifier must be present in IGP configuration. This router identifier acts as a stable IP address for the TE configuration. This stable IP address is flooded to all nodes. For all TE tunnels that originate at other nodes and end at this node, the tunnel destination must be set to the TE router identifier of the destination node, because that identifier is the address that the TE topology database at the tunnel head uses for its path calculation.

MPLS TE must be enabled under area configuration. To configure an OSPF area for MPLS TE, use the **mpls traffic-eng** command in the appropriate configuration mode.

To turn on MPLS TE for the indicated OSPF area on which MPLS TE is enabled, use the **mpls traffic-eng area** command in router configuration mode on Cisco IOS XE Software.

To specify the TE router identifier for the node that is to be the IP address that is associated with the given interface, use the **mpls traffic-eng router-id** command in router configuration mode.

mpls traffic-eng router-id *interface*

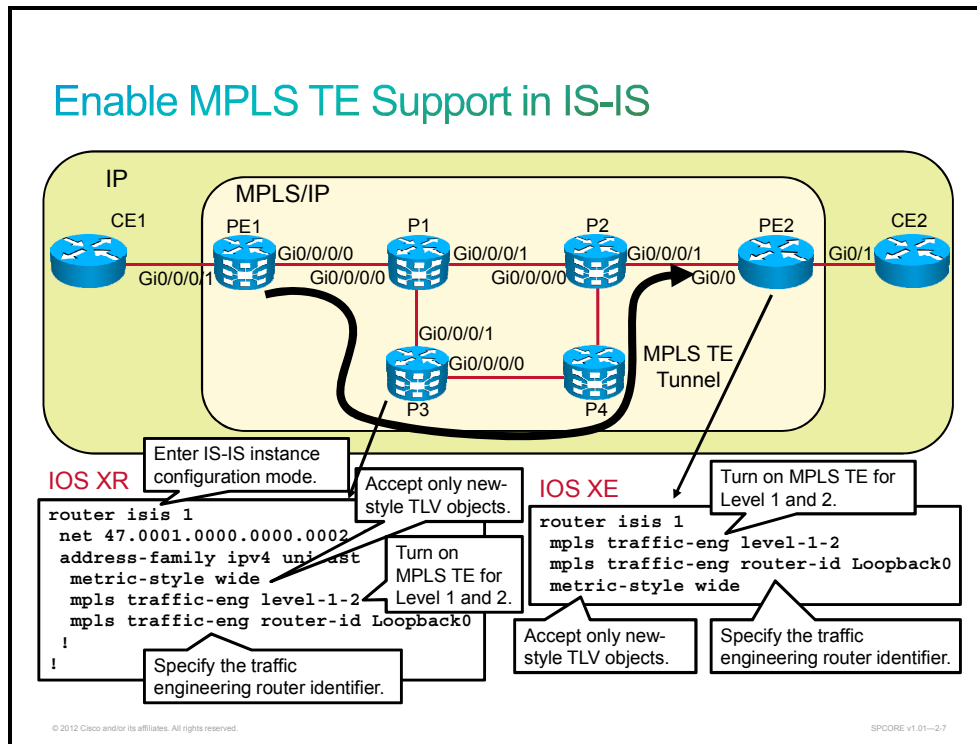
Syntax Description

Parameter	Description
<i>interface</i>	The MPLS TE router identifier is taken from the IP address of the supplied interface. This MPLS TE router identifier should be configured as the tunnel destination for tunnels that originate at other routers and terminate at this router. This interface should be a stable interface that will not go up and down, such as a loopback interface.

Note	MPLS TE support for IGP should be enabled on all routers on the path from headend to tail end of the MPLS TE tunnel.

IS-IS Configuration

This topic explains the commands to enable IS-IS to support MPLS TE.



To enable MPLS TE support for IS-IS routing protocol on Cisco IOS XR Software, enter address family configuration mode for configuring IS-IS routing and use the **mpls traffic-eng router-id interface** command to specify that the traffic engineering router identifier for the node is the IP address associated with a given OSPF interface.

mpls traffic-eng router-id {router-id | interface-type interface-instance}

To configure MPLS TE at IS-IS Level 1 and Level 2 on a router that is running IS-IS, use the **mpls traffic-eng level** command in address family configuration mode:

mpls traffic-eng level isis-level

To configure the IS-IS software to generate and accept only new-style type, length, and value (TLV) objects, use the **metric-style wide** command in address family configuration mode.

To turn on flooding of MPLS TE link information into the indicated IS-IS level, use the **mpls traffic-eng** command in router configuration mode on Cisco IOS XE Software. This command appears as part of the routing protocol tree and causes link resource information (for instance, the bandwidth available) for appropriately configured links to be flooded in the IS-IS link-state database:

mpls traffic-eng {level-1 | level-2}

To specify the TE router identifier for the node that is to be the IP address that is associated with the given interface, use the **mpls traffic-eng router-id** command in router configuration mode on Cisco IOS XE Software. To disable this feature, use the **no** form of this command.

mpls traffic-eng router-id *interface*

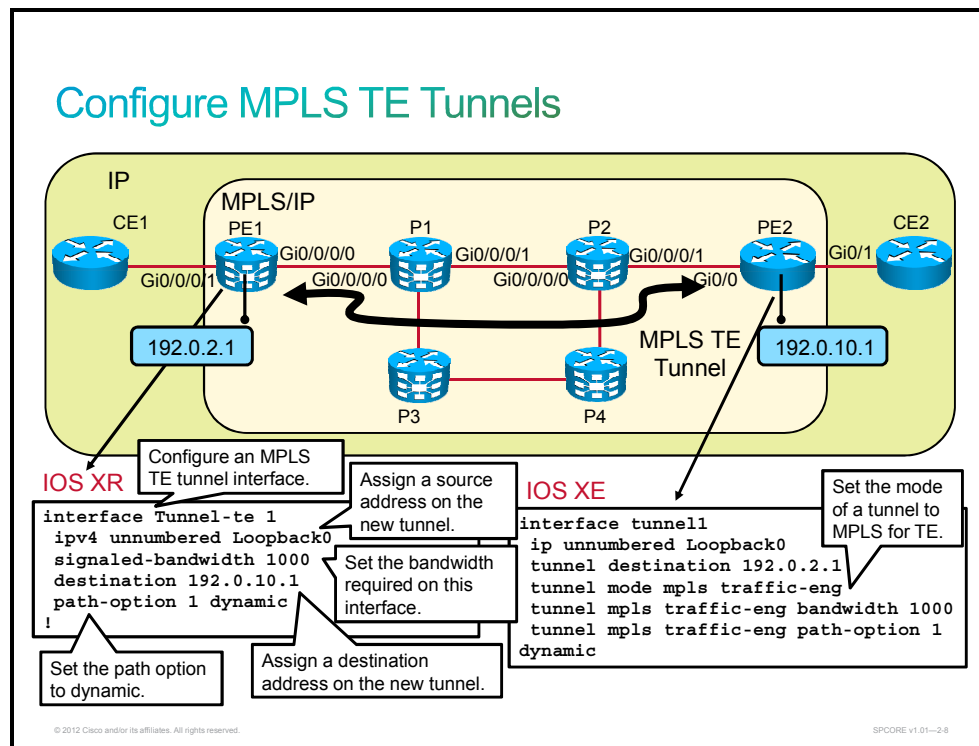
To configure a router running IS-IS so that it generates and accepts only new-style type, length, and value objects (TLVs), use the **metric-style wide** command in router configuration mode on Cisco IOS XE Software. To disable this function, use the **no** form of this command.

metric-style wide [**transition**] [**level-1** | **level-2** | **level-1-2**]

Note	MPLS TE support for IGP should be enabled on all routers on the path from headend to tail end of the MPLS TE tunnel.

MPLS TE Tunnels Configuration

This topic explains the commands to enable MPLS TE Tunnels.



This figure shows a typical implementation of a dynamic MPLS TE tunnel. Two tunnels are created: one from router PE1 to PE2 and one from PE2 to PE1. The actual path could be PE1-P1-P2-PE2 or PE1-P1-P3-P4-P2-PE2, as selected by IGP protocol, because the path option of the tunnel is set to dynamic. Alternatively, you can select the explicit path option, where you can manually specify the desired path of the MPLS TE tunnel.

Note The MPLS TE tunnel is unidirectional.

Use the **interface tunnel-te** *tunnel-id* command to configure an MPLS TE tunnel interface on Cisco IOS XR Software. You can set several MPLS TE tunnel parameters in interface tunnel-te configuration mode:

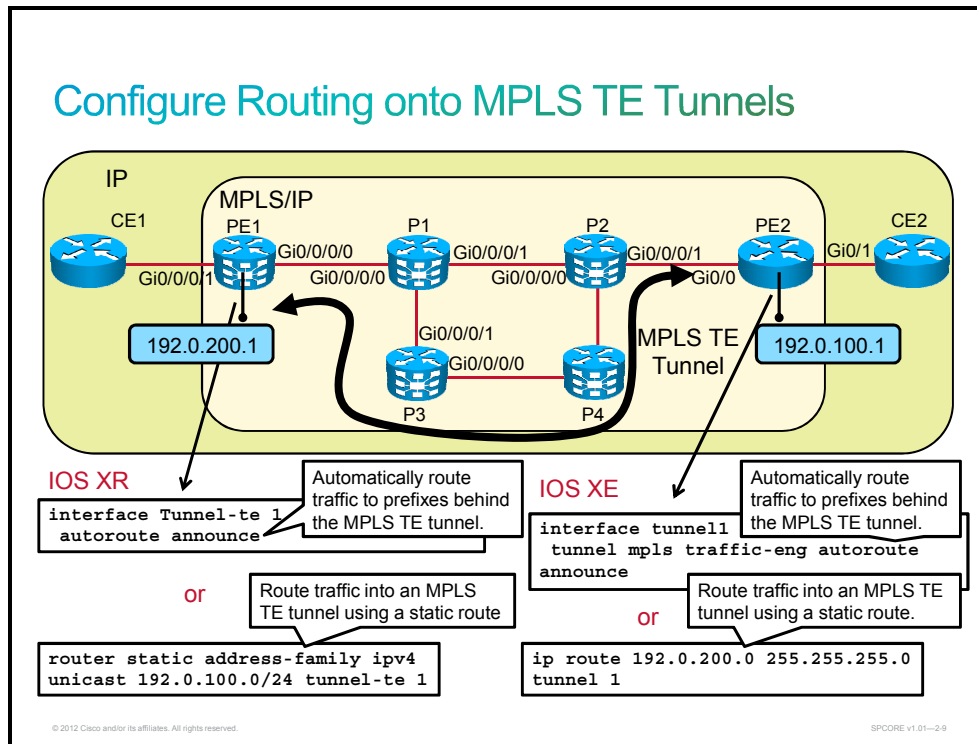
- Use the **ipv4 unnumbered** *interface* command in interface tunnel-te configuration mode to assign a source address, so that forwarding can be performed on the new tunnel. Loopback is commonly used as the interface type.
- Use the **destination** *ip-address* command to assign a destination address on the new tunnel. The destination address is the MPLS TE router ID of the remote node.
- Use the **signalled-bandwidth** *bandwidth* command to set the bandwidth that is required on this tunnel-te interface.
- Use the **path-option** *priority* **dynamic** command to set the path option to dynamic.

Use the **interface tunnel** command to declare a tag-switched path (TSP) tunnel interface on Cisco IOS XE Software. You can set several MPLS TE tunnel parameters in interface tunnel configuration mode:

- Use the **ip unnumbered interface** command in interface tunnel configuration mode to assign a source address, so that forwarding can be performed on the new tunnel. Loopback is commonly used as the interface type.
- Use the **tunnel destination ip-address** command to specify the destination for a tunnel interface.
- Use the **tunnel mode mpls traffic-eng** command to set the mode of a tunnel to MPLS for TE.
- Use the **tunnel mpls traffic-eng bandwidth** command to configure the bandwidth that is required for an MPLS TE tunnel. Bandwidth is specified in kb/s.
- Use the **tunnel mpls traffic-eng path-option** command to configure a path option for an MPLS TE tunnel.

Static Route and Autoroute Configurations

This topic explains the commands to enable Static Routing and Autoroute.



The autoroute feature evaluates MPLS TE tunnel interface with an IGP, and automatically routes traffic to prefixes behind the MPLS TE tunnel, based on Interior Gateway Protocol (IGP) metrics.

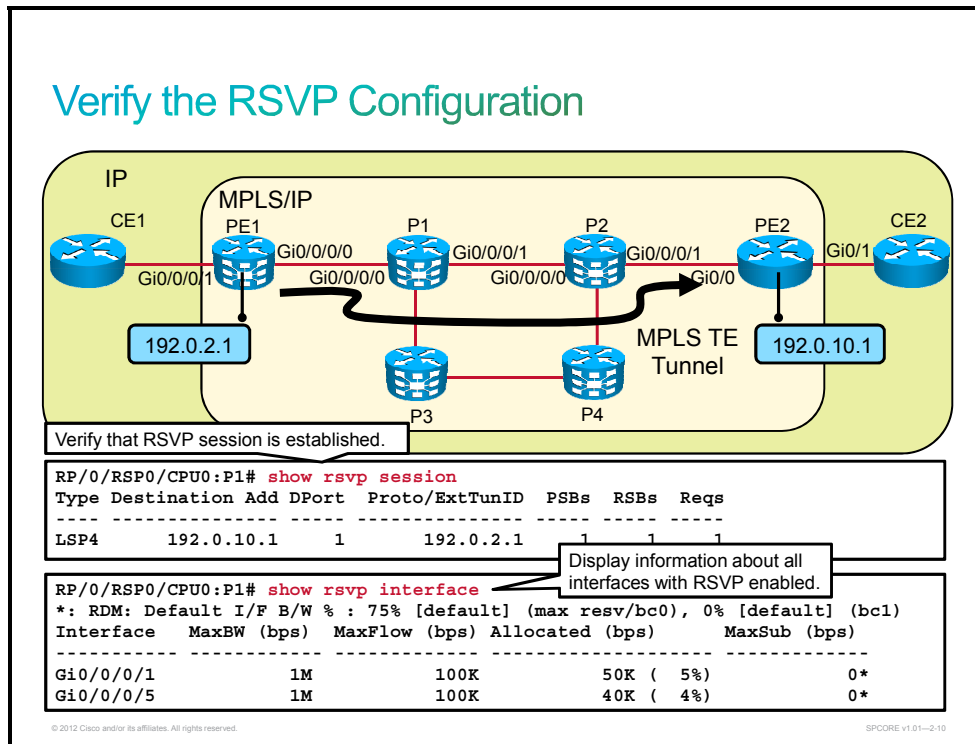
Use the **autoroute announce** command in interface configuration mode to specify that the IGP should use the tunnel (if the tunnel is up) in its enhanced shortest path first (SPF) calculation on Cisco IOS XR Software.

Another option to route traffic into an MPLS TE tunnel is by using a static route. You should route traffic to prefixes behind MPLS tunnel 1 to interface tunnel, in the example tunnel-te1. This configuration is used for static routes when the **autoroute announce** command is not used.

To instruct the IGP to use the tunnel in its SPF calculation (if the tunnel is up) on Cisco IOS XE Software, use the **tunnel mpls traffic-eng autoroute announce** command in interface configuration mode.

Monitoring MPLS TE Operations

This topic describes the show commands used to monitor MPLS TE operations.



The **show rsvp session** command verifies that all routers on the path of the LSP are configured with at least one Path State Block (PSB) and one Reservation State Block (RSB) per session. In the example, the output represents an LSP from ingress (head) router 192.0.2.1 to egress (tail) router 192.0.10.1.

To display information about all interfaces with RSVP enabled, use the **show rsvp interface** command in EXEC mode. You can also see information about allocated bandwidth to MPLS TE tunnels on RSVP-enabled interfaces.

Verify MPLS TE Tunnels

```
CPU0:PE7# show mpls traffic-eng tunnels
Fri Nov 11 08:35:02.316 UTC
Signalling Summary:
  LSP Tunnels Process: running
  RSVP Process: running
  Forwarding: enabled
  Periodic reoptimization: every 3600 seconds, next in 2412 seconds
  Periodic FRR Promotion: every 300 seconds, next in 85 seconds
  Auto-bw enabled tunnels: 0 (disabled)

Name: tunnel-te1 Destination: 192.0.10.1
Status:
  Admin: up Oper: up Path: valid Signalling: connected

Path info (OSPF 1 area 0):
  Node hop count: 3
  Hop0: 192.168.71.1
  Hop1: 192.168.2.1
  Hop2: 192.168.2.2
  Hop3: 192.168.82.2
  Hop4: 192.168.82.80
  Hop5: 192.0.10.1
```

Annotations in the image:

- Callout for "running" (LSP Tunnels Process): Status of the LSP tunnels process
- Callout for "running" (RSVP Process): Status of the RSVP process
- Callout for "up Oper: up": Configure up or down. / Operationally up or down
- Callout for the hop list: Hop list of current LSP

To display information about MPLS TE tunnels, use the **show mpls traffic-eng tunnels** command in EXEC mode. Some output in the figure is omitted. Here is the full output of the command:

```
RP/0/RSP0/CPU0:PE1# show mpls traffic-eng tunnels
Fri Nov 11 08:41:16.386 UTC
Signalling Summary:
  LSP Tunnels Process: running
  RSVP Process: running
  Forwarding: enabled
  Periodic reoptimization: every 3600 seconds, next in 2038 seconds
  Periodic FRR Promotion: every 300 seconds, next in 11 seconds
  Auto-bw enabled tunnels: 0 (disabled)

Name: tunnel-te1 Destination: 192.0.10.1
Status:
  Admin: up Oper: up Path: valid Signalling: connected

  path option 1, type dynamic (Basis for Setup, path weight 4)
  G-PID: 0x0800 (derived from egress interface properties)
  Bandwidth Requested: 50 kbps CT0
  Creation Time: Thu Oct 20 16:44:38 2011 (3w0d ago)
Config Parameters:
  Bandwidth: 50 kbps (CT0) Priority: 7 7 Affinity: 0x0/0xffff
  Metric Type: TE (default)
  Hop-limit: disabled
  AutoRoute: enabled LockDown: disabled Policy class: not set
  Forwarding-Adjacency: disabled
  Loadshare: 0 equal loadshares
  Auto-bw: disabled
  Fast Reroute: Disabled, Protection Desired: None
  Path Protection: Not Enabled
History:
  Tunnel has been up for: 20:45:45 (since Thu Nov 10 11:55:31 UTC 2011)
  Current LSP:
    Uptime: 20:01:23 (since Thu Nov 10 12:39:53 UTC 2011)
  Reopt. LSP:
    Last Failure:
      LSP not signalled, identical to the [CURRENT] LSP
```

```

    Date/Time: Thu Nov 10 14:15:14 UTC 2011 [18:26:02 ago]
  Prior LSP:
    ID: path option 1 [2]
    Removal Trigger: reoptimization completed

  Path info (OSPF 1 area 0):
  Node hop count: 3
  Hop0: 192.168.71.1
  Hop1: 192.168.2.1
  Hop2: 192.168.2.2
  Hop3: 192.168.82.2
  Hop4: 192.168.82.80
  Hop5: 10.8.1.1

  LSP Tunnel 10.8.1.1 1 [523] is signalled, connection is up
  Tunnel Name: PE8_t1 Tunnel Role: Tail
  InLabel: GigabitEthernet0/0/0/2, implicit-null
  Signalling Info:
    Src 10.8.1.1 Dst 10.7.1.1, Tun ID 1, Tun Inst 523, Ext ID 10.8.1.1
    Router-IDs: upstream 10.0.1.1
                  local 10.7.1.1
  Bandwidth: 40 kbps (CT0) Priority: 1 1 DSTE-class: no match
  Path Info:
    Incoming Address: 192.168.71.70
    Incoming:
    Explicit Route:
      Strict, 192.168.71.70
      Strict, 10.7.1.1
    Record Route: Disabled
    Tspec: avg rate=40 kbits, burst=1000 bytes, peak rate=40 kbits
    Session Attributes: Local Prot: Not Set, Node Prot: Not Set, BW Prot:
  Not Set
    Resv Info: None
    Record Route: Disabled
    Fspec: avg rate=40 kbits, burst=1000 bytes, peak rate=40 kbits
  Displayed 1 (of 1) heads, 0 (of 0) midpoints, 1 (of 1) tails
  Displayed 1 up, 0 down, 0 recovering, 0 recovered heads

```

Verify MPLS TE Tunnels (Cont.)

```
RP/0/RSP0/CPU0:PE1# show mpls traffic-eng topology
My_System_id: 10.7.1.1 (OSPF 1 area 0)
My_BC_Model_Type: RDM
Signalling error holddown: 10 sec Global Link Generation 37838

IGP Id: 0.0.0.1, MPLS TE Id: 10.0.1.1 Router Node (OSPF 1 area 0)
Link[0]:Broadcast, DR:192.168.2.2, Nbr Node Id:1, gen:37830
  Frag Id:8, Intf Address:192.168.2.1, Intf Id:0
  Nbr Intf Address:0.0.0.0, Nbr Intf Id:0
  TE Metric:2, IGP Metric:2, Attribute Flags:0x0
  BC Model ID:RDM
  Physical BW:1000000 (kbps), Max Reservable BW Global:1000 (kbps)
  Max Reservable BW Sub:0 (kbps)
```

Neighbor Interface address of this link	Total Allocated BW (kbps)	Global Pool Reservable BW (kbps)	Sub Pool Reservable BW (kbps)
bw[0]:	0	1000	0
bw[1]:	0	1000	0
bw[2]:	0	1000	0
bw[3]:	0	1000	0
bw[4]:	0	1000	0
bw[5]:	0	1000	0
bw[6]:	0	1000	0
bw[7]:	50	950	0

Bandwidth (in kb/s) allocated at that priority

Available bandwidth (in kb/s) reservable at that priority in global pool

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.12

To display the MPLS TE network topology currently known at this node, use the **show mpls traffic-eng topology** command in EXEC mode. Some output in the figure is omitted. Here is the full output of the command:

```
RP/0/RSP0/CPU0:PE7#show mpls traffic-eng topology
```

```
Fri Nov 11 09:07:05.630 UTC
```

```
My_System_id: 10.7.1.1 (OSPF 1 area 0)
```

```
My_BC_Model_Type: RDM
```

```
Signalling error holddown: 10 sec Global Link Generation 37838
```

```
IGP Id: 0.0.0.1, MPLS TE Id: 10.0.1.1 Router Node (OSPF 1 area 0)
```

```
Link[0]:Broadcast, DR:192.168.2.2, Nbr Node Id:1, gen:37830
```

```
  Frag Id:8, Intf Address:192.168.2.1, Intf Id:0
```

```
  Nbr Intf Address:0.0.0.0, Nbr Intf Id:0
```

```
  TE Metric:2, IGP Metric:2, Attribute Flags:0x0
```

```
  Attribute Names:
```

```
  Switching Capability:None, Encoding:unassigned
```

```
  BC Model ID:RDM
```

```
  Physical BW:1000000 (kbps), Max Reservable BW Global:1000 (kbps)
```

```
  Max Reservable BW Sub:0 (kbps)
```

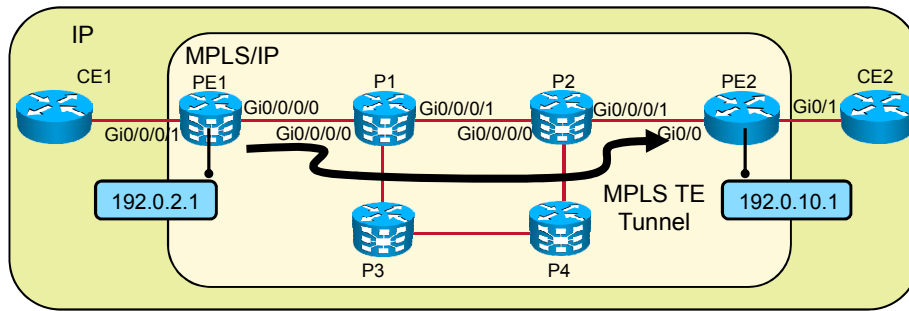
	Total Allocated BW (kbps)	Global Pool Reservable BW (kbps)	Sub Pool Reservable BW (kbps)
bw[0]:	0	1000	0
bw[1]:	0	1000	0
bw[2]:	0	1000	0

bw[3]:	0	1000	0
bw[4]:	0	1000	0
bw[5]:	0	1000	0
bw[6]:	0	1000	0
bw[7]:	50	950	0

Link[1]:Broadcast, DR:192.168.71.70, Nbr Node Id:6, gen:37831
 Frag Id:12, Intf Address:192.168.71.1, Intf Id:0
 Nbr Intf Address:0.0.0.0, Nbr Intf Id:0
 TE Metric:1, IGP Metric:1, Attribute Flags:0x0
 Attribute Names:
 Switching Capability:None, Encoding:unassigned
 BC Model ID:RDM
 Physical BW:1000000 (kbps), Max Reservable BW Global:1000 (kbps)
 Max Reservable BW Sub:0 (kbps)

	Total Allocated BW (kbps)	Global Pool Reservable BW (kbps)	Sub Pool Reservable BW (kbps)
	-----	-----	-----
bw[0]:	0	1000	0
bw[1]:	40	960	0
bw[2]:	0	960	0
bw[3]:	0	960	0
bw[4]:	0	960	0
bw[5]:	0	960	0
bw[6]:	0	960	0
bw[7]:	0	960	0

Verify a Routing Table with Autoroute Enabled



```
RP/0/RSP0/CPU0:PE1# show ip route | include tunnel
Fri Nov 11 10:09:01.339 UTC
O 192.0.10.1/32 [110/5] via 192.0.10.1, 22:12:51, tunnel-te1
O 10.8.10.1/32 [110/6] via 192.0.10.1, 22:12:51, tunnel-te1
O 192.168.108.0/24 [110/5] via 192.0.10.1, 22:12:51, tunnel-te1
```

Prefixes behind PE2 are reachable through the MPLS TE tunnel.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2-13

Verify that the prefixes that are behind the tail-end router of the MPLS TE tunnel are reachable through the MPLS TE tunnel interface, as shown in the figure.

MPLS TE Case Study: Dynamic MPLS TE Tunnel

This topic explains creating a dynamic MPLS TE tunnel using a case study.

MPLS TE Case Study

The service provider has these components:

- Core network:
 - Core routers in locations are heavily meshed
 - Two permanent connections to the Internet
- Local or remote access networks:
 - Connected to core in a resilient way, allowing a remote access via point of presence
- Access networks:
 - BGP customers

© 2012 Cisco and/or its affiliates. All rights reserved.

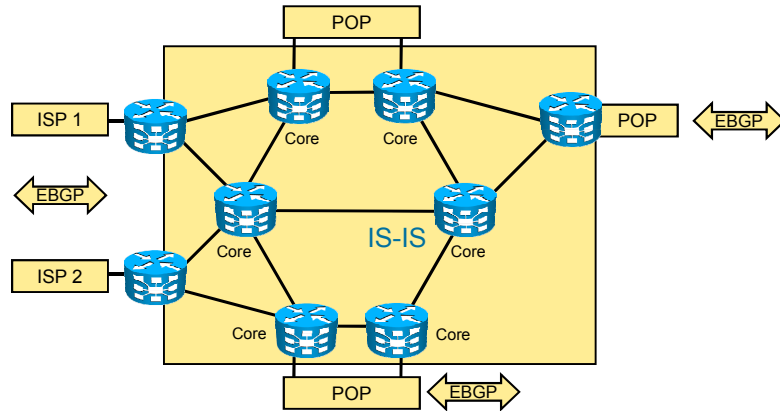
SPCORE v1.01--2-14

The example in the figure shows a classic ISP architecture based on three levels of hierarchy. The design should bring together some of the aspects of traffic engineering and routing design that are discussed in this module:

- Routing protocol choice and interaction between different routing protocols
- Support for MPLS TE tunnels

MPLS TE Case Study: IGP Design

- IS-IS is used for routing inside and between the core and POPs:
- A simple flat design using one IS-IS Level 2 backbone.
 - Loopbacks are advertised with a 32-bit network mask.
 - Use a new flavor of IS-IS TLVs (wide metric).



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.15

The network is based on a three-level hierarchy:

- **Core network:** Highly meshed central sites with high bandwidth requirements between them
- **Point of presence (POP) sites:** A distribution layer of regional sites, which are connected back to the core over redundant links, and which provide access for remote sites
- **Border Gateway Protocol (BGP) peers:** Upstream Internet providers and customer networks that are connected to the distribution sites via leased lines

The core and regional networks are a complex mesh of routers, and require efficient, scalable routing with fast convergence. A link-state protocol is ideally suited to this situation. Therefore, Integrated IS-IS is the choice.

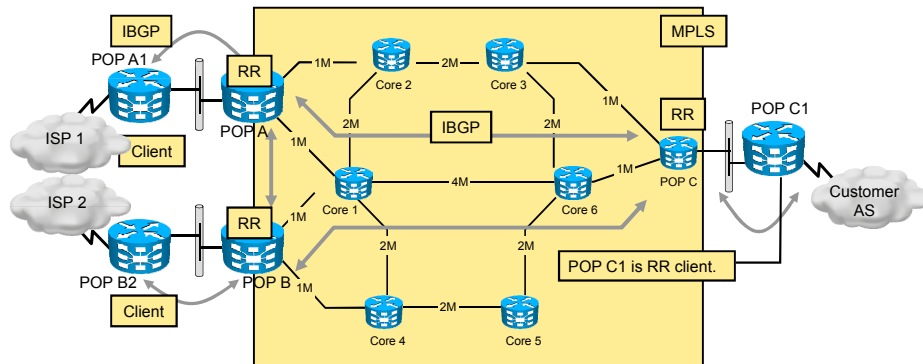
The proposed structure of the IS-IS protocol is as follows:

- Simply enabling Integrated IS-IS on Cisco routers sets their operation as Level 2 routers.
- All IP subnets would be visible individually to all routers on the IS-IS network from their Level 2 advertisements.
- A “wide metric” is used, which allows greater granularity of path selection.

MPLS TE Case Study: BGP Design

All POP sites are fully meshed with IBGP:

- MPLS is in the core, but not inside the POP.
- Route reflectors are used at POP sites to ease configuration of future POP routers.
- Route reflectors are connected in a full mesh.



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.18

The method of core transport uses MPLS as a transport mechanism in the core network. Packets are switched through the core of the network at Layer 2, bypassing the traditional Layer 3 routing process.

The issue with the edge-only BGP peer design is that the routers between the edges need to have routes to process the packets. If these packets pass through these routers with MPLS tags, they no longer need IP routes to these destinations.

In all cases, BGP relies on another protocol to resolve its next-hop address. An IGP is required to do this, so IS-IS must contain routes to the edge routers and to their attached (public) subnets.

To reduce the number of internal Border Gateway Protocol (IBGP) sessions that are required between POPs, route reflectors are used.

These tools allow the full-mesh requirement of traditional IBGP operation to be relaxed.

Note BGP configuration will not be shown in this case study.

Creating Dynamic Path Options for a Traffic Tunnel

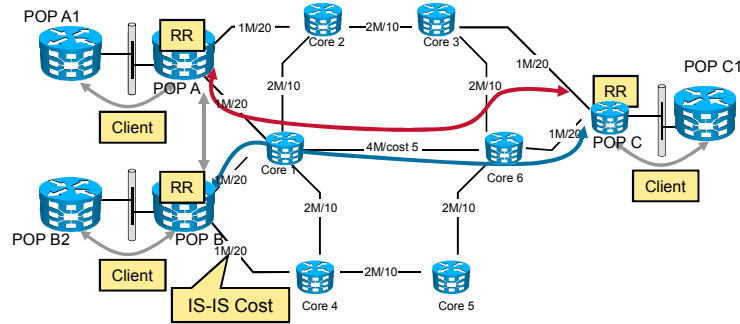
Objective: Engineer the traffic across the network.

Traffic originating from POP A and POP B destined for POP C is classified into two tunnels providing a guaranteed bandwidth of 250 kb/s.

Red tunnel: from POP A to POP C

Blue tunnel: from POP B to POP C

The traffic between POP A and POP B is not subject to MPLS TE.



This sample configuration demonstrates how to implement TE with an existing MPLS network.

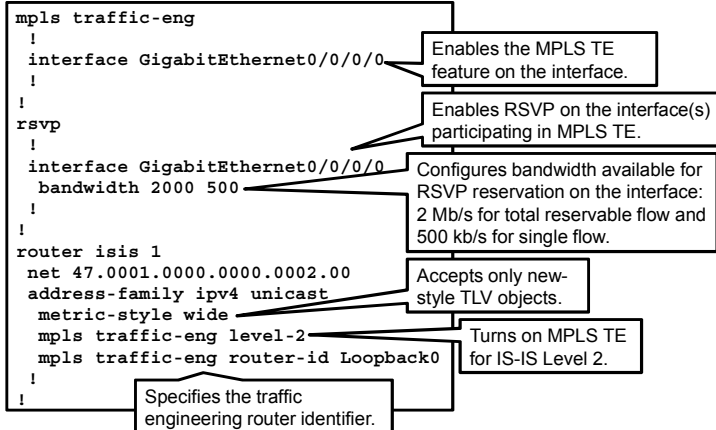
The figure shows the implementation of two TE tunnels of 250 kb/s. Either the tunnels are being automatically set up by the ingress label switch routers (LSRs), (POP A, POP B, and POP C routers), or they are manually configured with the explicit paths.

“TE” is a generic term that refers to the use of different technologies to optimize the utilization of the capacity and topology of a given backbone.

Creating Dynamic Path Options for a Traffic Tunnel (Cont.)

MPLS TE Platform: Sample Configuration of Core Router

Provide an underlying platform for MPLS TE by configuring extended IS-IS and RSVP for bandwidth assurance: 2 Mb/s of the bandwidth is reservable for traffic tunnels.



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.18

MPLS TE uses an extension to existing protocols such as RSVP, IS-IS, and OSPF to calculate and establish unidirectional tunnels that are set according to the network constraint. Traffic flows are mapped on the different tunnels depending on their destination.

With Cisco, MPLS TE is built on these mechanisms:

- A link-state IGP (such as IS-IS), with extensions for the global flooding of resource information and extensions for the automatic routing of traffic onto LSP tunnels, as appropriate
- An MPLS TE path calculation module (constraint-based routing [CBR]), which determines the paths to use for LSP tunnels
- Unidirectional LSP tunnels, which are signaled through RSVP
- An MPLS TE link management module that manages link admission and the bookkeeping of the resource information to be flooded
- Label switching forwarding, which provides routers with a Layer 2-like ability to direct traffic across multiple hops, as directed by the resource-based routing algorithm

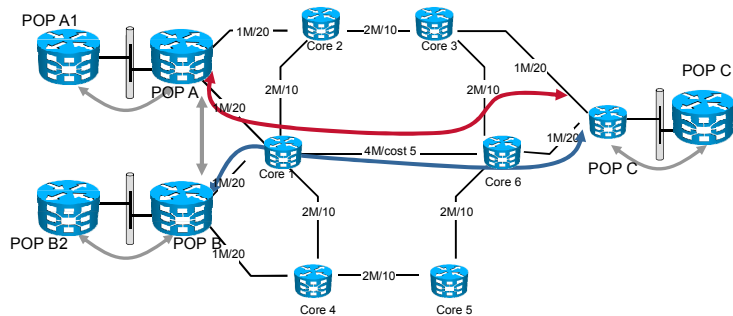
#Sample Cisco IOS Configuration of Core Router

```
ip cef
mpls ip
mpls traffic-eng tunnels
interface GigabitEthernet 0/0
mpls traffic-eng tunnels
ip rsvp bandwidth 1500 500
ip router isis
router isis
passive-interface Loopback0
net 49.0001.0000.0001.0001.00
is-type level-2-only
metric-style wide
mpls traffic-eng router-id Loopback0
mpls traffic-eng level-2
```

Creating Dynamic Path Options for a Traffic Tunnel (Cont.)

Sample path setup:

- Based on the default parameters and the assumption that the links are free, the best path for both would be the same as is computed by IS-IS.
- Therefore, the label switched paths for both tunnels use the middle Core 1 and Core 6 routers.



The example in the figure shows how the CBR algorithm proposes a path between tunnel endpoints that satisfies the initial requests at the headend of the tunnel.

Based on the assumption that all TE links are free, the traffic from POP A to POP C and from POP B to POP C is directed along the same least-cost path (Core 1-Core 6) because it is used by IS-IS for native IP routing.

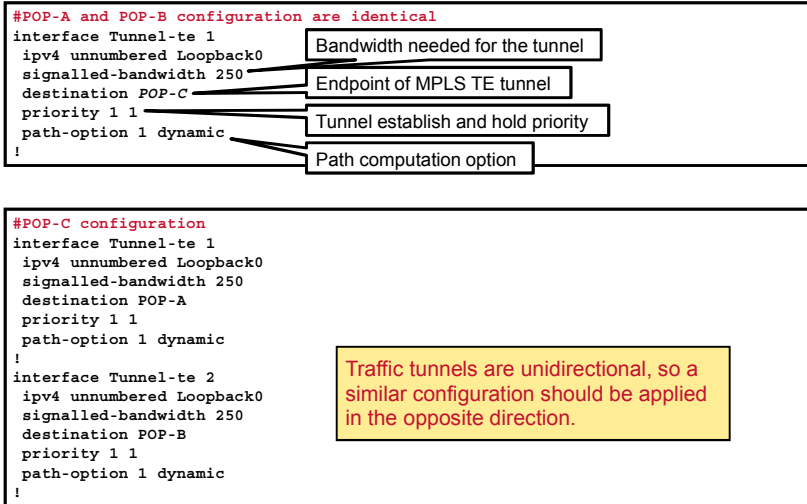
The reason is very simple. CBR is a routing process that takes into account these two considerations:

- The best route is the least-cost route with enough resources. CBR uses its own metric (administrative weight, or TE cost), which is by default equal to the IGP.

If there is a tie, the path with the highest minimum available bandwidth is selected. If a tie continues to exist, then the path with the smallest hop count is selected. Finally, if there is still a tie, a path is selected randomly.

- The result of CBR is an explicit route, which is used by RSVP to reserve resources and establish an LSP path.

Creating Dynamic Path Options for a Traffic Tunnel (Cont.)



To set up a dynamic TE tunnel (assuming that the IGP platform has been prepared), follow these command steps in the tunnel interface configuration mode:

- **ip unnumbered loopback0:** Gives the tunnel interface an IP address. An MPLS TE tunnel interface should be unnumbered because it represents a unidirectional link.
- **destination:** Specifies the destination for a tunnel. The destination of the tunnel must be the source of the tunnel in the opposite direction, usually a loopback address.
- **priority:** Configures the setup and reservation priority for an MPLS TE tunnel.
- **signaled- bandwidth:** Configures the bandwidth for the MPLS TE tunnel.
- **path-option number dynamic:** The LSP path is dynamically calculated.

#POP-A and POP-B Cisco IOS Configurations are the same

```

interface Tunnel1
  ip unnumbered Loopback0
  tunnel destination POP-C
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng priority 1 1
  tunnel mpls traffic-eng bandwidth 250
  tunnel mpls traffic-eng path-option 1 dynamic

```

#POP-C Cisco IOS Configuration

```

interface Tunnel1
  ip unnumbered Loopback0
  tunnel destination POP-A
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng priority 1 1
  tunnel mpls traffic-eng bandwidth 250
  tunnel mpls traffic-eng path-option 1 dynamic
interface Tunnel2
  ip unnumbered Loopback0
  tunnel source Loopback0
  tunnel destination POP-B
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng priority 1 1
  tunnel mpls traffic-eng bandwidth 250
  tunnel mpls traffic-eng path-option 1 dynamic

```

MPLS TE Case Study Continue: Explicit MPLS TE Tunnel

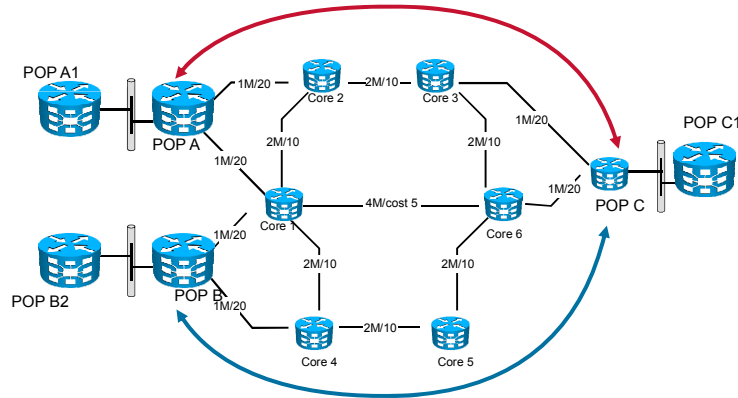
This topic explains creating an explicit MPLS TE tunnel using a case study.

Creating Explicit Path Options for a Traffic Tunnel

The second option is to explicitly define the desired path, as long as there are enough resources available along the path.

Example:

- The traffic is forwarded along the upper path (Core 2-Core 3) for the **red** tunnel and along the lower path (Core 4-Core 5) for the **blue** tunnel.



The example in the figure shows how to avoid the first step in the CBR algorithm by manually setting the explicit path between tunnel endpoints. This path might be derived from the least-cost path.

The best route might not be the least-cost route, given enough resources. The best route might be any sequence of next-hop routers that are configured at the headend of the tunnel.

Such a route, as proposed by the network administrator, is then checked against the extended link-state database that is carrying information on currently available resources.

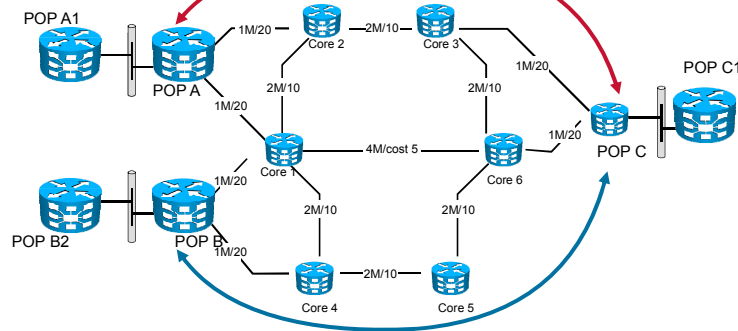
If this check is successful, CBR honors the route and RSVP is initiated to reserve some bandwidth and establish an LSP path. Otherwise, the tunnel stays down.

Creating Explicit Path Options for a Traffic Tunnel (Cont.)

#POP-A configuration

```
interface Tunnel-te 1
  ipv4 unnumbered Loopback0
  signalled-bandwidth 250
  destination POP-C
  priority 1 1
  path-option 1 explicit name Core2-3
!
explicit-path name Core2-3
  index 1 next-address ipv4 unicast Core-2
  index 2 next-address ipv4 unicast Core-3
  index 3 next-address ipv4 unicast POP-C
```

Explicit path represents the explicit route object, encoded as incoming IP addresses of the routers along the path.



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.22

To set up a static TE tunnel (assuming that the IGP platform has been prepared), use these additional steps:

- In tunnel interface configuration mode:
 - **path-option *number* explicit**: This command enables you to configure the tunnel to use a named IP explicit path from the TE topology database.
- In global configuration mode:
 - **explicit-path**: An IP explicit path is a list of IP addresses, each representing a node or link (incoming IP address) in the explicit path.
 - To include a path entry at a specific index, use the **index next-address** command in explicit path configuration mode. To return to the default behavior, use the **no** form of this command.
 - **index *index-id* {next-address ipv4 unicast *A.B.C.D.*}**

#POP-A Cisco IOS Configuration

```
interface Tunnel1
  ip unnumbered Loopback0
  tunnel destination POP-C
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng priority 1 1
  tunnel mpls traffic-eng bandwidth 250
  tunnel mpls traffic-eng path-option 1 explicit name Core2-3
!
ip explicit-path name Core2-3 enable
  next-address {strict} Core-2
  next-address {strict} Core-3
  next-address {strict} POP-C
```

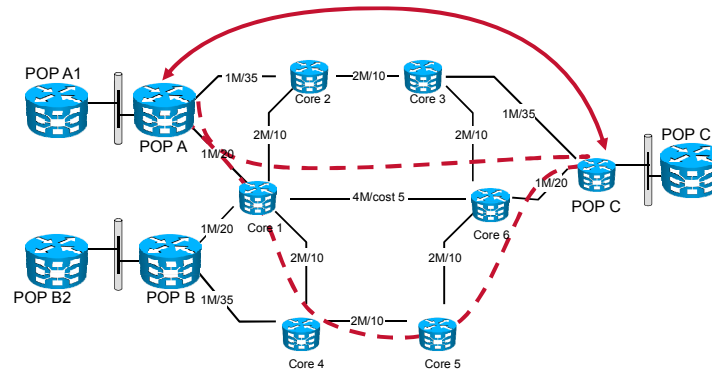
MPLS TE Case Study Continue: Periodic Tunnel Optimization

This topic explains enabling periodic tunnel optimization using a case study.

Configuring Periodic Optimization

A static tunnel is used to forward traffic. A dynamic backup path is configured for use if the static path is broken or if an intermediate router refuses to honor the reservation. If a link goes down between Core 2 and Core 3, the LSP path for the red tunnel (as directed by constraint-based routing) might use one of these two paths:

- POP A-Core 1-Core 6-POP C
- POP A-Core 1-Core 4-Core 5-Core 6-POP C



The LSP path is constantly monitored to maintain the network traffic tunnel in a desired state. When the path is broken and the tunnel had been set up dynamically, the headend router tries to find an alternative solution. This process is referred to as rerouting.

Reoptimization occurs when a device examines tunnels with established LSPs to see if better LSPs are available. If a better LSP seems to be available, the device attempts to signal the better LSP and, if successful, replaces the old and inferior LSP with the new and better LSP.

This reoptimization might be triggered manually, or it might occur at configurable intervals (the default is 1 hour). Instability and oscillations can result if the reoptimization interval is set too small. However, the network will not react to unexpected shifts in traffic if the interval is too great. One hour is a reasonable compromise. With reoptimization, traffic is routed so that it sees the lightest possible loads on the links that it traverses.

Unfortunately, reoptimization does not bring any improvements for a tunnel that has been established statically. In this instance, the path is explicitly determined, which compels the headend router to strictly follow the explicit path.

Configuring Periodic Optimization (Cont.)

```
#POP-A and POP-B configuration are identical
!
mpls traffic-eng
  reoptimize 300
!
interface Tunnel-te 1
  ipv4 unnumbered Loopback0
  signalled-bandwidth 250
  destination POP-C
  priority 1 1
  path-option 1 explicit name Core2-3
  path-option 2 dynamic
!
explicit-path name Core2-3
  index 1 next-address ipv4 unicast Core-2
  index 2 next-address ipv4 unicast Core-3
  index 3 next-address ipv4 unicast POP-C
```

The automatic LSP reoptimization is triggered periodically every 5 minutes to check for better paths.

The relative preference of this option: a lower path option number is tried before a higher path option number.

A second option will be used if the first path fails. **The second path is not pre-established.**

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.24

The example in the figure shows how traffic can be engineered across a path in the network and how a backup route for that traffic-engineered path can be established.

The primary path is manually specified, so it is explicit. If this path suddenly cannot be followed, the MPLS TE engine uses the next path option, which in this example is a dynamic route.

The drawback to this solution is the time that is needed to establish a backup TE route for the lost LSP path, and the time that is needed to revert to the primary path once it becomes available again.

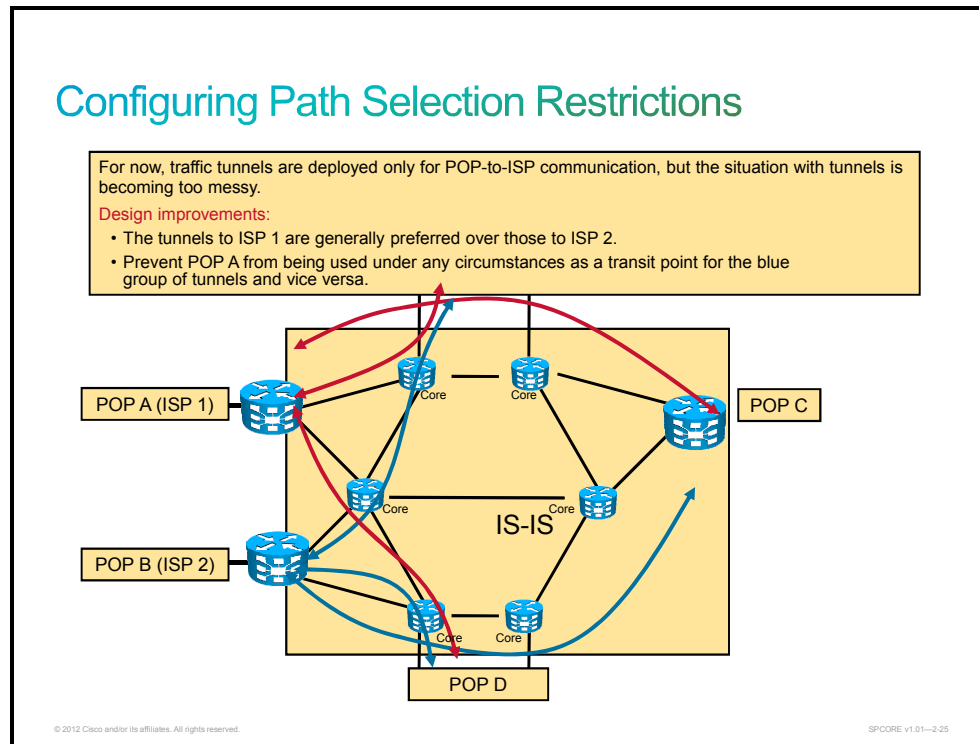
Though the search for an alternate path is periodically triggered, there is still downtime while the alternate path is being built.

#POP-A nad POP-B Cisco IOS Configuration

```
mpls traffic-eng reoptimize timers frequency 300
interface Tunnel1
  ip unnumbered Loopback0
  tunnel source Loopback0
  tunnel destination POP C
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng priority 1 1
  tunnel mpls traffic-eng bandwidth 250
  tunnel mpls traffic-eng path-option 1 explicit name Core2-3
  tunnel mpls traffic-eng path-option 2 dynamic
!
ip explicit-path name Core2-3 enable
  next-address strict Core-2
  next-address strict Core-3
  next-address strict POP-C
```


MPLS TE Case Study Continue: Path Selection Restrictions

This topic explains enabling Path Selection Restrictions using a case study.



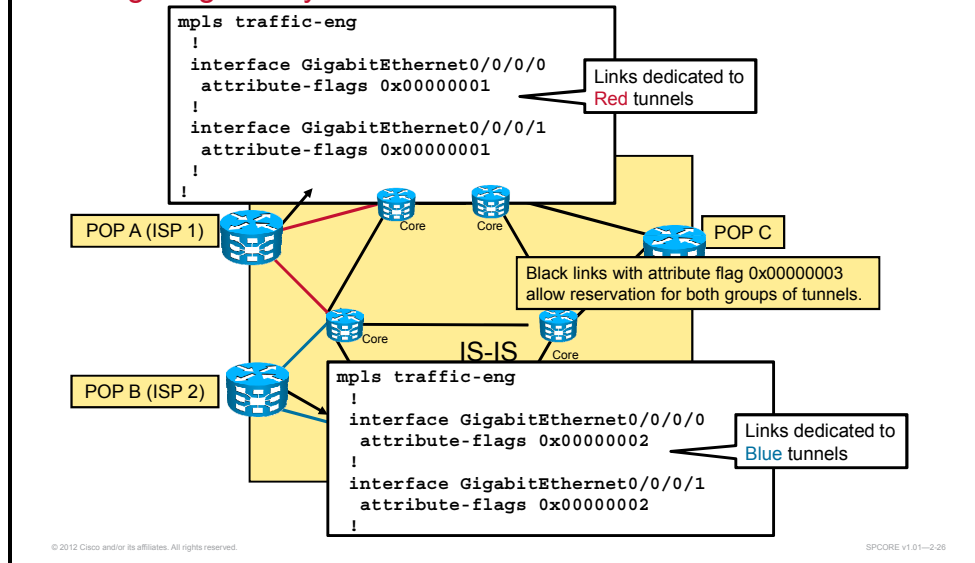
In many cases, some links will need to be excluded from the constraint-based SPF computation. This exclusion can be implemented by using the resource class affinity bits of the traffic tunnel and the resource class bits of the links over which the tunnel should pass (following the computed LSP path).

A 32-bit resource class affinity string that is accompanied by a corresponding resource class mask characterizes each traffic tunnel. The “0” bits in the mask exclude the respective link resource class bits from being checked.

Each link is characterized by its resource class 32-bit string, which is set to 0 by default. The matching of the tunnel resource class affinity string with the resource class string of the link is performed during the LSP path computation.

Configuring Path Selection Restrictions (Cont.)

Configuring Affinity Bits on Links



The figure shows a sample network with the tunnel resource class affinity bits and link resource bits. The main goal is to force the CBR algorithm to use only links that are explicitly dedicated to certain tunnels for its path computation.

Because it is desirable to move all blue tunnels away from POP A interfaces and red tunnels from POP B interfaces, different link resource class bits are set: 0x00000001 for red interfaces and 0x00000002 for blue interfaces. Those link resource class attribute bits then become a part of the LSP advertisements, which allows all participants to include this information when they compute paths for TE tunnels.

#POP-A Cisco IOS Configuration

```
interface GigabitEthernet 0/0
 mpls traffic-eng attribute-flags 0x00000001
interface GigabitEthernet 0/1
 mpls traffic-eng attribute-flags 0x00000001
```

#POP-B Cisco IOS Configuration

```
interface GigabitEthernet 0/0
 mpls traffic-eng attribute-flags 0x00000002
interface GigabitEthernet 0/1
 mpls traffic-eng attribute-flags 0x00000002
```

Configuring Path Selection Restrictions (Cont.)

Configuring Affinity Bits on Tunnels

```
#POP-C configuration
interface Tunnel-te 1
description Red tunnel
ipv4 unnumbered Loopback0
destination POP-A
priority 1 1
affinity 0x00000001 mask 0x00000001
!
interface Tunnel-te 2
description Blue tunnel
ipv4 unnumbered Loopback0
destination POP-B
priority 2 2
affinity 0x00000002 mask 0x00000002
!
```

Link attributes for Red tunnel;
Mask 1 means it must match.

Link attributes for Blue tunnel

In the example of the first tunnel (Red tunnel), the tunnel requires a match only on the last bit, whereas second tunnel (Blue tunnel) checks the setting of the next-to-last bit. With tunnel resource class affinity bits and link resource class bits set, the constraint-based path computation considers only the paths where the match is found.

For Red tunnel, the affinity bits are set at 0x00000001, and the mask is set at 0x00000001; the attributes of the red links (POP A interfaces) match (attributes 0x00000001). The black links (core interfaces) also match (attributes 0x00000011). The blue links (POP B interfaces) that are marked with the attribute 0x00000002 are excluded from constraint-based SPF computation because they do not match.

```
#POP-C Cisco IOS configuration
interface Tunnel1
description Red tunnel
ip unnumbered Loopback0
tunnel destination POP-A
tunnel mpls traffic-eng priority 1 1
tunnel mpls traffic-eng affinity 0x00000001 mask 0x00000001
.....
interface Tunnel2
description Blue tunnel
ip unnumbered Loopback0
tunnel destination POP-B
tunnel mpls traffic-eng priority 2 2
tunnel mpls traffic-eng affinity 0x00000002 mask 0x00000002
.....
```

Configuring Path Selection Restrictions (Cont.)

Excluding Links or Nodes for Traffic Tunnels

```
#POP-C configuration
interface Tunnel-te 1
  description Red tunnel
  ipv4 unnumbered Loopback0
  destination POP-A
  priority 1 1
  path-option 1 explicit name ex1
  !
  explicit-path name ex1
    index 1 exclude-address ipv4 unicast POP-B
  !
interface Tunnel-te 2
  description Blue tunnel
  ipv4 unnumbered Loopback0
  destination POP-B
  priority 2 2
  path-option 1 explicit name ex2
  !
  explicit-path name ex2
    index 1 exclude-address ipv4 unicast POP-A
  !
```

Address is excluded
from the path.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.28

Link or node exclusion is accessible using the **explicit-path** command, which enables you to create an IP explicit path and enter a configuration submode for specifying the path. Link or node exclusion uses the **exclude-address** submode command to specify addresses that are excluded from the path.

If the excluded address for an MPLS TE LSP identifies a flooded link, the constraint-based SPF routing algorithm does not consider that link when computing paths for the LSP. If the excluded address specifies a flooded MPLS TE router ID, the constraint-based SPF routing algorithm does not allow paths for the LSP to traverse the node that is identified by the router ID.

Addresses are not excluded from an IP explicit path unless they are explicitly excluded by the **exclude-address** command.

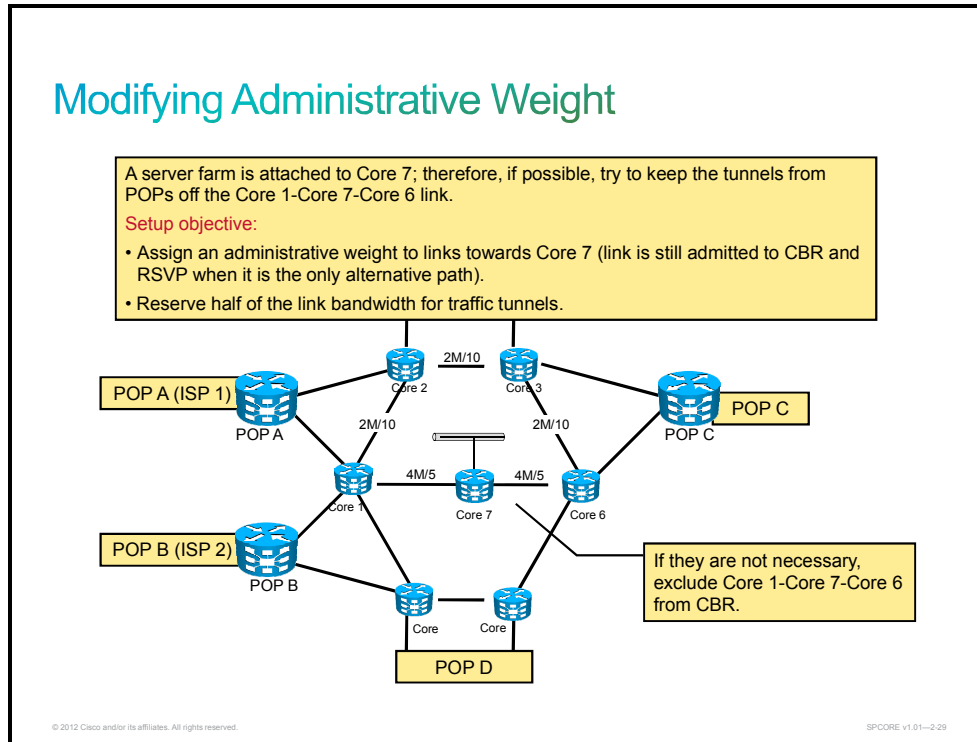
Note MPLS TE will accept an IP explicit path that is composed of either all exclude addresses configured by the **exclude-address** command, or all include addresses configured by the **next-address** command, but not a combination of both.

In a previous example, affinity bits were used to restrict the possible paths that could be used for tunnel creation. This example shows the use of the IP exclude address feature to accomplish the same function. By excluding the desired nodes, tunnels will not use any links that lead to those nodes.

Note POP A and POP B in our example are the router IDs of the specified nodes. Typically, those addresses conform to the loopback address of the node.

MPLS TE Case Study Continue: Modifying the Administrative Weight

This topic explains modifying the Administrative Weight using a case study.



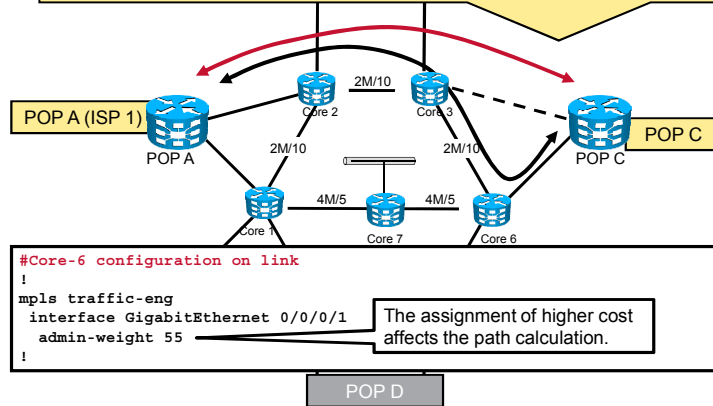
The constrained-based path computation selects the path that the dynamic traffic tunnel will take, based on the administrative weight (TE cost) of each individual link.

This administrative weight is, by default, equal to the IGP link metric (cost). Increase the TE cost on the link if it is desirable to exclude a certain link from any path computation, while keeping the link available if that link represents the only available path.

Modifying Administrative Weight (Cont.)

Configuring Administrative Weight

In the absence of a static path, due to a failed POP C-to-Core 3 link, from the POP C perspective, the preferred path for a dynamic tunnel should go through Core 6-Core 3-Core 2.



In the example in the figure, the TE cost of the link between Core 1 and Core 7 is increased to 55, which makes links providing alternative paths more economical and more attractive for backup purposes.

```
int GigabitEthernet0/0
mpls traffic-eng administrative-weight 55
```

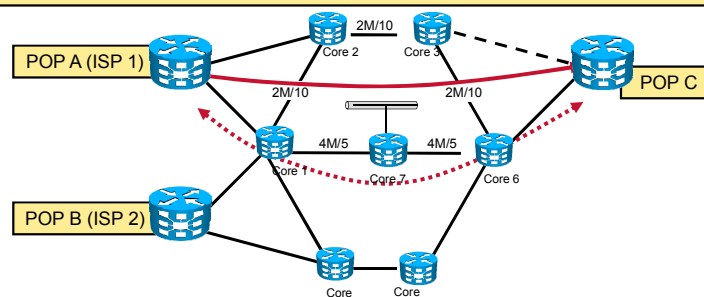
MPLS TE Case Study Continue: Autoroute and Forwarding Adjacency

This topic explains enabling Autoroute and Forwarding Adjacency using a case study.

Autoroute Forwarding Engine

Autoroute refresher:

- The tunnel metric is the IGP cost to the tunnel endpoint, regardless of the actual tunnel path (CBR LSP computation).
- Tune the tunnel metric to prefer one tunnel over the other (**absolute** - a positive metric value; **relative** - a positive, negative, or zero value to the IGP metric).
- The tunnel metric must be equal to or lower than the native IP path, to replace the existing next hops.
- Traffic that is directed to prefixes beyond the tunnel tail end is pushed onto the tunnel (if MPLS is tunneled into the TE tunnel, a stack of labels is used).



The autoroute feature enables the headend routers to see the MPLS TE tunnel as a directly connected interface and to use it in their modified SPF computations. The MPLS TE tunnel is used only for normal IGP route calculation (at the headend only) and is not included in any constraint-based path computation.

With the autoroute feature, the traffic tunnel does this:

- Appears in the routing table
- Has an associated IP metric (cost equal to the best IGP metric to the tunnel endpoint), and is also used for forwarding the traffic to destinations behind the tunnel endpoint

The autoroute feature enables all the prefixes that are topologically behind the MPLS TE tunnel endpoint (tail end) to be reachable via the tunnel itself (unlike static routing where only statically configured destinations are reachable via the tunnel).

Even with the autoroute feature, the tunnel itself is not used in link-state updates and the remainder of the network does not have any knowledge of it.

Autoroute Forwarding Engine (Cont.)

Changing Autoroute Metrics

```
#POP-C configuration
interface Tunnel-te 1
  ipv4 unnumbered Loopback0
  signalled-bandwidth 250
  destination POP-A
  priority 1 1
  path-option 1 explicit name Core3-2
  autoroute announce
  autoroute metric absolute 1
!
explicit-path name Core3-3
  next-address Core-3
  next-address Core-2
  next-address POP-A
!
interface Tunnel-te 2
  ipv4 unnumbered Loopback0
  signalled-bandwidth 125
  destination POP-A
  priority 1 1
  path-option 1 dynamic
  autoroute announce
  autoroute metric absolute 2
!
```

Announce the presence of the tunnel to the routing protocol, and select the primary path.

The metric determines the secondary path.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2-32

Because the autoroute feature includes the MPLS TE tunnel in the modified SPF path calculation, the metric of the tunnel plays a significant role. The cost of the tunnel is equal to the best IGP metric to the tunnel endpoint, regardless of the LSP path.

The tunnel metric is tunable using either relative or absolute metrics, as in the example. When the routing process is selecting the best paths to the destination, the tunnel metric is compared to other existing tunnel metrics and to all the native IGP path metrics. The lower metric is better, and if the MPLS TE tunnel has a lower metric, it is installed as a next hop to the respective destinations.

If there are tunnels with equal metrics, they are installed in the routing table and they provide load balancing. The load balancing is done proportionally to the configured bandwidth of the tunnel.

```
#POP-C Cisco IOS Configuration
interface Tunnel1
  ip unnumbered Loopback0
  tunnel destination POP-A
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng autoroute announce
  tunnel mpls traffic-eng autoroute metric absolute 1
  tunnel mpls traffic-eng priority 1 1
  tunnel mpls traffic-eng bandwidth 250
  tunnel mpls traffic-eng path-option 1 explicit name Core3-2
interface Tunnel2
  ip unnumbered Loopback0
  tunnel destination POP-A
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng autoroute announce
  tunnel mpls traffic-eng autoroute metric absolute 2
  tunnel mpls traffic-eng priority 2 2
  tunnel mpls traffic-eng bandwidth 125
  tunnel mpls traffic-eng path-option 1 dynamic
!
ip explicit-path name Core3-2 enable
  next-address Core-3
  next-address Core-2
  next-address POP-A
```


Autoroute Forwarding Engine (Cont.)

Configuring Forwarding Adjacency

```
#POP-C configuration
interface Tunnel-te 12
  ipv4 unnumbered Loopback0
  destination POP-A
  forwarding-adjacency
!
#POP-A configuration
interface Tunnel-te 21
  ipv4 unnumbered Loopback0
  destination POP-C
  forwarding-adjacency
!
```

Advertise a TE tunnel as a link to the IS-IS network.

For the point-to-point link between POP A and POP C to be announced into IS-IS, a tunnel must exist from POP A to POP C with forwarding adjacency enabled.

To advertise a TE tunnel as a link in an IGP network, use the forwarding adjacency feature. You must configure a forwarding adjacency on two LSP tunnels bidirectionally, from A to B and from B to A. Otherwise, the forwarding adjacency is advertised but not used in the IGP network.

#POP-C Cisco IOS Configuration

```
interface Tunnel12
  ip unnumbered Loopback0
  tunnel destination POP-A
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng forwarding-adjacency
...
```

##POP-C Cisco IOS Configuration

```
interface Tunnel21
  ip unnumbered Loopback0
  tunnel destination POP-C
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng forwarding-adjacency
...
```

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- Several steps are required to configure MPLS TE tunnels
- MPLS TE functionality should be enabled on all routers on the path from headend to tail end of the MPLS TE tunnel
- RSVP support should be enabled on all routers on the path from headend to tail end of the MPLS TE tunnel
- MPLS TE support for OSPF should be enabled on all routers on the path from headend to tail end of the MPLS TE tunnel
- MPLS TE support for IS-IS should be enabled on all routers on the path from headend to tail end of the MPLS TE tunnel
- The MPLS TE tunnel is unidirectional
- The autoroute feature automatically routes traffic to prefixes behind the MPLS TE tunnel
- To display information about MPLS TE tunnels, use the **show mpls traffic-eng tunnels** command

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.34

Summary (Cont.)

- Verify that the prefixes that are behind the tail-end router of the MPLS TE tunnel are reachable through the MPLS TE tunnel interface
- The best route might be any sequence of next-hop routers that are configured at the headend of the tunnel
- Reoptimization occurs when a device examines tunnels with established LSPs to see if better LSPs are available
- In many cases, some links will need to be excluded from the constraint-based SPF computation
- This administrative weight is, by default, equal to the IGP link metric (cost)
- The autoroute feature enables the headend routers to see the MPLS TE tunnel as a directly connected interface and to use it in their modified SPF computations

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.35

Protecting MPLS TE Traffic

Overview

This lesson describes the advanced Multiprotocol Label Switching (MPLS) traffic engineering (TE) commands for link protection and the advanced MPLS TE commands for bandwidth control. The configuration commands are accompanied by usage guidelines and examples.

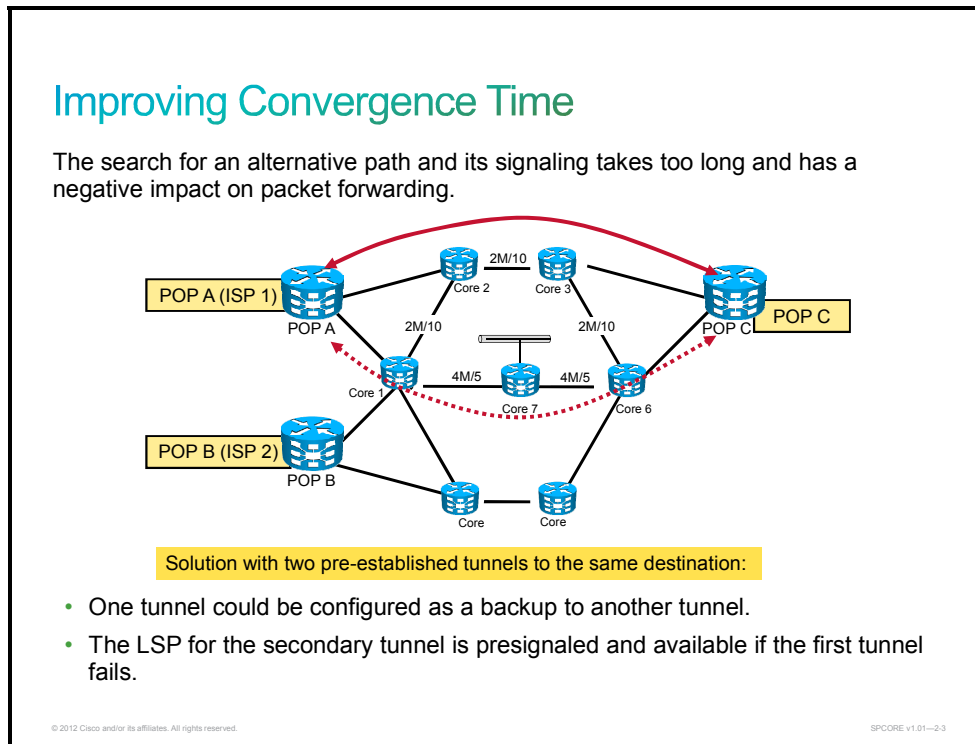
Objectives

Upon completing this lesson, you will be able to describe the MPLS TE commands for the implementation of MPLS traffic tunnels. You will be able to meet these objectives:

- Describe using a backup TE tunnel to improve convergence time
- Explain backup MPLS TE tunnel configurations
- Discuss the drawbacks of using a backup MPLS TE tunnels
- Describe the use of Fast Reroute in a case study
- Explain the use of Link Protection with Fast Reroute using a case study
- Explain the use of Node Protection with Fast Reroute using a case study
- Explain the Fast Reroute Link Protection Configurations using a case study
- Explain the Automatic Bandwidth Adjustment feature and configuration
- Describe the basic DiffServ-Aware MPLS TE Tunnels concept and configuration

Improving MPLS TE Convergence Time

This topic describes using a backup TE tunnel to improve convergence time.



At first, the need for fast convergence is not obvious. Historically, applications were designed to recover from network outages. However, with the current increased usage of voice and video applications, network down time must be kept at a minimum. The amount of time that it previously took for a network to converge after a link failure is now unacceptable.

For example, a flapping link can result in headend routers being constantly involved in constraint-based computations. Because the time that elapses between link failure detection and the establishment of a new label-switched path (LSP) can cause delays for critical traffic, there is a need for pre-established alternative paths (backups).

Here, two tunnels are used between the same endpoints at the same time. The main requirement in this scenario is that preconfigured tunnels between the same endpoints must use diverse paths. As soon as the primary tunnel fails, the traffic is transitioned to the backup tunnel. The traffic is returned to the primary tunnel if conditions provide for the reestablishment of traffic.

Having two pre-established paths is the simplest form of MPLS TE path protection. Several steps must be taken in preparation for effective switching between the tunnels. These steps include routing to the proper tunnel.

Configuring Backup MPLS TE tunnels

This topic explains backup MPLS TE tunnel configurations.

Configuring Parallel Tunnels

```
#POP-C configuration
interface Tunnel-te 1
  ipv4 unnumbered Loopback0
  destination POP-A
  signalled-bandwidth 250
  priority 1 1
  path-option 1 explicit name Core2-3
!
explicit-path name Core3-2
  index 1 next-address ipv4 unicast Core-3
  index 2 next-address ipv4 unicast Core-2
  index 3 next-address ipv4 unicast POP-A
!
interface Tunnel-te 2
  ipv4 unnumbered Loopback0
  destination POP-A
  priority 2 2
  signalled-bandwidth 125
  path-option 1 dynamic
!
router static address-family ipv4 unicast POP-A/1 mask tunnel-te 1 10
router static address-family ipv4 unicast POP-A/1 mask tunnel-te 2 11
```

A lower priority is used, and due to the double counting of reservations, one-half the initial bandwidth is requested by the backup tunnel.

A pair of floating static routes can be used for primary/backup selection.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.4

The example shows two configured tunnels: Tunnel-te 1 (following the LSP path Core 3-Core 2-POP A) and Tunnel-te 2 (using a dynamic path).

In the presence of two tunnels, static routing is deployed with two floating static routes pointing to the tunnels.

As soon as the primary tunnel (Tunnel-te 1) fails, the static route is gone and the traffic is transitioned to the secondary tunnel. The traffic is returned to the primary tunnel if the conditions support the reestablishment of traffic.

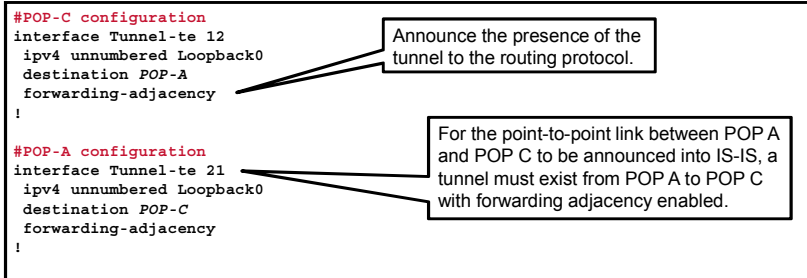
Other options include spreading the load proportionally to the requested bandwidth using the Cisco Express Forwarding mechanism, load balancing, or by having one group of static routes pointing to Tunnel-te 1 and another to Tunnel-te 2.

#POP-C Cisco IOS Configuration

```
interface Tunnel1
  ip unnumbered Loopback0
  tunnel destination POP-A
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng priority 1 1
  tunnel mpls traffic-eng bandwidth 250
  tunnel mpls traffic-eng path-option 1 explicit name Core3-2
interface Tunnel2
  ip unnumbered Loopback0
  tunnel destination POP-A
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng priority 2 2
  tunnel mpls traffic-eng bandwidth 125
  tunnel mpls traffic-eng path-option 1 dynamic
!
ip explicit-path name Core3-2 enable
  next-address Core-3
  next-address Core-2
  next-address POP-A
ip route POP-A/1 mask tunnel 1 10
ip route POP-A/1 mask tunnel 2 11
```

Configuring Parallel Tunnels (Cont.)

- Any of the available methods can be used to advertise the tunnel.
- In this example, forwarding adjacency is used to advertise the TE tunnel as a link in an IGP network.



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.6

Any of the available methods can be used to advertise the tunnel. In this example, forwarding adjacency is used to advertise the TE tunnel as a link in an interior gateway protocol (IGP) network.

You must configure a forwarding adjacency on two LSP tunnels bidirectionally, from A to B and from B to A. Otherwise, the forwarding adjacency is advertised but not used in the IGP network.

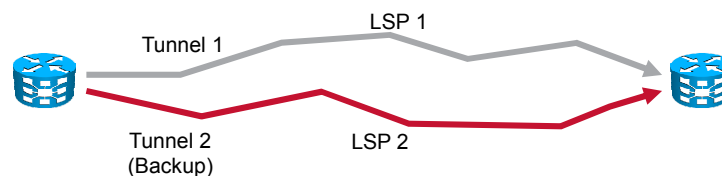
Drawbacks of Backup MPLS TE tunnels

This topic discusses the drawbacks of using a backup MPLS TE tunnels.

Drawbacks of Parallel Tunnels

Path Protection with Preconfigured Tunnels

- Preconfigured tunnels speed up recovery by moving the traffic on a preinstalled LSP as soon as the headend learns that the primary LSP is down.
- Drawbacks:
 - A backup tunnel allocates labels and reserves bandwidth over the entire path.
 - There is double counting of reservations via RSVP over the entire path.



Because the time that elapses between link failure detection and the establishment of a new LSP path can cause delays for critical traffic, there is a need for alternative pre-established paths (backups). Therefore, two tunnels are used between the same endpoints at the same time.

Note Preconfigured tunnels between the same endpoints must use diverse paths.

As soon as the primary tunnel fails, traffic is transitioned to the backup tunnel. The traffic is returned to the primary tunnel if conditions provide for the re-establishment of traffic.

Note Having two pre-established paths is the simplest form of MPLS TE path protection. Another option is to use the precomputed path only and establish the LSP path on demand. In the latter case, there is no overhead in resource reservations.

The figure shows two preconfigured tunnels: Tunnel 1 (LSP 1) is a primary tunnel, and Tunnel 2 (LSP 2) is a backup tunnel. Their physical paths are diverse.

The switchover to the backup tunnel is done at the headend as soon as the primary tunnel failure is detected, via Resource Reservation Protocol (RSVP) or via IGP.

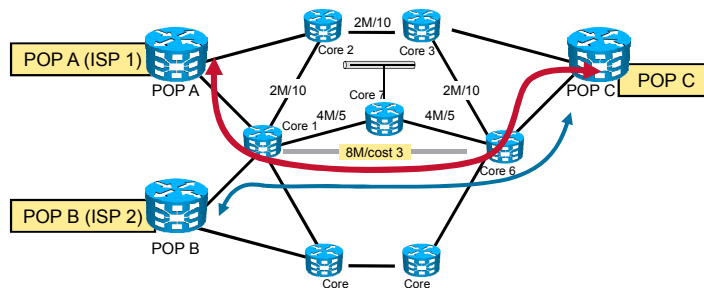
There is an obvious benefit to having a preconfigured backup tunnel. However, the solution presents some drawbacks as well:

- The backup tunnel requires all the mechanisms that the primary one requires. The labels must be allocated and bandwidth must be reserved for the backup tunnel as well.
- From the RSVP perspective, the resource reservations (bandwidth) are counted twice.

Fast Reroute Case Study

This topic describes the use of Fast Reroute in a case study.

Fast Reroute: Case Study



- The company decided to retain only dynamic tunnels. A new high-speed link was introduced between Core 1 and Core 6 to influence CBR and native path selection and to speed up transport across the network.
- The new high-speed link is now heavily used by traffic tunnels and may cause a serious disruption.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.7

In this case study, a company has decided to use only dynamic tunnels. A new high-speed link has been introduced between Core 1 and Core 6. This link influences CBR and native path selection, and speeds up transport across the network. The result, however, is that this new high-speed link is now heavily used by traffic tunnels and may cause a serious disruption if it fails.

Fast Reroute

FRR allows for temporary routing around a **failed link** or a **failed node** while the headend is rerouting the LSP:

- FRR is controlled by the routers with preconfigured backup tunnels around the protected link or node (**link** or **node** protection).
- The headend is notified of the failure through the IGP and through RSVP.
- The headend then attempts to establish a new LSP that bypasses the failure (LSP rerouting).

© 2012 Cisco and/or its affiliates. All rights reserved.

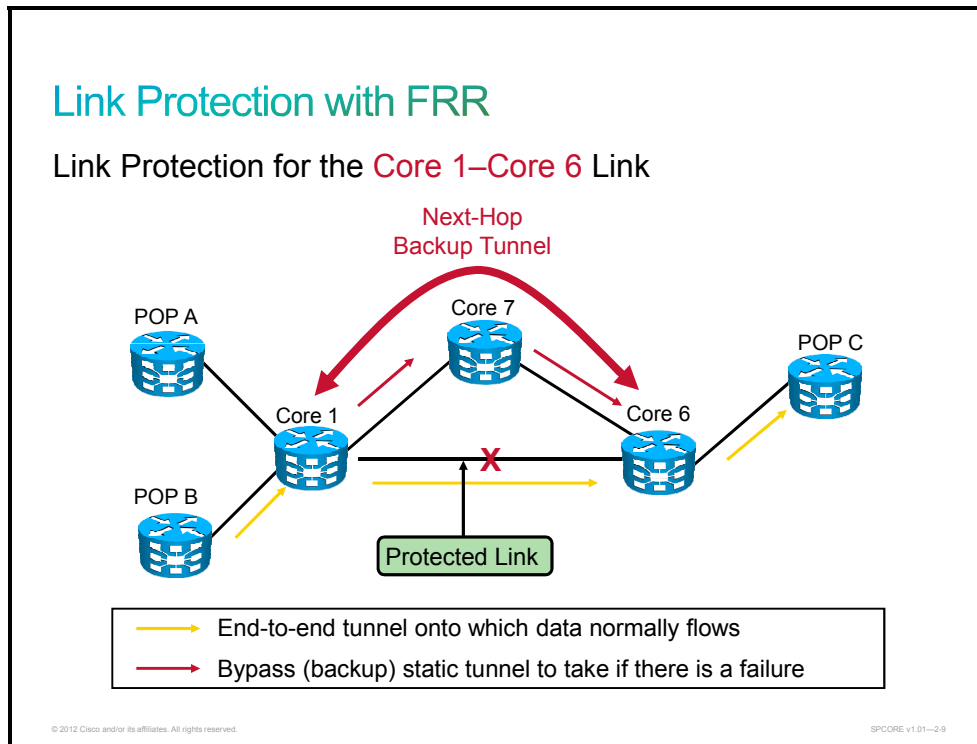
SPCORE v1.01—2-8

Fast Reroute (FRR) is a mechanism for protecting a MPLS TE LSP from link and node failures by locally repairing the LSP at the point of failure. The FRR mechanism allows data to continue to flow while the headend router attempts to establish a new end-to-end LSP that bypasses the failure. FRR locally repairs any protected LSPs by rerouting them over backup tunnels that bypass failed links or nodes.

The headend is notified of the failure through the IGP and through RSVP. The headend then attempts to establish a new LSP that bypasses the failure (LSP rerouting).

Fast Reroute Case Study Continue: Link Protection

This topic explains the use of Link Protection with Fast Reroute using a case study.



Backup tunnels that bypass only a single link of the LSP path provide link protection. They protect LSPs if a link along their path fails by rerouting the LSP traffic to the next hop (bypassing the failed link). These tunnels are referred to as next-hop backup tunnels because they terminate at the next hop of the LSP beyond the point of failure.

This process gives the headend of the tunnel time to reestablish the tunnel along a new, optimal route.

Link Protection with FRR (Cont.)

“Link Down” Event

- The router realizes that the link is down.
 - The router issues an IGP advertisement.
 - The router Issues an RSVP message with session attribute flag 0x01=ON . (This means, “Do not break the tunnel; you may continue to forward packets during the reoptimization.”)
- In the event of a failure, an LSP is intercepted and locally rerouted using a backup tunnel.
 - The original LSP is nested within a protection LSP.
 - There is minimum disruption of LSP flow (under 50 ms, time to detect and switch).
- The headend is notified by RSVP PathErr and by IGP.
 - A special flag in RSVP PathErr (reservation in place) indicates that the path states must not be destroyed, so that the LSP flow is not interrupted.
 - The headend of the tunnel smoothly reestablishes the tunnel along a new route.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2-10

Paths for LSPs are calculated at the LSP headend. Under failure conditions, the headend determines a new route for the LSP. Recovery at the headend provides for the optimal use of resources. However, because of messaging delays, the headend cannot recover as fast as making a repair at the point of failure.

To avoid packet flow disruptions while the headend is performing a new path calculation, the FRR option of MPLS TE is available to provide protection from link or node failures (failure of a link or an entire router).

The function is performed by routers that are directly connected to the failed link, because they reroute the original LSP to a preconfigured tunnel and therefore bypass the failed path.

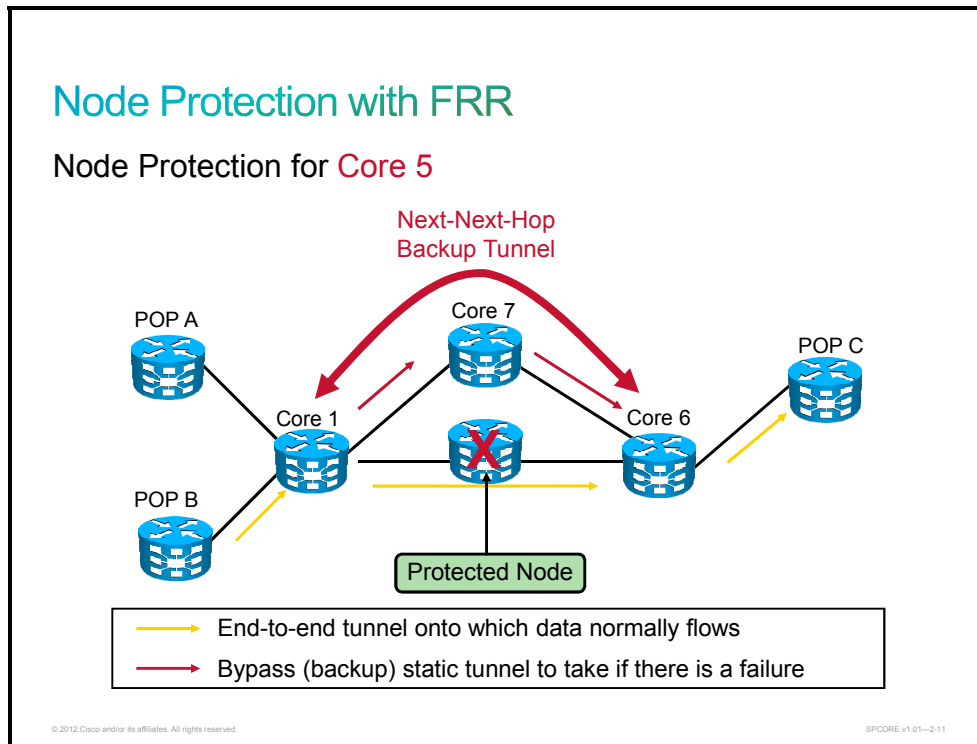
Note In terms of forwarding, it can be said that the original LSP is nested within the protection LSP.

The reaction to a failure, with such a preconfigured tunnel, is almost instant. The local rerouting takes less than 50 ms, and a delay is only caused by the time that it takes to detect the failed link and to switch the traffic to the link protection LSP.

When the headend of the tunnel is notified of the path failure through the IGP or through RSVP, it attempts to establish a new LSP.

Fast Reroute Case Study Continue: Node Protection

This topic explains the use of Node Protection with Fast Reroute using a case study.



FRR provides node protection for LSPs.

Backup tunnels that bypass next-hop nodes along LSP paths are called next-next-hop backup tunnels because they terminate at the node following the next-hop node of the LSP path, thereby bypassing the next-hop node. They protect LSPs, if a node along their path fails, by enabling the node upstream of the failure to reroute the LSPs and their traffic around the failed node to the next-next hop.

FRR supports the use of RSVP hellos to accelerate the detection of node failures. Next-next-hop backup tunnels also provide protection from link failures, because they bypass the failed link as well as the node.

Node Protection with FRR (Cont.)

- Router node fails; the router detects this failure by an “interface down” notification.
 - It switches LSPs going out that interface onto their respective backup tunnels (if any).
- RSVP hellos can also be used to trigger FRR.
 - Messages are periodically sent to the neighboring router.
 - If no response is received, hellos declare that the neighbor is down.
 - This causes any LSPs going out that interface to be switched to their respective backup tunnels.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01—2-12

When a router link or a neighboring node fails, the router often detects this failure by an “interface down” notification. On a gigabit switch router (GSR) packet over SONET (POS) interface, this notification is very fast. When a router notices that an interface has gone down, it switches LSPs going out that interface onto their respective backup tunnels (if any).

RSVP hellos can also be used to trigger FRR. If RSVP hellos are configured on an interface, messages are periodically sent to the neighboring router. If no response is received, the hellos declare that the neighbor is down. This action causes any LSPs going out that interface to be switched to their respective backup tunnels.

Fast Reroute Case Study Continue: Fast Reroute Link Protection Configurations

This topic explains the Fast Reroute Link Protection Configurations using a case study.

FRR Configuration

- Enable FRR on LSPs.
- Create a backup tunnel to the next hop or to the next-next hop.
- Assign backup tunnels to a protected interface.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.13

This section assumes that you want to add FRR protection to a network in which MPLS TE LSPs are configured.

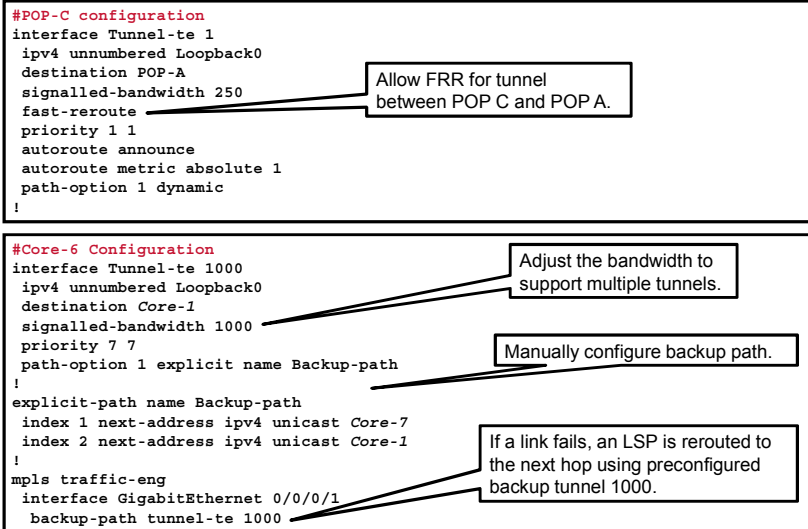
Before performing the configuration tasks in this topic, it is assumed that you have done these tasks:

- Enabled MPLS TE on all relevant routers and interfaces
- Configured MPLS TE tunnels

Here are the tasks that are required to use FRR to protect LSPs in your network from link or node failures:

- Enable FRR on LSPs.
- Create a backup tunnel to the next hop or to the next-next hop.
- Assign backup tunnels to a protected interface.

FRR Configuration: Link Protection



The example in the figure lists both sets of configuration commands that are needed when you are provisioning a backup for a link over a tunnel:

- Configuration of Core 6 to provide a backup tunnel around the protected link
- POP C configuration of a tunnel and FRR assignment

On Cisco IOS XR Software, use the **fast-reroute** interface configuration command to enable an MPLS TE tunnel to use a backup tunnel in the event of a link failure (if a backup tunnel exists).

On Cisco IOS XR Software, to configure the interface to use a backup tunnel in the event of a detected failure, use the **backup-path tunnel-te** command in the appropriate mode.

POP-C Ciso IOS Software Configuration

```
interface Tunnell
  ip unnumbered Loopback0
  tunnel destination POP-A
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng autoroute announce
  tunnel mpls traffic-eng autoroute metric absolute 1
  tunnel mpls traffic-eng fast-reroute
  tunnel mpls traffic-eng priority 1 1
  tunnel mpls traffic-eng bandwidth 250
  tunnel mpls traffic-eng path-option 1 dynamic
```

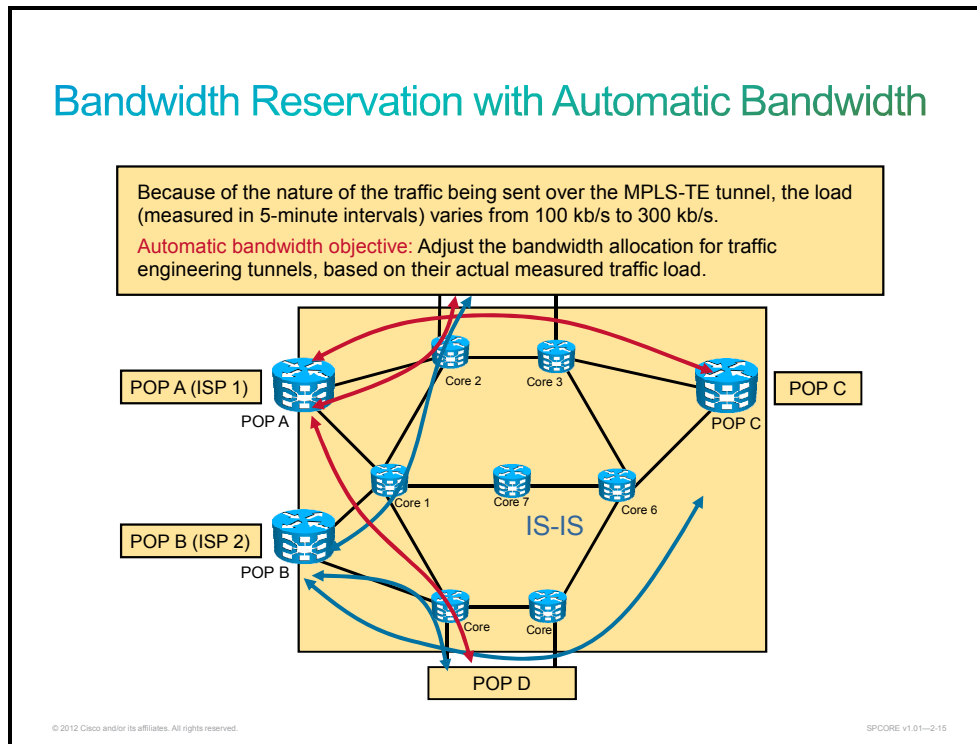
#Core-6 Ciso IOS Software Configuration

```
interface Tunnel 1000
  ip unnumbered Loopback0
  tunnel destination Core-1
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng priority 7 7
  tunnel mpls traffic-eng bandwidth 1000
```

```
NO tunnel mpls traffic-eng autoroute announce
tunnel mpls traffic-eng path-option 1 explicit name Backup-path
!
ip explicit-path name Backup-path enable
next-address strict Core-7
next-address strict Core-1
interface ser0/0
mpls traffic-eng backup tunnel 1000
```


MPLS TE Bandwidth Control

This topic explains the Automatic Bandwidth Adjustment feature and configuration.



After initial TE is complete, network administrators may need an effective way to continually adjust tunnel routes and bandwidth reservations without doing any redesigning.

Both Cisco IOS Software and Cisco IOS XR Software have an automatic bandwidth adjustment feature that measures utilization averages and dynamically adjusts tunnel bandwidth reservations to meet actual application resource requirements.

This powerful feature creates “self-tuning tunnels” that relieve network administrators of many of the daily hands-on management tasks that are necessary with other TE techniques.

The MPLS TE automatic bandwidth feature measures the traffic in a tunnel and periodically adjusts the signaled bandwidth for the tunnel. MPLS TE automatic bandwidth is configured on individual LSPs at every headend. MPLS TE monitors the traffic rate on a tunnel interface. Periodically, MPLS TE resizes the bandwidth on the tunnel interface to align it closely with the traffic in the tunnel. MPLS TE automatic bandwidth can perform these functions:

- Monitor periodic polling of the tunnel output rate
- Resize the tunnel bandwidth by adjusting the highest rate observed during a given period

Bandwidth Reservation with Automatic Bandwidth (Cont.)

The traffic engineering automatic bandwidth feature adjusts the bandwidth allocation for TE tunnels based on their measured traffic load:

- This feature periodically changes tunnel bandwidth reservation based on the traffic output of the tunnel.
- The average output rate is sampled for each tunnel.
- The allocated bandwidth is periodically adjusted to be the largest sample for the tunnel since the last adjustment.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.18

TE automatic bandwidth adjustment provides the means to automatically adjust the bandwidth allocation for TE tunnels based on their measured traffic load.

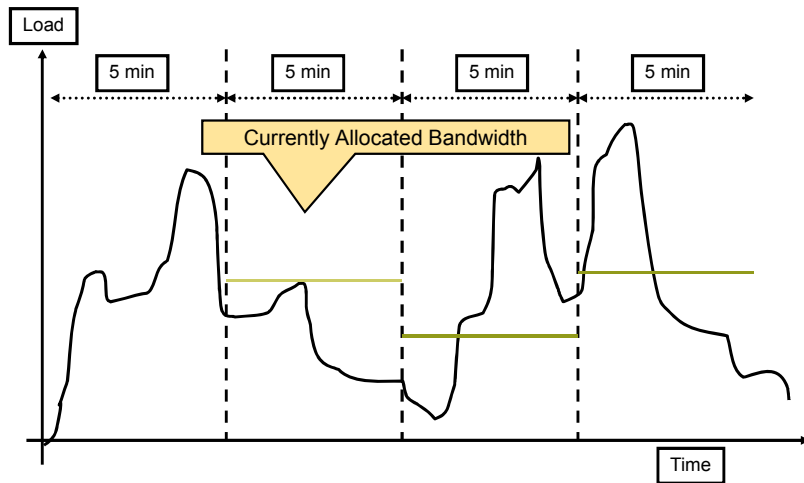
TE automatic bandwidth adjustment samples the average output rate for each tunnel that is marked for automatic bandwidth adjustment. For each marked tunnel, it periodically (for example, once per day) adjusts the allocated bandwidth of the tunnel to be the largest sample for the tunnel since the last adjustment.

The frequency with which tunnel bandwidth is adjusted and the allowable range of adjustments are configurable on a per-tunnel basis. In addition, both the sampling interval and the interval used to average the tunnel traffic to determine the average output rate are user-configurable on a per-tunnel basis.

The benefit of the automatic bandwidth feature is that it makes it easy to configure and monitor the bandwidth for MPLS TE tunnels. If automatic bandwidth is configured for a tunnel, TE automatically adjusts the bandwidth of the tunnel.

The automatic bandwidth adjustment feature treats each enabled tunnel independently. In other words, it adjusts the bandwidth for each such tunnel according to the adjustment frequency that is configured for the tunnel and the sampled output rate for the tunnel since the last adjustment, without regard for any adjustments previously made or pending for other tunnels.

Bandwidth Reservation with Automatic Bandwidth (Cont.)



The diagram shows the load on the tunnel and intervals of measurement. The input and output rates on the tunnel interfaces are averaged over a predefined interval (load interval). In the example, the interval is the previous 5 minutes.

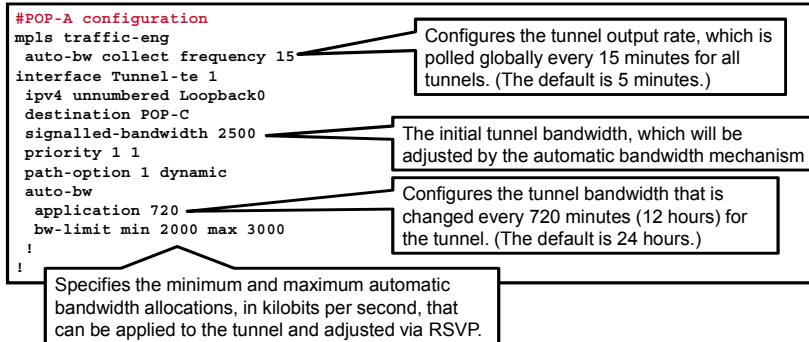
The automatic bandwidth adjustments are done periodically, for example, once per day. For each tunnel for which automatic bandwidth adjustment is enabled, the platform maintains information about sampled output rates and the time remaining until the next bandwidth adjustment.

When the adjustments are done, the currently allocated bandwidth (shown as horizontal solid lines in the diagram) is reset to the maximum of one of the following:

- The largest average rate that has been sampled since the last bandwidth adjustment
- The configured maximum value

If the new bandwidth is not available, the previously allocated bandwidth is maintained.

Automatic Bandwidth Configuration Commands



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2-18

The example in the figure shows the setting of MPLS traffic-engineered tunnels that can actually tune their own bandwidth requirements to increase or decrease their RSVP reservations, as warranted by changing network conditions.

When readjusting bandwidth constraint on a tunnel, a new RSVP TE path request is generated, and if the new bandwidth is not available, the last good LSP will continue to be used. The network experiences no traffic interruptions.

For every MPLS TE tunnel that is configured for automatic bandwidth adjustment, the average output rate is sampled, based on various configurable parameters. The tunnel bandwidth is then readjusted automatically based on the largest average output rate that was noticed during a certain interval or a configured maximum bandwidth value.

Automatic bandwidth allocation monitors the X minutes (default = 5 minutes) average counter, keeping track of the largest average over some configurable interval Y (default = 24 hours), and then readjusting a tunnel bandwidth based upon the largest average for that interval.

The automatic bandwidth feature is implemented with the following commands on Cisco IOS XR Software:

- **auto-bw collect frequency *minutes***
Configures the automatic bandwidth collection frequency, and controls the manner in which the bandwidth for a tunnel collects output rate information but does not adjust the tunnel bandwidth. By default, this value is set to 5 minutes.
- **application *minutes***
Configures the application frequency in minutes for the applicable tunnel. By default, the frequency is 24 hours.
- **bw-limit {*min bandwidth*} {*max bandwidth*}**
Configures the minimum and maximum automatic bandwidth in kilobits per second, set on a tunnel.

The automatic bandwidth feature is implemented with these commands on Cisco IOS Software:

- **mpls traffic-eng auto-bw timers** [*frequency seconds*]

This is a global command to define the interval during which to sample the X average for each tunnel. By default, this value is set to 300 seconds (5 minutes).

- **clear mpls traffic-eng auto-bw timers**

This command is used to clear the timers that were defined by the previous command.

- **tunnel mpls traffic-eng auto-bw** {*frequency seconds*} {*max-bw kbs*} {*min-bw kbs*}

By default, the frequency is 24 hours.

The last command controls the Y interval between bandwidth readjustments and is tunnel-specific. Setting the **max-bw** value limits the maximum bandwidth that a tunnel can adjust to. Similarly, setting the **min-bw** value provides the smallest bandwidth that the tunnel can adjust to. When both **max-bw** and **min-bw** values are specified, the tunnel bandwidth remains between these values.

#POP-A Cisco IOS configuration

```
mpls traffic-eng auto-bw timers frequency 300
```

```
interface tunnel1
```

```
ip unnumbered loopback 0
```

```
tunnel destination POP-C
```

```
tunnel mode mpls traffic-eng
```

```
tunnel mpls traffic-eng bandwidth 2500
```

```
tunnel mpls traffic-eng priority 1 1
```

```
tunnel mpls traffic-eng path-option 1 dynamic
```

```
tunnel mpls traffic-eng auto bw frequency 3600 max-bw 3000 min-bw 1000
```

DiffServ-Aware MPLS TE Tunnels

This topic describes the basic DiffServ-Aware MPLS TE Tunnels concept and configuration.

DiffServ-Aware Service

Further enhancement to satisfy stricter requirements:

- Separate pools to allow for different limits for the amount of traffic admitted on any given link:
 - One subpool, for tunnels that carry traffic requiring strict bandwidth guarantees or delay guarantees
 - The global pool, for best-effort or DiffServ traffic

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.19

MPLS DiffServ-Aware Traffic Engineering (DS-TE) is an extension of the regular MPLS TE feature. Regular traffic engineering does not provide bandwidth guarantees to different traffic classes. A single bandwidth constraint is used in regular TE that is shared by all traffic. To support various classes of service (CoS), users can configure multiple bandwidth constraints. These bandwidth constraints can be treated differently, based on the requirement for the traffic class using that constraint.

MPLS DS-TE enables you to configure multiple bandwidth constraints on an MPLS-enabled interface. Available bandwidths from all configured bandwidth constraints are advertised using IGP. The TE tunnel is configured with bandwidth value and class-type requirements. Path calculation and admission control take the bandwidth and class-type into consideration. RSVP is used to signal the TE tunnel with bandwidth and class-type requirements.

Providing Strict QoS Guarantees Using DiffServ-Aware TE Subpool Tunnels

A tunnel using the subpool bandwidth can satisfy the stricter quality of service (QoS) requirements if you do all of the following:

1. Select a queue, or in differentiated services (DiffServ) terminology, select a per-hop behavior (PHB), to be used exclusively by the strict guarantee traffic.

If delay or jitter guarantees are sought, the DiffServ Expedited Forwarding PHB (EF PHB) is used. You must configure the bandwidth of the queue to be at least equal to the bandwidth of the subpool.

If only bandwidth guarantees are sought, the DiffServ Assured Forwarding PHB (AF PHB) is used.

2. Ensure that the guaranteed traffic that is sent through the subpool tunnel is placed in the queue *at the outbound interface of every tunnel hop*, and that no other traffic is placed in this queue.

You do this by marking the traffic that enters the tunnel with a unique value in the mpls exp bits field, and steering only traffic with that marking into the queue.

3. Ensure that this queue is never oversubscribed; that is, see that no more traffic is sent into the subpool tunnel than the queue can manage.

You do this by rate-limiting the guaranteed traffic before it enters the subpool tunnel. The aggregate rate of all traffic entering the subpool tunnel should be less than or equal to the bandwidth capacity of the subpool tunnel. Excess traffic can be dropped (in the case of delay or jitter guarantees) or can be marked differently for preferential discard (in the case of bandwidth guarantees).

4. Ensure that the amount of traffic that is entering the queue is limited to an appropriate percentage of the total bandwidth of the corresponding outbound link. The exact percentage to use depends on several factors that can contribute to accumulated delay in your network: your QoS performance objective, the total number of tunnel hops, the amount of link fan-in along the tunnel path, the burstiness of the input traffic, and so on.

You do this by setting the subpool bandwidth of each outbound link to the appropriate percentage of the total link bandwidth.

Providing Differentiated Service Using DiffServ-Aware TE Global Pool Tunnels

You can configure a tunnel using global pool bandwidth to carry best-effort as well as several other classes of traffic. Traffic from each class can receive DiffServ service if you do all of the following:

1. Select a separate queue (a distinct DiffServ PHB) for each traffic class. For example, if there are three classes (gold, silver, and bronze), there must be three queues (DiffServ AF2, AF3, and AF4).
2. Mark each class of traffic using a unique value in the MPLS experimental bits field (for example, gold = 4, silver = 5, bronze = 6).
3. Ensure that packets marked as “gold” are placed in the gold queue, “silver” in the silver queue, and so on. The tunnel bandwidth is set based on the expected aggregate traffic across all classes of service.

To control the amount of DiffServ tunnel traffic that you intend to support on a given link, adjust the size of the global pool on that link.

DiffServ-Aware Service (Cont.)

- An extension of MPLS-TE (basically, the signaling feature)
- Allows CBR of tunnels to use more restrictive bandwidth constraints
 - Dual-bandwidth pool traffic engineering
 - Global pool
 - Subpool
- DiffServ ensures that bandwidth for DiffServ-aware TE tunnels is set aside on each link in the network.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.20

DS-TE extends the current MPLS TE capabilities to introduce the awareness of a particular class of traffic, which is the guaranteed bandwidth traffic. DS-TE enables the service provider to perform a separate admission control and route computation of the guaranteed bandwidth traffic. DS-TE is another signaling feature of IGP and RSVP.

With only a single bandwidth pool on the link in traditional MPLS TE, when the bandwidth is reserved for the tunnel, the traffic within the tunnel is considered as a single class. For example, when voice and data are intermixed within the same tunnel, QoS mechanisms cannot ensure better service for the voice. Usually, class-based weighted fair queuing (CBWFQ) can be performed for the tunnel.

The idea behind DS-TE is to *guarantee* the bandwidth for DS-TE tunnels across the network. For critical applications (for example, voice), a separate DS-TE tunnel is created. Thus, two bandwidth pools are used, one for traditional MPLS TE tunnels and one for DS-TE tunnels. The DiffServ QoS mechanisms (low latency queuing, or LLQ) ensure that bandwidth is dedicated for DS-TE tunnels. In the initial phase, the DS-TE supports a single class of bandwidth. It is expected that subsequent phases of DS-TE will provide new capabilities, such as the support of multiple classes of bandwidth and the dynamic reprogramming of queuing or scheduling mechanisms.

DiffServ-Aware Service (Cont.)

MPLS-TE has these extensions for DS-TE:

- Two types of **bandwidth** limits per interface
- IGP advertising both types of **bandwidth**
- Tunnel configured with appropriate **bandwidth** type
- Appropriate **bandwidth** type considered in path calculations
- Tunnel signaled (via RSVP) with the appropriate **bandwidth** type

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2.21

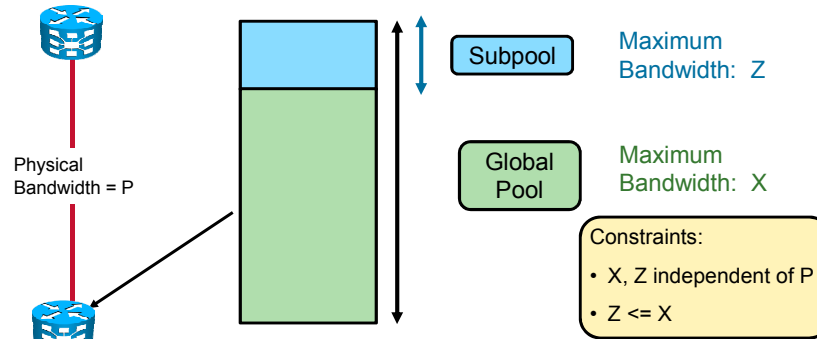
DS-TE tunnels are similar to regular TE tunnels. To support DS-TE, the following modifications to regular MPLS TE mechanisms have been made:

- There are two types of bandwidth per each link in the network (two bandwidth pools: the global pool and the subpool).
- Both of these bandwidths are announced in the link-state updates that carry resource information.
- The traffic tunnel parameters include the bandwidth type that the tunnel will use.
- The constraint-based path calculation (PCALC) is done with respect to the type of the bandwidth that the tunnel requires. In RSVP messages, it is always indicated whether the LSP to be set up is a regular MPLS TE tunnel or a DS-TE tunnel. Intermediate nodes perform admission control and bandwidth allocation (“locking” for the DS-TE) for the appropriate bandwidth pool.

DiffServ-Aware Service (Cont.)

DS-TE Dual-Bandwidth Pools

- **Global pool tracks the true available bandwidth** (taking into account the bandwidth used by both types of tunnels).
- **Subpool keeps track of only the constraint for the DS-TE.**



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.22

On each link in the network, two bandwidth pools are established:

- The global (main) pool keeps track of the true available bandwidth. The pool takes into account the bandwidth that is used by both of the tunnels.
- The subpool (DS-TE) tracks only the bandwidth for the DS-TE tunnels.

The bandwidths that are specified for both pools are independent of the actual physical bandwidth of the link (providing for oversubscription). The same situation applies to traditional MPLS TE with one bandwidth pool.

The only constraint for the two pools is that the bandwidth of the subpool (dedicated to DS-TE tunnels) must not exceed the bandwidth of the global pool.

DiffServ-Aware Service Configuration

```
router(config-if)#
```

```
ip rsvp bandwidth interface-kbps single-flow-kbps [sub-pool kbps]
```

- The sum of the bandwidth that is used by all tunnels on this interface cannot exceed the *interface-kbps* value, and the sum of bandwidth used by all subpool tunnels cannot exceed the *sub-pool kbps* value.

```
router(config-if)#
```

```
tunnel mpls traffic-eng bandwidth {sub-pool | [global]}  
bandwidth
```

- Configure the bandwidth of the tunnel and assign it to either the subpool or the global pool.

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01--2.23

DiffServ-Aware Service Configuration on Cisco IOS XR Software

To configure RSVP bandwidth on an interface using prestandard DS-TE mode, use the **bandwidth** command in RSVP interface configuration mode. To reset the RSVP bandwidth on that interface to its default value, use the **no** form of this command.

bandwidth [*total reservable bandwidth*] [**bc0** *bandwidth*] [**global-pool** *bandwidth*] [**sub-pool** *reservable-bw*]

Syntax Description

Parameter	Description
<i>total reservable bandwidth</i>	(Optional) Total reservable bandwidth (in kilobits per second) that RSVP accepts for reservations on this interface. The range is 0 to 4294967295 kb/s.
bc0 <i>bandwidth</i>	Configures the total reservable bandwidth in the bc0 pool (in kilobits, megabits, or gigabits per second). The default is kilobits per second. The range is 0 to 4294967295.
global-pool <i>bandwidth</i>	(Optional) Configures the total reservable bandwidth in the global pool. The range is 0 to 4294967295 kb/s.
sub-pool <i>kbps</i>	Amount of bandwidth (in kilobits per second) on the interface that is to be reserved to a portion of the total. The range is from 1 to the value of <i>interface-kbps</i> .

To configure the bandwidth that is required for an MPLS TE tunnel, use the **signalled-bandwidth** command in interface configuration mode. To return to the default behavior, use the **no** form of this command.

signalled-bandwidth {*bandwidth* [*class-type* *ct*] | **sub-pool** *bandwidth*}

no signalled-bandwidth {*bandwidth* [*class-type* *ct*] | **sub-pool** *bandwidth*}

tunnel mpls traffic-eng bandwidth {sub-pool | [global]} *bandwidth*

Syntax Description

Parameter	Description
<i>bandwidth</i>	Bandwidth that is required for an MPLS TE tunnel. Bandwidth is specified in kilobits per second. By default, bandwidth is reserved in the global pool. The range is 0 to 4294967295.
class-type <i>ct</i>	(Optional) Configures the class type of the tunnel bandwidth request. The range is 0 to 1. Class-type 0 is strictly equivalent to global-pool. Class-type 1 is strictly equivalent to the subpool.
sub-pool <i>bandwidth</i>	Reserves the bandwidth in the subpool instead of the global pool. The range is 1 to 4294967295. A subpool bandwidth value of 0 is not allowed.

DiffServ-Aware Service Configuration on Cisco IOS Software

To enable RSVP for IP on an interface, use the **ip rsvp bandwidth** interface configuration command.

ip rsvp bandwidth *interface-kbps single-flow-kbps [sub-pool kbps]*

Syntax Description

Parameter	Description
<i>interface-kbps</i>	Amount of bandwidth (in kilobits per second) on the interface to be reserved. The range is 1 to 10000000.
<i>single-flow-kbps</i>	Amount of bandwidth (in kilobits per second) that is allocated to a single flow (ignored in DS-TE). The range is 1 to 10000000.
sub-pool <i>kbps</i>	Amount of bandwidth (in kilobits per second) on the interface to be reserved to a portion of the total. The range is from 1 to the value of <i>interface-kbps</i> .

To configure the bandwidth that is required for an MPLS traffic engineering tunnel, use the **tunnel mpls traffic-eng bandwidth** command in interface configuration mode. To disable this bandwidth configuration, use the **no** form of this command.

tunnel mpls traffic-eng bandwidth {sub-pool | [global]} *bandwidth*

Syntax Description

Parameter	Description
sub-pool	(Optional) Indicates a subpool tunnel.
global	(Optional) Indicates a global pool tunnel. Entering this keyword is not necessary, for all tunnels are "global pool" in the absence of the keyword sub-pool .
<i>bandwidth</i>	The bandwidth, in kilobits per second, that is set aside for the MPLS TE tunnel. The range is between 1 and 4294967295.

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- Having two pre-established paths is the simplest form of MPLS TE path protection
- Any of the available methods can be used to advertise the tunnel
- Solution with backup tunnels presents some drawbacks (allocation of labels, resource reservation)
- Fast Reroute (FRR) is a mechanism for protecting a MPLS TE LSP from link and node failures by locally repairing the LSP at the point of failure
- Backup tunnels that bypass only a single link of the LSP path provide link protection
- Backup tunnels that bypass next-hop nodes along LSP paths are called next-next-hop backup tunnels
- On Cisco IOS XR Software, use the **fast-reroute** interface configuration command to enable an MPLS TE tunnel
- The MPLS TE automatic bandwidth feature measures the traffic in a tunnel and periodically adjusts the signaled bandwidth for the tunnel
- MPLS DiffServ-Aware Traffic Engineering (DS-TE) is an extension of the regular MPLS TE feature

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-2-24

Module Summary

This topic summarizes the key points that were discussed in this module.

Module Summary

- Traffic tunnels are configured with a set of resource requirements, such as bandwidth and priority. Traffic tunnel attributes affect how the path is set up and maintained.
- CSPF augments the link cost by considering other factors such as bandwidth availability or link latency when choosing a path. RSVP with TE extensions is used for establishing and maintaining LSPs.
- Dynamic constraint-based path computation is triggered by the headend of the tunnel.
- Fast reroute provides link protection to LSPs by establishing a backup LSP tunnel for a troubled link.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--2.1

This module discussed the requirement for traffic engineering (TE) in modern service provider networks that must attain optimal resource utilization. The traffic-engineered tunnels provide a means of mapping traffic streams onto available networking resources in a way that prevents the overuse of subsets of networking resources while other subsets are underused.

All the concepts and mechanics that support TE were presented, including tunnel path discovery with link-state protocols and tunnel path signaling with Resource Reservation Protocol (RSVP). Some of the advanced features of TE, such as automatic bandwidth allocation and guaranteed bandwidth, are introduced as well. Label-switched path (LSP) setup is always initiated at the headend of a tunnel. TE tunnels can be used for IP routing only if the tunnels are explicitly specified for routing.

This module explained the configuration of routers to enable basic traffic tunnels, the assignment of traffic to a tunnel, the control of path selection, and the performance of tunnel protection and tunnel maintenance. Configurations were shown for various Cisco platforms.

Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

- Q1) Which two situations can result in network congestion? (Choose two.) (Source: Introducing MPLS Traffic Engineering Components)
- A) The least-cost network path is being used to route traffic.
 - B) Traffic streams are efficiently mapped onto available resources.
 - C) Network resources are insufficient to accommodate the offered load.
 - D) Traffic streams are inefficiently mapped onto available resources.
- Q2) When you are using TE with a Layer 2 overlay model, which two options transport traffic across a network? (Choose two.) (Source: Introducing MPLS Traffic Engineering Components)
- A) DVCs
 - B) PVCs
 - C) RVCs
 - D) SVCs
- Q3) When you are using a traffic-engineered Layer 3 model, which two of the following are limitations? (Choose two.) (Source: Introducing MPLS Traffic Engineering Components)
- A) using static routes
 - B) using policy routing
 - C) unavailability of explicit routing
 - D) path computation that is based on just the IGP metric
- Q4) A set of data flows that share some common feature, attribute, or requirement is called _____. (Source: Introducing MPLS Traffic Engineering Components)
- A) a static route
 - B) a policy route
 - C) a TE tunnel
 - D) TE LSP
- Q5) Which two options can be used to advertise a traffic tunnel so that it will appear in the IP routing table? (Choose two.) (Source: Introducing MPLS Traffic Engineering Components)
- A) the autoroute feature
 - B) a TE mapping statement
 - C) static routes
 - D) the mpls traffic-eng tunnels command
- Q6) What is the role of RSVP in an MPLS TE implementation? (Source: Introducing MPLS Traffic Engineering Components)
- A) It identifies the best path for the tunnel.
 - B) It reserves the bandwidth that is required by the tunnel.
 - C) It performs the CBR calculations for the tunnel setup.
 - D) It assigns the label for the MPLS LSP.

- Q7) Which three options affect how a path is set up? (Choose three.) (Source: Introducing MPLS Traffic Engineering Components)
- A) priority
 - B) bandwidth
 - C) affinity attributes
 - D) MPLS label stack
 - E) MPLS EXP bit
- Q8) Which statement is correct concerning the RSVP sessions in an MPLS-TE application? (Source: Introducing MPLS Traffic Engineering Components)
- A) The sessions are run between the routers at the tunnel endpoints.
 - B) The sessions are run between hosts.
 - C) Two set of sessions are run, one between the hosts and their tunnel end routers, and another between the tunnel end routers.
- Q9) Admission control is invoked by the _____ message. (Source: Introducing MPLS Traffic Engineering Components)
-
- Q10) If there is a network failure, traffic tunnels are rerouted by _____. (Source: Introducing MPLS Traffic Engineering Components)
- A) the headend router
 - B) the upstream router that is nearest to the point of failure
 - C) the downstream router that is nearest to the point of failure
 - D) the tunnel end router
- Q11) During path reoptimization, which router first attempts to identify a better LSP? (Source: Introducing MPLS Traffic Engineering Components)
- A) the headend router
 - B) any router that has identified that it has new resources available
 - C) the tunnel end router
- Q12) Which option solves the problem of setting up two tunnels and having the resources counted twice? (Source: Introducing MPLS Traffic Engineering Components)
- A) path reuse
 - B) path monitoring
 - C) path rerouting
 - D) path reoptimization
- Q13) Which method is used to calculate the LSP? (Source: Introducing MPLS Traffic Engineering Components)
- A) CBR
 - B) DUAL algorithm
 - C) SPF algorithm
 - D) no calculation used
- Q14) Do statically configured destinations list their tunnels in the routing table? (Source: Introducing MPLS Traffic Engineering Components)
- A) Yes, they are listed as incoming interfaces.
 - B) Yes, they are listed as loopback interfaces.
 - C) Yes, they are listed as outgoing interfaces.
 - D) No, statically configured tunnels are not listed in the routing table.

- Q15) Which two terms are used when the IGP metric is modified? (Choose two.) (Source: Introducing MPLS Traffic Engineering Components)
- A) absolute
 - B) negative
 - C) positive
 - D) relative
- Q16) The feature that enables headend routers to see the MPLS-TE tunnel as a directly connected interface is called _____. (Source: Introducing MPLS Traffic Engineering Components)
-
- Q17) The cost of the TE tunnel is equal to the shortest _____ to the tunnel endpoint. (Source: Introducing MPLS Traffic Engineering Components)
-
- Q18) The IGP metric is 50; the tunnel metric has been set to a relative +2. Each path contains six routers. Which path will be used for routing? (Source: Introducing MPLS Traffic Engineering Components)
- A) The IGP path will be used.
 - B) The LSP path will be used.
 - C) Both paths will be used, and load balancing will be implemented.
 - D) It is not possible to answer from the information that has been provided.
- Q19) The mechanism that enables the announcement of established tunnels via IGP to all nodes within an area is called _____. (Source: Introducing MPLS Traffic Engineering Components)
-
- Q20) Which option is *not* a component of the maximum allocation multiplier attribute? (Source: Running MPLS Traffic Engineering)
- A) maximum bandwidth
 - B) unreserved bandwidth
 - C) minimum available bandwidth
 - D) maximum reservable bandwidth
- Q21) The MPLS-TE tunnel attribute _____ allows the network administrator to apply path selection policies. (Source: Running MPLS Traffic Engineering)
-
- Q22) Which two options about the tunnel resource class affinity mask are true? (Choose two.) (Source: Running MPLS Traffic Engineering)
- A) If bit is 0, do care.
 - B) If bit is 1, do care.
 - C) If bit is 0, do not care.
 - D) If bit is 1, do not care.

- Q23) Which two protocols can be used to propagate MPLS-TE link attributes? (Choose two.)
(Source: Running MPLS Traffic Engineering)
- A) BGP
 - B) OSPF
 - C) EIGRP
 - D) IS-IS
- Q24) In the case of a tie after CBR has been run, which two values are used to break the tie?
(Choose two.) (Source: Running MPLS Traffic Engineering)
- A) larger hop count
 - B) smallest hop count
 - C) highest minimum bandwidth
 - D) highest maximum bandwidth
- Q25) To enable MPLS-TE tunnel signaling on a Cisco IOS Software device, you must use the _____ command. (Source: Implementing MPLS TE)
- _____
- Q26) Which command enables MPLS-TE in an OSPF implementation on a Cisco IOS Software device? (Source: Implementing MPLS TE)
- A) mpls-te enable
 - B) mpls traffic-eng
 - C) metric-style wide
 - D) mpls traffic-eng area
- Q27) The Cisco IOS XR Software command that is used to instruct the IGP to use the tunnel in its SPF or next-hop calculation is the _____ command. (Source: Implementing MPLS TE)
- _____
- Q28) Engineered tunnels can be used for IP routing only if the tunnel is explicitly specified for routing via _____ and _____. (Source: Implementing MPLS TE)
- _____
- _____
- Q29) Links can be excluded from the constraint-based SPF computation by using the _____ and _____ over which the tunnel should pass. (Source: Implementing MPLS TE)
- _____
- _____
- Q30) The constraint-based path computation selects the path that the dynamic traffic tunnel will take, based on the administrative weight (TE cost), which is, by default, equal to the _____. (Source: Implementing MPLS TE)
- _____

- Q31) With the autoroute feature enabled, the traffic tunnel can do which two things? (Choose two.) (Source: Implementing MPLS TE)
- A) appear in the routing table
 - B) have an associated IP metric
 - C) be used in link-state updates
 - D) allow the rest of the network to be aware of the traffic tunnel
- Q32) The requirement for two tunnels between the same endpoints is that they must use _____ paths. (Source: Protecting MPLS TE Traffic)
- _____
- Q33) If conditions are corrected and provide for the re-establishment of traffic, the traffic is _____ to the primary tunnel. (Source: Protecting MPLS TE Traffic)
- _____
- Q34) If you do not configure forwarding adjacency on two LSP tunnels bidirectionally, from A to B and from B to A, the _____ is advertised but not used in the IGP network. (Source: Protecting MPLS TE Traffic)
- _____
- Q35) The _____ feature provides link protection to LSPs by establishing a backup LSP tunnel for the troubled link. (Source: Protecting MPLS TE Traffic)
- _____
- Q36) The Cisco IOS feature that measures utilization averages and dynamically adjusts tunnel bandwidth reservations is _____. (Source: Protecting MPLS TE Traffic)
- _____
- Q37) The _____ is used for tunnels that carry traffic that requires strict bandwidth guarantees or delay guarantees. (Source: Protecting MPLS TE Traffic)
- _____

Module Self-Check Answer Key

Q1)	C, D
Q2)	B, D
Q3)	C, D
Q4)	C
Q5)	A, C
Q6)	C
Q7)	A, B, C
Q8)	A
Q9)	Path
Q10)	A
Q11)	A
Q12)	D
Q13)	A
Q14)	C
Q15)	A, D
Q16)	autoroute
Q17)	IGP metric
Q18)	A
Q19)	forwarding adjacency
Q20)	C
Q21)	resource class affinity (or tunnel resource class affinity)
Q22)	B, C
Q23)	B, D
Q24)	B, C
Q25)	mpls traffic-eng tunnels
Q26)	D
Q27)	autoroute announce
Q28)	via policy routing that sets a next-hop interface to the tunnel, and via static routes that point to the tunnel
Q29)	resource class affinity bits of the traffic tunnel and resource class bits of the links
Q30)	IGP link metric (cost)
Q31)	A, B
Q32)	diverse
Q33)	returned
Q34)	forwarding adjacency
Q35)	Fast Reroute
Q36)	automatic bandwidth
Q37)	subpool

QoS in the Service Provider Network

Overview

Modern applications with differing network requirements create a need for administrative policies mandating how individual applications are to be treated by the network. Quality of service (QoS) is a crucial element of any administrative policy that mandates the management of application traffic on a network. This module introduces the concept of QoS in service provider networks, explains key issues of networked applications, and describes different methods for implementing QoS.

To facilitate true end-to-end QoS on an IP network, the IETF has defined two models: integrated services (IntServ) and differentiated services (DiffServ). IntServ follows the signaled QoS model, in which the end hosts signal their QoS needs to the network. DiffServ works on the provisioned QoS model, in which network elements are set up to service multiple classes of traffic with varying QoS requirements.

This module describes the implementation of the DiffServ model in service provider networks. The lessons describe the problems that could lead to inadequate QoS and the methods for implementing QoS and QoS mechanisms. The concepts and features of Multiprotocol Label Switching (MPLS) QoS are discussed, and techniques are presented that network administrators can apply to help the service providers meet the SLA of their customers.

Module Objectives

Upon completing this module, you will be able to understand the concept of QoS and explain the need to implement QoS. This ability includes being able to meet these objectives:

- Identify problems that could lead to poor quality of service and provide solutions to those problems
- Explain the IntServ and DiffServ QoS models and how they are used in converged networks
- List different QoS mechanisms and describe how to apply them in the network
- Explain how to use different QoS mechanisms in IP NGN networks and describe DiffServ support in MPLS networks

Understanding QoS

Overview

When packet transport networks were first used, IP was designed to provide best effort service for any type of traffic. To support the ever- increasing demands for speed and quality for these applications, two models for implementing quality of service (QoS) were designed: the differentiated services model (DiffServ) and the integrated services model (IntServ).

This lesson describes typical QoS problems: delay, jitter, bandwidth, availability, and packet loss. In addition to best effort service, it also describes differences between the DiffServ and IntServ models and provides information about the functionality of each model and how each fits into service provider networks.

Objectives

Upon completing this lesson, you will be able to identify problems that could lead to inadequate QoS. You will be able to meet these objectives:

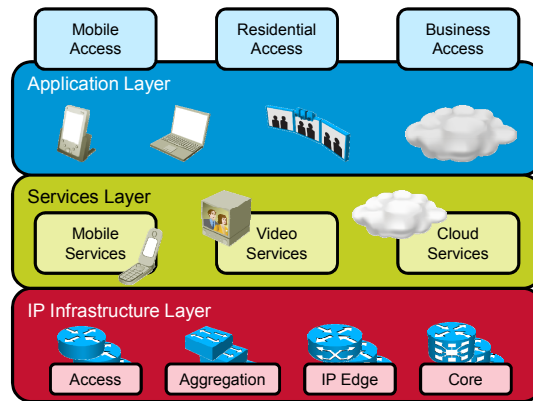
- Describe the Cisco IP NGN Architecture
- Describe the QoS Issues in Converged Networks
- Describe the need to classify traffic into different Traffic Classes
- Describe applying QoS Policies on the Traffic Classes
- Describe the service level agreement concept
- Describe the service level agreement measuring points in the network
- Describe best-effort and IntServ and DiffServ models
- Describes the IntServ model
- Describe the DiffServ model
- Describe the DSCP field in the IP header
- Describe the different QoS mechanisms that can be applied to an interface based on the DiffServ Model
- Describe using MQC to enable the QoS mechanisms

Cisco IP NGN Architecture

This topic describes the Cisco IP NGN Architecture.

Cisco IP NGN Architecture

- The Cisco IP NGN is a next-generation service provider infrastructure for video, mobile, and cloud or managed services.
- The Cisco IP NGN provides an all-IP network for services and applications, regardless of access type.



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3.3

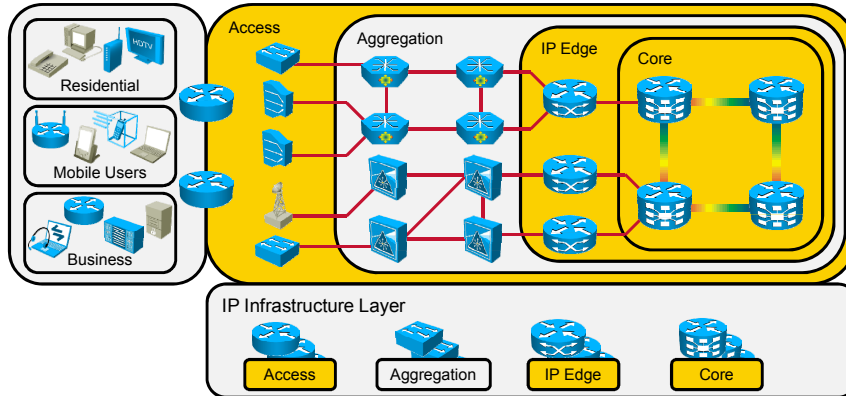
Historically, service providers were specialized for different types of services, such as telephony, data transport, and Internet service. The popularity of the Internet through telecommunications convergence has evolved, and now the Internet is used for all types of services. The development of interactive mobile applications, increasing video and broadcasting traffic, and the adoption of IPv6 have pushed service providers to adopt new architecture to support new services on the reliable IP infrastructure with a good level of performance and quality.

Cisco IP Next-Generation Network (NGN) is the next-generation service provider architecture for providing voice, video, mobile, and cloud or managed services to users. Cisco NGN networks are designed to provide all-IP transport for all services and applications, regardless of access type. IP infrastructure, service, and application layers are separated in NGN networks, thus enabling addition of new services and applications without any changes in the transport network.

To deliver any type of service with the required quality and performance, NGN uses QoS-enabled transport technologies to provide services for various applications, independent of the underlying transport technology.

Cisco IP NGN Infrastructure Layer

- The IP infrastructure layer provides connectivity between customer and service provider.
- End-to-end QoS must be implemented to satisfy the requirements for the most demanding services.



The IP infrastructure layer is responsible for providing a reliable infrastructure for running upper layer services. It comprises these parts:

- Core network
- IP edge network
- Aggregation network
- Access network

The IP infrastructure layer provides the reliable, high speed, and scalable foundation of the network. End users are connected to service providers through customer premises equipment (CPE), devices using any possible technology. Access and aggregation network devices are responsible for enabling connectivity between customer equipment and service provider edge equipment. The core network is used for fast switching packets between edge devices.

To provide the highest level of service quality, QoS must be implemented across all areas of the network. For an existing network, it is said that QoS is only as strong as the weakest link. Therefore, different QoS tools must be implemented in all parts of the IP infrastructure layer. Optimally, every device (host, server, switch, or router) that manages the packet along its network path should employ QoS to ensure that the packet is not unduly delayed or lost between endpoints.

QoS Issues in Converged Networks

This topic describes the QoS Issues in Converged Networks.

Quality Issues in Converged Networks

- **Lack of bandwidth:** Multiple flows compete for a limited amount of bandwidth. The maximum available bandwidth equals the bandwidth of the weakest link.
- **End-to-end delay:** Packets must traverse many network devices and links that add to the overall delay. The delay is the sum of propagation, processing, and queuing delays in the network.
- **Variation of delay (jitter):** Subsequent packets may have different delays and that difference can lead to quality issues.
- **Packet loss:** Packets may have to be dropped when a link is congested. Packet loss is commonly caused by tail drop when the output queue on the router is full.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3-6

There are four main problems facing converged networks:

Bandwidth capacity: Large graphics files, multimedia uses, and increasing use of voice and video cause bandwidth capacity problems over data networks. The best way to increase bandwidth is to increase the link capacity to accommodate all applications and users, with some extra bandwidth to spare. Although this solution sounds simple, increasing bandwidth is expensive and takes time to implement. There are often technological limitations in upgrading to a higher bandwidth.

End-to-end delay (both fixed and variable): Delay is the time that it takes for a packet to reach the receiving endpoint after being transmitted from the sending endpoint. This period of time is called the “end-to-end delay,” and consists of two components:

- **Fixed network delay:** Two types of fixed delays are serialization and propagation delays. Serialization is the process of placing bits on the circuit. The higher the circuit speed, the less time it takes to place the bits on the circuit. Therefore, the higher the speed of the link, the less serialization delay is incurred. Propagation delay is the time that it takes for frames to transit the physical media.
- **Variable network delay:** A processing delay is a type of variable delay; it is the time that is required by a networking device to look up the route, change the header, and complete other switching tasks. Sometimes, the packet must also be manipulated, for example, when the encapsulation type or the Time to Live (TTL) must be changed. Each of these steps can contribute to the processing delay.

Each hop in the network adds to the overall delay:

- Propagation delay is caused by the speed of light traveling in the media; for example, the speed of light traveling in fiber optics or copper media.
- Serialization delay is the time that it takes to clock all the bits in a packet onto the wire. This is a fixed value that is a function of the link bandwidth.
- There are processing and queuing delays within a router, which can be caused by a wide variety of conditions.

Variation of delay (also called jitter): Jitter is the delta, or difference, in the total end-to-end delay values of two voice packets in the voice flow.

Packet loss: Loss of packets is usually caused by congestion in the WAN, resulting in speech dropouts or a stutter effect if the playout side tries to accommodate by repeating previous packets. Most applications that use TCP do experience slowdowns because TCP adjusts to the network resources. Dropped TCP segments cause TCP sessions to reduce their window sizes.

There are some other applications that do not use TCP and cannot manage drops.

You can follow these approaches to prevent drops in sensitive applications:

- Increase link capacity to ease or prevent congestion.
- Guarantee enough bandwidth and increase buffer space to accommodate the bursts of fragile applications. There are several QoS mechanisms available in Cisco IOS and Cisco IOS XR Software that can guarantee bandwidth and provide prioritized forwarding to drop-sensitive applications. These mechanisms are listed:
 - Priority queuing (PQ)
 - IP Real-Time Transport Protocol (RTP) priority
 - Class-based weighted fair queuing (CBWFQ)
 - Low latency queuing (LLQ)
- Prevent congestion by randomly dropping packets before congestion occurs. You can use weighted random early detection (WRED) to selectively drop lower-priority traffic first, before congestion occurs.

These are some other mechanisms that you can use to prevent congestion:

Traffic shaping: Traffic shaping delays packets instead of dropping them, and includes generic traffic shaping, Frame Relay traffic shaping (FRTS), and class-based shaping.

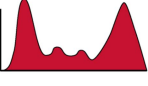
Traffic policing: Traffic policing, including committed access rate (CAR) and class-based policing, can limit the rate of less-important packets to provide better service to drop-sensitive packets.

QoS and Traffic Classes

This topic describes the need to classify traffic into different Traffic Classes.

QoS and Traffic Classes

- Not all applications require the same treatment.
- The network needs the ability to provide better or “special” service to a set of users and applications (to the detriment of other users and applications).
- QoS policy must comply with traffic-specific requirements.

Voice	Video	Data
		
		
Smooth Benign Drop-Sensitive Delay-Sensitive UDP Priority	Bursty Greedy Drop-Sensitive Delay-Sensitive UDP Priority	Smooth or Bursty Benign or Greedy Drop-Insensitive Delay-Insensitive TCP Retransmits

© 2012 Cisco and/or its affiliates. All rights reserved. SPCORE v1.01--3-8

Before networks converged, network engineering was mainly focused on connectivity. However, the rates at which data came onto the network resulted in bursty data flows. Data, arriving in packets, tried to grab as much bandwidth as it could at any given time. Access was on a first-come, first-served basis. The data rate available to any one user varied, depending on the number of users accessing the network at any given time.

The protocols that have been developed have adapted to the bursty nature of data networks, and brief outages are survivable. For example, when you retrieve email, a delay of a few seconds is generally not noticeable. A delay of minutes is annoying, but not serious.

In converged network, voice, video, and data traffic use the same network facilities. Merging these different traffic streams with dramatically differing requirements can lead to a number of problems.

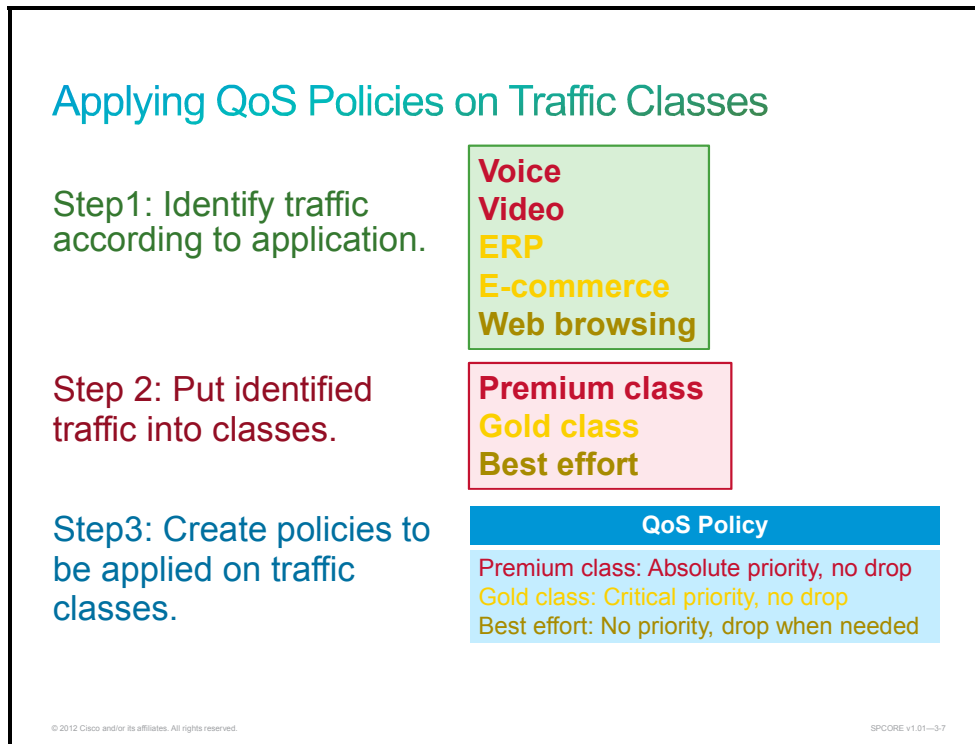
Voice traffic has extremely stringent QoS requirements. Voice traffic usually generates a smooth demand on bandwidth and has minimal impact on other traffic as long as the voice traffic is managed. While voice packets are typically small (60 to 120 bytes), they cannot tolerate delay or drops. The result of delays and drops are poor, and often unacceptable, voice quality. Because drops cannot be tolerated, User Datagram Protocol (UDP) is used to package voice packets, because TCP retransmit capabilities have no value. Voice packets can tolerate no more than a 150-ms delay (one-way requirement) and less than 1 percent packet loss. A typical voice call will require 17 to 106 kb/s of guaranteed priority bandwidth plus an additional 150 b/s per call for voice-control traffic.

Videoconferencing applications also have stringent QoS requirements very similar to voice requirements. But videoconferencing traffic is often bursty and greedy in nature and, as a result, can impact other traffic. Therefore, it is important to understand the videoconferencing requirements for a network and to provision carefully for it. The minimum bandwidth for a videoconferencing stream would require the actual bandwidth of the stream (dependent upon the type of videoconferencing codec being used) plus some overhead. For example, a 384-kb/s video stream would actually require a total of 460 kb/s of priority bandwidth.

Data traffic QoS requirements vary greatly. Different applications may make very different demands on the network (for example, a human resources application versus an automated teller machine application). Even different versions of the same application may have varying network traffic characteristics. In enterprise networks, important (business-critical) applications are usually easy to identify. Most applications can be identified based on TCP or UDP port numbers. Some applications use dynamic port numbers that, to some extent, make classifications more difficult.

Applying QoS Policies on Traffic Classes

This topic describes applying QoS Policies on the Traffic Classes.

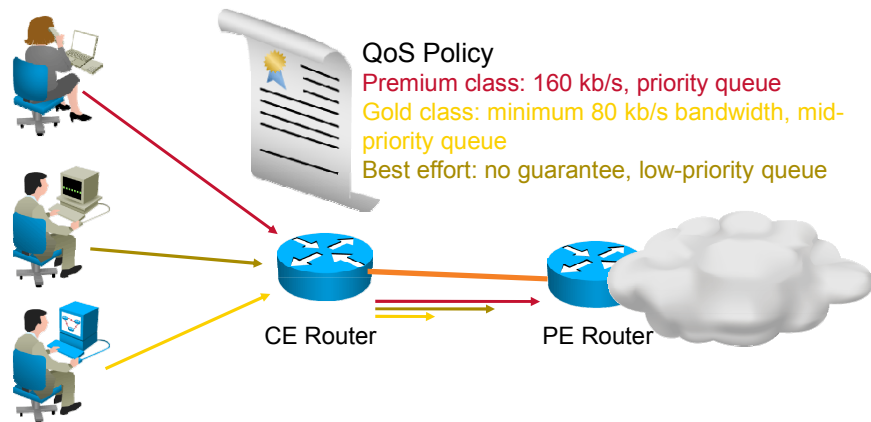


There are three basic steps that are involved in implementing QoS on a network:

- Step 1** Identify the traffic on the network and its requirements. Study the network to determine the type of traffic that is running on the network and then determine the QoS requirements for the different types of traffic.
- Step 2** Group the traffic into classes with similar QoS requirements. For example, three classes of traffic can be defined: voice and video (premium class), high priority (gold class) and best effort.
- Step 3** Define QoS policies that will meet the QoS requirements for each traffic class.

Applying QoS Policies on Traffic Classes Example:

Voice, database, and web browsing are classified into premium, gold, and best effort classes, respectively:



Using the three previously defined traffic classes, you can determine QoS policies:

- **Voice and video (premium class):** Minimum bandwidth: 160 kb/s. Use QoS marking to mark voice packets as a high priority; use priority queue to minimize delay.
- **Business applications (gold class):** Minimum bandwidth: 80 kb/s. Use QoS marking to mark critical data packets as medium-high priority; use medium-priority queue.
- **Web traffic (best effort):** Use QoS marking to mark these data packets as a low priority. Use queuing mechanism to prioritize best-effort traffic flows that are below the premium and gold classes.

You can apply similar QoS actions on service provider routers to meet the expectations of users. These expectations can be formalized through service level agreements (SLAs).

Service Level Agreement

This topic describes the service level agreement concept.

Service Level Agreement

- Service quality offered by service provider can be measured with statistics such as throughput, usage, percentage of loss, and uptime.
- Service providers formalize these expectations within service level agreements (SLAs), which clearly state the acceptable bounds of network performance.

Example of SLAs offered by an ISP:

Class	RTT (ms)	Delay (ms)	Availability (%)	Jitter (ms)
Premium	40	<150	99.99	<8
Gold	45	<200	99.9	<15
Silver	50	<500	99	<30

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3-9

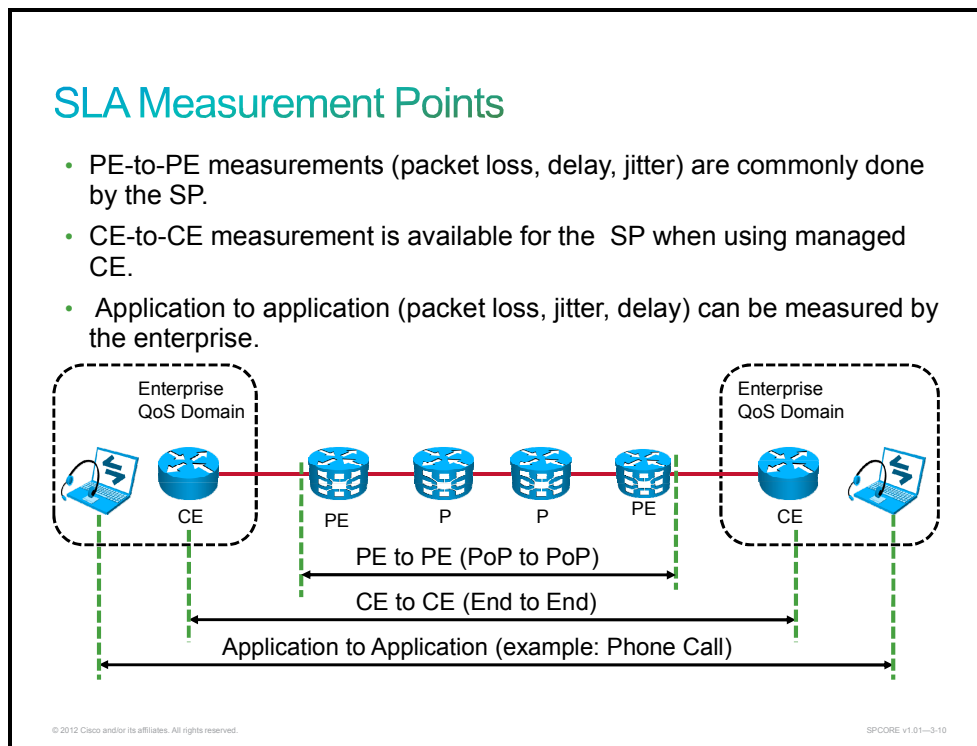
IT management perceives the network quality through management statistics such as throughput, usage, percentage of loss, and user complaints. Expectations and “quality problems” from an IT perspective tend to be more absolute and measurable. (For example, the one-way latency of more than 150 ms is unacceptable for voice calls, or a transaction refresh time for a bank teller application must be fewer than 5 seconds.)

Service providers formalize these expectations within SLAs, which clearly state the acceptable bounds of network performance. Corporate enterprise networks typically do not have such formal SLAs, but they nevertheless manage their networks on some form of measurement and monitoring. Some enterprise networks might indeed use SLAs of various levels of formality between the IT department and customer departments that they serve.

User complaints might result even though the network met the SLA. For example, the packet loss statistics over a certain period of time might be within bounds, but a user whose file transfer session timed out or whose print job was lost during a period of momentary congestion would most likely perceive it as a “network problem.”

Service Level Agreement Measuring Points

This topic describes the service level agreement measuring points in the network.



IP service levels can be assured, network operation can be verified proactively, and network performance can be accurately measured. Active monitoring continuously measures the network performance between multiple paths in the network, providing ongoing performance baseline information.

Cisco IOS IP SLA is a network performance measurement and diagnostic tool that uses active monitoring, which includes the generation of traffic in a continuous, reliable, and predictable manner.

There are several points in the network where SLA measurements can take place:

- **PE to PE:** The most common points for measurements of SLA parameters
- **CE to CE:** Measurements of SLA parameters from the customer site, available from the service provider when the CE router is managed by the service provider
- **Application to application:** Measurements of SLA parameters for end-to-end applications, available only from the enterprise. Some types of applications have the ability to measure SLA parameters such as delay, jitter, and packet loss. For example, an IP telephony phone call between Cisco IP phones can provide these statistics on the display of the IP phone.

Models for Implementing QoS

This topic describes best-effort and IntServ and DiffServ models.

Models for Implementing QoS

There are three models for providing the differentiated levels of network service required in converged networks: IntServ, DiffServ, and best effort.

- **Best-effort** service is the model inherent in IP networks from the beginning of packet transport networks. By default, no QoS is applied on any packets.
- **Integrated services (IntServ)** is based on signaling events from endpoints requesting special treatment for packets that are delivered to a network.
- **Differentiated services (DiffServ)** divides traffic into classes and applies a different level of service for each class.

© 2012 Cisco and/or its affiliates. All rights reserved.

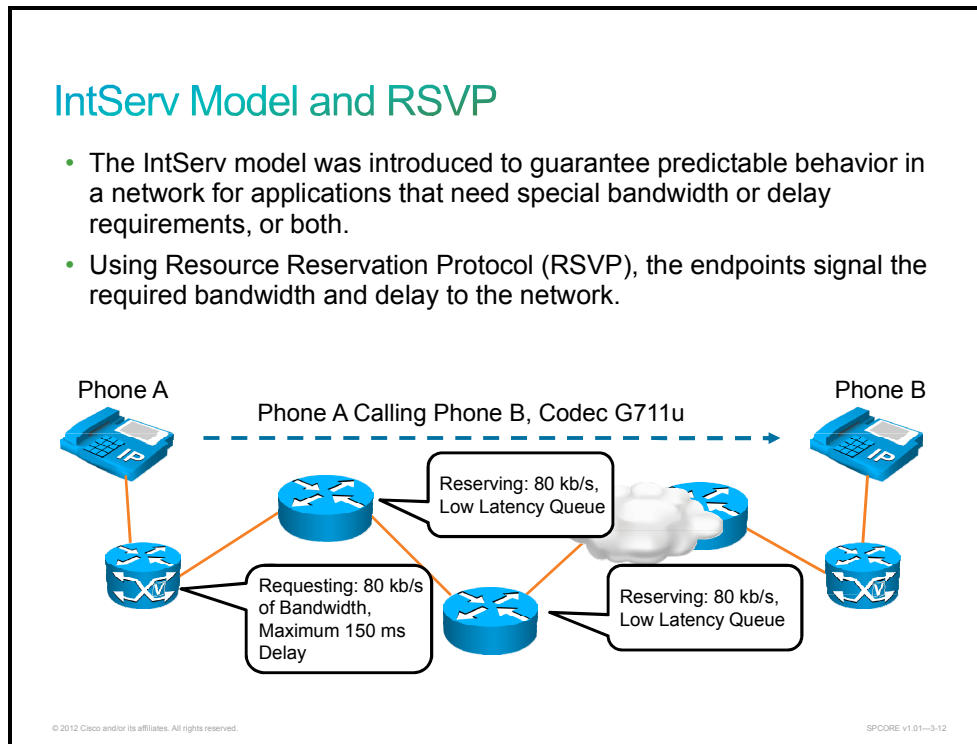
SPCORE v1.01--3-11

The following three models exist for implementing QoS in a network:

- **Best effort:** With the best-effort model, QoS is not applied to packets. If it is not important when or how packets arrive, the best-effort model is appropriate. If QoS policies are not implemented, traffic is forwarded using the best-effort model. All network packets are treated the same; an emergency voice message is treated like a digital photograph that is attached to an email. Without QoS, the network cannot tell the difference between packets and, as a result, cannot treat packets preferentially.
- **IntServ:** IntServ can provide very high QoS to IP packets. Essentially, applications signal to the network that they will require special QoS for a period of time, so that bandwidth is reserved. With IntServ, packet delivery is guaranteed. However, IntServ can severely limit network scalability. IntServ is similar to a concept known as “hard QoS.” With hard QoS, traffic characteristics such as bandwidth, delay, and packet-loss rates are guaranteed end to end. Predictable and guaranteed service is ensured for mission-critical applications. There will be no impact on traffic when guarantees are made, regardless of additional network traffic.
- **DiffServ:** DiffServ provides the greatest scalability and flexibility in implementing QoS in a network. Network devices recognize traffic classes and provide different levels of QoS to different traffic classes. DiffServ is similar to a concept known as “soft QoS.” With soft QoS, QoS mechanisms are used without prior signaling. In addition, QoS characteristics (bandwidth and delay, for example), are managed on a hop-by-hop basis by policies that are established independently at each intermediate device in the network. The soft QoS approach is not considered an end-to-end QoS strategy because end-to-end guarantees cannot be enforced. However, soft QoS is a more scalable approach to implementing QoS than hard QoS, because many (hundreds or potentially thousands) of applications can be mapped into a small set of classes upon which similar sets of QoS behaviors are implemented.

IntServ Model and RSVP

This topic describes the IntServ model.



In the IntServ model, the application requests a specific kind of service from the network before sending data. The application informs the network of its traffic profile and requests a particular kind of service that can encompass its bandwidth and delay requirements. The application is expected to send data only after it gets a confirmation from the network. The application is also expected to send data that lies within its described traffic profile.

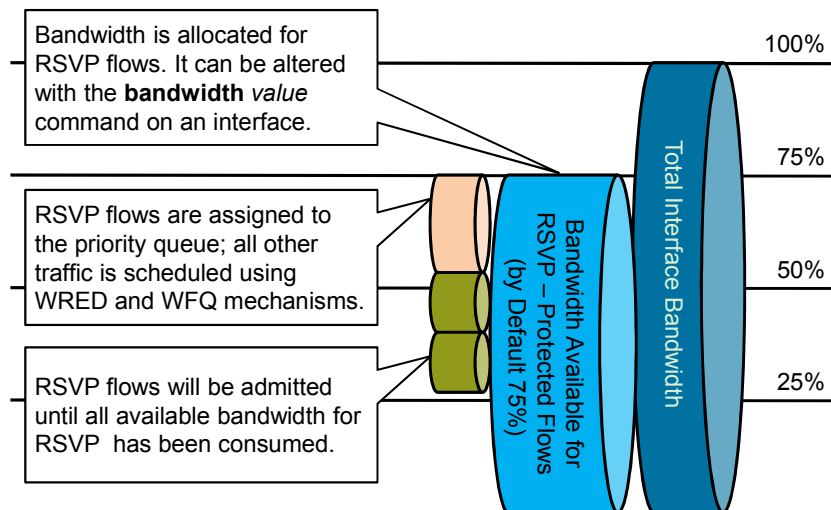
The network performs admission control that is based on information from the application and available network resources. The network commits to meeting the QoS requirements of the application as long as the traffic remains within the profile specifications. The network fulfills its commitment by maintaining the per-flow state, and then performing packet classification, policing, and intelligent queuing based on that state.

In this model, Resource Reservation Protocol (RSVP) can be used by applications to signal their QoS requirements to the router. RSVP is an IP service that allows end systems or hosts on either side of a router network to establish a reserved-bandwidth path between them to predetermine and ensure QoS for their data transmission. RSVP is currently the only standard signaling protocol that is designed to guarantee network bandwidth from end to end for IP networks.

RSVP is an IETF standard (RFC 2205) protocol for allowing an application to dynamically reserve network bandwidth. RSVP enables applications to request a specific QoS for a data flow (shown in the figure). The Cisco implementation also allows RSVP to be initiated within the network, using a configured proxy RSVP. Network managers can take advantage of RSVP benefits in the network, even for non-RSVP-enabled applications and hosts.

Hosts and routers use RSVP to deliver QoS requests to the routers along the paths of the data stream. Hosts and routers also use RSVP to maintain the router and host state to provide the requested service, usually bandwidth, and latency. RSVP uses LLQ or WRED QoS mechanisms, setting up the packet classification and scheduling that is required for the reserved flows. LLQ and WRED will be covered later in another lesson.

RSVP Interface Queuing



The figure outlines how RSVP data flows are allocated when RSVP is configured on an interface. The maximum bandwidth available on any interface is 75 percent of the line speed; the rest is used for control plane traffic. When RSVP is configured on an interface, the option is to use the entire usable bandwidth or a certain configured amount of bandwidth. The default is for RSVP data flows to use up to 75 percent of the available bandwidth.

Differentiated Services Model

This topic describes the DiffServ model.

Differentiated Services Model

- Network traffic is identified by the appropriate class.
- There is different treatment regarding traffic for each traffic class.
- This model can be very scalable.
- Traffic classification is performed closest to the source of traffic; classified packets are marked using the **DSCP** field in IP packets.
- Having traffic divided into classes becomes simple to manage in the core; every hop in the core can perform actions based on traffic classes.
- QoS actions that are based hops are called per-hop behavior (**PHB**).

© 2012 Cisco and/or its affiliates. All rights reserved.

SPICORE v1.01-3-14

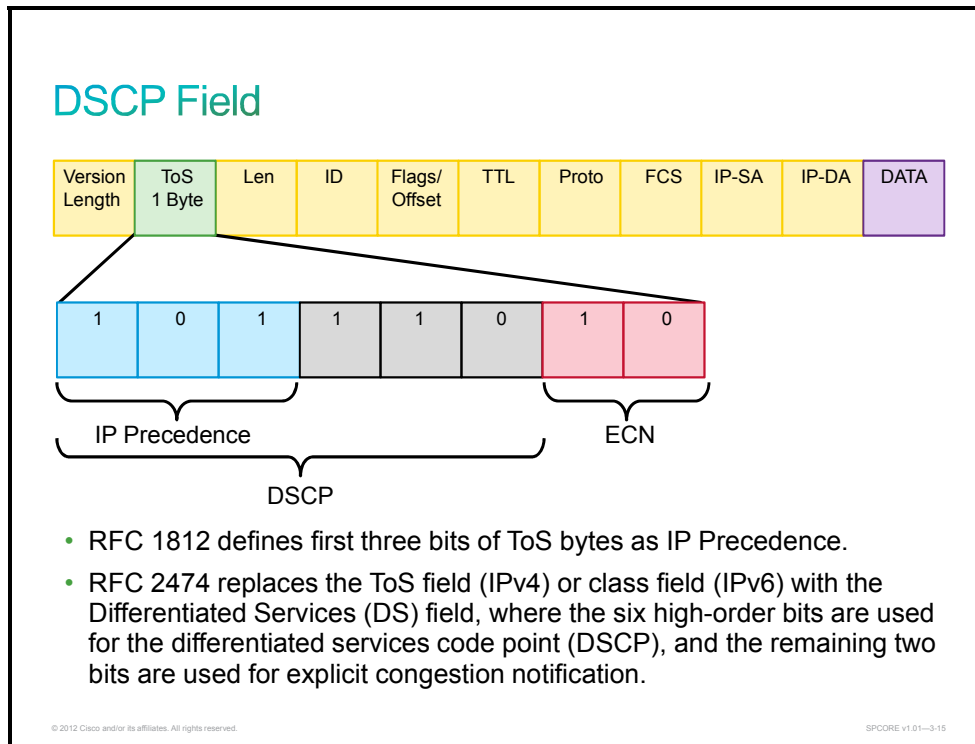
DiffServ was designed to overcome the limitations of both the best-effort and IntServ models. DiffServ can provide an “almost guaranteed” QoS, while still being cost-effective and scalable.

DiffServ is similar to a concept known as “soft QoS,” in which QoS mechanisms are used without prior signaling. In addition, QoS characteristics (bandwidth and delay, for example), are managed on a hop-by-hop basis by policies that are established independently at each intermediate device in the network. The soft QoS approach is not considered an end-to-end QoS strategy because end-to-end guarantees cannot be enforced but it is a more scalable approach. Many applications can be mapped into a small set of classes upon which similar sets of QoS behaviors are implemented. Although QoS mechanisms in this approach are enforced and applied on a hop-by-hop basis, uniformly applying global meaning to each traffic class provides both flexibility and scalability.

With DiffServ, network traffic is divided into classes that are based on business requirements. Each of the classes can then be assigned a different level of service. As the packets traverse a network, each of the network devices identifies the packet class and services the packet according to that class. You can choose many levels of service with DiffServ. For example, voice traffic from IP phones is usually given preferential treatment over all other application traffic. Email is generally given best-effort service. Nonbusiness traffic can either be given very poor service or blocked entirely.

DSCP Field

This topic describes the DSCP field in the IP header.



DiffServ uses the Differentiated Services (DS) field in the IP header to mark packets according to their classification into behavior aggregates (BAs). A BA is the collection of packets traversing a DiffServ node with the same differentiated services code point (DSCP) marking. The DS field occupies the same 8 bits of the IP header that were previously used for the Type of Service (ToS) byte.

Three IETF standards describe the purpose of the 8 bits of the DS field:

- RFC 791 includes specification of the ToS field, where the high-order 3 bits are used for IP precedence. The other bits are used for delay, throughput, reliability, and cost.
- RFC 1812 modifies the meaning of the ToS field by removing meaning from the five low-order bits (those bits should all be 0). This practice gained widespread use and became known as the original IP precedence.
- RFC 2474 replaces the ToS field with the DS field, where the six high-order bits are used for the DSCP. The remaining 2 bits are used for explicit congestion notification (ECN). RFC 3260 ("New Terminology and Clarifications for DiffServ") updates RFC 2474 and provides terminology clarifications.

IP version 6 (IPv6) also provides support for QoS marking via a field in the IPv6 header. Similar to the ToS (or DS) field in the IPv4 header, the Traffic Class field (8 bits) is available for use by originating nodes and forwarding routers to identify and distinguish between different classes or priorities of IPv6 packets. The Traffic Class field can be used to set specific precedence or DSCP values, which are used the same way that they are used in IPv4.

IPv6 also has a 20-bit field that is known as the Flow Label field. The flow label enables per-flow processing for differentiation at the IP layer. It can be used for special sender requests and is set by the source node. The flow label must not be modified by an intermediate node. The main benefit of the flow label is that transit routers do not have to open the inner packet to identify the flow, which aids with identification of the flow when using encryption and in other scenarios. The Flow Label field is described in RFC 3697.

Per-Hop Behavior

Per-Hop Behavior	Value	Service
Default	000 XXX	Best effort
Expedited Forwarding (EF)	101 110	Low delay
Assured Forwarding	XXX XX0	Guaranteed bandwidth

Example of ToS byte (EF):

1	0	1	1	1	0	1	0

- IP Precedence field = 101 (bits 5-7)
- DSCP field = 101110 (bits 2-7)
- ECN field = 10 (bits 0-1)

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-3-18

These per-hop behaviors (PHBs) are defined by IETF standards:

- **Default PHB:** Used for best-effort service (bits 5 to 7 of DSCP = 000)
- **Expedited Forwarding (EF) PHB:** Used for low-delay service (bits 2 to 7 of DSCP = 101110)
- **Assured Forwarding (AF) PHB:** Used for guaranteed bandwidth service (bits 5 to 7 of DSCP = 001, 010, 011, or 100)
- **Class-Selector PHB:** Used for backward compatibility with non-DiffServ-compliant devices (RFC 1812-compliant devices [bits 2 to 4 of DSCP = 000])

For example, if ToS byte equals 10111010, then IP precedence field is 101, DSCP field is 101110, and ECN field is 10. DSCP 101110 is recommended for the EF PHB.

The EF PHB is identified based on the following:

- The EF PHB ensures a minimum departure rate. The EF PHB provides the lowest possible delay for delay-sensitive applications.
- The EF PHB guarantees bandwidth. The EF PHB prevents starvation of the application if there are multiple applications using EF PHB.
- The EF PHB polices bandwidth when congestion occurs. The EF PHB prevents starvation of other applications or classes that are not using this PHB.
- Packets requiring EF should be marked with DSCP binary value 101110 (46 or 0x2E).

Per-Hop Behavior (Cont.)

Class	Value			Drop Probability (dd)	Value	AF Value
AF1	001	dd	0	Low	01	AF11
AF2	010	dd	0	Medium	10	AF12
AF3	011	dd	0	High	11	Low
AF4	100	dd	0			

- **Assured Forwarding (AF) PHB** is used for guaranteed bandwidth service (bits 5 to 7 of DSCP = 001, 010, 011, or 100).
- Packets requiring AF PHB should be marked with DSCP value **aaadd0**, where **aaa** is the number of the class and **dd** is the drop probability.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3-17

The AF PHB is identified based on the following:

- The AF PHB guarantees a certain amount of bandwidth to an AF class.
- The AF PHB allows access to extra bandwidth, if available.
- Packets requiring AF PHB should be marked with DSCP value aaadd0, where aaa is the number of the class and dd is the drop probability.

There are four standard, defined AF classes. Each class should be treated independently and should have allocated bandwidth that is based on the QoS policy. Each AF class is assigned an IP precedence and has three drop probabilities: low, medium, and high.

AF_{xy}: Assured Forwarding (RFC 2597), where *x* corresponds to the IP precedence value (only 1–4 are used for AF classes), and *y* corresponds to the drop preference value (1, 2, or 3).

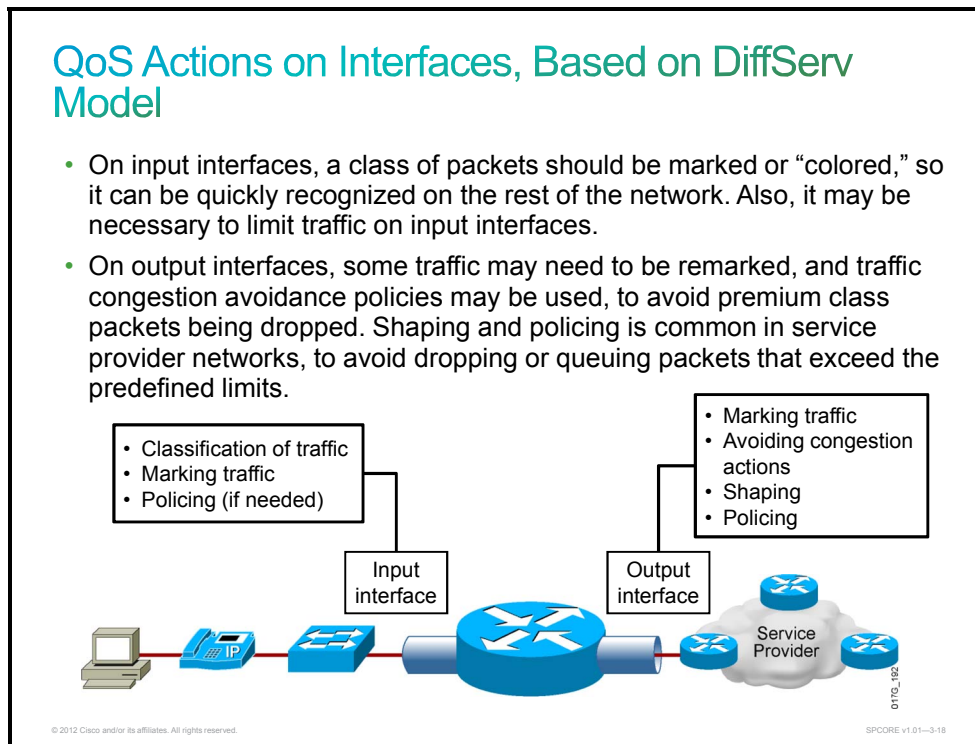
This table maps the binary and decimal representations of DSCP, IP precedence value, and PHB for all DSCP values.

DSCP (Binary)	DSCP (Decimal)	IP Precedence	Per-Hop Behavior
000000	0	0	Default (Class Selector 0)
001000	8	1	Class Selector 1 (CS1)
001010	10	1	AF11
001100	12	1	AF12
001110	14	1	AF13
010000	16	2	Class Selector 2 (CS2)
010010	18	2	AF21
010100	20	2	AF22
010110	22	2	AF23

DSCP (Binary)	DSCP (Decimal)	IP Precedence	Per-Hop Behavior
011000	24	3	Class Selector 3 (CS3)
011010	26	3	AF31
011100	28	3	AF32
011110	30	3	AF33
100000	32	4	Class Selector 4 (CS4)
100010	34	4	AF41
100100	36	4	AF42
100110	38	4	AF43
101000	40	5	Class Selector 5 (CS5)
101110	46	5	Expedited Forwarding (EF)
110000	48	6	Class Selector 6 (CS6)
111000	56	7	Class Selector 7 (CS7)

QoS Actions on Interfaces

This topic describes the different QoS mechanisms that can be applied to an interface based on the DiffServ Model.



In a QoS-enabled network, classification is performed on every input interface. Marking should be performed as close to the network edge as possible, in the originating network device, if possible. Devices farther from the edge of the network, such as routers and switches, can be configured to trust or distrust the markings that are made by devices on the edge of the network. An IP phone, for example, will not trust the markings of an attached PC, while a switch will generally be configured to trust the markings of an attached IP phone.

It only makes sense to use congestion management, congestion avoidance, and traffic-shaping mechanisms on output interfaces, because these mechanisms help maintain smooth operation of links by controlling how much and which type of traffic is allowed on a link. On some router and switch platforms, congestion management mechanisms can be applied on the input interface.

Congestion avoidance is typically employed on an output interface wherever there is a chance that a high-speed link or aggregation of links feeds into a slower link (such as a LAN feeding into a WAN).

Policing and shaping are typically employed on output interfaces to control the flow of traffic from a high-speed link to lower-speed links to prevent premium class packets from being dropped. Policing is also employed on input interfaces to control the flow into a network device from a high-speed link by dropping excess low-priority packets. With shaping and policing QoS actions, some packets that exceed predefined limits are discarded or delayed.

MQC Introduction

This topic describes using MQC to enable the QoS mechanisms.

Applying QoS Policies to Traffic Classes: Cisco IOS XR Software

Divide traffic into classes:

```
class-map Premium
  match protocol rtp
class-map Gold
  match access group ipv4 100
class-map class-default
  match protocol http
```

Create policy and apply to the interface:

```
policy-map policy1
  class Premium
    priority 160
  class Gold
    bandwidth 80
  class class-default
  !
interface GigabitEthernet 0/1/0/9
  service-policy output policy1
```

In Cisco IOS XR Software, QoS features are enabled through the Modular QoS Command-Line Interface (MQC) feature. The MQC is a CLI structure that allows you to create policies and attach these policies to interfaces. A traffic policy contains a traffic class and one or more QoS features. A traffic class is used to classify traffic, whereas the QoS features in the traffic policy determine how to treat the classified traffic. One of the main goals of MQC is to provide a platform-independent interface for configuring QoS across Cisco platforms. MQC will be covered in detail later in the course.

Applying traffic policies in Cisco IOS XR Software is accomplished via the MQC mechanism. Consider this example of configuring MQC on a network with voice telephony:

1. Classify traffic into classes. In this example, traffic is divided into three classes: premium, gold, and best-effort (the class default). To create a traffic class containing match criteria, use the **class-map** command to specify the traffic class name, and then use appropriate match commands in class-map configuration mode, as needed.
2. Build a single policy map that defines three different traffic policies (different bandwidth and delay requirements for each traffic class): NoDelay, BestService, and Whenever, and assign the already defined classes of traffic to the policies. Premium traffic is assigned to NoDelay. Gold traffic is assigned to BestService. Best-effort traffic is assigned to Whenever. To create a traffic policy, use the **policy-map** global configuration command to specify the traffic policy name. The traffic class is associated with the traffic policy when the **class** command is used. The **class** command must be issued after you enter policy map configuration mode. After entering the **class** command, the router is automatically in policy-map class configuration mode, which is where the QoS policies for the traffic policy are defined.
3. Assign the policy map to selected router (or switch) interfaces. After the traffic class and traffic policy are created, you must use the **service-policy** interface configuration command to attach a traffic policy to an interface and to specify the direction in which the policy should be applied (either on packets coming into the interface or packets leaving the interface).

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- To provide the highest level of service quality, QoS must be implemented across all areas of the service provider network
- Each hop in the network adds to the overall delay
- Not all applications require the same treatment
- There are three basic steps that are involved in implementing QoS on a network
- Service providers formalize customer's expectations within SLA
- There are several points in the network where SLA measurements can take place
- There are three models for implementing QoS in a network
- In the IntServ model, the application requests a specific kind of service from the network before sending data
- DiffServ was designed to overcome the limitations of both the best-effort and IntServ models
- DiffServ uses the Differentiated Services (DS) field in the IP header to mark packets according to their classification
- Marking should be performed as close to the network edge as possible
- The MQC is a CLI structure that allows you to create policies and attach these policies to interfaces

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3-20

Implementing QoS in the SP Network

Overview

Converged IP networks must provide secure, predictable, measurable, and sometimes guaranteed services. Quality of service (QoS) provides network administrators and architects with a set of techniques that are used to manage network resources. Cisco IOS XR Software provides a rich set of tools that enables network administrators to adapt the Cisco IP Next-Generation Networks (NGN) infrastructure layer to provide predictable behavior for new services and applications.

The moment an IP packet enters the network, it is classified and is usually marked with its class identification. From that point on, the packet is treated by various QoS mechanisms according to the packet classification. Depending upon the mechanisms it encounters, the packet could be expedited, delayed, or even dropped.

This lesson describes methods for QoS implementation, QoS implementation techniques, and how to properly apply QoS techniques into the service provider network.

Objectives

Upon completing this lesson, you will be able to list and describe methods for implementing QoS and QoS mechanisms. You will be able to meet these objectives:

- List the different QoS mechanisms
- Describe traffic classification
- Describe traffic marking
- Describe Congestion Management (Queuing)
- Describe Congestion Avoidance (RED and WRED)
- Describe traffic policing
- Describe traffic shaping
- Compare traffic shaping vs. traffic policing
- Lists and describes methods for implementing QoS
- Explains the MQC method to implement QoS

- Describe QoS requirements on the different devices in the service provider environment
- Describe the Service Provider Trust Boundary
- Describe the QoS requirements on the PE routers
- Describe the QoS requirements on the P routers
- Describe Hierarchical QoS policies

QoS Mechanisms

This topic lists the different QoS mechanisms.

QoS Mechanisms

- **Classification:** Supported by a class-oriented QoS mechanism
- **Marking:** Used to mark packets based on classification, metering, or both
- **Congestion management:** Used to prioritize the transmission of packets, with a queuing mechanism on each interface
- **Congestion avoidance:** Used to drop packets early to avoid congestion later in the network
- **Policing:** Used to enforce a rate limit by dropping or marking down packets
- **Shaping:** Used to enforce a rate limit by delaying packets, using buffers

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3.3

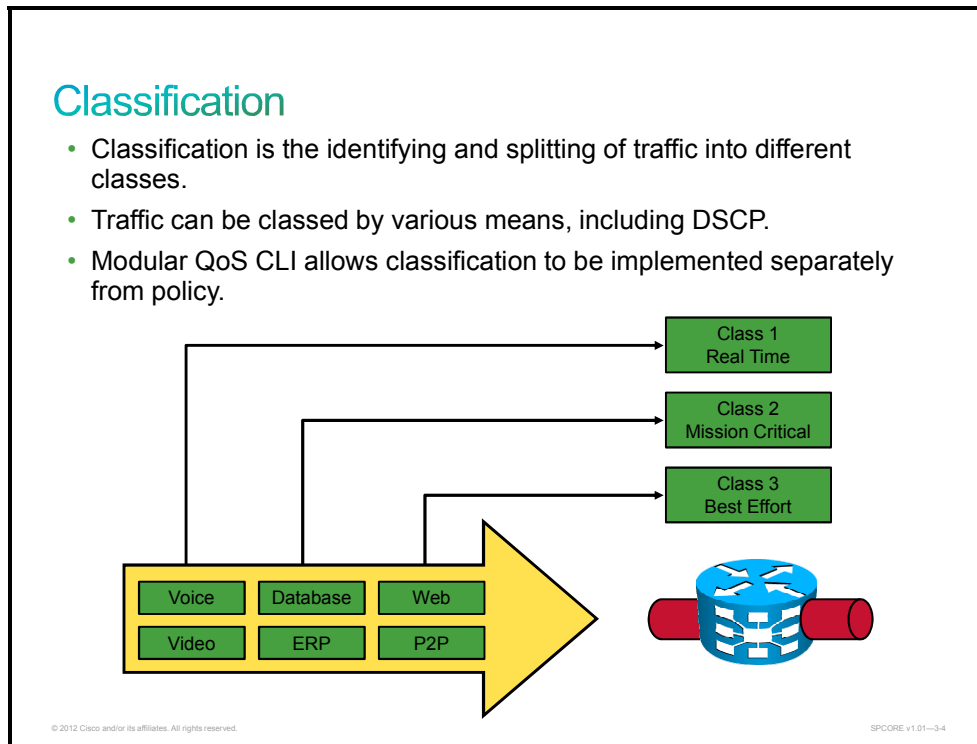
The main categories of tools that are used to implement QoS in a network are as follows:

- **Classification and marking:** The identifying and splitting of traffic into different classes and the marking of traffic according to behavior and business policies
- **Congestion management:** The prioritization, protection, and isolation of traffic that is based on markings
- **Congestion avoidance:** The discarding of specific packets, based on markings, to avoid network congestion
- **Policing and shaping:** Traffic conditioning mechanisms that police traffic by dropping misbehaving traffic to maintain network integrity. These mechanisms also shape traffic to control bursts by queuing excess traffic.

Packet classification identifies the traffic flow and marking identifies traffic flows that require congestion management or congestion avoidance on a data path. The Modular QoS CLI (MQC) is used to define the traffic flows that should be classified; each traffic flow is called a class of service or class. Later, a traffic policy is created and applied to a class. All traffic that is not identified by defined classes falls into the category of a default class.

Classification

This topic describes traffic classification.



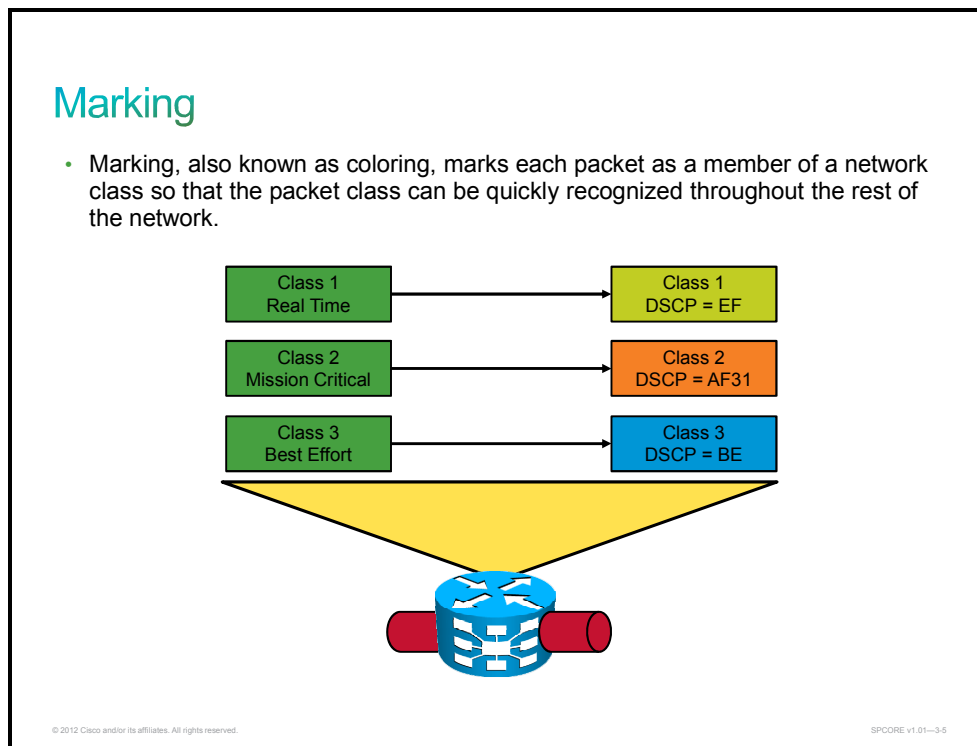
Classification is the identifying and splitting of traffic into different classes. In a QoS-enabled network, all traffic is classified at the input interface of every QoS-aware device.

The concept of trust is very important for deploying QoS. When an end device (such as a workstation or an IP phone) marks a packet with class of service (CoS) or differentiated services code point (DSCP), a switch or router has the option of accepting or not accepting the QoS marking values from the end device. If the switch or router chooses to accept the QoS marking values, the switch or router trusts the end device. If the switch or router trusts the end device, it does not need to do any reclassification of packets coming from that interface. If the switch or router does not trust the interface, it must perform a reclassification to determine the appropriate QoS value for the packets that are coming in from that interface. Switches and routers are generally set to not trust end devices, and must be specifically configured to trust packets coming from an interface.

Identification of a traffic flow can be performed by using several methods within a router, such as matching traffic using access control lists (ACLs), using protocol match, or matching the IP precedence, IP DSCP, Multiprotocol Label Switching (MPLS) EXP bit, or class of service (CoS).

Marking

This topic describes traffic marking.



Marking, also known as coloring, involves marking each packet as a member of a network class so that devices throughout the rest of the network can quickly recognize the packet class. Marking is performed as close to the network edge as possible, and is typically done using MQC.

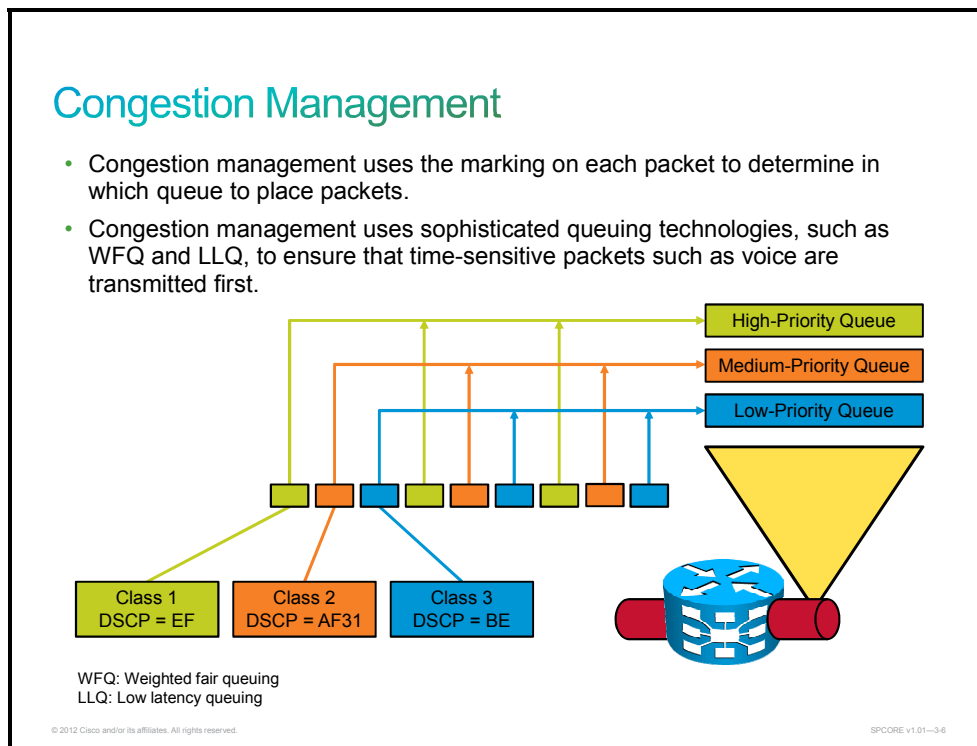
QoS mechanisms set bits in the IP, MPLS, or Ethernet header according to the class of the packet. Other QoS mechanisms use these bits to determine how to treat the packets when they arrive. If the packets are marked as high-priority voice packets, the packets will generally not be dropped by congestion avoidance mechanisms, and will be given immediate preference by congestion management queuing mechanisms. However, if the packets are marked as low-priority file transfer packets, they will have a higher drop probability when congestion occurs, and will generally be moved to the end of the congestion management queues.

Marking of a traffic flow is performed in these ways:

- Setting IP precedence or DSCP bits in the IP Type of Service (ToS) byte
- Setting CoS bits in the Layer 2 headers
- Setting EXP bits within the imposed or the topmost MPLS label
- Setting qos-group and discard-class bits

Congestion Management

This topic describes Congestion Management (Queuing)



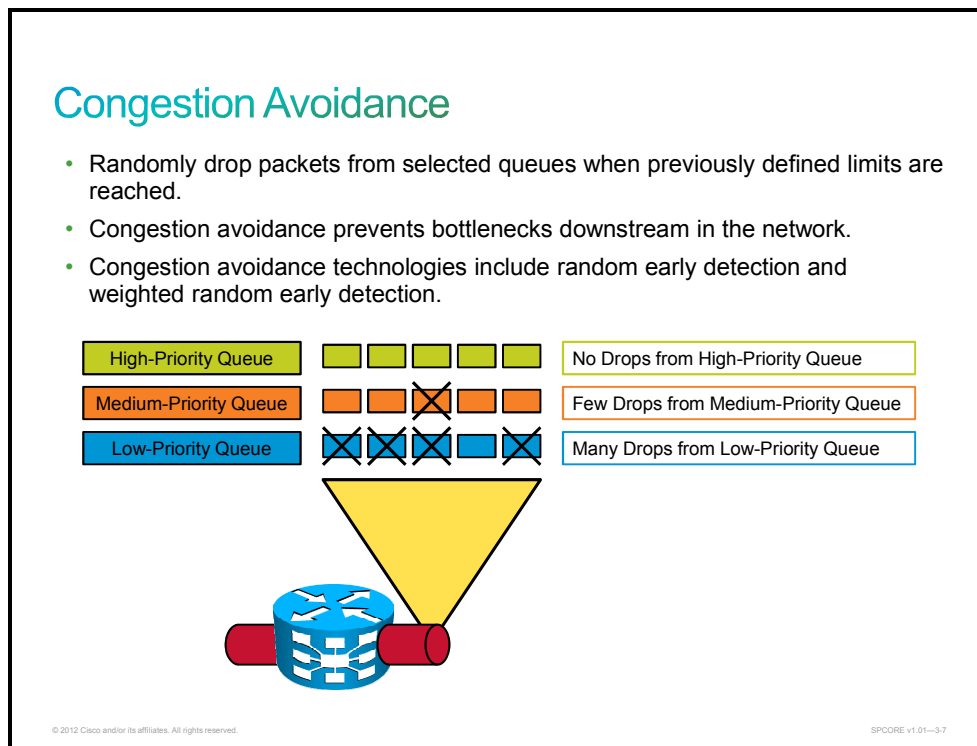
Congestion management mechanisms (queuing algorithms) use the marking on each packet to determine in which queue to place the packets. Different queues are given different treatment by the queuing algorithm, based on the class of packets in the queue. Generally, queues with high-priority packets receive preferential treatment.

Congestion management is implemented on all output interfaces in a QoS-enabled network by using queuing mechanisms to manage the outflow of traffic. Each queuing algorithm was designed to solve a specific network traffic problem, and each has a particular effect on network performance.

Cisco IOS XR Software implements the low latency queuing (LLQ) feature, which brings strict priority queuing (PQ) to the modified deficit round robin (MDRR) scheduling mechanism. LLQ with strict PQ allows delay-sensitive data, such as voice, to be dequeued and sent before packets in other queues are dequeued.

Congestion Avoidance

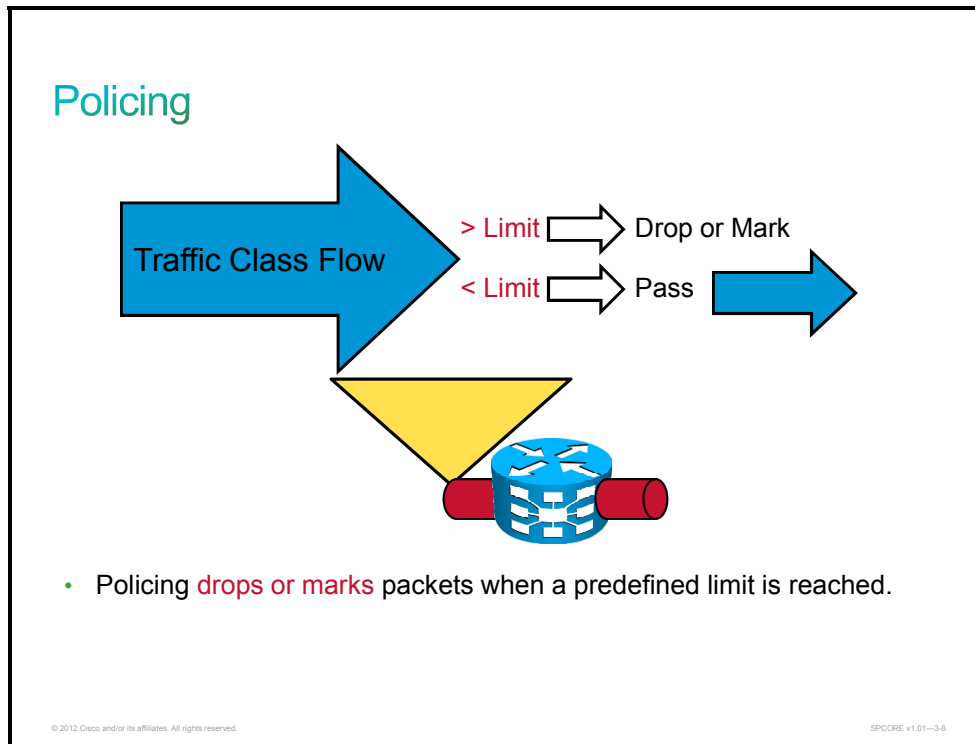
This topic describes Congestion Avoidance (RED and WRED).



Congestion-avoidance techniques monitor network traffic flows to anticipate and avoid congestion at common network and internetwork bottlenecks, before problems occur. They are typically implemented on output interfaces wherever a high-speed link or set of links feeds into a lower-speed link. These techniques are designed to provide preferential treatment for traffic (such as a video stream) that has been classified as real-time critical under congestion situations, while concurrently maximizing network throughput and capacity utilization and minimizing packet loss and delay. Cisco IOS XR Software supports the random early detection (RED), weighted RED (WRED), and tail-drop QoS congestion-avoidance features.

Policing

This topic describes traffic policing.



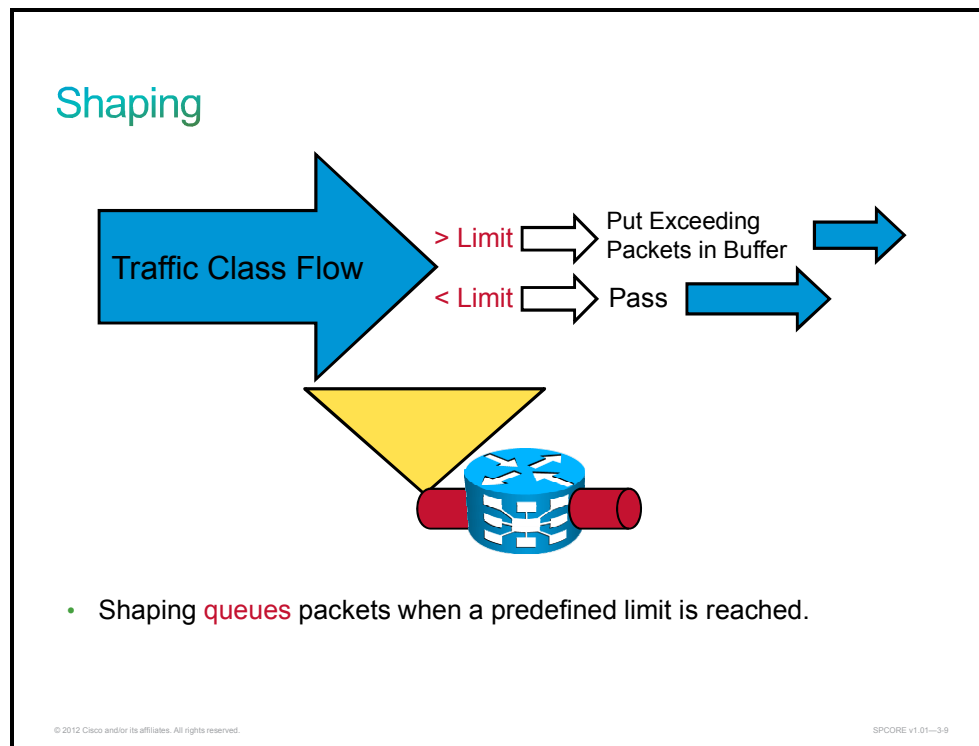
The traffic policing feature limits the input or output transmission rate of a class of traffic based on user-defined criteria, and can mark packets by setting values such as IP precedence, QoS group, or DSCP value. Policing mechanisms can be set to drop traffic classes that have lower QoS priority markings first.

Policing is the ability to control bursts and conform traffic to ensure that certain types of traffic get certain types of bandwidth.

Policing mechanisms can be used at either input or output interfaces. These mechanisms are typically used to control the flow into a network device from a high-speed link by dropping excess low-priority packets. A good example would be the use of policing by a service provider to slow down a high-speed inflow from a customer that was in excess of the service agreement. In a TCP environment, this policing would cause the sender to slow its packet transmission.

Shaping

This topic describes traffic shaping.

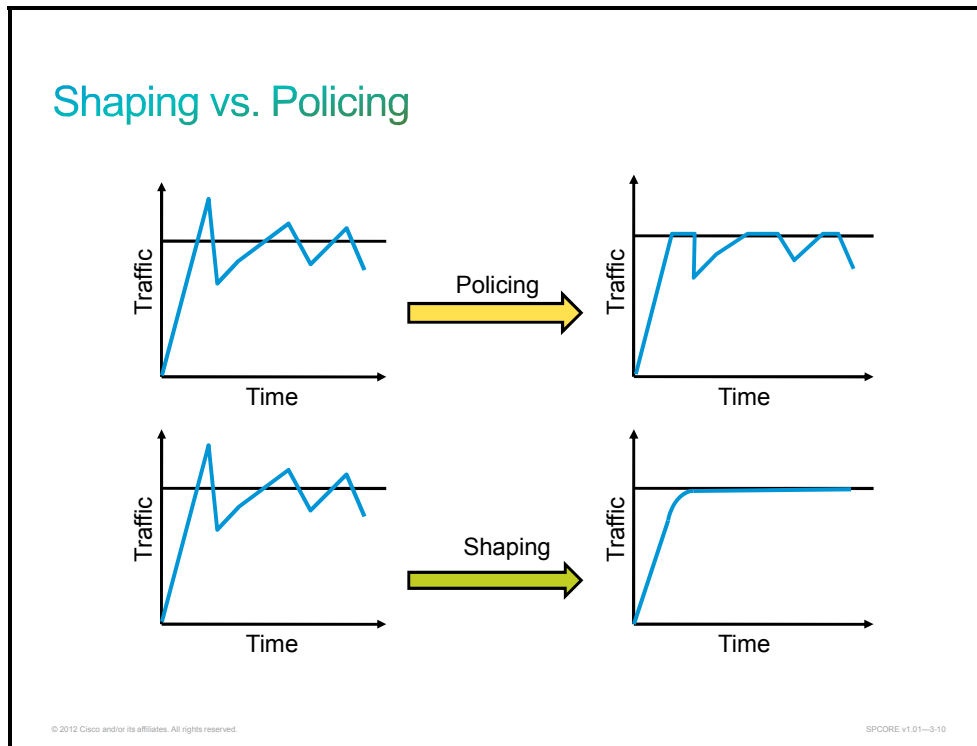


Traffic shaping allows control over the traffic that leaves an interface, to match its flow to the speed of the remote target interface and ensure that the traffic conforms to the policies contracted for it. Thus, traffic adhering to a particular profile can be shaped to meet downstream requirements, thereby eliminating bottlenecks in topologies with data-rate mismatches.

Cisco IOS XR Software supports a class-based traffic shaping method through a CLI mechanism in which parameters are applied per class.

Shaping vs. Policing

This topic compares traffic shaping vs. traffic policing.



This diagram illustrates the main difference between shaping and policing. Traffic policing propagates bursts. When the traffic rate reaches the configured maximum rate, excess traffic is dropped (or remarked). The result is an output rate that appears as a saw-tooth with crests and troughs. In contrast to policing, traffic shaping retains excess packets in a queue and then schedules the excess for later transmission over increments of time. The result of traffic shaping is a smoothed packet output rate.

Shaping implies the existence of a queue and of sufficient memory to buffer delayed packets, while policing does not. Queuing is an outbound concept; packets going out an interface get queued and can be shaped. Only policing can be applied to inbound traffic on an interface. Ensure that you have sufficient memory when you enable shaping. In addition, shaping requires a scheduling function for later transmission of any delayed packets. This scheduling function allows you to organize the shaping queue into different queues. Examples of scheduling functions are class-based weighted fair queuing (CBWFQ) and LLQ.

Implementing QoS

This topic lists and describes methods for implementing QoS.

Methods for Implementing QoS

- **CLI** can be used to individually configure QoS policy on each interface.
- **Cisco AutoQoS** (VoIP or Enterprise) is a best-practice QoS configuration that automatically generates QoS commands.
- **MQC** (Modular QoS CLI) allows the creation of modular QoS policies and attachment of these policies on the interface.
- **CiscoWorks QPM** (QoS Policy Manager) is CiscoWorks tool that allows a network administrator to create, control, and monitor QoS policies.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-3-11

Several years ago, the only way to implement QoS in a network was by using the CLI to individually configure QoS policies at each interface. This was a time-consuming, tiresome, and error-prone task that involved cutting and pasting configurations from one interface to another.

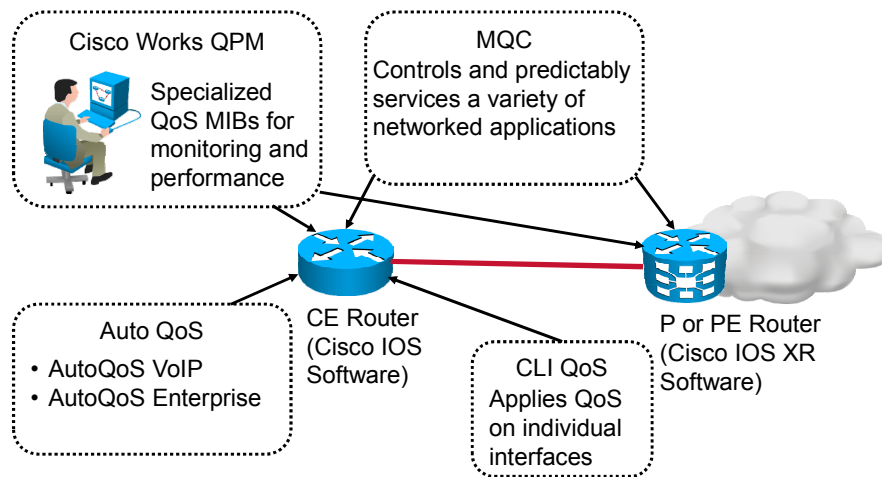
Cisco introduced the MQC to simplify QoS configuration by making configurations modular. Using MQC, you can configure QoS with a building-block approach, using a single module repeatedly to apply policy to multiple interfaces.

Cisco AutoQoS VoIP or AutoQoS for the Enterprise can be implemented with QoS features that support VoIP traffic and data traffic, without an in-depth knowledge of these underlying technologies.

CiscoWorks QoS Policy Manager (QPM) provides a scalable platform for defining, applying, and monitoring QoS policy on a systemwide basis for Cisco devices, including routers and switches. QPM enables the baselining of profile network traffic, creates QoS policies at an abstract level, controls the deployment of policies, and monitors QoS to verify the intended results. As a centralized tool, CiscoWorks QPM is used to monitor and provision QoS for groups of interfaces and devices.

Configuring QoS Mechanisms

Predominant method of configuring QoS is MQC



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-3.12

Based on the type of QoS configuration technique, QoS mechanisms are deployed differently.

At one time, the CLI was the only way to implement QoS in a network. It was a painstaking task, involving copying one interface configuration, and then pasting it into other interface configurations.

MQC is a CLI structure that allows you to create traffic policies and then attach these policies to interfaces. A traffic policy contains one or more traffic classes and one or more QoS features. A traffic class is used to classify traffic; the QoS features in the traffic policy determine how to treat the classified traffic. MQC offers excellent modularity and the ability to fine-tune complex networks. This lesson will focus on the MQC method.

AutoQoS is an intelligent macro that enables you to enter one or two simple AutoQoS commands to enable all the appropriate features for the recommended QoS setting for an application on a specific interface. Cisco AutoQoS was introduced in Cisco IOS Software Release 12.2(15)T and Cisco IOS XE Software Release 3.1.0 SG. AutoQoS discovery (enterprise) was introduced in Cisco IOS Software Release 12.3(7)T, and is not available in Cisco IOS XE Software. AutoQoS is not supported on Cisco IOS XR Software. There are two versions of AutoQoS:

- **AutoQoS VoIP:** In its initial release, AutoQoS VoIP provided best-practice QoS configuration for VoIP on both Cisco switches and routers. This was accomplished by entering one global or interface command. Depending on the platform, the AutoQoS macro would then generate commands into the recommended VoIP QoS configurations, along with class maps and policy maps, and would apply those to a router interface or switch port.
- **AutoQoS for the Enterprise:** AutoQoS for the Enterprise relies on network-based application recognition (NBAR) to gather statistics and detect ten traffic types, resulting in the provisioning of class maps and policy maps for these traffic types.

CiscoWorks QPM allows the analyzing of traffic throughput by application or service class. This analysis leverages that information to configure QoS policies to differentiate traffic and define the QoS functions that are applied to each type of traffic flow. QPM uses MIBs to generate statistics about the performance of the network. Specialized QoS MIBs enable CiscoWorks QPM to graphically display key QoS information in the form of reports. These reports can graphically illustrate the overall input traffic flow divided by traffic class, the traffic that was actually sent, and the traffic that was dropped because of QoS policy enforcement.

The latest QPM version (4.1.6) is supported on Cisco IOS devices from the 12.0 release. On Cisco IOS XR devices, QPM is supported from the 3.3 release for Cisco Carrier Routing System devices and from version 3.6.1 for Cisco 12000 Series Gigabit Switch Routers. On Cisco 1000 Series Aggregation Services Routers, it is supported in Cisco IOS XE Software from release 2.2(33).

MQC

This topic explains the MQC method to implement QoS.

Modular QoS CLI

Predominant methodology on Cisco IOS, IOS XE, and IOS XR Software

- Great scalability of method
- Uniform CLI structure for all QoS features
- Separates classification engine from the policy
- Steps to configure QoS using MQC:
 1. Define traffic classes using the **class-map** command.
 2. Define policies for traffic classes using the **policy-map** command.
 3. Apply service policy on interface (inbound or outbound) using the **service-policy** command.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3-13

The MQC was introduced to allow any supported classification to be used with any QoS mechanism.

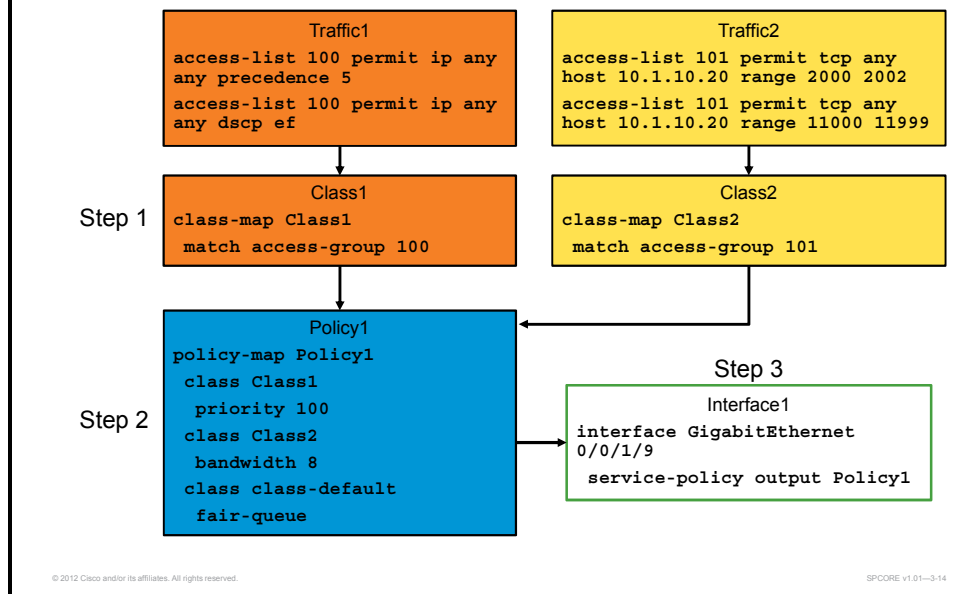
The separation of classification from the QoS mechanism allows new Cisco software versions to introduce new QoS mechanisms and reuse all available classification options. Also, older QoS mechanisms can benefit from new classification options.

Another important benefit of the MQC is the reusability of a configuration. MQC allows the same QoS policy to be applied to multiple interfaces. The MQC, therefore, is a consolidation of all the QoS mechanisms that have so far been available only as standalone mechanisms.

Implementing QoS by using the MQC consists of three steps:

- Step 1** Configure classification by using the **class-map** command.
- Step 2** Configure traffic policy by associating the traffic class with one or more QoS features using the **policy-map** command.
- Step 3** Attach the traffic policy to inbound or outbound traffic on interfaces, subinterfaces, or virtual circuits (VCs) by using the **service-policy** command.

MQC Configurations Steps: Example



In general, provisioning QoS policies requires these steps:

- Step 1** Specify traffic classes.
- Step 2** Associate actions with each traffic class to formulate policies.
- Step 3** Activate the policies.

The specification of a classification policy—that is, the definition of traffic classes—is separate from the specification of the policies that act on the results of the classification.

The **class-map** command defines a named object representing a class of traffic, specifying the packet matching criteria that identify packets that belong to this class. This is the basic form of the command:

```

class-map class-map-name-1
    match match-criteria-1
class-map class-map-name-n
    match match-criteria-n
  
```

The **policy-map** command defines a named object that represents a set of policies to be applied to a set of traffic classes. An example of such a policy is policing the traffic class to some maximum rate. The basic form of the command is as follows:

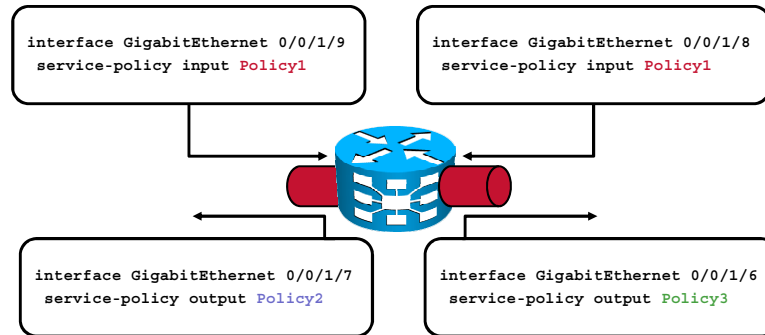
```

policy-map policy-map-name
    class class-map-name-1
        policy-1
    class class-map-name-n
        policy-m
        policy-m+1
  
```

The **service-policy** command attaches a policy map and its associated policies to a target, a named interface.

Applying QoS Policy on Interfaces

- One policy per direction can be used on an interface.
- One policy can be applied on multiple interfaces.



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3-15

A service policy associates a policy with a particular target and direction within a device. The **policy-map** command must have defined the policy previously. The separation of the policy definition from the policy invocation reduces the complexity of the QoS configuration.

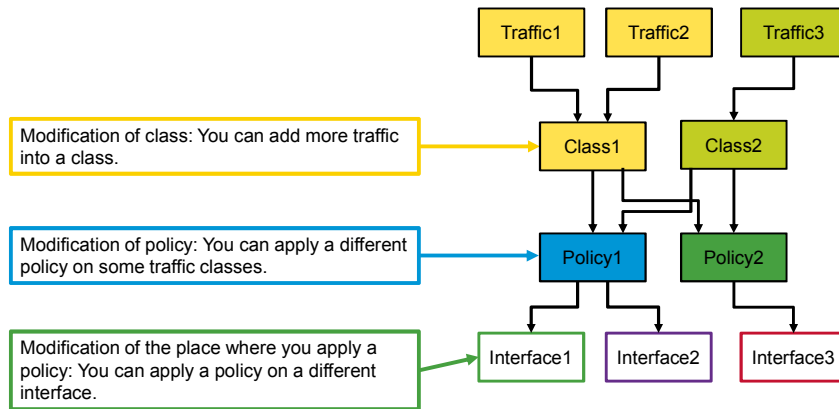
The configuration of the **service-policy** command determines both the direction and the attachment point of the QoS policy. You can attach a policy to an interface (physical or logical), to a permanent virtual circuit (PVC), or to special points to control route processor traffic. Examples of logical interfaces include these:

- MFR (Multilink Frame Relay)
- Multilink (Multilink PPP)
- Port channel (Ethernet channel of interfaces)
- POS channel (packet-over-SONET/SDH channel of interfaces)
- Virtual template

Two directions are possible for a policy: input and output. The policy direction is relative to the attachment point. The attachment point and direction influence the type of actions that a policy supports (for example, some interfaces may not support input queuing policies).

MQC Policy Modification

- QoS policy can be modified at any step during MQC configuration.



Modification of the class: QoS traffic class can be modified in different ways. You can create new classes, edit an existing class by adding more traffic into it, add conditional matching statements on existing classes, or remove classes that are no longer being used by any policy.

Modification of policy: Policy can be modified in many ways. You can apply a different policy on traffic class, apply a child policy, or simply add a new per-hop behavior (PHB) to an existing traffic class. Change of policy is immediately reflected on traffic that is passing through the interface.

Modification of attachment point or direction: The same policy can be applied on multiple interfaces. You can disable or enable policy on an interface by entering one command.

MQC on Cisco IOS, IOS XR, and IOS XE Software

- MQC configuration differs on Cisco IOS, IOS XE, and IOS XR Software.
- Cisco IOS XE and Cisco IOS QoS configurations are the same.

Cisco IOS and IOS XE

```
class-map match-all premium
match dscp ef
class-map match-all gold
match dscp af31
!
policy-map Policy1
class premium
bandwidth 15000
class gold
bandwidth 10000
interface Gigabit 0/1/0
service-policy output Policy1
```



© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco IOS XR

```
class-map match-any premium
match dscp ef
end-class-map
!
class-map match-any gold
match dscp af31
end-class-map
!
policy-map Policy1
class premium
bandwidth 15 mbps
!
class gold
bandwidth 10 mbps
!
class class-default
!
end-policy-map
!
interface GigabitEthernet0/0/0/1
service-policy output Policy1
```



SPCORE v1.01-3-17

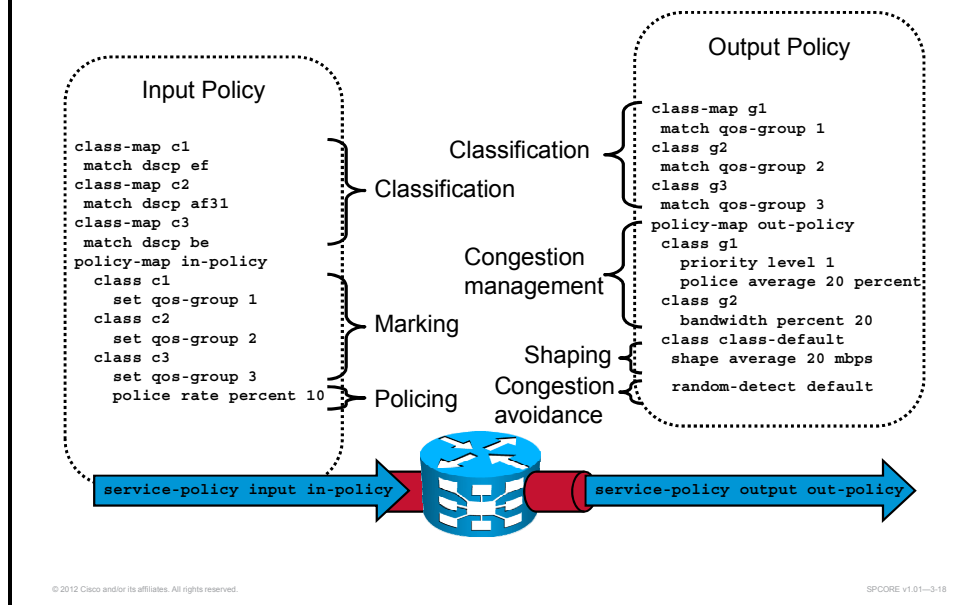
Cisco IOS XR Software supports a differentiated service, a multiple-service model that can satisfy different QoS requirements.

MQC QoS commands on Cisco IOS and Cisco IOS XE Software are identical. Each QoS technique has slightly different capabilities between Cisco IOS and IOS XE and Cisco IOS XR Software.

In Cisco IOS XR Software, features are generally disabled by default and must be explicitly enabled. There are some differences in the default syntax values used in MQC. For example, if you create a traffic class with the **class-map** command in Cisco IOS and Cisco IOS XR Software, Cisco IOS Software creates by default a traffic class that must match all statements under the service class that is defined. Cisco IOS XR Software creates a traffic class that matches any of the statements under the service class. Another difference is the available set of capabilities in different types of software. The Cisco IOS XR QoS features enable networks to control and predictably service various networked applications and traffic types. Implementing Cisco IOS XR QoS offers these benefits:

- Control over resources
- Tailored services
- Coexistence of mission-critical applications

Applying QoS Mechanisms on an Interface



In a QoS-enabled network, classification is performed on every input interface. Marking should be performed as close to the network edge as possible—in the originating network device, if possible. Devices farther from the edge of the network, such as routers and switches, can be configured to trust or not trust the markings that are made by devices on the edge of the network. An IP phone, for example, will not trust the markings of an attached PC, while a switch will generally be configured to trust the markings of an attached IP phone.

It is wise to use congestion management, congestion avoidance, and traffic-shaping mechanisms on output interfaces, because these mechanisms help maintain the smooth operation of links by controlling how much and which type of traffic is allowed on a link. On some router and switch platforms, congestion management mechanisms, such as weighted round robin (WRR) and modified deficit round robin (MDRR) can be applied on the input interface.

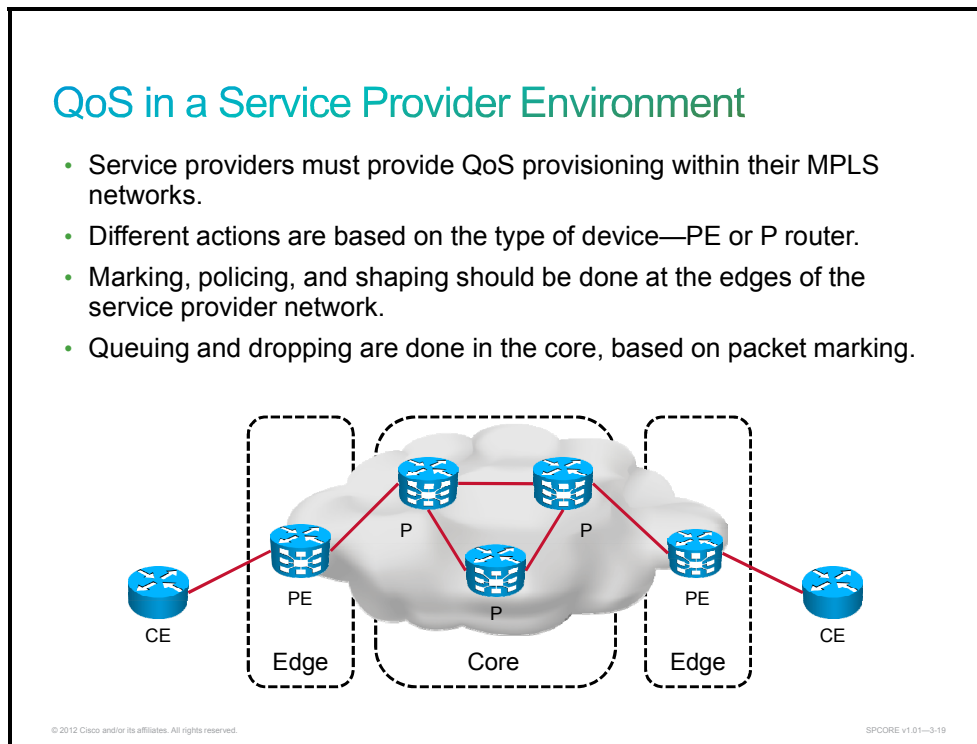
Congestion avoidance is typically employed on an output interface wherever there is a chance that a high-speed link, or aggregation of links, feeds into a slower link (such as a LAN feeding into a WAN).

Policing and shaping are typically employed on output interfaces to control the flow of traffic from a high-speed link to lower-speed links. Policing is also employed on input interfaces to control the flow into a network device from a high-speed link by dropping excess low-priority packets.

Detailed QoS configurations will be covered in other lessons. This figure provides only an overview of ways that the different QoS mechanisms are applied.

QoS in Service Provider Environment

This topic describes QoS requirements on the different devices in the service provider environment.



To support enterprise-subscriber voice, video, and data networks, service providers must include QoS provisioning within their MPLS VPN service offerings. To face that challenge, service providers must do these things:

- Support enterprises with diverse QoS policies
- Ensure contracted rates per service level agreements (SLAs)
- Ensure loss, latency, and jitter, per class and per SLA
- Maintain QoS transparency for customers
- Measure and report SLA metrics
- Plan capacity that is based on SLA metric measurements

The service provider IP core is used to provide high-speed packet transport. Therefore, all the markings, policing, and shaping should be performed only at the provider edge (PE) router on the PE-to-customer edge (CE) link, and not at the core. Using the differentiated services (DiffServ) model, only the edge requires a complex QoS policy. At the core, only queuing and dropping are required. The operation of queuing and dropping will be based on the markings that are done at the PE.

The reason for these procedures is the any-to-any and full-mesh nature of MPLS VPNs, where enterprise subscribers depend on their service providers to provision PE-to-CE QoS policies that are consistent with their CE-to-PE policies.

In addition to these PE-to-CE policies, service providers will likely implement ingress policers on their PEs to identify whether the traffic flows from the customer are in- or out-of-contract.

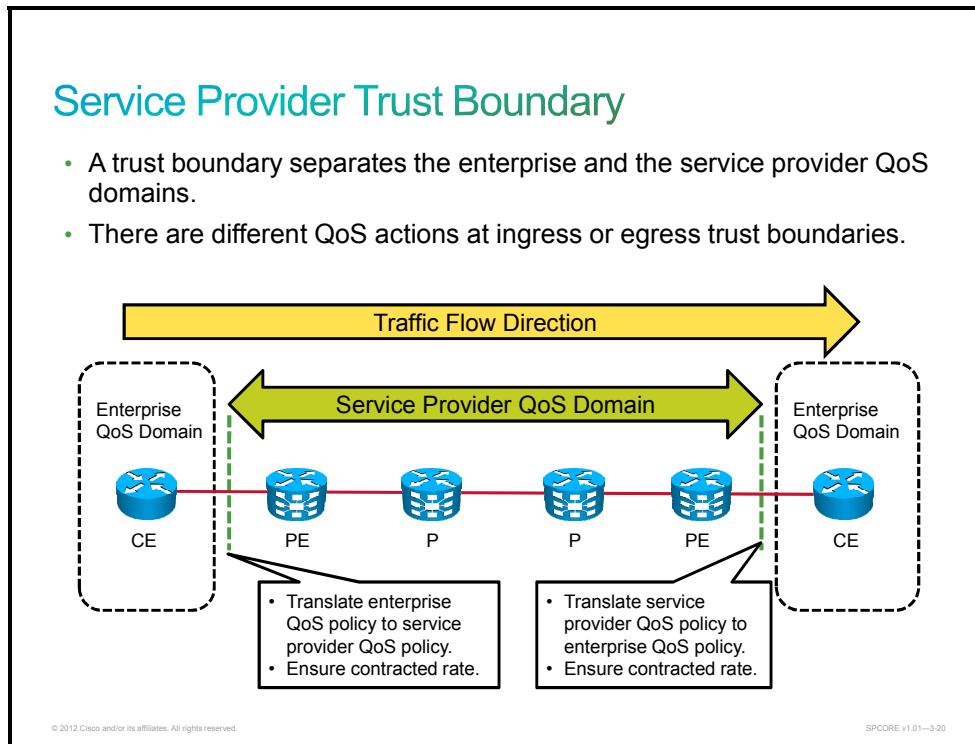
Optionally, service providers may also provision QoS policies within their core networks, using DiffServ or MPLS traffic engineering (TE).

To guarantee end-to-end QoS, enterprises must manage QoS with their MPLS VPN service providers; their policies must be both consistent and complementary.

Service providers can mark at Layer 2 (MPLS EXP) or at Layer 3 (DSCP). Marking will be covered in more detail in other lessons.

Service Provider Trust Boundary

This topic describes the Service Provider Trust Boundary.



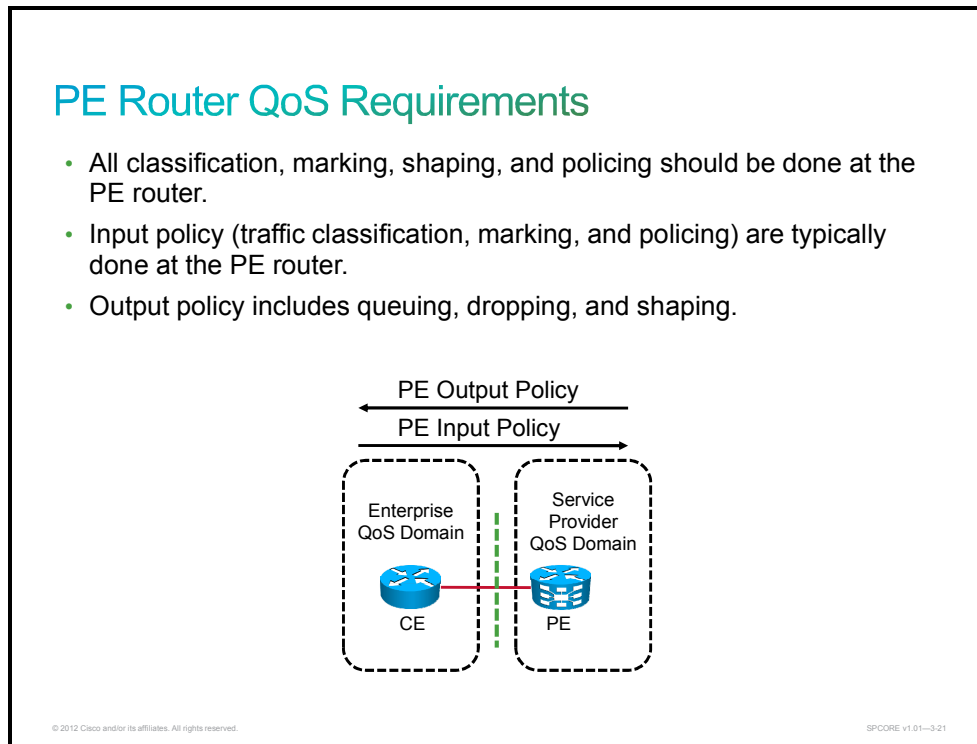
There are many places in the network in which the application of QoS, either marking or classification, occurs. A primary function of provider edge policies is to establish and enforce trust boundaries. A trust boundary is the point within the network where markings begin to be accepted. Markings that were previously set by the enterprise are overridden at the trust boundary.

The concept of trust is important and integral to implementing QoS. As soon as the end devices have set their marking, the switch can either trust them or not trust them. If the device at the edge trusts the settings, it does not need to do any reclassification. If it does not trust the settings, it must perform reclassification for the appropriate QoS.

Enterprise QoS policies are applied on the CE router and must comply with available bandwidth and application requirements. On the other end, the service provider can ensure the contracted rate by using traffic shaping and policing tools. Because the service provider can mark packets in a different manner than the enterprise can, the service provider needs to apply classification and marking policies at the PE routers. To achieve end-to-end service levels, enterprise and service-provider QoS designs must be consistent and complimentary. The only way to guarantee service levels in such a scenario is for the service provider to provision QoS scheduling that is compatible with the enterprise policies on all PE links to the CE devices.

PE router QoS Requirements

This topic describes the QoS requirements on the PE routers.



The QoS requirements on the CE and PE routers will differ, depending on whether the CE is managed by the service provider.

For unmanaged CE service, the WAN edge output QoS policy on the CE will be managed and configured by the enterprise customer.

For managed CE, the service provider will implement QoS policy on the PE router. At the PE input interface, the service provider will have a policy to classify, mark, or map the traffic. The service provider also typically implements traffic policing to rate-limit the input traffic rate from the enterprise customer, so that the traffic rate does not exceed the contractual rate as specified in the SLA.

The service provider can enforce the SLA for each traffic class by using the output QoS policy on the PE. For example, queuing mechanisms are used to give a maximum bandwidth guarantee to the real-time voice and video traffic class, give a minimum bandwidth guarantee to the data traffic classes, and use class-based shaping to provide a maximum rate limit to each data traffic class.

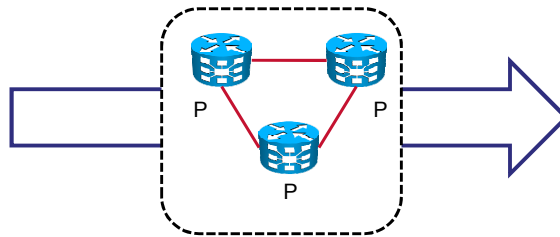
For both managed and unmanaged CE service, the service provider typically has an output policy on the PE router using congestion management and congestion avoidance mechanisms. To compensate for a speed mismatch or oversubscription, traffic shaping may be required.

P Router QoS Requirements

This topic describes the QoS requirements on the P routers.

Core P Router QoS Requirements

- The service provider IP core is used to provide high-speed packet transport.
- Queuing and dropping are done in the core, based on packet marking done at the edge.
- There are two methods for QoS design:
 - Best effort with overprovisioning (expensive)
 - DiffServ backbone (commonly used)



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3.22

Two of the IP backbone design methods include a best-effort backbone with overprovisioning and a DiffServ backbone.

The more traditional approach is to use a best-effort backbone with overprovisioning. However, to meet increasing application needs (VoIP, videoconferencing, e-learning, and so on), deploying a DiffServ backbone and offering different SLAs for the different traffic classes can greatly reduce the cost, improve delay, jitter, and packet loss, and meet network QoS requirements.

Congestion avoidance and congestion management are commonly used on the provider (P) router output interface. The P router input interface does not need to have any QoS policy applied.

QoS policies on P routers are optional. Such policies are optional because some service providers overprovision their MPLS core networks, and therefore do not require any additional QoS policies within their backbones; however, other providers might implement simplified DiffServ policies within their cores, or might even deploy MPLS TE to manage congestion scenarios within their backbones.

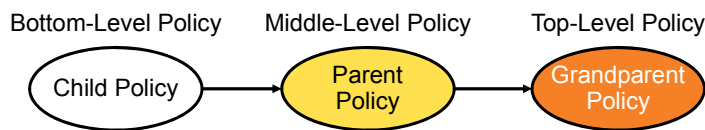
Hierarchical QoS Policies

This topic describes Hierarchical QoS policies.

Hierarchical QoS

- Specifies QoS behavior at different policy levels
- Provides a high degree granularity in traffic management
- Uses the **service-policy** command to apply policy to another policy and a policy to an interface
- Applies a child policy to a class of parent policy

Three-level hierarchical policy:



Hierarchical QoS allows you to specify QoS behavior at multiple policy levels, which provides a high degree of granularity in traffic management. A hierarchical policy is a QoS model that enables you to specify QoS behavior at multiple levels of hierarchy. You can use hierarchical policies to do these things:

- Allow a parent class to shape multiple queues in a child policy.
- Apply specific policy map actions on the aggregate traffic.
- Apply class-specific policy map actions.
- Restrict the maximum bandwidth of a VC, while allowing policing and marking of traffic classes within the VC.

The **service-policy** command is used to apply a policy to another policy, and a policy to an interface, subinterface, VC, or VLAN.

For example, in a three-level hierarchical policy, use the **service-policy** command to apply these policies:

- Bottom-level policy to a middle-level policy
- Middle-level policy to a top-level policy
- Top-level policy to an interface, subinterface, VC, or VLAN

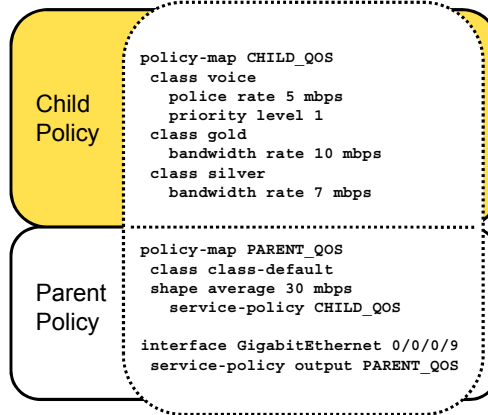
Depending on the type of hierarchical QoS policy you configure, you can do these things:

- Shape multiple queues to a single rate
- Divide a single class of traffic into one or more subclasses
- Specify the maximum transmission rate of a set of traffic classes that are queued separately, which is essential for virtual interfaces such as Frame Relay PVCs and IEEE 802.1Q virtual VLANs
- Configure minimum-bandwidth queues on VCs
- Shape the aggregate traffic of queues on a physical interface

Hierarchical QoS Example:

- Two levels of hierarchy:
 - Child policy enforces an LLQ mechanism on three traffic classes.
 - Parent policy shapes all traffic to 30 mb/s.
 - Child policy is applied on default class of parent policy.

IOS XR Example:



In this example, a two-level QoS policy is configured. In the CHILD_QOS policy, LLQ is configured so that the voice class has priority over other classes. Policing must be configured for the voice traffic class to limit priority traffic and prevent low-priority traffic from starving. The PARENT_QOS policy is used to shape all traffic passing through the GigabitEthernet 0/0/0/9 interface in the outbound direction to 30 Mb/s to enable queuing of excess bursts of traffic. For all traffic belonging to the class-default class and passing through the pipe of 30 Mb/s of bandwidth, the CHILD_QOS policy is applied.

As you configure hierarchical QoS, consider these guidelines:

- When you are defining policies, start at the bottom level of the hierarchy. For example, for a two-level hierarchical policy, define the bottom-level policy, and then define the top-level policy. For a three-level hierarchical policy, define the bottom-level policy, the middle-level policy, and then the top-level policy.
- Do not specify the **input** or **output** keyword in the **service-policy** command when you are configuring a bottom-level policy within a top-level policy.
- Configure bottom-level policies only in middle-level and top-level policies.

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- There are several QoS mechanisms
- Classification is the identifying and splitting of traffic into different classes
- Marking colors each packet as a member of a network class so that devices throughout the rest of the network can quickly recognize the packet class
- Congestion management mechanisms (queuing algorithms) use the marking on each packet to determine in which queue to place the packets
- Congestion avoidance techniques monitor network traffic flows to anticipate and avoid congestion at common network and internetwork bottlenecks, before problems occur
- The traffic policing feature limits the input or output transmission rate of a class of traffic based on user-defined criteria, and can mark packets
- Traffic shaping allows control over the traffic that leaves an interface, to match its flow to the speed of the remote target interface and ensure that the traffic conforms to the policies contracted for it

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3-28

Summary (Cont.)

- Traffic shaping allows control over the traffic that leaves an interface, to match its flow to the speed of the remote target interface and ensure that the traffic conforms to the policies contracted for it
- Shaping implies the existence of a queue and of sufficient memory to buffer delayed packets, while policing does not
- MQC simplifies QoS configuration by making configurations modular
- The MQC was introduced to allow any supported classification to be used with any QoS mechanism
- All the markings, policing, and shaping should be performed only at the provider edge (PE) router on the PE-to-customer edge (CE) link
- There are many places in the network in which the application of QoS, either marking or classification, occurs
- For unmanaged CE service, the WAN edge output QoS policy on the CE will be managed and configured by the enterprise customer.
- Congestion avoidance and congestion management are commonly used on the provider (P) router output interface
- A hierarchical policy is a QoS model that enables you to specify QoS behavior at multiple levels of hierarchy.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3-28

Implementing MPLS Support for QoS

Overview

Quality of service (QoS) has become an integral part of a multiservice, converged network and service implementation. QoS for IP Multiprotocol Label Switching (MPLS) networks offers network engineers a single source of information for the design, deployment, and implementation of QoS-enabled services on an MPLS network, using Cisco IOS, IOS XE, and IOS XR Software. You will learn about the technology behind the MPLS QoS differentiated services (DiffServ) approach and related technologies. You will learn the different design options that are available to build a multiservice MPLS network.

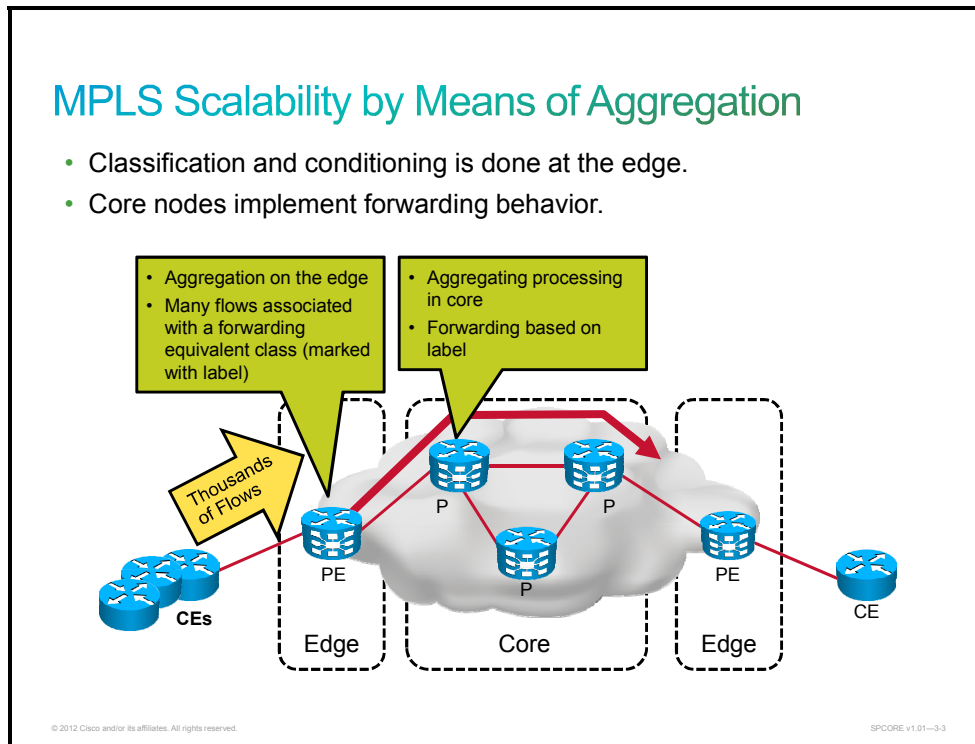
Objectives

Upon completing this lesson, you will understand how MPLS marks frames and how an MPLS network performs per-hop behavior (PHB) to offer predictable QoS classes. You will be able to meet these objectives:

- Describe basic MPLS QoS concepts
- Describe the MPLS EXP field
- Describe the QoS Group (internal router QoS marker)
- Explain MPLS QoS configurations on a PE Router
- Explain MPLS QoS configurations on a P Router
- Describe the show commands used for monitoring MPLS QoS
- Describe Point-to-Cloud Service Model in MPLS VPNs
- Describes Point-to-Point Service Model in MPLS VPNs
- Lists the three MPLS DiffServ Modes
- Describe the MPLS DiffServ Pipe Mode
- Describe the MPLS DiffServ Short-Pipe Mode
- Describe the MPLS DiffServ Uniform Mode
- Describes MPLS DiffServ Traffic Engineering

MPLS QoS

This topic describes basic MPLS QoS concepts.



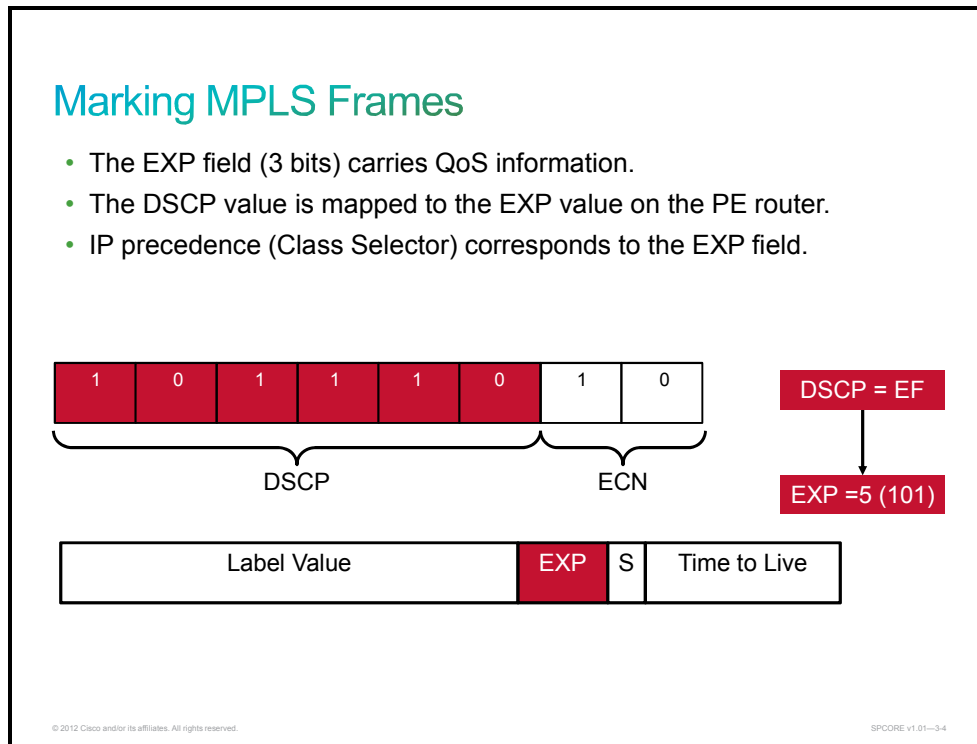
The main goals of the DiffServ model are to provide scalability and a similar level of QoS as the integrated services (IntServ) model, without the need for a per-flow basis. The network simply identifies a class (not an application) and applies the appropriate PHB (a QoS mechanism).

DiffServ offers application-level QoS and traffic management in an architecture that incorporates mechanisms to control bandwidth, delay, jitter, and packet loss. Cisco DiffServ complements the Cisco IntServ offering by providing a more scalable architecture for an end-to-end QoS solution. MPLS does not define a new QoS architecture. MPLS QoS has focused on supporting current IP QoS architectures.

DiffServ defines a QoS architecture that is based on flow aggregates; traffic must be conditioned and marked at the network edges and at internal nodes to provide different QoS treatment to packets, based on their markings. MPLS packets need to carry the packet marking in their headers because label switch routers (LSRs) do not examine the IP header during forwarding. A three-bit field in the MPLS shim header is used for this purpose. The DiffServ functionality of an LSR is almost identical to the functionality that is provided by an IP router and the QoS treatment that is given to packets (or PHB, in DiffServ terms).

MPLS EXP

This topic describes the MPLS EXP field.



Marking with the MPLS experimental (EXP) bit value, in addition to the standard IP QoS information, ensures these results:

- Standard IP QoS policies are followed before the packets enter the MPLS network.
- At the ingress router to the MPLS network (the provider edge [PE] device), the differentiated services code point (DSCP) or IP Precedence value of the packet is mapped to the MPLS EXP field. These mappings are part of the QoS policy.
- The per-hop behavior (PHB) for the packets in the MPLS backbone is based on the MPLS EXP field.
- The DSCP or IP Precedence value in the IP header continues to be the basis for IP QoS when the packet leaves the MPLS network.

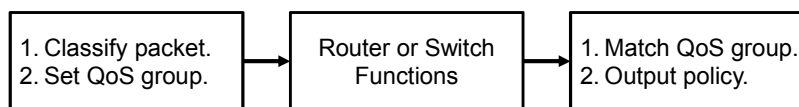
Packet behavior for the QoS provisioning components, congestion management, and congestion avoidance are derived from the MPLS EXP bits.

QoS Group

This topic describes the QoS Group (internal router QoS marker).

Classification Based on QoS Group

- QoS group is the internal label used by the router or switch to identify packets as a member of specific class.
- This label is *not* part of the packet header and is local to the router or switch.
- The label provides a way to tag a packet for subsequent QoS action.
- The QoS group label is identified at ingress and used at egress.



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3-5

A QoS group is an internal label that is used by the switch or the router to identify packets as a member of a specific class. The label is not part of the packet header and is restricted to the switch that sets the label and is not communicated between devices. QoS groups provide a way to tag a packet for subsequent QoS action, without explicitly marking (changing) the packet.

A QoS group is identified at ingress and that information is used at egress. It is assigned in an input policy to identify packets in an output policy.

You use QoS groups to aggregate different classes of input traffic for a specific action in an output policy. For example, you can classify an ACL on ingress by using the **set qos-group** command and then use the **match qos-group** command in an output policy. This Cisco IOS XR configuration example shows how to use QoS group markings (the configuration for Cisco IOS, IOS XE, and IOS XR Software is similar):

Class map:

```
class-map acl
  match access-group name acl
exit
```

Input policy map:

```
policy-map set-qos-group
  class acl
    set qos-group 5
exit
```

Output policy map:

```
policy-map shape
  class qos-group 5
```

```
shape average 10 mbps
exit
```

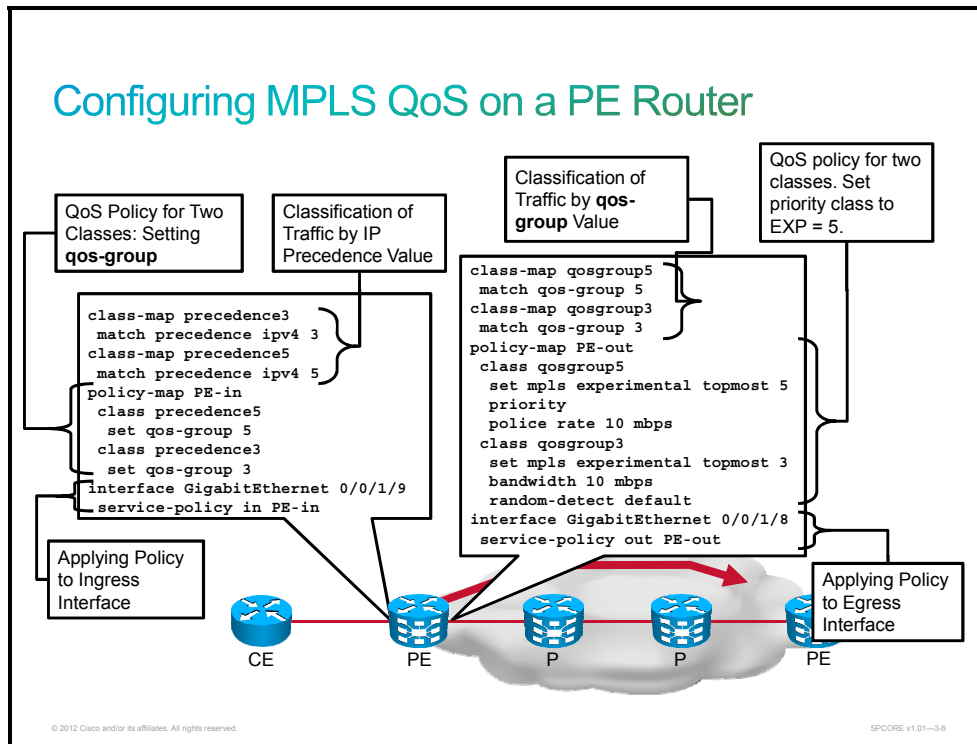
QoS groups can be used to aggregate multiple input streams across input classes and policy maps to have the same QoS treatment on the egress port. Assign the same QoS group number in the input policy map to all streams that require the same egress treatment, and match the QoS group number in the output policy map to specify the required queuing and scheduling actions.

QoS groups are also used to implement the MPLS tunnel mode. In this mode, the output per-hop behavior of a packet is determined by the input EXP bits, but the packet remains unmodified. You match the EXP bits on input, set a QoS group, and then match that QoS group on output to obtain the required QoS behavior.

The **set qos-group** command is used only in an input policy. The assigned QoS group identification is then used in an output policy with no mark or change to the packet. The command **match qos-group** is used in the output policy. The command **match qos-group** cannot be used for an input policy map.

Configuring MPLS QoS on a PE Router

This topic explains MPLS QoS configurations on a PE Router.



In the following configuration example, traffic that is sourced from the customer edge (CE) router has different IP precedence values. Also, the CE router generally uses some congestion avoidance and congestion management mechanisms to protect high-priority traffic from being dropped. Classification of ingress traffic on the PE router is based on DiffServ PHB marking (IP precedence). This Cisco IOS XR configuration shows how to configure MPLS QoS on a PE router (Cisco IOS and IOS XE configuration is similar).

Class maps that are configured on the PE router match the packets that are based on IP precedence.

```
class-map precedence3
match precedence ipv4 3
class-map precedence5
match precedence ipv4
```

The input policy applies the appropriate QoS group to packets belonging to a specific class.

```
policy-map PE-in
class precedence5
set qos-group 5
class precedence3
set qos-group 3
```

The input policy is applied to the input interface of the PE router:

```
interface GigabitEthernet 0/0/1/9
service-policy in PE-in
```


Classification of egress traffic on the PE router is based on QoS group marking. Class maps for the output policy that is configured on the PE router match the packets that are based on QoS group value.

```
class-map qosgroup5
  match qos-group 5
class-map qosgroup3
  match qos-group 3
```

The output policy applies the appropriate congestion management and congestion avoidance mechanisms to packets belonging to a specific class. It also sets the appropriate MPLS EXP marking for that specific class.

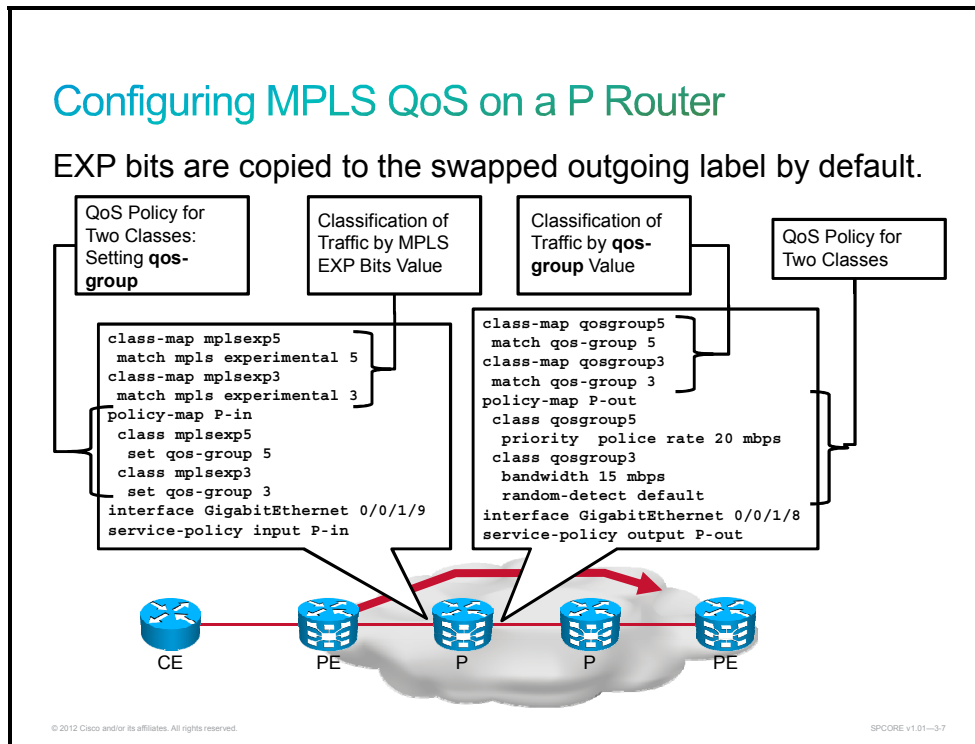
```
policy-map PE-out
  class qosgroup5
    set mpls experimental topmost 5
    priority
    police rate 10 mbps
  class qosgroup3
    set mpls experimental topmost 3
    bandwidth 10 mbps
    random-detect default
```

Finally, the output policy is applied to the output interface of the PE router:

```
interface GigabitEthernet 0/0/1/8
  service-policy out PE-out
```

Configuring MPLS QoS on a P Router

This topic explains MPLS QoS configurations on a P Router.



Provider (P) router MPLS QoS configuration is somewhat different from PE router configuration. Traffic that is sourced from a PE router has different MPLS EXP markings. Classification of ingress traffic on a P router is based on MPLS EXP bits. This Cisco IOS XR configuration shows an example of the way that you can configure MPLS QoS on a P router (Cisco IOS and IOS XE configuration is similar).

Class maps that are configured on a P router match the packets that are based on MPLS EXP markings.

```
class-map mplsexp5
match mpls experimental 5
class-map mplsexp3
match mpls experimental 3
```

The input policy applies the appropriate QoS group to packets belonging to a specific class.

```
policy-map P-in
class mplsexp5
set qos-group 5
class mplsexp3
set qos-group 3
```

The input policy is applied to the input interface of the P router:

```
interface GigabitEthernet 0/0/1/9
service-policy input P-in
```

Classification of egress traffic on the P router is based on QoS group markings. Class maps for output policy that is configured on the PE router match the packets that are based on QoS group value.

```
class-map qosgroup5
  match qos-group 5
class-map qosgroup3
  match qos-group 3
```

The output policy applies appropriate congestion management and congestion avoidance mechanisms to packets belonging to a specific class.

```
policy-map P-out
  class qosgroup5
    priority police rate 20 mbps
  class qosgroup3
    bandwidth 15 mbps
    random-detect default
```

Finally, the output policy is applied to the output interface of the P router:

```
interface GigabitEthernet 0/0/1/8
  service-policy out PE-out
```

Monitoring MPLS QoS

This topic describes the show commands used for monitoring MPLS QoS.

MPLS QoS Monitoring

```
RP/0/RSP0/CPU0:P2#show policy-map interface gigabitethernet 0/0/1/8 output
Tue Oct 4 20:56:56.757 UTC

GigabitEthernet0/0/1/8 output: P-out

Class qosgroup3
  Classification statistics      (packets/bytes)      (rate - kbps)
  Matched                      :      0/0              0
  Transmitted                  :      0/0              0
  Total Dropped                :      0/0              0

  Queueing statistics
    Queue ID                    : 268435850
    High watermark (Unknown)    :
    Inst-queue-len (packets)    : 0
    Avg-queue-len (Unknown)     :
    Taildropped(packets/bytes)  : 0/0
    Queue(conform)              :      0/0              0
    Queue(exceed)              :      0/0              0
    RED random drops(packets/bytes) : 0/0

  WRED profile for Default WRED Curve
    RED Transmitted (packets/bytes) : N/A
    RED random drops(packets/bytes) : 0/0
    RED maxthreshold drops(packets/bytes): N/A

! ... Output omitted for brevity...
```

What You Can Observe

Drops within class qosgroup3

Queueing within class qosgroup3

Drops using WRED within class qosgroup3

© 2012 Cisco and/or its affiliates. All rights reserved. SPCORE v1.01-3-8

Monitoring MPLS QoS on a specific router is based on observing the statistics about congestion on that specific interface. The Cisco IOS XR **show policy-map interface** command displays the packet statistics for classes on the specified interface. The same command, with similar output, exists for Cisco IOS and IOS XE Software.

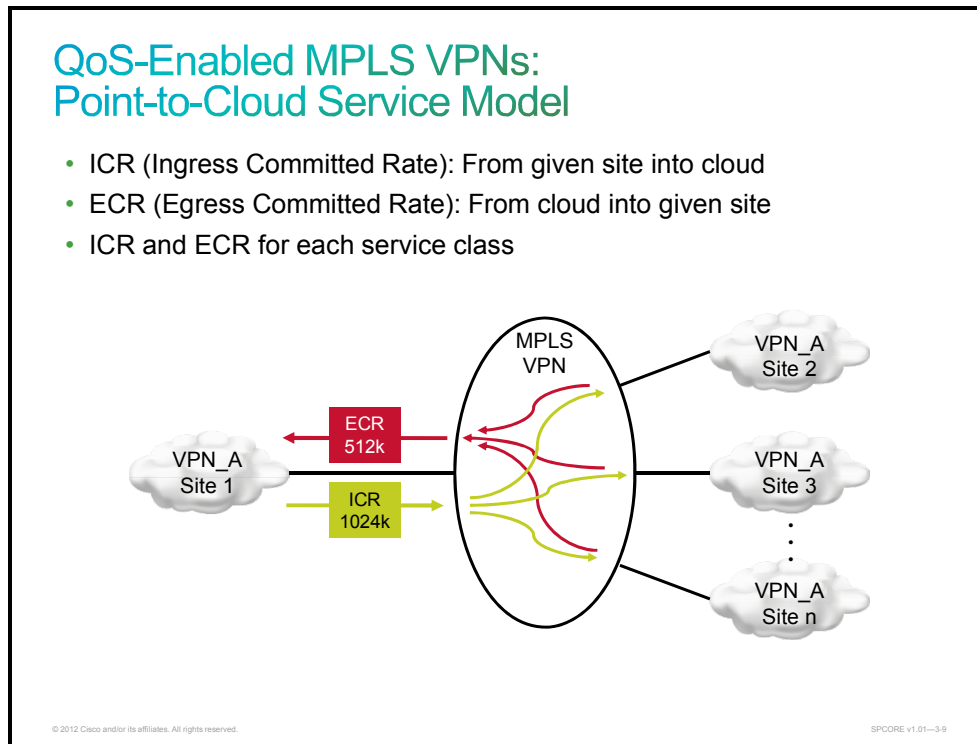
Conceptually, congestion is defined by the Cisco IOS, IOS XE, and IOS XR Software configuration guide: “During periods of transmit congestion at the outgoing interface, packets arrive faster than the interface can send them.”

In other words, congestion typically occurs when a fast ingress interface feeds a relatively slow egress interface. A common congestion point is a branch office router with an Ethernet port facing the LAN and a serial port facing the WAN. Users on the LAN segment generate 10 Mb/s of traffic, which is fed into a T1 with 1.5 Mb/s of bandwidth.

Congestion is observed through matched, transmitted, and dropped packets. Queuing and congestion avoidance mechanisms prevent high priority packets from being dropped when congestion occurs. Thus, different classes of traffic can be observed, and the number of dropped or transmitted packets can be viewed for each service class.

QoS-Enabled MPLS VPNs: Point-to-Cloud Service Model

This topic describes Point-to-Cloud Service Model in MPLS VPNs



With Cisco IOS, IOS XE, and IOS XR MPLS, service providers can use either or both of two approaches to implement QoS guarantees to customers: the point-to-cloud model and the point-to-point model.

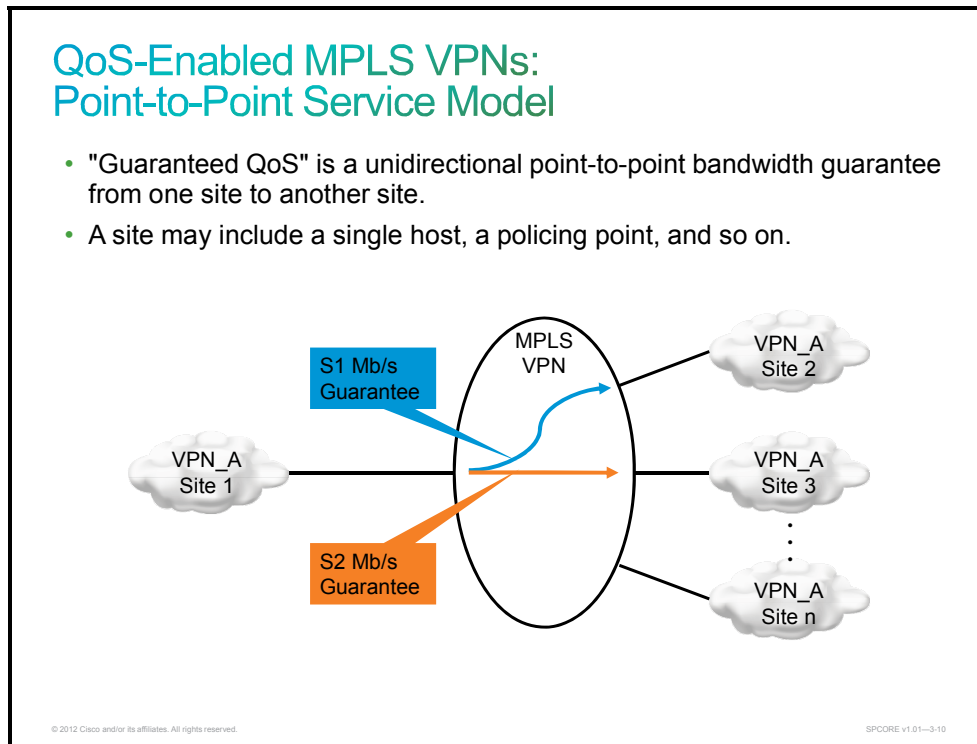
Service providers offering QoS services will want to provide an ingress committed rate (ICR) guarantee and an egress committed rate (ECR) guarantee, possibly for each service class offered. ICR refers to the traffic rate coming into the service provider network, which is given a particular treatment. ECR refers to the traffic rate that is given a particular treatment from the service provider to the customer site. As long as traffic does not exceed ICR and ECR limits, the network provides bandwidth and delay guarantees.

For example, as long as HTTP traffic does not exceed 1 Mb/s (into the network and out of the network to the customer site), the bandwidth and low delay are guaranteed. This is the point-to-cloud model because, for QoS purposes, the service provider need not keep track of traffic destinations, as long as the destinations are within the ICR and ECR bounds. (This model is also sometimes called the hose model).

With Cisco IOS and IOS XE and IOS XR MPLS, the QoS guarantees of a service provider can be transparent to customers. That is, a service provider can provide these guarantees in a nonintrusive way. Customer sites can deploy a consistent, end-to-end DiffServ implementation without having to adapt to a service provider QoS implementation. A service provider can prioritize traffic for a customer without remarking the DSCP field of the IP packet. A separate marking is used to provide QoS within the MPLS network, and it is discarded when the traffic leaves the MPLS domain. The QoS marking that is delivered to the destination network corresponds to the marking that is received when the traffic entered the MPLS network.

QoS-Enabled MPLS VPNs: Point-to-Point Service Model

This topic describes Point-to-Point Service Model in MPLS VPNs.



For the more stringent applications, where the customer desires a point-to-point guarantee, a virtual data pipe needs to be constructed to deliver the highly critical traffic.

For example, an enterprise may want two hub sites or data centers that are connected with high service level agreement guarantees. DiffServ-Aware Traffic Engineering (DS-TE) engages, automatically choosing a routing path that satisfies the bandwidth constraint for each service class that is defined. DS-TE also relieves the service provider from having to compute the appropriate path for each customer, and each service class per customer. This model is referred to as the point-to-point model (sometimes also called the pipe model).

MPLS DiffServ QoS Models

This topic lists the three MPLS DiffServ Modes.

DiffServ Tunneling Modes

- **Uniform mode:**
 - The EXP value is changed in the provider core.
 - At the egress PE, the subscriber DSCP or ToS field values are altered.
 - The subscriber will need to reset the original value on the customer edge (CE) device.
- **Pipe mode:**
 - The provider uses its own EXP values, including an egress PE-CE link, but does not alter subscriber DSCP or ToS values.
 - Subscribers receive traffic with their original DSCP or ToS marked value.
- **Short-pipe mode:**
 - The provider changes the EXP values in the core, but honors the subscriber DSCP or ToS values on the egress PE-CE link.
 - Subscribers receive traffic that is marked with the original DSCP or ToS value.

© 2012 Cisco and/or its affiliates. All rights reserved.SPCORE v1.01-3.11

In many instances, it is preferable for the service provider to maintain its own QoS service policies and customer service-level agreements (SLAs) without overriding the DSCP or IP Precedence values of the enterprise customer. MPLS can be used to tunnel the QoS markings of a packet and create QoS transparency for the customer. It is possible to mark the MPLS EXP field independently of the PHB marked in the IP Precedence or DSCP fields. A service provider may choose from an existing array of classification criteria, including or excluding the IP PHB marking, to classify those packets into a different PHB. The PHB behavior is then marked only in the MPLS EXP field during label imposition. This marking is useful to a service provider that requires SLA enforcement of the customer packets by promoting or demoting the PHB of a packet, without regard to the QoS marking scheme and without overwriting the IP PHB markings of the customer. The service provider SLA enforcement can be thought of in terms of adding a layer of PHB to a packet or encapsulating the PHB of the packet with a different tunnel PHB layer.

Some service providers re-mark packets at Layer 3 to indicate whether traffic is in contract or out-of-contract. Although this practice conforms to DiffServ standards, such as RFC 2597, it is not always desirable from the standpoint of the enterprise customer. Because MPLS labels include 3 bits that are commonly used for QoS marking, it is possible to tunnel DiffServ, that is, to preserve Layer 3 DiffServ markings through a service provider MPLS VPN cloud, while still performing re-marking (via MPLS EXP bits) within the cloud to indicate in- or out-of-contract traffic. RFC 3270 defines three distinct modes of MPLS DiffServ tunneling:

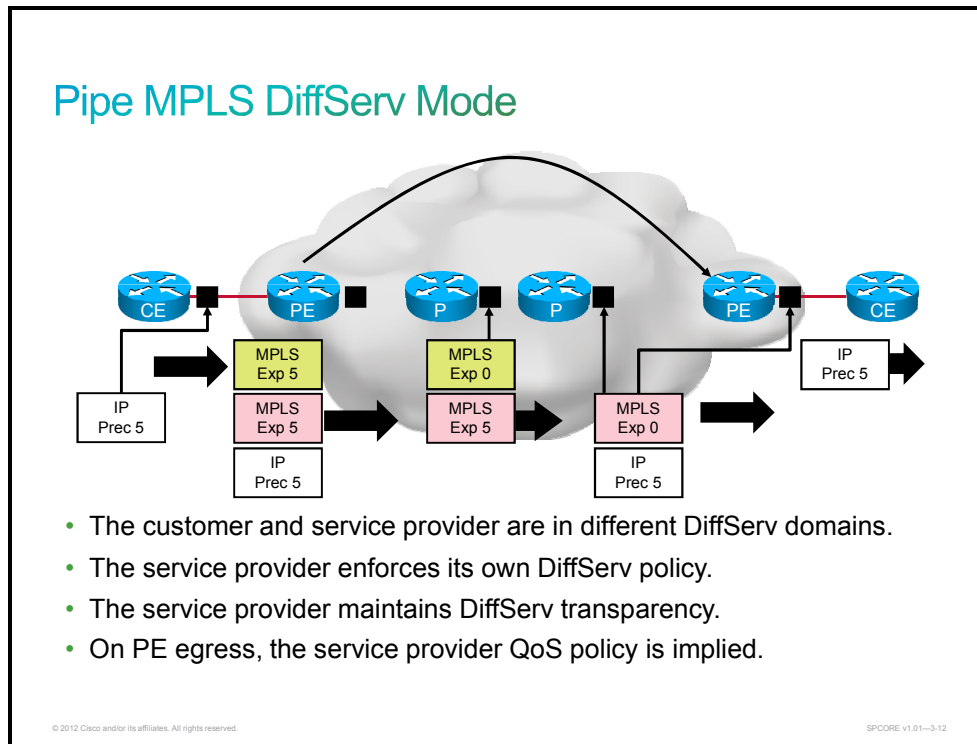
- Uniform mode
- Short-pipe mode
- Pipe mode

The default behavior of the DSCP MPLS EXP bits as a packet travels from one CE router to another CE router across an MPLS core is as follows:

- Imposition of the label (IP to label)
 - The IP precedence of the incoming IP packet is copied to the MPLS EXP bits of all pushed label(s).
 - The first three bits of the DSCP bit are copied to the MPLS EXP bits of all pushed labels.
 - This technique is also known as ToS reflection.
- MPLS forwarding (label to label)
 - The EXP is copied to the new labels that are swapped and pushed during forwarding or imposition.
 - At label imposition, the underlying labels are not modified with the value of the new label that is being added to the current label stack.
 - At label disposition, the EXP bits are not copied to the newly exposed label EXP bits.
- Disposition of the label (label to IP)
 - At label disposition, the EXP bits are not copied to the IP precedence or DSCP field of the newly exposed IP packet.

MPLS DiffServ Pipe Mode

This topic describes the MPLS DiffServ Pipe Mode.



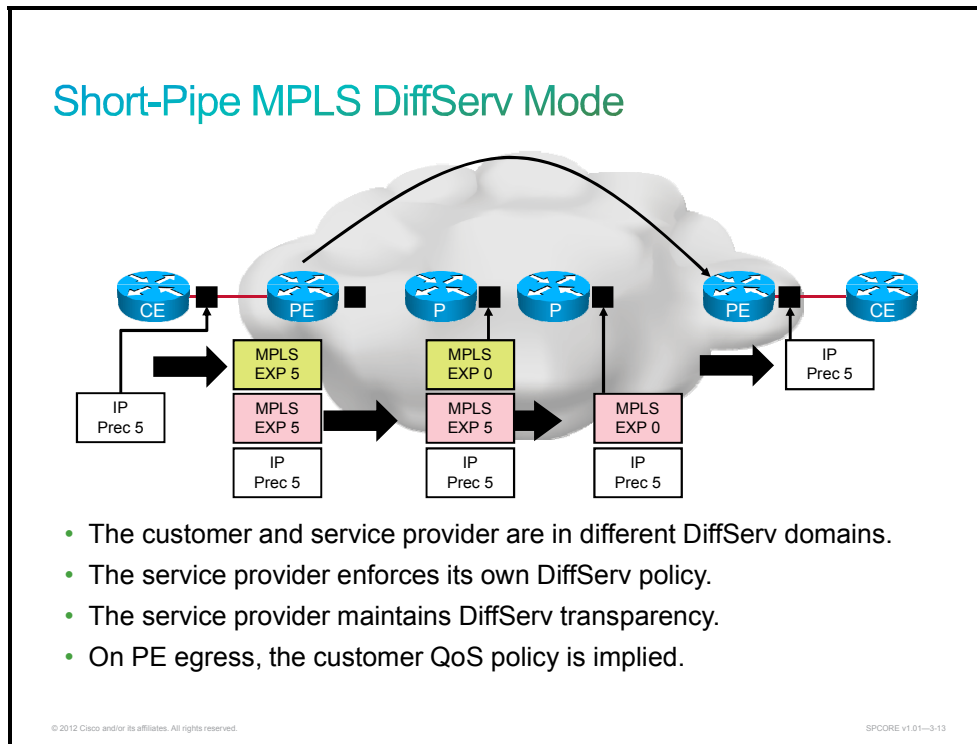
The pipe model conceals the tunneled PHB marking between the label-switched path (LSP) ingress and egress nodes. This model guarantees that there are no changes to the tunneled PHB marking through the LSP, even if a label switch router (LSR) along the path performs traffic conditioning and re-marks the traffic. All LSRs that the LSP traverses use the LSP PHB marking and ignore the tunneled PHB marking. This model proves useful when an MPLS network connects other DiffServ domains. The MPLS network can implement DiffServ and can also be transparent for the connected domains. RFC 3270 defines this model as mandatory for MPLS networks that are supporting DiffServ.

Pipe mode is very like short-pipe mode, since the customer and service provider are in different DiffServ domains. The difference between the two is that with pipe mode, the service provider derives the outbound classification for weighted random early detection (WRED) and weighted fair queuing (WFQ), based on the DiffServ policy of the service provider. This classification affects how the packet is scheduled on the egress PE before the label is popped. This implementation avoids the additional operational overhead of per-customer configurations on each egress interface on the egress PE.

When a packet reaches the edge of the MPLS core, the egress PE router classifies the newly exposed IP packets for outbound queuing, based on the MPLS PHB from the EXP bits of the recently removed label.

MPLS DiffServ Short-Pipe Mode

This topic describes the MPLS DiffServ Short-Pipe Mode.



The short-pipe model represents a small variation of the pipe model. The short-pipe model also guarantees that there are no changes to the tunneled PHB marking, even if an LSR re-marks the LSP PHB marking. The short-pipe model shares the same ability of the pipe model to allow an MPLS network to be transparent from the DiffServ point of view. The short-pipe model differs, however, on how the LSP egress infers the packet PHB. The LSP egress uses the tunneled PHB marking to infer the packet PHB and consequently, serve the packet. Given this difference between the short-pipe model and the pipe model, an MPLS network may implement LSPs using the short-pipe model, regardless of whether the LSRs perform penultimate hop-popping (PHP).

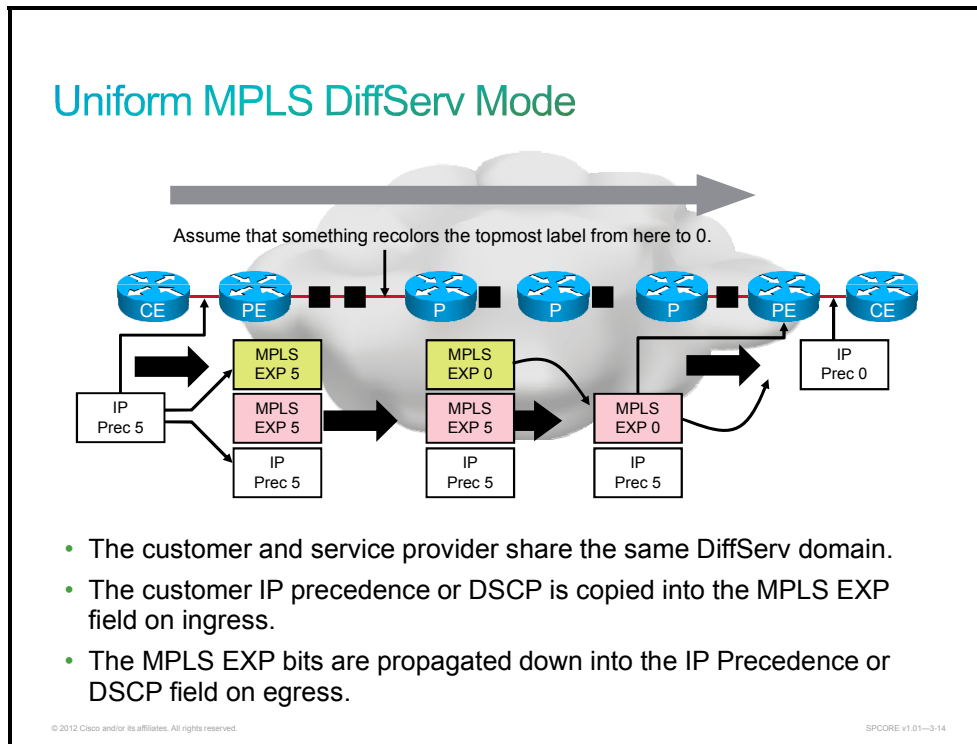
Short-pipe mode is used when the customer and service provider are in different DiffServ domains, which is typical. Short-pipe mode is useful when the service provider wants to enforce its own DiffServ policy, while maintaining DiffServ transparency. The outmost label is utilized as the single meaningful information source as it relates to the QoS PHB of the service provider. On MPLS label imposition, the IP classification is not copied into the EXP of the outermost label. Rather, based on the QoS policy of the service provider, an appropriate value for the MPLS EXP is set on the ingress PE. The MPLS EXP value could be different from the original IP precedence or the DSCP. The MPLS EXP will accomplish the class of service (CoS) marking on the topmost label, but preserve the underlying IP DSCP. If the service provider reclassifies the traffic in the MPLS cloud for any reason, the EXP value of the topmost label is changed. On egress of the service provider network, when the label is popped, the PE router will not affect the value of the underlying DSCP information. In this way, the MPLS EXP is not propagated to the DSCP field. Therefore, the DSCP transparency is maintained.

Note that the egress PE, in short-pipe mode, uses the original IP precedence or DSCP to classify the packet it sends to the enterprise network. The enterprise set the original IP precedence per its own QoS policy. The service provider may apply enterprise QoS policy at egress PE for traffic going towards the CE. In this example, the PE implements per-customer egress QoS policies for traffic towards the CE.

When a packet reaches the edge of the MPLS core, the egress PE router classifies the newly exposed IP packets for outbound queuing, based on the IP PHB from the DSCP value of this IP packet.

MPLS DiffServ Uniform Mode

This topic describes the MPLS DiffServ Uniform Mode.



The uniform model makes the LSP an extension of the DiffServ domain of the encapsulated packet. In this model, a packet has only a single meaningful PHB marking (which resides in the most recent encapsulation). LSRs propagate the packet PHB to the exposed encapsulation when they perform a pop operation. This propagation implies that any packet re-marking is reflected on the packet marking when it leaves the LSP. The LSP becomes an integral part of the DiffServ domain of the packet, unlike the transparent transport that the pipe and short-pipe models provided. This model proves useful when an MPLS network connects other DiffServ domain and all networks (including the MPLS network) need to behave as a single DiffServ domain.

Uniform mode is utilized when the customer and service provider share the same DiffServ domain. The outmost header is always used as the single meaningful information source about the QoS PHB. On MPLS label imposition, the IP precedence classification is copied into the outermost experimental field of the label. On egress of the service provider network, when the label is popped, the router propagates the EXP bits down into the IP precedence or the DSCP field.

So, if a P router in the service provider network changes the topmost EXP value, the changed EXP gets propagated to the original IP precedence or DSCP. The change could be the result of anything, downgrading traffic class or congestion, for example. This behavior results in the loss of QoS transparency and it is the default.

DiffServ tunneling uniform mode has only one layer of QoS, which reaches end to end. The ingress PE router copies the DSCP from the incoming IP packet into the MPLS EXP bits of the imposed labels. As the EXP bits travel through the core, they may or may not be modified by intermediate P routers. At the egress P router, the EXP bits are copied to the EXP bits of the newly exposed label, after the PHP. Finally, at the egress PE router, the EXP bits are copied to the DSCP bits of the newly exposed IP packet.

MPLS DS-TE

This topic describes MPLS DiffServ Traffic Engineering.

MPLS DS-TE

- Constraint-based routing selects a routing path, satisfying the service constraint.
- Bandwidth that is reservable on each link for CBR is managed through two pools:
 - Global pool (regular TE tunnel bandwidth)
 - Subpool (bandwidth for "guaranteed" traffic)
- There is a separate bandwidth reservation for different traffic classes.
- The subpool enables the PHB class bandwidth allocation.
- The link bandwidth is distributed in pools or with bandwidth constraints.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-3-15

MPLS TE allows constraint-based routing of IP traffic. One of the constraints that is satisfied by constraint-based routing (CBR) is the availability of the required bandwidth over a selected path. DS-TE extends MPLS traffic engineering to enable you to perform constraint-based routing for guaranteed traffic, which satisfies a more restrictive bandwidth constraint than one that satisfies by CBR for regular traffic. The more restrictive bandwidth is termed a subpool, while the regular TE tunnel bandwidth is called the global pool. (The subpool is a portion of the global pool.) This ability to satisfy a more restrictive bandwidth constraint translates into an ability to achieve higher QoS performance (in terms of delay, jitter, or loss) for the guaranteed traffic.

For example, DS-TE can be used to ensure that traffic is routed over the network so that, on every link, no more than 40 percent (or any assigned percentage) of the link capacity is reserved for guaranteed traffic (for example, voice), while there can be up to 100 percent of the link capacity reserved for regular traffic. Assuming that QoS mechanisms are also used on every link to queue guaranteed traffic separately from regular traffic, it then becomes possible to enforce separate “overbooking” ratios for guaranteed and regular traffic. (In fact, for the guaranteed traffic it becomes possible to enforce no overbooking at all—or even an underbooking—so that very high QoS can be achieved end-to-end for that traffic, even while for the regular traffic a significant overbooking continues to be enforced.)

Also, the ability to enforce a maximum percentage of guaranteed traffic on any link enables the network administrator to directly control the end-to-end QoS performance parameters without having to rely on overengineering or on expected shortest path routing behavior. This ability is essential for transport of applications that have very high QoS requirements (such as real-time voice, a virtual IP leased line, or bandwidth trading), where overengineering cannot be assumed everywhere in the network.

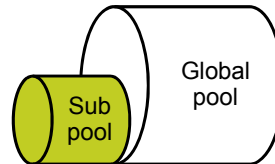
Cisco IOS XR DS-TE Implementation

- Two types of tunnels can be defined:
 - Data tunnels that are constrained by the RSVP global pool bandwidth
 - Voice tunnels that are only constrained by the RSVP subpool bandwidth
- The entire global pool is available when a subpool is not used.

The **bandwidth** command sets the available RSVP bandwidth that is available on the interface.

The **signaled bandwidth** command sets the available bandwidth for the subpool or the global pool of the LSP.

```
rsvp interface pos0/6/0/0
  bandwidth 100 150 sub-pool 50
interface tunnel-te2
  signaled bandwidth 20
interface tunnel-te1
  signaled bandwidth sub-pool 10
```



© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-3-18

MPLS DS-TE enables per-class TE across an MPLS network. DS-TE provides more granular control to minimize network congestion and improve network performance. DS-TE retains the same overall operation framework of MPLS TE (link information distribution, path computation, signaling, and traffic selection). However, it introduces extensions to support the concept of multiple classes and to make per-class constraint-based routing possible.

DS-TE must keep track of the available bandwidth for each class of traffic. For this reason, class types are defined. TE LSPs can have different preemption priorities, regardless of their class type. Class types represent the concept of a class for DS-TE in a similar way that PHB scheduling class (PSC) represents it for DiffServ. Note that flexible mappings between class types and PSCs are possible. You can define a one-to-one mapping between class types and PSCs. Alternatively, a class type can map to several PSCs, or several class types can map to one PSC.

Suppose a network supports voice and data traffic with voice being EF PHB (EF queue) and data being best-effort (BE queue), class type 1 (CT1) can be mapped to the EF queue, while CT0 can be mapped to the BE queue. Separate TE LSPs are established with separate bandwidth requirements from CT0 and from CT1.

All aggregate (known as **bandwidth global pool**) MPLS TE traffic is mapped to CT0 by default. Cisco allows only two class types to be defined—CT0 and CT1. CT1 is known as **bandwidth subpool**.

To configure basic DS-TE on Cisco IOS XR Software, use the commands that are described in the table.

Cisco IOS XR Commands

Command	Description
rsvp interface <i>type interface-path-id</i> Example: RP/0/RP0/CPU0:router(config)# rsvp interface pos0/6/0/0	Enters RSVP configuration mode and selects an RSVP interface.
bandwidth [<i>total reservable bandwidth</i>] [bc0 <i>bandwidth</i>] [global-pool <i>bandwidth</i>] [sub-pool <i>reservable-bw</i>] Example: RP/0/RP0/CPU0:router(config-rsvp-if)# bandwidth 100 150 subpool 50	Sets the reserved RSVP bandwidth that is available on this interface by using the prestandard DS-TE mode. The range for the <i>total reservable bandwidth</i> argument is 0 to 4294967295.
interface tunnel-te <i>tunnel-id</i> Example: RP/0/RP0/CPU0:router(config)# interface tunnel-te 2	Configures an MPLS-TE tunnel interface.
signalled-bandwidth { <i>bandwidth</i> [class-type <i>ct</i>] <i>subpool bandwidth</i> } Example: RP/0/RP0/CPU0:router(config-if)# signalled-bandwidth subpool 10	Sets the bandwidth that is required on this interface. Because the default tunnel priority is 7, tunnels use the default TE class map (class type 1, priority 7).

To configure basic DS-TE on Cisco IOS and IOS XE Software, use the commands that are described in the table.

Cisco IOS and IOS XE Commands

Command	Description
Router(config)# interface <i>interface-id</i> Example: Router(config)# interface FastEthernet 0/1	Moves configuration to the interface level, directing subsequent configuration commands to the specific interface that is identified by the interface ID.
Router(config-if)# ip rsvp bandwidth <i>interface-kbps single-flow-kbps</i> [subpool <i>kbps</i>] Example: Router(config-if)# ip rsvp bandwidth 150000 subpool 45000	Enables RSVP for IP on an interface and specifies the amount of interface bandwidth (in kb/s) allocated for RSVP flows (for example, TE tunnels)
Router(config)# interface tunnel-if <i>num</i>	Enters tunnel interface configuration mode.

Command	Description
Example: Router(config)# interface tunnel 1	
Router(config-if)# tunnel mpls traffic-eng bandwidth [<i>kbps</i> subpool <i>kbps</i>] Example: Router(config-if)# tunnel mpls traffic-eng bandwidth subpool 50000	Indicates that the tunnel should use bandwidth from the subpool or the global pool.

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- MPLS packets need to carry the packet marking in their headers because label switch routers (LSRs) do not examine the IP header during forwarding
- Marking is done with the MPLS experimental (EXP) bit value
- A QoS group is an internal label that is used by the switch or the router to identify packets as a member of a specific class
- CE router generally uses some congestion avoidance and congestion management mechanisms to protect high-priority traffic from being dropped
- Classification of ingress traffic on a P router is based on MPLS EXP bits
- Monitoring MPLS QoS on a specific router is based on observing the statistics about congestion on that specific interface
- When the customer desires a point-to-point guarantee, a virtual data pipe needs to be constructed to deliver the highly critical traffic

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-3-17

Summary (Cont.)

- In many instances, it is preferable for the service provider to maintain its own QoS service policies and customer service-level agreements (SLAs) without overriding the DSCP or IP Precedence values of the enterprise customer
- The pipe model conceals the tunneled PHB marking between the label-switched path (LSP) ingress and egress nodes
- In short-pipe mode egress PE uses the original IP precedence or DSCP to classify the packet it sends to the enterprise network
- Uniform mode is utilized when the customer and service provider share the same DiffServ domain
- MPLS DS-TE enables per-class TE across an MPLS network

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01-3-18

Module Summary

This topic summarizes the key points that were discussed in this module.

Module Summary

- Two QoS architectures were defined for IP: IntServ (provides granular QoS guarantees with explicit resource reservation) and DiffServ (provides a QoS approach based on aggregates, or classes, of traffic).
- MQC provides the user interface to the QoS behavioral mode. Three commands define configuration components: **class-map**, **policy-map**, and **service-policy**.
- Depending on the DiffServ domains that are wanted and from which header the PHB marking is derived, there are three DiffServ tunneling modes: pipe, short pipe and uniform.

© 2012 Cisco and/or its affiliates. All rights reserved.

SPCORE v1.01--3-1

Two quality of service (QoS) architectures have been defined for IP: integrated services (IntServ) and differentiated services (DiffServ). IntServ provides granular QoS guarantees with explicit resource reservation. IntServ uses Resource Reservation Protocol (RSVP) as a signaling protocol. DiffServ provides a coarse QoS approach based on aggregates (classes) of traffic. Cisco QoS uses a behavioral model that abstracts the QoS implementation details.

The Modular QoS CLI (MQC) provides the user interface for the QoS behavioral model. Three commands define the configuration components: **class-map**, **policy-map**, and **service-policy**. The **class-map** commands control traffic classification and correspond to the classification component of the Telecommunications Management Network (TMN). The **policy-map** command defines a policy template that groups QoS actions (including marking, policing, shaping, congestion management, active queue management, and so on). The **service-policy** command instantiates a previously defined QoS policy and defines its direction. The MQC provides a template-based, hardware-independent configuration model for QoS across different Cisco platforms.

Depending on the DiffServ domains that are wanted and from which header the PHB marking is derived, there are three DiffServ tunneling modes: pipe, short pipe, and uniform.

Multiprotocol Label Switching (MPLS) does not define new QoS architectures. Currently, MPLS provides support for DiffServ. MPLS does not introduce any modifications to the traffic-conditioning and PHB concepts that are defined in DiffServ. A label switch router (LSR) uses the same traffic-management mechanisms (metering, marking, shaping, policing, queuing, and so on) to condition and implement the different PHBs for MPLS traffic. An MPLS network may use traffic engineering (TE) to complement its DiffServ implementation.

An MPLS network may implement DiffServ to support a diverse range of QoS requirements and services in a scalable manner. MPLS DiffServ is not specific to the transport of IP traffic over an MPLS network. An MPLS DiffServ implementation is concerned only with supporting the PHBs that can satisfy the QoS requirements of all the types of traffic it carries. In addition, an MPLS network can grow without having to introduce major changes to its DiffServ design as the number of label switched paths (LSPs) in the network increases. These characteristics play an important role in the implementation of large MPLS networks that can transport a wide spectrum of traffic.

MPLS provides native TE capabilities that can improve network efficiency and service guarantees. These MPLS TE capabilities bring explicit routing, constraint-based routing (CBR), and bandwidth reservation to MPLS networks.

Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

- Q1) How much one-way delay can a voice packet tolerate? (Source: Understanding QoS)
- A) 15 ms
 - B) 150 ms
 - C) 300 ms
 - D) 200 ms
- Q2) Which three options are advantages of using MQC? (Choose three.) (Source: Understanding QoS)
- A) reduction in time to configure a complex policy
 - B) ability to apply one policy to multiple interfaces
 - C) separation of classification from policy definition
 - D) automatic generation of CLI commands from MQC macros
- Q3) How many bits constitute the DSCP field of the IP header? (Source: Understanding QoS)
- A) 3
 - B) 4
 - C) 6
 - D) 8
- Q4) What is the binary representation of the DSCP value of EF? (Source: Understanding QoS)
-
- Q5) Which QoS mechanism is used on both input and output interfaces? (Source: Implementing Cisco QoS and QoS Mechanisms)
- A) classification
 - B) traffic policing
 - C) traffic shaping
 - D) congestion management
- Q6) The QoS requirements on the CE and PE routers differ depending on which factor? (Source: Implementing Cisco QoS and QoS Mechanisms)
- A) whether the PE router is managed by the service provider
 - B) whether the CE router is managed by the service provider
 - C) whether the service provider is using an MPLS core
 - D) the number of traffic classes that are supported by the service provider
- Q7) Which option is a Layer 2 QoS marking? (Source: Implementing Cisco QoS and QoS Mechanisms)
- A) CoS
 - B) DSCP
 - C) EXP
 - D) QoS group

- Q8) Which option is a congestion-avoidance mechanism? (Source: Implementing Cisco QoS and QoS Mechanisms)
- A) LFI
 - B) QPM
 - C) MRF
 - D) WRED
- Q9) Which two QoS mechanisms are used in the service provider core on P routers? (Choose two.) (Source: Implementing Cisco QoS and QoS Mechanisms)
- A) policing
 - B) marking
 - C) queuing
 - D) dropping
- Q10) What is the purpose of the QoS group on Cisco switches and routers? (Source: Implementing MPLS Support for QoS)
-
- Q11) Which command is used to display the statistics of applied QoS policy on an interface? (Source: Implementing MPLS Support for QoS)
-
- Q12) “ECR” refers to the traffic rate that is given a particular treatment from the service provider to the customer site. (Source: Implementing MPLS Support for QoS)
- A) true
 - B) false

Module Self-Check Answer Key

- Q1) B
- Q2) A, B, C
- Q3) C
- Q4) 101110
- Q5) B
- Q6) B
- Q7) A
- Q8) D
- Q9) C, D
- Q10) A QoS group is an internal label that is used by the switch or router to identify packets as members of a specific class.
- Q11) **show policy-map interface** *interface_number*
- Q12) A

