Juniper NETWORKS

Traffic Engineering with MPLS

Dan Lockwood
Chris Summers

Agenda

- Introduction to traffic engineering
  - Brief history
  - Vocabulary
  - Requirements for Traffic Engineering
  - Basic Examples
- Signaling LSPs with RSVP
  - RSVP signaling protocol
  - RSVP objects
  - Extensions to RSVP

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 2

Agenda

- Constraint-based traffic engineering
  - Extensions to IS-IS and OSPF
  - Traffic Engineering Database
  - User defined constraints
  - Path section using CSPF algorithm
- Traffic protection
  - Secondary LSPs
  - Hot-standby LSPs
  - Fast Reroute

Agenda

- Advanced traffic engineering features
  - Circuit cross connect (CCC)
  - IGP Shortcuts
  - Configuring for transit traffic
  - Configuring for internal destinations

Introduction to Traffic Engineering

Why Engineer Traffic?

What problem are we trying to solve with Traffic Engineering?

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

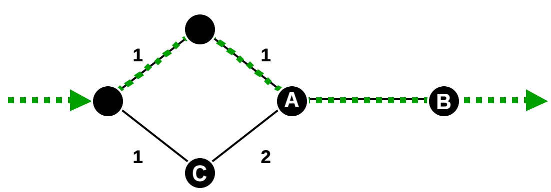
Slide 6

Brief History

- Early 1990's
  - Internet core was connected with T1 and T3 links between routers
  - Only a handful of routers and links to manage and configure
  - Humans could do the work manually
  - IGP (Interior Gateway Protocol) Metric-based traffic control was sufficient

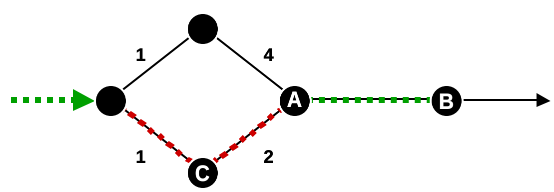
IGP Metric-Based Traffic Engineering

- Traffic sent to A or B follows path with lowest metrics



IGP Metric-Based Traffic Engineering

- **Drawbacks**
  - Redirecting traffic flow to A via C causes traffic for B to move also!
  - Some links become underutilized or overutilized



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 9

Metric-Based Traffic Engineering

- **Drawbacks**
  - Only serves to move problem around
    - Some links underutilized
    - Some links overutilized
  - Lacks granularity
    - All traffic follows the IGP shortest path
  - Continuously adjusting IGP metrics adds instability to the network

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 10

Discomfort Grows

- Mid 1990's
  - ISPs became uncomfortable with size of Internet core
  - Large growth spurt imminent
  - Routers too slow
  - IGP metric engineering too complex
  - IGP routing calculation was topology driven, not traffic driven
  - Router based cores lacked predictability

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 11

Why Traffic Engineering?

- There is a need for a more granular and deterministic solution

"A major goal of Internet Traffic Engineering is to facilitate efficient and reliable network operations while simultaneously optimizing network resource utilization and performance."

RFC 2702
Requirements for Traffic Engineering over MPLS

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 12

Overlay Networks are Born

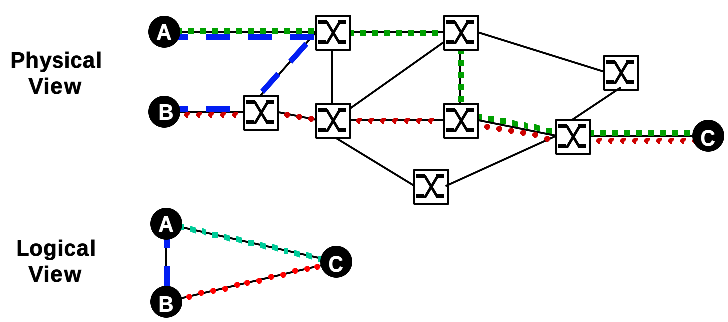
- ATM switches offered performance and predictable behavior
- ISPs created "overlay" networks that presented a virtual topology to the edge routers in their network
- Using ATM virtual circuits, the virtual network could be reengineered without changing the physical network
- Benefits
  - Full traffic control
  - Per-circuit statistics
  - More balanced flow of traffic across links

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 13

Overlay Networks

- ATM core ringed by routers
- PVCs overlaid onto physical network



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 14

Path Creation

- Off-line path calculation tool uses
  - Link utilization
  - Historic traffic patterns
- Produces virtual network topology
  - Primary and backup PVCs
- Generates switch and router configurations

Overlay Network Drawbacks

- Growth in full mesh of ATM PVCs stresses everything
  - With 5 routers, adding 1 requires only 10 new PVCs
  - With 200 routers, adding 1 requires 400 new PVCs
    - From 39,800 to 40,200 PVCs total
  - Router IGP runs out of steam
  - Practical limitation of atomically updating configurations in each switch and router
- Not well integrated
  - Network does not participate in path selection and setup

Overlay Network Drawbacks

- ATM cell overhead
  - Approximately 20% of bandwidth
  - OC-48 link wastes 498 Mbps in ATM cell overhead
  - OC-192 link wastes 1.99 Gbps
- ATM SAR speed
  - OC-48 SAR
    - Trailing behind the router curve
    - Very difficult to build
  - OC-192 SAR?

Slide 17

Routers Caught Up

- Current generation of routers have
  - High speed, wire-rate interfaces
  - Deterministic performance
  - Software advances
- Solution
  - Fuse best aspects of ATM PVCs with high-performance routing engines
  - Use low-overhead circuit mechanism
  - Automate path selection and configuration
  - Implement quick failure recovery

Slide 18

Benefits of MPLS

- Low-overhead virtual circuits for IP
  - Originally designed to make routers faster
    - Fixed label lookup faster than longest match used by IP routing
  - Not true anymore!
- Value of MPLS is now in traffic engineering
- One, integrated network
- Same forwarding mechanism can support multiple applications
  - Traffic Engineering, VPNs, etc....

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 19

What are the fundamental requirements?

- RFC 2702
  - Requirement for Traffic Engineering over MPLS
- Requirements
  - Control
  - Measure
  - Characterize
  - Integrate routing and switching
  - All at a lower cost

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 20

Fundamental Requirements

- Need the ability to:
  - Map traffic to an LSP
  - Monitor and measure traffic
  - Specify explicit path of an LSP
    - Partial explicit route
    - Full explicit route
  - Characterize an LSP
    - Bandwidth
    - Priority/ Preemption
    - Affinity (Link Colors)
  - Reroute or select an alternate LSP

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

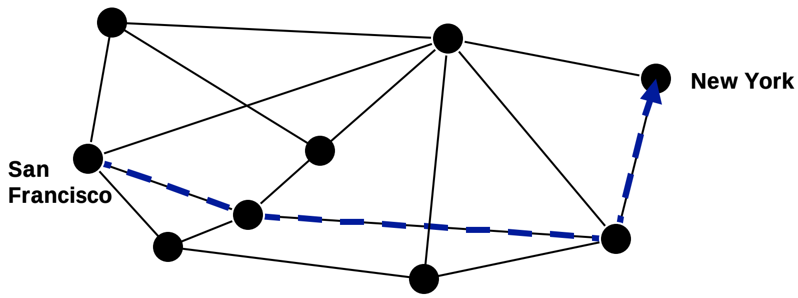
Slide 21

Juniper NETWORKS

MPLS Fundamentals

MPLS Vocabulary

- Label-switched path (LSP)
  - Unidirectional path through interior network

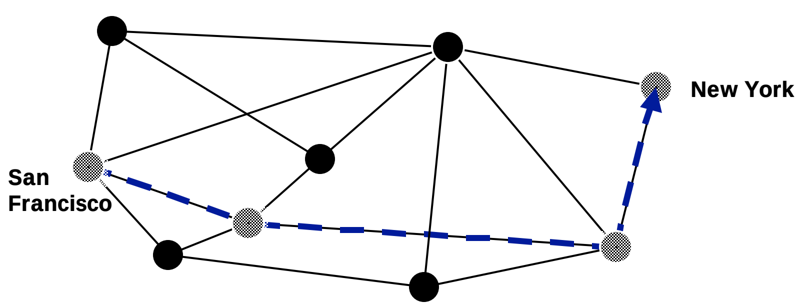


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 23

MPLS Vocabulary

- Label-switching router (LSR) performs
  - MPLS packet forwarding
  - LSP setup
  - In this course, “router” implies LSR capability

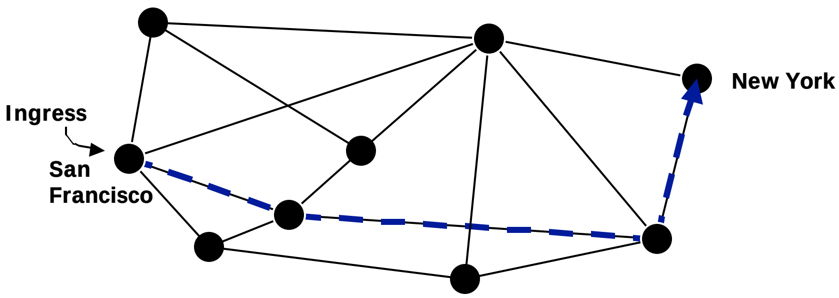


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 24

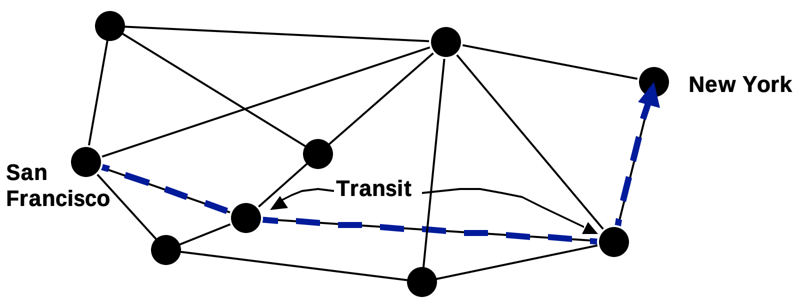
MPLS Vocabulary

- Ingress router
  - Packets enter LSP here
  - Also called a “head-end” router
  - Upstream from other routers



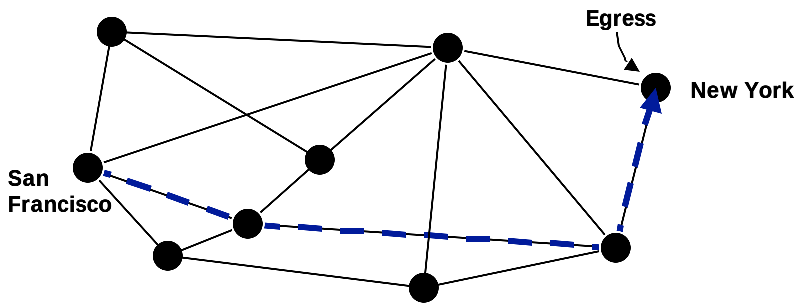
MPLS Vocabulary

- Transit router
  - Zero or more transit routers
  - Swaps MPLS label
  - Sends traffic to next hop in LSP



MPLS Vocabulary

- Egress router
  - Packets exit LSP here
  - Also called “tail-end” router
  - Downstream from other routers

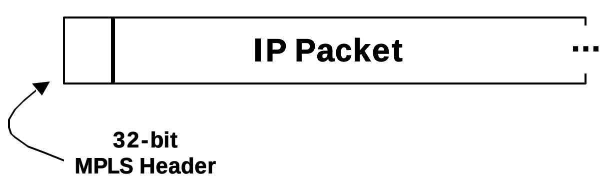


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 27

MPLS Header

- IP packet is encapsulated in MPLS header and sent down LSP

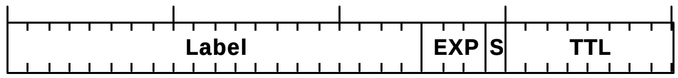


- IP packet is restored at end of LSP by egress router
  - TTL is adjusted by default

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 28

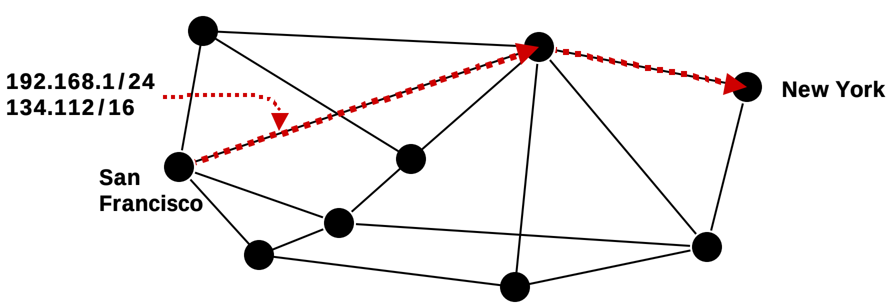
MPLS Header



- Label
  - Used to match packet to LSP
- Experimental bits
  - Carries packet queuing priority (CoS)
- Stacking bit
- Time to live
  - Copied from IP TTL

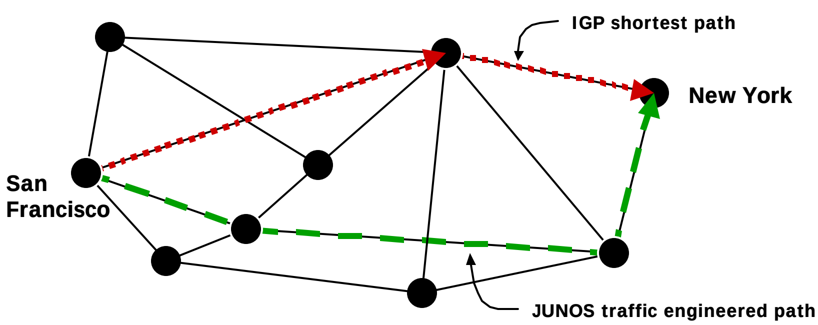
Router Based Traffic Engineering

- Standard IGP routing
- IP prefixes bound to physical next hop
  - Typically based on IGP calculation



Router Based Traffic Engineering

◆ Engineer unidirectional paths through your network without using the IGP's shortest path calculation

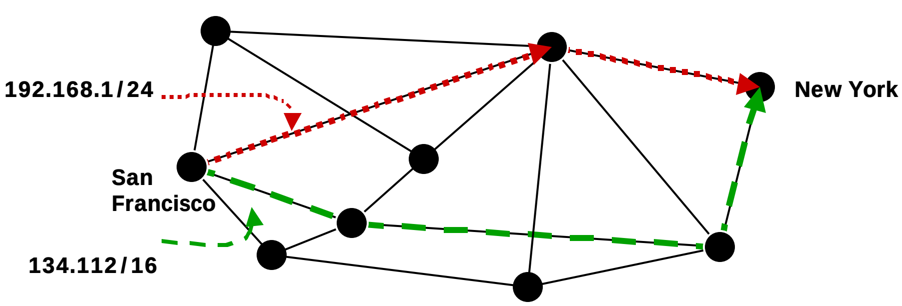


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 31

Router Based Traffic Engineering

◆ IP prefixes can now be bound to LSPs



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 32

MPLS Labels

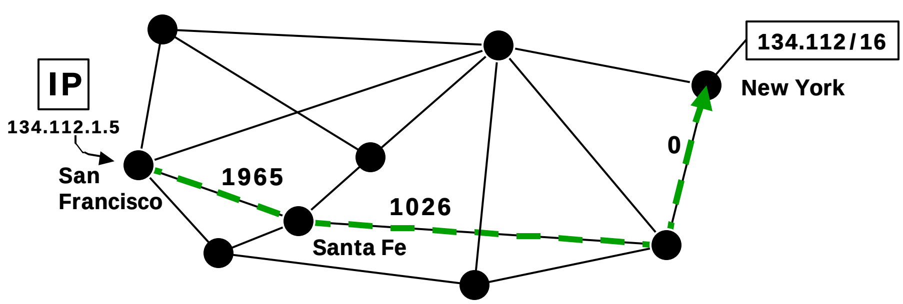
- Assigned manually or by a signaling protocol in each LSR during path setup
- Labels change at each segment in path
- LSR swaps incoming label with new outgoing label
- Labels have “local significance”

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 33

JUNOS MPLS Example

- An IP packet destined to 134.112.1.5/32 arrives in SF
- San Francisco has route for 134.112/16
  - Next hop is the LSP to New York

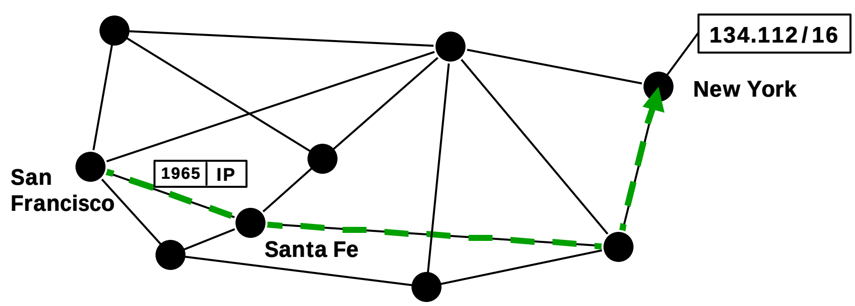


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 34

JUNOS MPLS Example

- San Francisco prepends MPLS header onto IP packet and sends packet to first transit router in the path

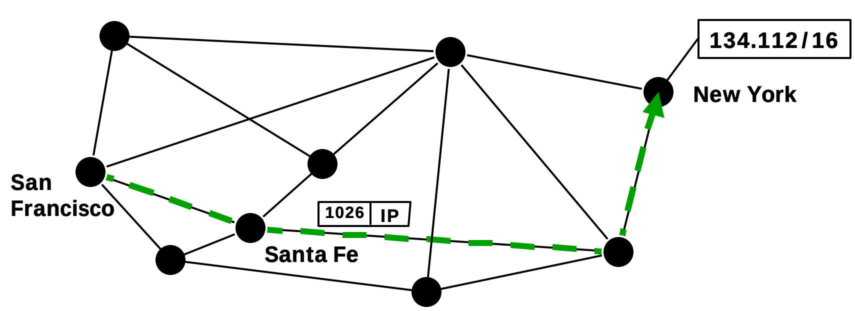


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 35

JUNOS MPLS Example

- Because the packet arrived at Santa Fe with an MPLS header, Santa Fe forwards it using the MPLS forwarding table
- MPLS forwarding table derived from mpls.0 switching table

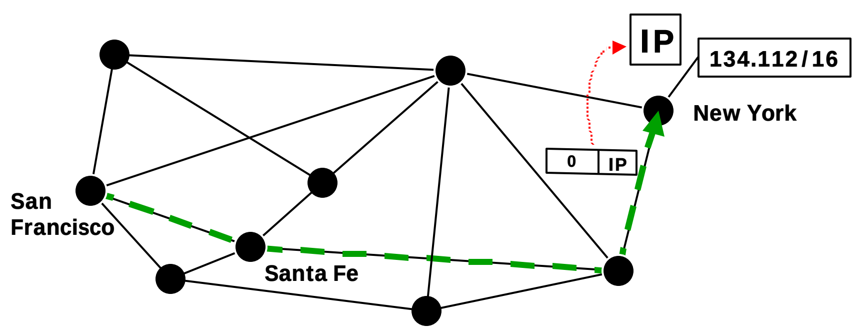


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 36

JUNOS MPLS Example

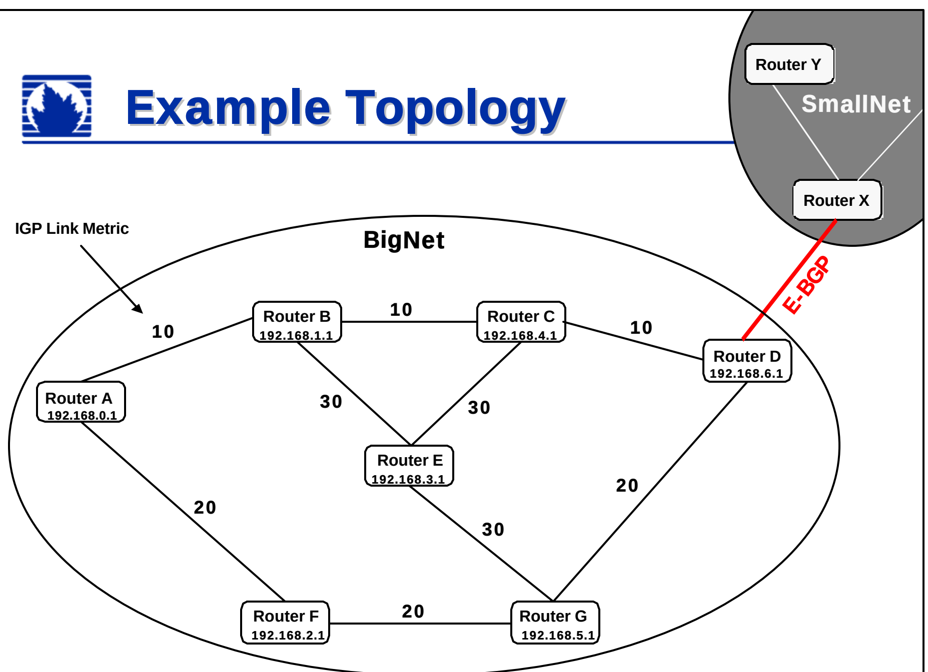
- Packet arrives from penultimate router with label 0
- Egress router sees label 0 and strips MPLS header
- Egress router performs standard IP forwarding decision



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

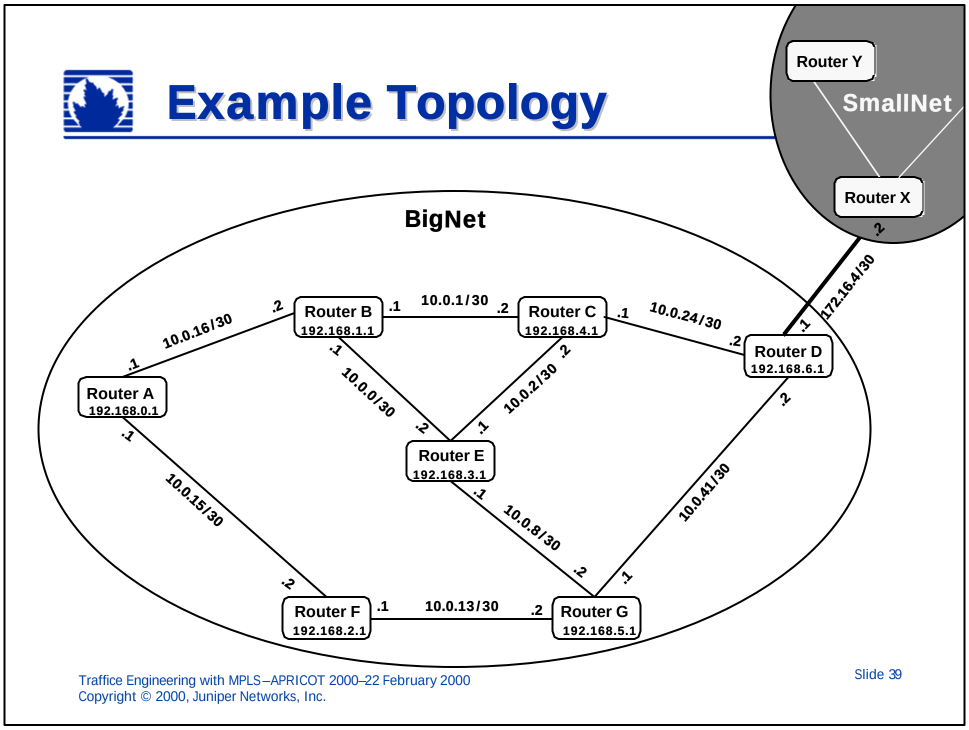
Slide 37

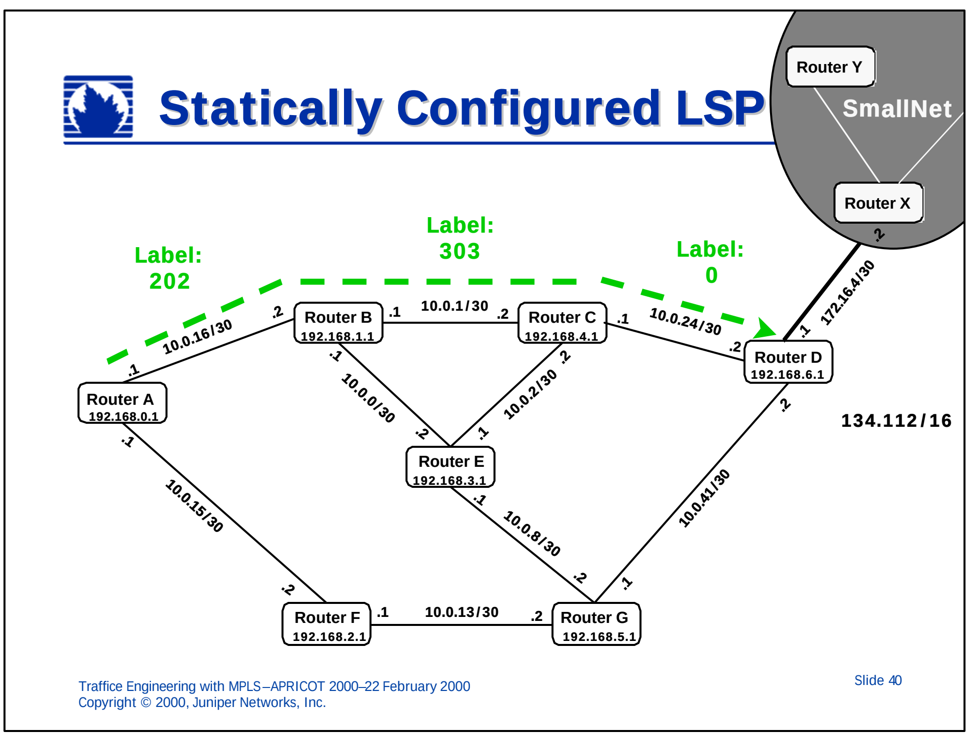
Example Topology



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 38


Example Topology
Router Y
SmallNet
Router X
BigNet
Router A
192.168.0.1
Router B
192.168.1.1
Router C
192.168.4.1
Router D
192.168.6.1
Router E
192.168.3.1
Router F
192.168.2.1
Router G
192.168.5.1
10.0.16/30
10.0.1/30
10.0.24/30
172.16.4/30
10.0.0/30
10.0.2/30
10.0.15/30
10.0.8/30
10.0.41/30
10.0.13/30
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.
Slide 39


Statically Configured LSP
Router Y
SmallNet
Router X
Label: 202
Label: 303
Label: 0
Router A
192.168.0.1
Router B
192.168.1.1
Router C
192.168.4.1
Router D
192.168.6.1
Router E
192.168.3.1
Router F
192.168.2.1
Router G
192.168.5.1
134.112/16
10.0.16/30
10.0.1/30
10.0.24/30
172.16.4/30
10.0.0/30
10.0.2/30
10.0.15/30
10.0.8/30
10.0.41/30
10.0.13/30
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.
Slide 40

Juniper NETWORKS

Traffic Engineering
Signaled LSPs

Static vs Signaled LSPs

- Static LSPs
  - Are 'nailed up' manually
  - Have manually assigned MPLS labels
  - Needs configuration on each router
- Signaled LSPs
  - Signaled by RSVP
  - Have dynamically assigned MPLS labels
  - Configured on ingress router only

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 42

Signaled Label-Switched Paths

- Configured at ingress router only
  - RSVP sets up transit and egress routers automatically
  - Path through network chosen at each hop using routing table
  - Intermediate hops can be specified as “transit points”
    - Strict—Must use hop, must be directly connected
    - Loose—Must use hop, but use routing table to find it
- Advantages over static paths
  - Performs “keepalive” checking
  - Supports fail-over to unlimited secondary LSPs
  - Excellent visibility

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 43

RSVP Path Signaling

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.

Slide 44

Path Signaling

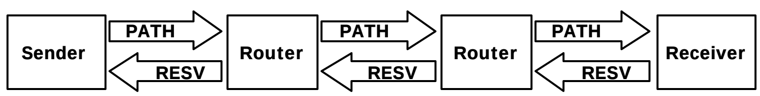
- JUNOS uses RSVP for Traffic Engineering
  - Internet standard for reserving resources
  - Extended to support
    - Explicit path configuration
    - Path numbering
    - Route recording
  - Provides keepalive status
    - For visibility
    - For redundancy

RSVP

- A generic QoS signaling protocol
- An Internet control protocol
  - Uses IP as its network layer
- Originally designed for host-to-host
- Uses the IGP to determine paths
- RSVP is not
  - A data transport protocol
  - A routing protocol
- RFC 2205

Basic RSVP Path Signaling

- Simplex flows
- Ingress router initiates connection
- "Soft" state
  - Path and resources are maintained dynamically
  - Can change during the life of the RSVP session
- Path message sent downstream
- Resv message sent upstream



Other RSVP Message Types

- PathTear
  - Sent to egress router
- ResvTear
  - Sent to ingress router
- PathErr
  - Sent to ingress router
- ResvErr
  - Sent to egress router
- ResvConf

Extended RSVP

- Extensions added to support establishment and maintenance of LSPs
  - Maintained via “hello” protocol
- Used now for router-to-router connectivity
- Includes the distribution of MPLS labels

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 49

MPLS Extensions to RSVP

- Path and Resv message objects
  - Explicit Route Object (ERO)
  - Label Request Object
  - Label Object
  - Record Route Object
  - Session Attribute Object
  - Tspec Object
- For more detail on contents of objects:

  daft-ietf-mpls-rsvp-lsp-tunnel-04.txt

  Extensions to RSVP for LSP Tunnels

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 50

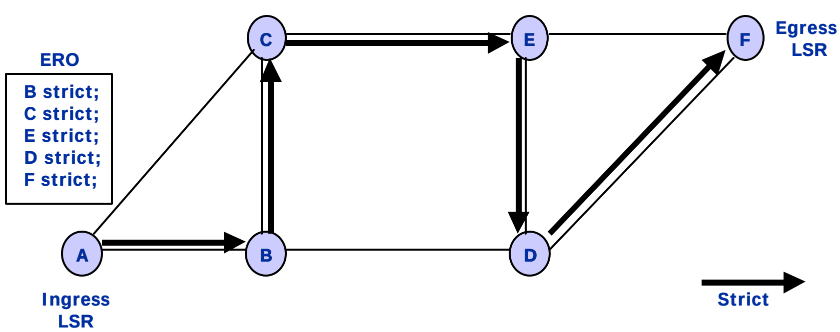
Explicit Route Object

- Used to specify the route RSVP Path messages take for setting up LSP
- Can specify loose or strict routes
  - Loose routes rely on routing table to find destination
  - Strict routes specify the directly-connected next hop
- A route can have both loose and strict components

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000 Slide 51

ERO: Strict Route

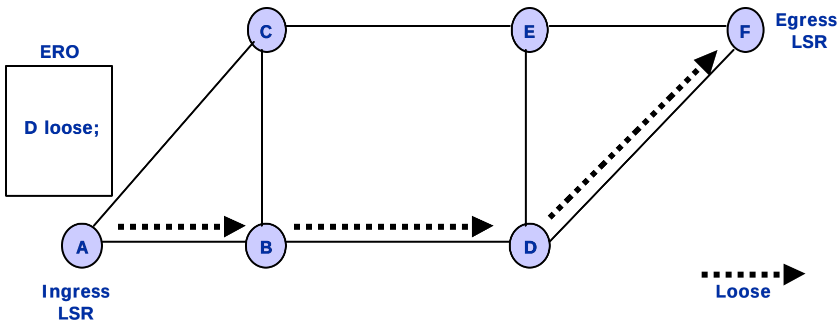
- Next hop must be directly connected to previous hop



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000 Slide 52

ERO: Loose Route

- Consult the routing table at each hop to determine the best path

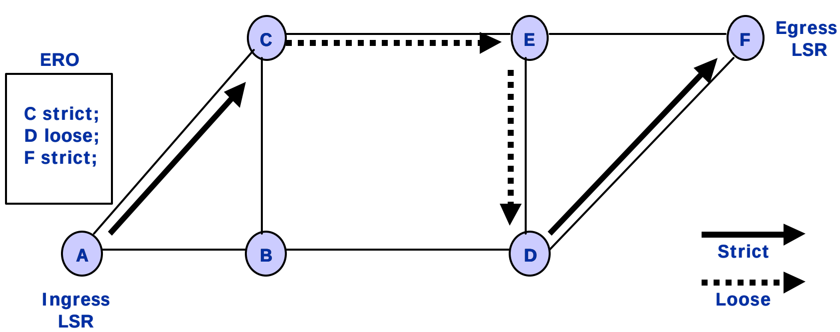


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 53

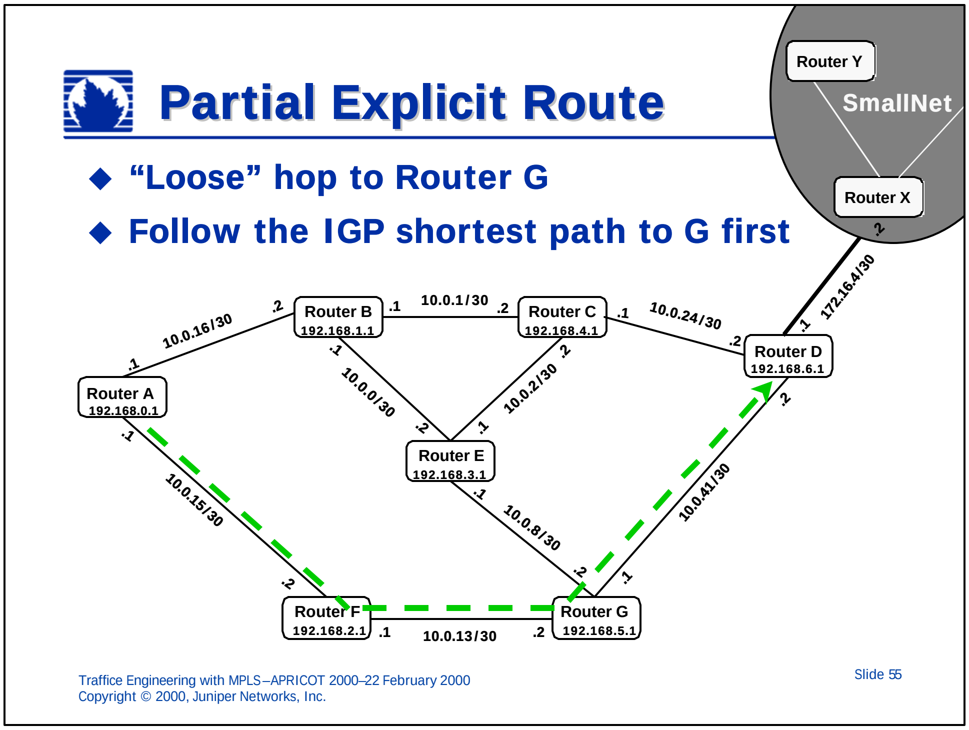
ERO: Strict / Loose Path

- Strict and loose routes can be mixed

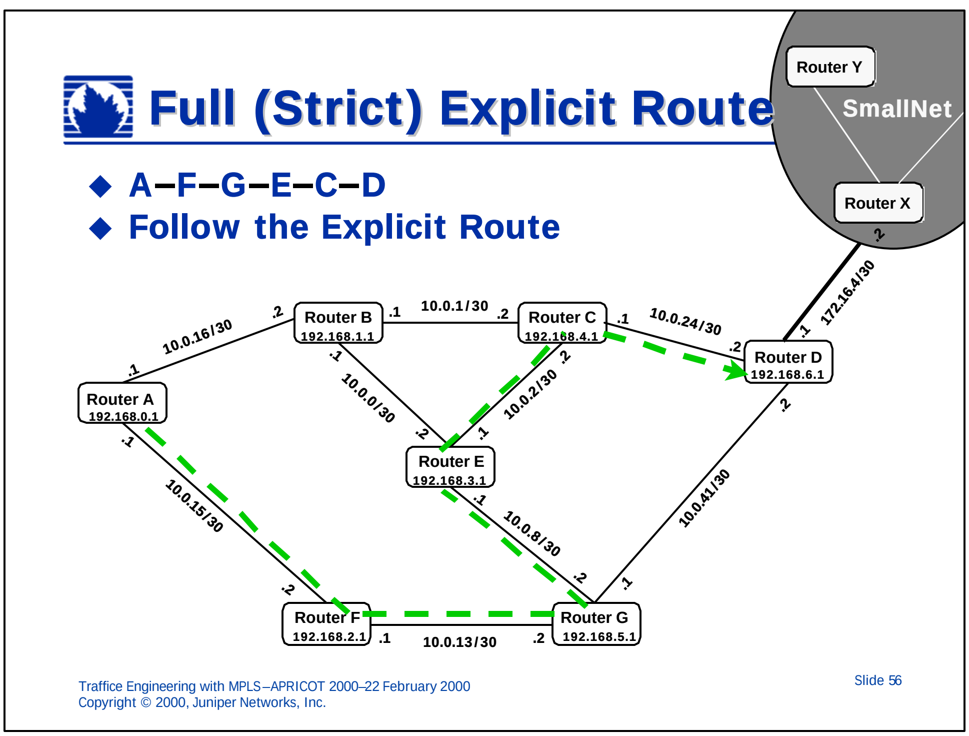


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 54



Partial Explicit Route
"Loose" hop to Router G
Follow the IGP shortest path to G first
Router Y
SmallNet
Router X
172.16.4/30
Router A
192.168.0.1
Router B
192.168.1.1
Router C
192.168.4.1
Router D
192.168.6.1
Router E
192.168.3.1
Router F
192.168.2.1
Router G
192.168.5.1
10.0.16/30
10.0.1/30
10.0.24/30
10.0.0/30
10.0.2/30
10.0.15/30
10.0.8/30
10.0.41/30
10.0.13/30
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.
Slide 55



Full (Strict) Explicit Route
A–F–G–E–C–D
Follow the Explicit Route
Router Y
SmallNet
Router X
172.16.4/30
Router A
192.168.0.1
Router B
192.168.1.1
Router C
192.168.4.1
Router D
192.168.6.1
Router E
192.168.3.1
Router F
192.168.2.1
Router G
192.168.5.1
10.0.16/30
10.0.1/30
10.0.24/30
10.0.0/30
10.0.2/30
10.0.15/30
10.0.8/30
10.0.41/30
10.0.13/30
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.
Slide 56

Hop-by-Hop ERO Processing

- If Destination Address of RSVP message belongs to your router
  - You are the egress router
  - End ERO processing
  - Send RESV message along reverse path to ingress
- Otherwise, examine next object in ERO
  - Consult routing table
  - Determine physical next hop
- If ERO object is strict
  - Verify physical next hop is directly connected
- Forward to physical next hop

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 57

Label Objects

- Label Request Object
  - Added to PATH message at ingress LSR
  - Requests that each LSR provide label to upstream LSR
- Label Object
  - Carried in RESV messages along return path upstream
  - Provides label to upstream LSR

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 58

Record Route Object—PATH Message

- Added to PATH message by ingress LSR
- Adds outgoing IP address of each hop in the path
  - In downstream direction
- Loop detection mechanism
  - Sends "Routing problem, loop detected" PathErr message
  - Drops PATH message

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 59

Record Route Object — RESV Message

- Added to RESV message by egress LSR
- Adds outgoing IP address of each hop in the path
  - In upstream direction
- Loop detection mechanism
  - Sends "Routing problem, loop detected" ResvErr message
  - Drops RESV message

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 60

Session Attribute Object

- Added to PATH message by ingress router
- Controls LSP
  - Priority
  - Preemption
  - Fast-reroute
- Identifies session
  - ASCII character string for LSP name

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 61

Tspec Object

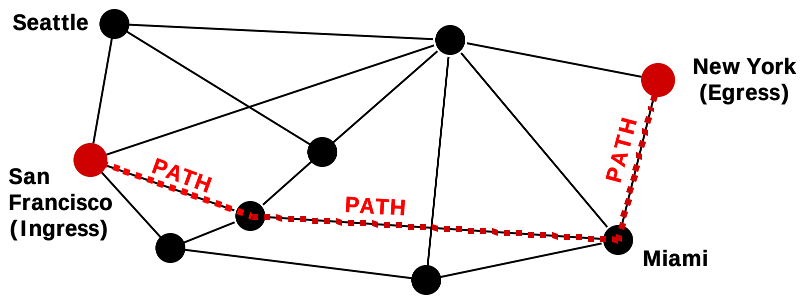
- Contains link management configuration
  - Requested bandwidth
  - Minimum and maximum LSP packet size

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 62

Path Signaling Example

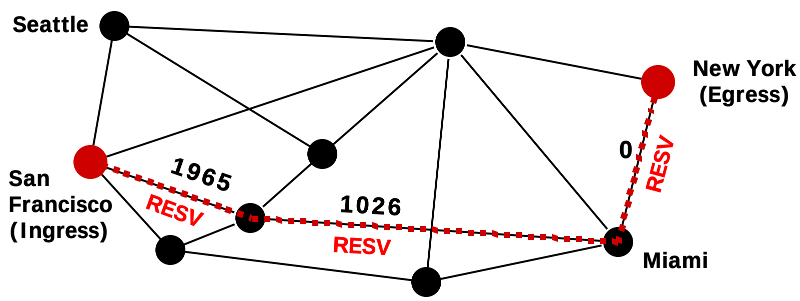
- Signaling protocol sets up path from San Francisco to New York, reserving bandwidth along the way



Slide 63

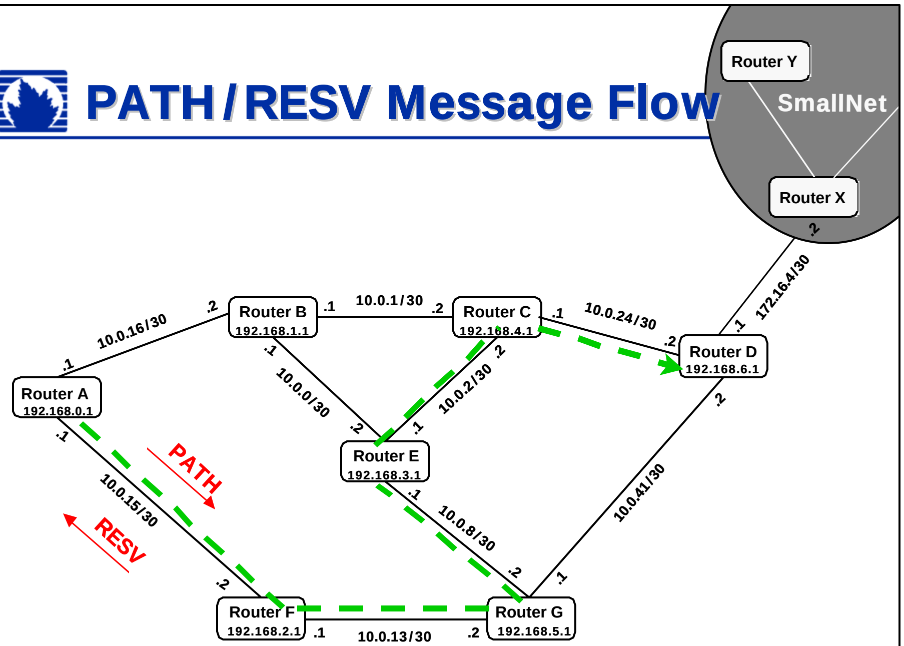
Path Signaling Example

- Once path is established, signaling protocol assigns label numbers in reverse order from New York to San Francisco



Slide 64

PATH/RESV Message Flow



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 65

Adjacency Maintenance—Hello Message

- New RSVP extension
  - Hello message
  - Hello Request
  - Hello Acknowledge
- Rapid node to node failure detection
  - Asynchronous updates
  - 3 second default update timer
  - 12 second default dead timer

```
[edit protocols rsvp interface interface-name]
user@host# set hello-interval seconds
```

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 66

Path Maintenance—Refresh Messages

- Maintains reservation of each LSP
- Sent every 30 seconds by default
- Consists of PATH and RESV messages
- Node to node, not end to end

```
[edit protocols rsvp]
user@host# set refresh-time seconds
```


RSVP Message Aggregation

- Bundles up to 30 RSVP messages within single PDU
- Controls
  - Flooding of PathTear or PathErr messages
  - Periodic refresh messages (PATH and RESV)
- Enhances protocol efficiency and reliability
- Disabled by default

```
[edit protocols rsvp interface interface-name]
user@host# set aggregate
```

Enable RSVP Authentication

- **If desired, enable authentication of RSVP messages on selected interfaces**

```
[edit protocols rsvp]
user@host# set interface interface-name
                                authentication-key "key"
```

- **Example:**

```
[edit protocols rsvp interface so-1/2/3]
user@host# set authentication-key "KnockKnock"
```

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 69

Juniper NETWORKS

Mapping Transit Traffic into LSP

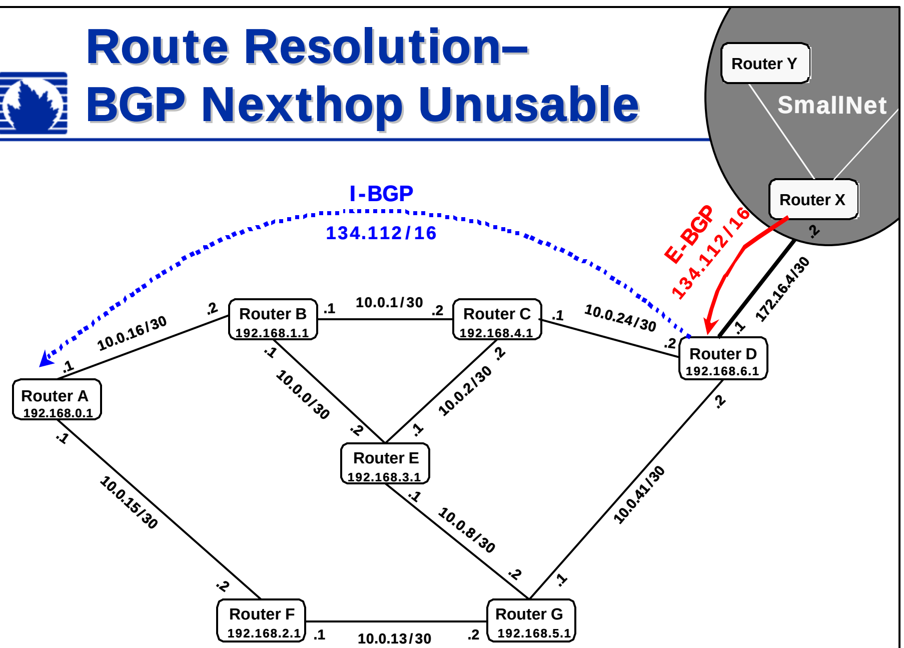
Mapping Transit Traffic Flows

- Mapping BGP Routes to LSPs
  - Used for BGP Next Hop resolution
  - Now BGP Next Hops can resolve through LSPs
  - BGP can install LSP as the physical next hop

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

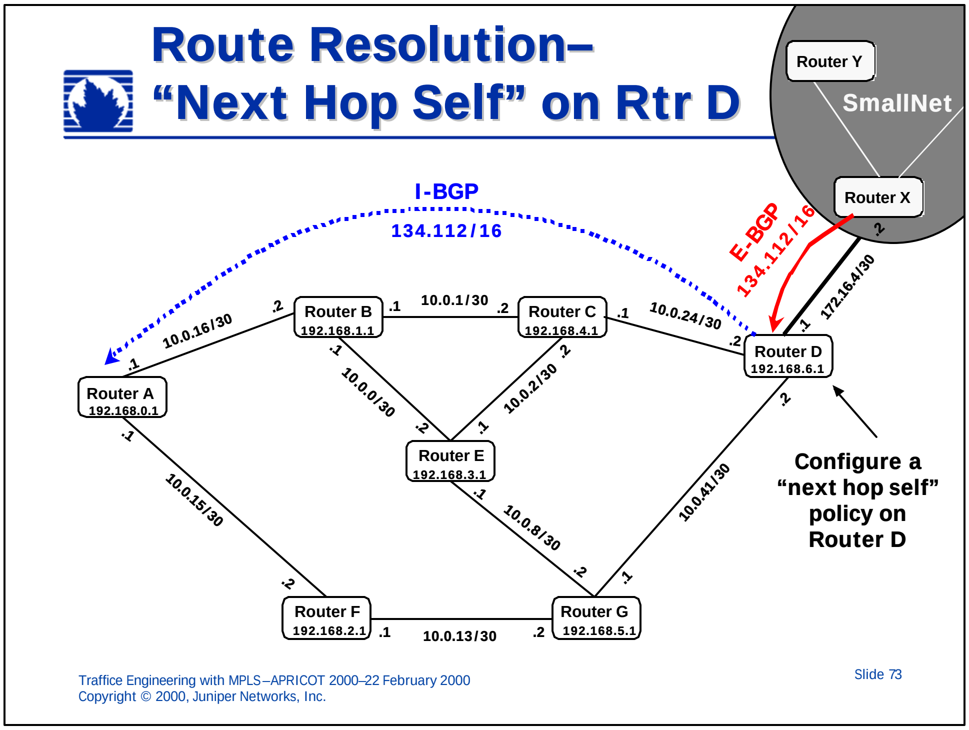
Slide 71

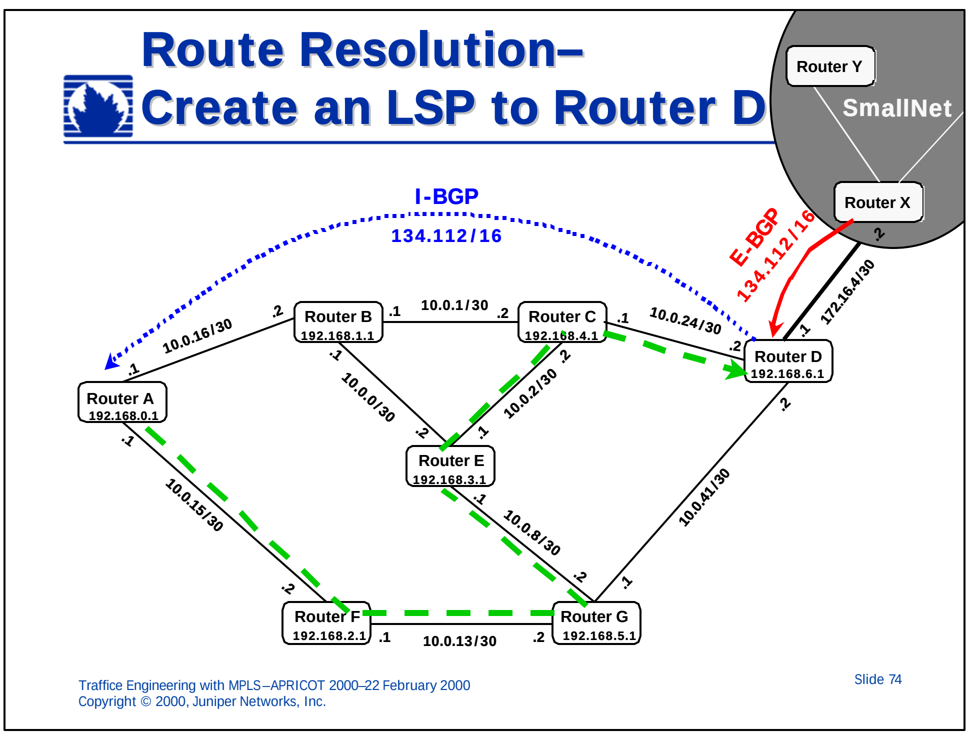
Route Resolution– BGP Nexthop Unusable



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 72


Route Resolution–
“Next Hop Self” on Rtr D
Router Y
SmallNet
Router X
I-BGP
134.112/16
E-BGP
134.112/16
172.16.4/30
Router A
192.168.0.1
Router B
192.168.1.1
Router C
192.168.4.1
Router D
192.168.6.1
Router E
192.168.3.1
Router F
192.168.2.1
Router G
192.168.5.1
10.0.16/30
10.0.1/30
10.0.24/30
10.0.0/30
10.0.2/30
10.0.15/30
10.0.8/30
10.0.41/30
10.0.13/30
Configure a
“next hop self”
policy on
Router D
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.
Slide 73


Route Resolution–
Create an LSP to Router D
Router Y
SmallNet
Router X
I-BGP
134.112/16
E-BGP
134.112/16
172.16.4/30
Router A
192.168.0.1
Router B
192.168.1.1
Router C
192.168.4.1
Router D
192.168.6.1
Router E
192.168.3.1
Router F
192.168.2.1
Router G
192.168.5.1
10.0.16/30
10.0.1/30
10.0.24/30
10.0.0/30
10.0.2/30
10.0.15/30
10.0.8/30
10.0.41/30
10.0.13/30
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.
Slide 74

Route Resolution—Summary

- LSPs appear at ingress router as next hop addresses
  - Points to egress router, as if it were directly connected
  - When LSP is up, next hop is usable and attractive to BGP
  - When LSP is down, LSP next hop is unusable
    - Can still use normal IP routing to reach next hop
- Only BGP pays attention to LSP next hops when resolving routes
  - Other protocols can use this information with additional configuration

Juniper NETWORKS

Traffic Engineering Constrained Routing

Signaled vs Constrained LSPs

- Common Features
  - Signaled by RSVP
  - MPLS labels automatically assigned
  - Configured on ingress router only
- Signaled LSPs
  - CSPF not used
  - User configured ERO handed to RSVP for signaling
  - RSVP consults routing table to make next hop decision
- Constrained LSPs
  - CSPF used
  - Full path computed by CSPF at ingress router
  - Complete ERO handed to RSVP for signaling

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 77

Constrained Shortest Path First Algorithm

- Modified “shortest path first” algorithm
- Finds shortest path based on IGP metric while satisfying additional constraints
- Integrates TED (Traffic Engineering Database)
  - IGP topology information
  - Available bandwidth
  - Link color
- Modified by administrative constraints
  - Maximum hop count
  - Bandwidth
  - Strict or loose routing
  - Administrative groups

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 78

CSPF Algorithm

1. Compute LSPs one at a time,
   - Start with the highest "setup" priority LSP
   - If equal priority, start with LSPs with highest bandwidth
2. Prune the topology database (TED) of all the links that are not full duplex and do not have sufficient reservable bandwidth
3. If "include" statement configured, prune all links that do not share any included colors
4. If "exclude" statement configured, prune all links that contain excluded colors and do not contain a color

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000 Slide 79

CSPF Algorithm, continued

5. Find the shortest path towards the LSP's egress router, taking into account explicit-path constraints
   - For example, if the path must pass through Router A, two separate SPFs are computed–one from the ingress router to Router A, the other from Router A to the egress router
6. If several paths have equal cost, choose the one whose last hop address is the same as the LSP's destination
7. If several equal-cost paths remain, select the one with the fewest number of hops
8. If several equal-cost paths remain, apply the CSPF load-balancing rule configured on the LSP

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000 Slide 80

Computing the ERO

- Ingress LSR passes user defined restrictions to CSPF
  - Strict and loose hops
  - Bandwidth constraints
  - Admin Groups
- CSPF algorithm
  - Factors in user defined restrictions
  - Runs computation against the TED
  - Determines the shortest path
- CSPF hands full ERO to RSVP for signaling

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 81

Traffic Engineering Database

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 82

Traffic Engineering Database

- CSPF uses TED to calculate explicit paths across the physical topology
- Similar to IGP link-state database
- Relies on extensions to IGP
  - Network link attributes
  - Topology information
- Separate from IGP database

TE Extensions to ISIS/OSPF

- Describes traffic engineering topology
  - Traffic engineering database (TED)
    - Bandwidth
    - Administrative groups
  - Does not necessarily match regular routed topology
    - Subset of IGP domain
  - ISIS Extensions
  - IP reachability TLV
  - IS reachability TLV
- OSPF Extension
  - Type 10 Opaque LSA

ISIS TE Extensions

- IP Reachability TLV
  - IP prefixes that are reachable
  - IP link default metric
    - Extended to 32 bits (wide metrics)
  - Up/down bit
    - Avoids loops in L1/L2 route leaking

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 85

ISIS TE Extensions

- IS Reachability TLV
  - IS neighbors that are reachable
  - ID of adjacent router
    - IP addresses of interface (/32 prefix length)
  - Sub-TLVs describe the TE topology

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 86

ISIS IS Reachability TLV

- Sub-TLVs contain
  - Local interface IP address
  - Remote interface IP address
  - Maximum link bandwidth
  - Maximum reservable link bandwidth
  - Unreserved link bandwidth
  - Traffic engineering metric
  - Administrative group
  - Reserved TLVs for future expansion

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

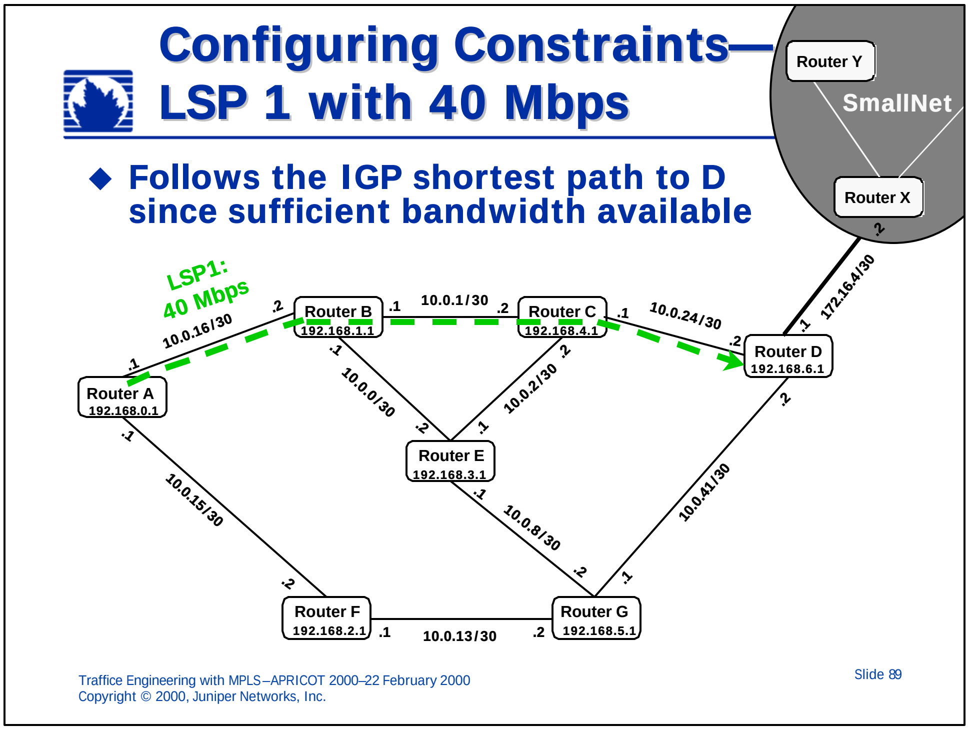
Slide 87

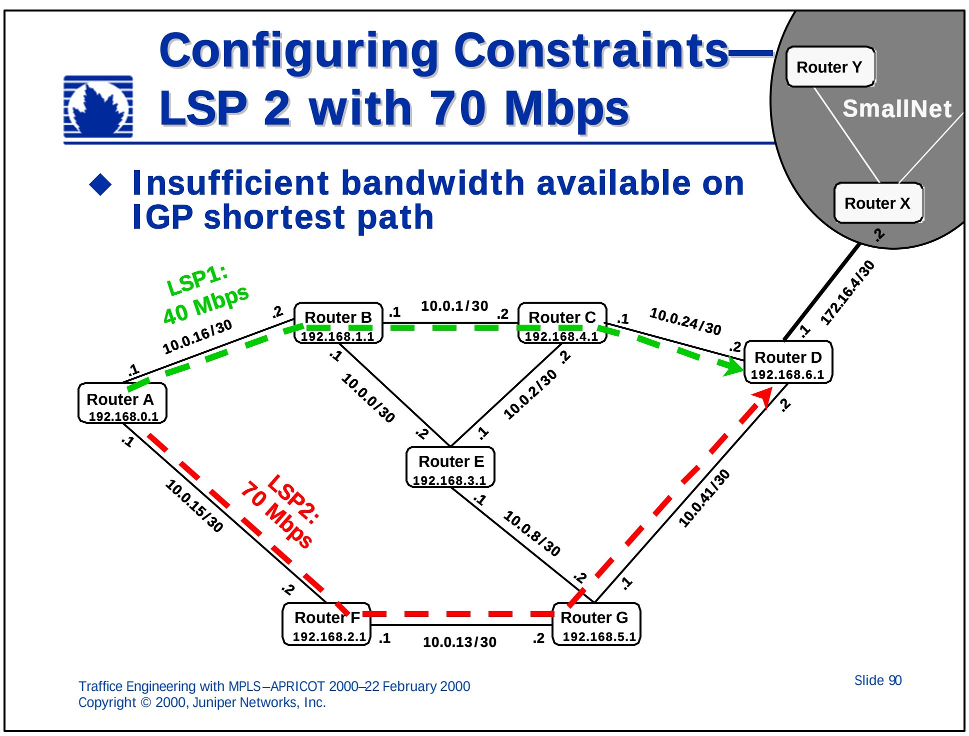
OSPF TE Extensions

- Opaque LSA
  - Original Router LSA not extensible
  - Type 10 LSA
  - Area flooding scope
  - Standard LSA header (20 bytes)
  - TE capabilities
- Traffic Engineering LSA
  - Work in progress

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 88


Configuring Constraints—
LSP 1 with 40 Mbps
Follows the IGP shortest path to D since sufficient bandwidth available
Router Y
SmallNet
Router X
LSP1:
40 Mbps
Router A
192.168.0.1
Router B
192.168.1.1
Router C
192.168.4.1
Router D
192.168.6.1
Router E
192.168.3.1
Router F
192.168.2.1
Router G
192.168.5.1
10.0.16/30
10.0.1/30
10.0.24/30
172.16.4/30
10.0.0/30
10.0.2/30
10.0.15/30
10.0.8/30
10.0.41/30
10.0.13/30
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.
Slide 89

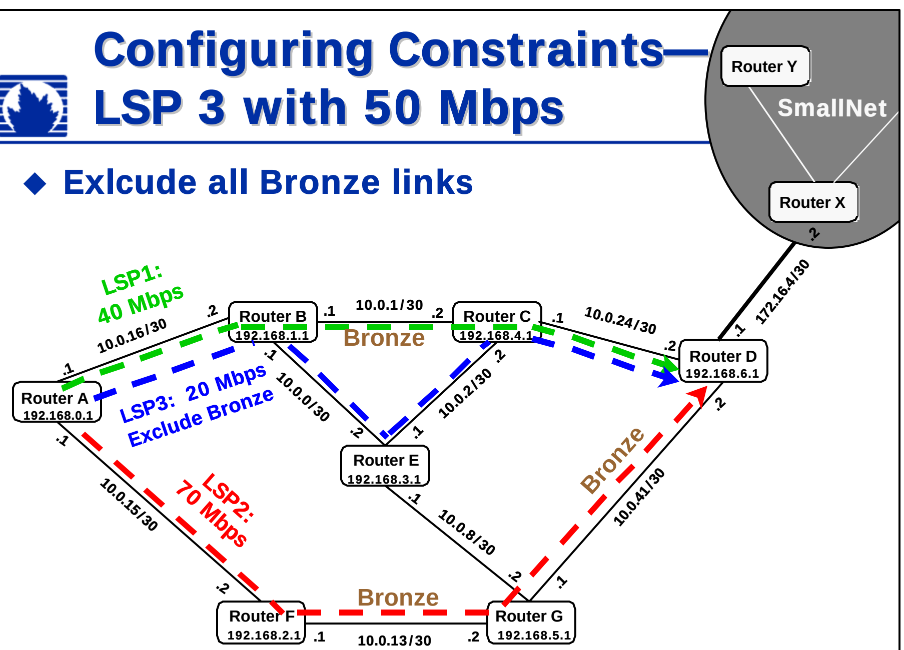

Configuring Constraints—
LSP 2 with 70 Mbps
Insufficient bandwidth available on IGP shortest path
Router Y
SmallNet
Router X
LSP1:
40 Mbps
LSP2:
70 Mbps
Router A
192.168.0.1
Router B
192.168.1.1
Router C
192.168.4.1
Router D
192.168.6.1
Router E
192.168.3.1
Router F
192.168.2.1
Router G
192.168.5.1
10.0.16/30
10.0.1/30
10.0.24/30
172.16.4/30
10.0.0/30
10.0.2/30
10.0.15/30
10.0.8/30
10.0.41/30
10.0.13/30
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.
Slide 90

Affinity (Link Colors)

- Ability to assign a color to each link
  - Gold
  - Silver
  - Bronze
- Up to 32 colors available
- Can define an affinity relationship
  - Include
  - Exclude

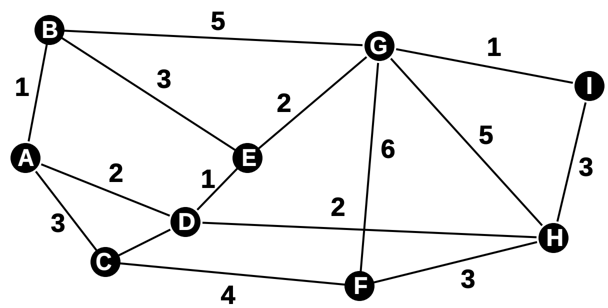
Configuring Constraints—LSP 3 with 50 Mbps

- Exlcude all Bronze links



Test Your Knowledge (I)

- Choose the path from A to I the way an IGP would

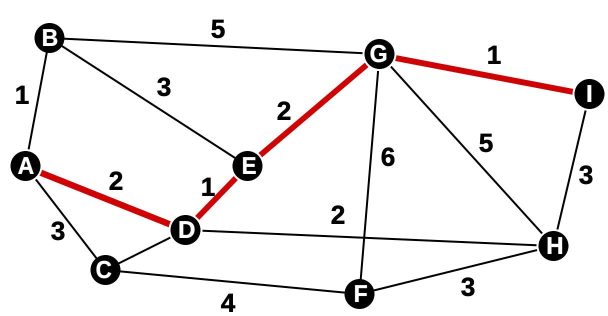


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 93

Test Your Knowledge (I)

- A-D-E-G-I has the lowest metric—6



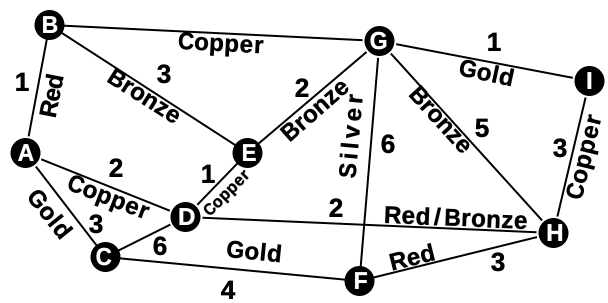
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 94

Test Your Knowledge (II)

- **Choose the path from A to I using:**

```
admin group {
    include [gold sliver];
}
```

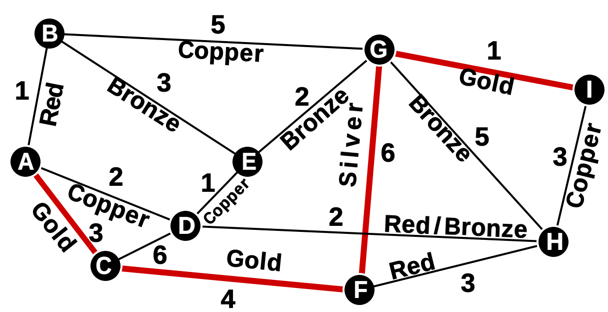


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 95

Test Your Knowledge (II)

- **A-C-F-G-I uses only gold or silver links**



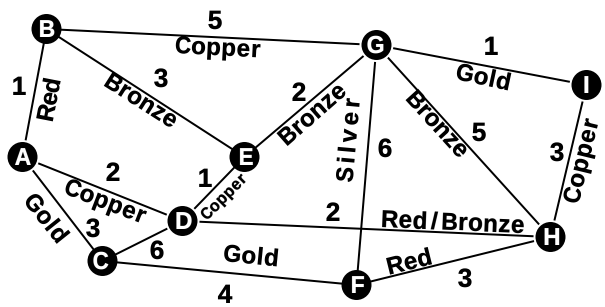
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 96

Test Your Knowledge (III)

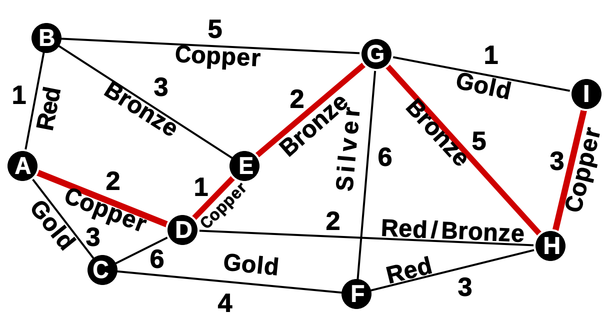
- Choose the path from A to I using:

```
admin group {
    include [copper bronze];
    exclude red;
}
```



Test Your Knowledge (III)

- A-D-E-G-H-I is the shortest path excluding the admin class

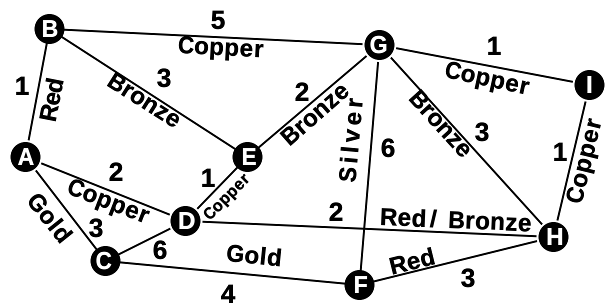


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.

Test Your Knowledge (IV)

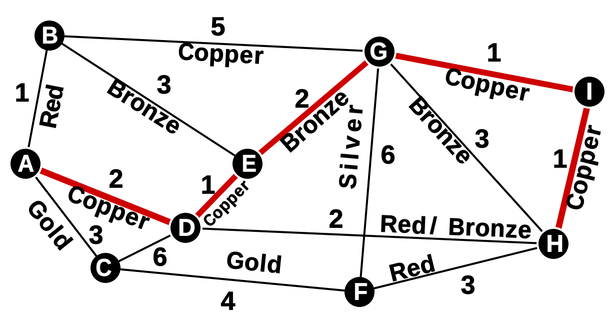
- **Choose the path from A to H using:**

```
admin group {
    include [copper bronze];
    exclude red;
}
```



Test Your Knowledge (IV)

- **A-D-E-G-I-H is the shortest path excluding the admin class**



Preemption

- Defines relative importance of LSPs on same ingress router
- CSPF uses priority to optimize paths
- Higher priority LSPs
  - Are established first
  - Offer more optimal path selection
  - May tear down lower priority LSPs when rerouting
- Default configuration makes all LSPs equal

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 101

Preemption

- Controlled by two settings
  - Setup priority and hold (reservation) priority
    - New LSP compares its setup priority with hold priority of existing LSP
    - If setup priority is less than hold priority, existing LSP is rerouted to make room
  - Priorities from 0 (strong) through 7 (weak)
  - Defaults
    - Setup priority is 7 (do not preempt)
    - Reservation priority is 0 (do not allow preemption)
- Use with caution
  - No large scale experience with this feature

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 102

LSP Reoptimization

- Reroutes LSPs that would benefit from improvements in the network
  - Special rules apply
- Disabled by default
- Enable with

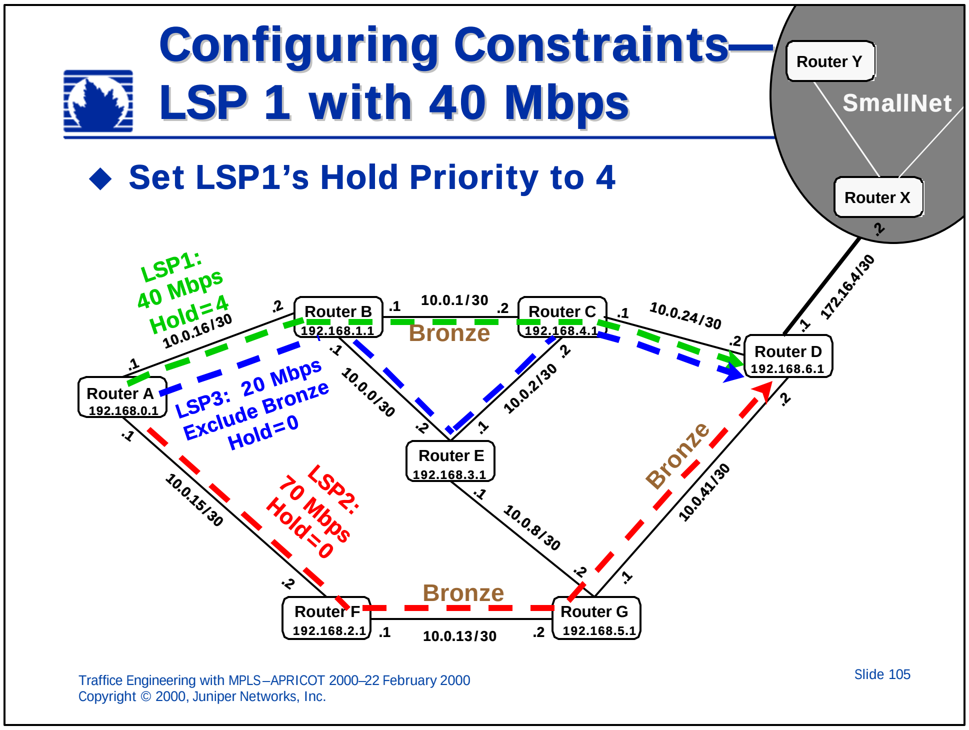
```
[edit protocols mpls label-switched path lsp-path-name]
user@host# set optimize-timer seconds
```

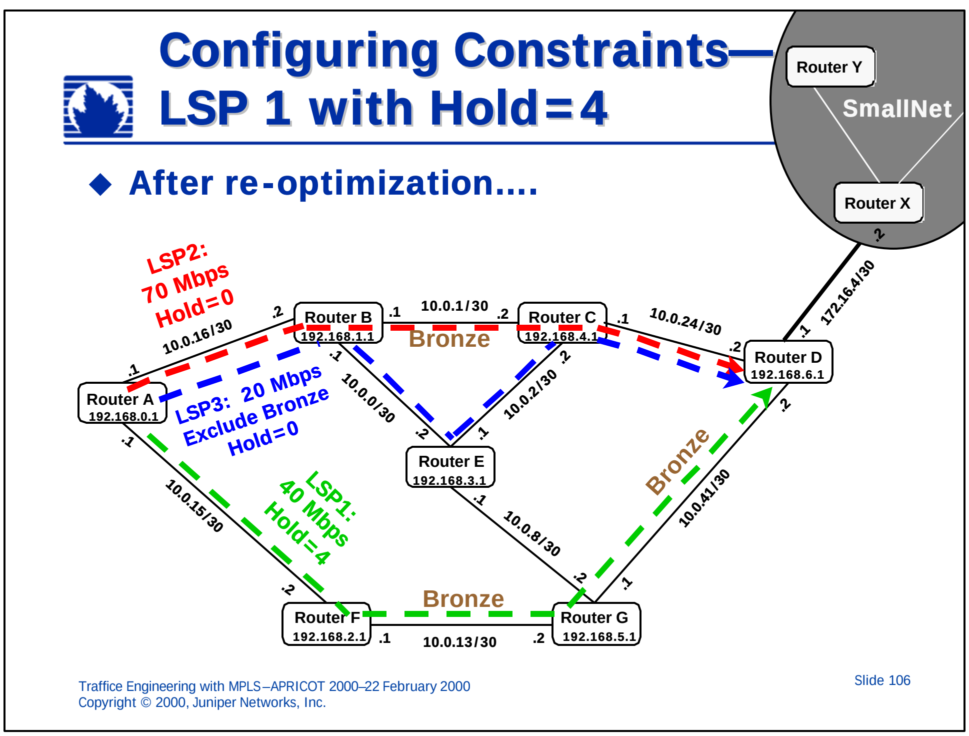
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000 Slide 103

LSP Reoptimization Rules

- Reoptimize if new path can be found that meets all of the following
  - Has lower IGP metric
  - Has fewer hops
  - Does not cause preemption
  - Reduces congestion by 10%
    - Compares aggregate available bandwidth of new and old path
- Intentionally conservative rules, use with care

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000 Slide 104


Configuring Constraints—
LSP 1 with 40 Mbps
◆ Set LSP1's Hold Priority to 4
Router Y
SmallNet
Router X
LSP1:
40 Mbps
Hold=4
LSP3: 20 Mbps
Exclude Bronze
Hold=0
LSP2:
70 Mbps
Hold=0
Router A
192.168.0.1
Router B
192.168.1.1
Router C
192.168.4.1
Router D
192.168.6.1
Router E
192.168.3.1
Router F
192.168.2.1
Router G
192.168.5.1
10.0.16/30
10.0.1/30
10.0.24/30
172.16.4/30
10.0.0/30
10.0.2/30
10.0.15/30
10.0.8/30
10.0.41/30
10.0.13/30
Bronze
Bronze
Bronze
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.
Slide 105


Configuring Constraints—
LSP 1 with Hold=4
◆ After re-optimization....
Router Y
SmallNet
Router X
LSP2:
70 Mbps
Hold=0
LSP3: 20 Mbps
Exclude Bronze
Hold=0
LSP1:
40 Mbps
Hold=4
Router A
192.168.0.1
Router B
192.168.1.1
Router C
192.168.4.1
Router D
192.168.6.1
Router E
192.168.3.1
Router F
192.168.2.1
Router G
192.168.5.1
10.0.16/30
10.0.1/30
10.0.24/30
172.16.4/30
10.0.0/30
10.0.2/30
10.0.15/30
10.0.8/30
10.0.41/30
10.0.13/30
Bronze
Bronze
Bronze
Traffice Engineering with MPLS–APRICOT 2000–22 February 2000
Copyright © 2000, Juniper Networks, Inc.
Slide 106

Class of Service (CoS)

- Each packet is assigned class-of-service value at ingress
- CoS is inherited from precedence bits in the IP header
- Can be overridden by configuration statement

```
primary path-name {
   [...]
   class-of-service cos-value;
}
```

- CoS field determines output queue on transit and egress routers

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 107

Class of Service (CoS)

- CoS determined prior to prefix or label lookup
  - Packet entering router contains CoS values
  - CoS values copied into Packet Notification cell
- CoS queuing done on output port

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 108

LSP Load Balancing

- Two categories
  - Selecting path for each LSP
    - Multiple equal cost IP paths to egress are available
    - Random
    - Least-fill
    - Most-fill
  - Balance traffic over multiple LSP
    - Multiple equal cost LSPs to egress are available
    - BGP can load balance prefixes over 8 LSPs

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 109

LSP Load Balancing

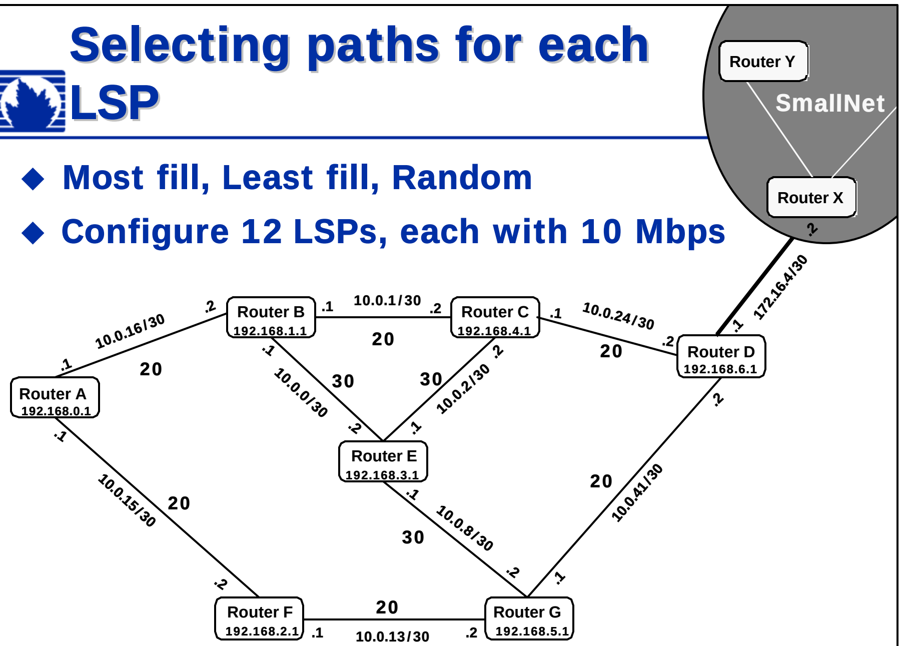
- Selecting path for each LSP
  - Random is default
    - Distributes LSPs randomly over available equal cost paths
  - Least-fill
    - Distributes LSPs over available equal cost paths based on available link bandwidth
  - Most-fill
    - LSPs fill one link first, then next

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 110

Selecting paths for each LSP

- Most fill, Least fill, Random
- Configure 12 LSPs, each with 10 Mbps

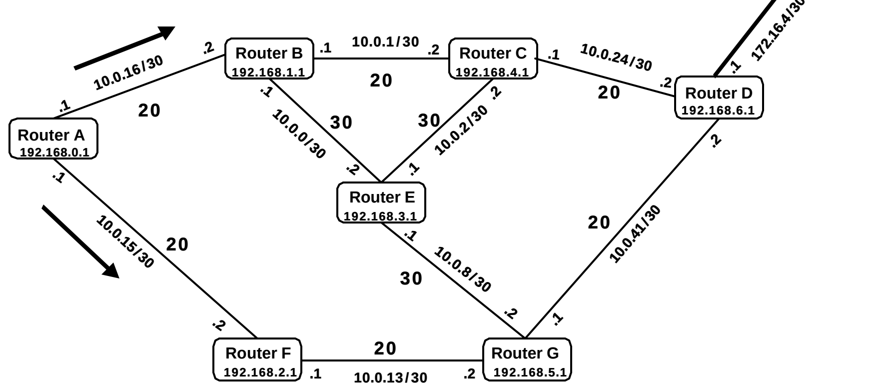


Load Balancing

- Balancing traffic over multiple LSPs
  - 8 equal cost paths for BGP
  - JUNOS default is per-prefix
  - Per-packet knob available

Balancing traffic over equal cost IGP paths

- Without LSPs configured, prefixes are distributed over equal cost IGP paths

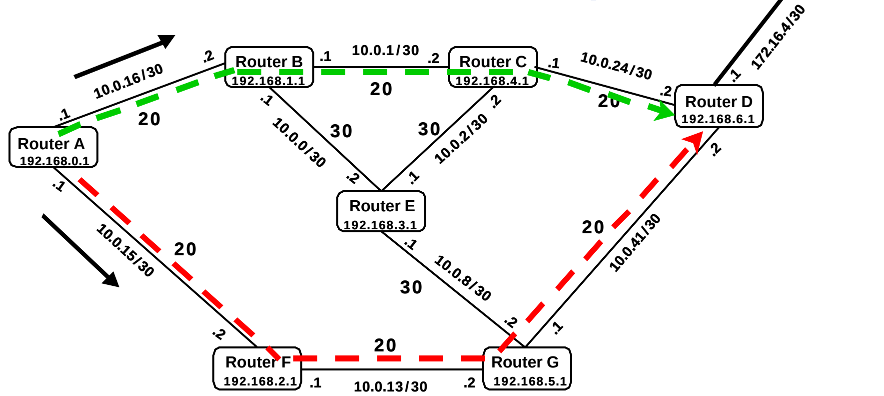


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 113

Balancing traffic over equal cost LSPs

- Same behavior, now over LSPs
- Prefixes distributed over multiple LSPs



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 114

Advanced Traffic Engineering Features

Traffic Protection

Traffic Protection

- Primary LSP
  - Retry timer
  - Retry limit
- Secondary LSPs
  - Standby option
- Fast Reroute
- Adaptive mode

Primary LSP

- Optional
  - If configured, becomes preferred path for LSP
- If no primary configured
  - LSR makes all decisions to reach egress
- Zero or one primary path
- Revertive capability
  - Revertive behavior can be modified

Primary LSP

- Revertive Capability
  - Retry timer
    - Time between attempts to bring up failed primary path
    - Default is 30 seconds
    - Primary must be stable two times (2x) retry timer before reverts back
  - Retry limit
    - Number of attempts to bring up failed primary path
    - Default is 0 (unlimited retries)
    - If limit reached, human intervention then required

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 119

Secondary LSP

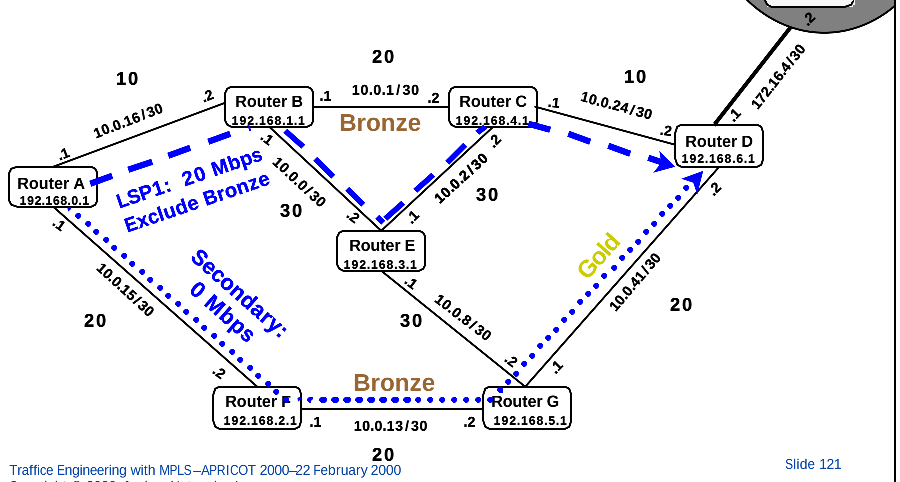
- Optional
- Zero or more secondary paths
- All secondary paths are equal
  - Selection based on listed order of configuration
- Standby knob
  - Maintains secondary path in ‘up’ condition
  - Eliminates call-setup delay of secondary LSP
  - Additional state information must be maintained

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 120

Secondary Paths— LSP 1, exclude Bronze

- Secondary – avoid primary if possible



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 121

Adaptive Mode

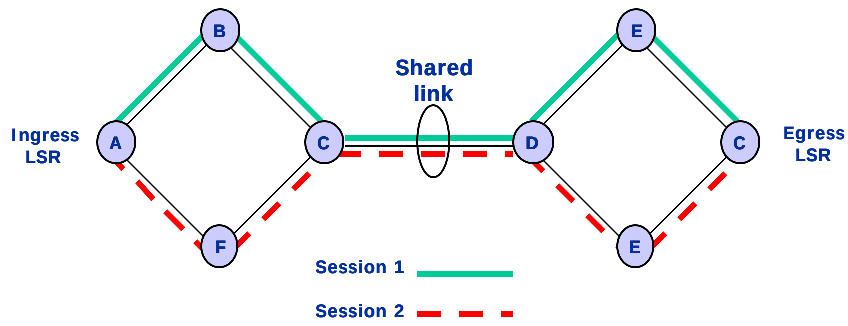
- Applies to
  - LSP rerouting
  - Primary & secondary sharing links
- Avoids double counting
- SE Reservation style

```
[edit protocols mpls label-switched-path lsp-path-name]
user@host# set adaptive
```

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 122

Shared Links

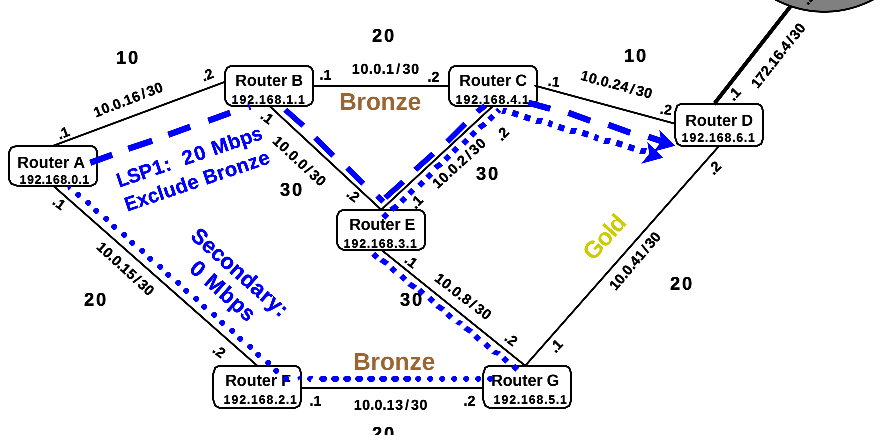


- FF reservation style:
  - Each session has its own identity
  - Each session has its own bandwidth reservation
- SE Reservation style:
  - Each session has its own identity
  - Sessions share a single bandwidth reservation

Slide 123

Secondary Paths— LSP 1, exclude Bronze

- Secondary – in Standby mode, exclude Gold



Slide 124

Fast Reroute

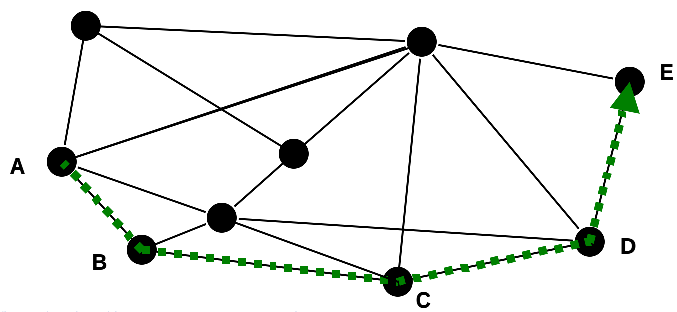
- Configured on ingress router only
- Detours around node or link failure
  - ~100s of ms reroute time
- Detour paths immediately available
- Crank-back to node, not ingress router
- Uses TED to calculate detour

Fast Reroute

- Short term solution to reduce packet loss
- If node or link fails, upstream node
  - Immediately detours
  - Signals failure to ingress LSR
- Only ingress LSR knows policy constraints
  - Ingress computes alternate route
    - Based on configured secondary paths
  - Initiates long term reroute solution

Fast Reroute Example

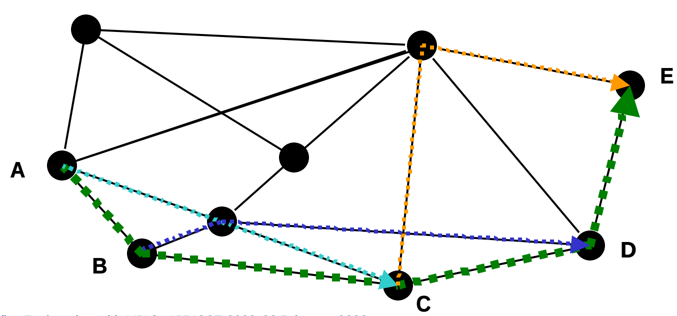
- Primary LSP from A to E



Slide 127

Fast Reroute Example

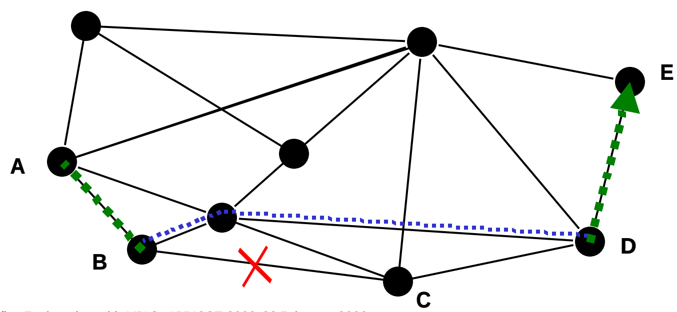
- Enable fast reroute on ingress
  - A creates detour around B
  - B creates detour around C
  - C creates detour around D



Slide 128

Fast Reroute Example - Short Term Solution

- B to C link fails
  - B immediately detours around C
  - B signals to A that failure occurred

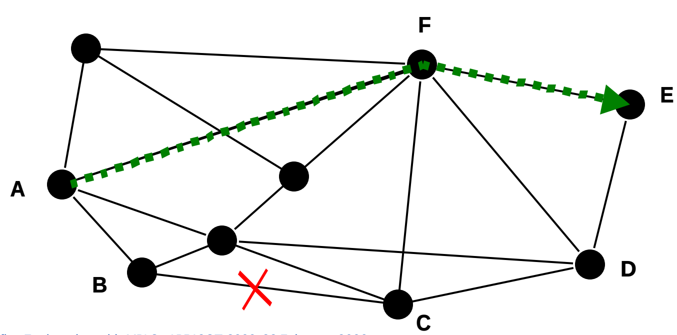


Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 129

Fast Reroute Example – Long Term Solution

- A calculates and signals new primary path



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 130

LSP Rerouting

- Initiated by ingress LSR
  - Exception is fast reroute
- Conditions that trigger reroute
  - More optimal route becomes available
  - Failure of a resource along the LSP path
  - Preemption occurs
  - Manual configuration change
- Make before break (if adaptive)
  - Establish new LSP with SE style
  - Transfer traffic to new LSP
  - Tear down old LSP

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 131

CSPF Algorithm

1. Compute LSPs one at a time,
   - Start with the highest "setup" priority LSP
   - If equal priority, start with LSPs with highest bandwidth
2. Prune the topology database (TED) of all the links that are not full duplex and do not have sufficient reservable bandwidth
3. If "include" statement configured, prune all links that do not share any included colors
4. If "exclude" statement configured, prune all links that contain excluded colors and do not contain a color

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 132

CSPF Algorithm, continued

5. Find the shortest path towards the LSP's egress router, taking into account explicit-path constraints
   - For example, if the path must pass through Router A, two separate SPFs are computed–one from the ingress router to Router A, the other from Router A to the egress router
6. If several paths have equal cost, choose the one whose last hop address is the same as the LSP's destination
7. If several equal-cost paths remain, select the one with the fewest number of hops
8. If several equal-cost paths remain, apply the CSPF load-balancing rule configured on the LSP

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 133

Juniper NETWORKS

Advanced Route Resolution

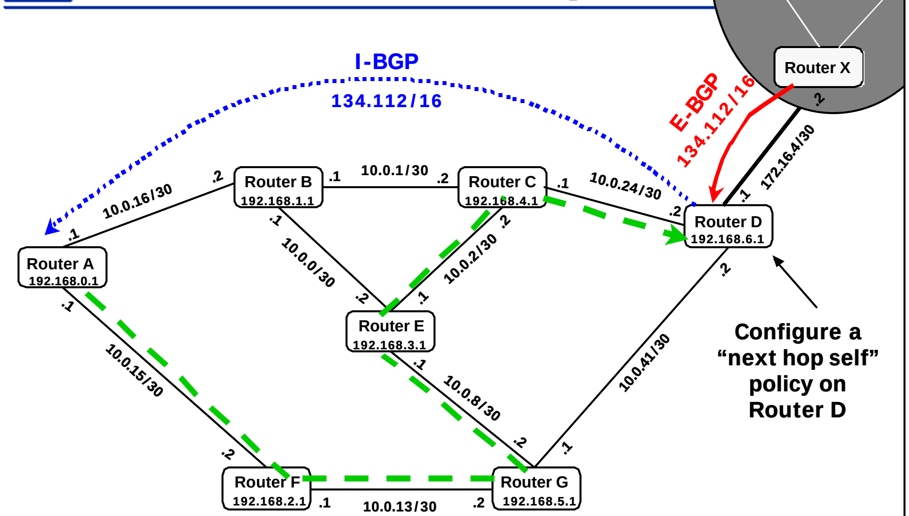
Mapping Transit Traffic

- Mapping transit destinations
  - Default mode
  - Only BGP can use LSPs for its recursive route calculations
  - Only BGP prefixes that have the LSP destination address as the BGP next-hop are resolvable through the LSP

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 135

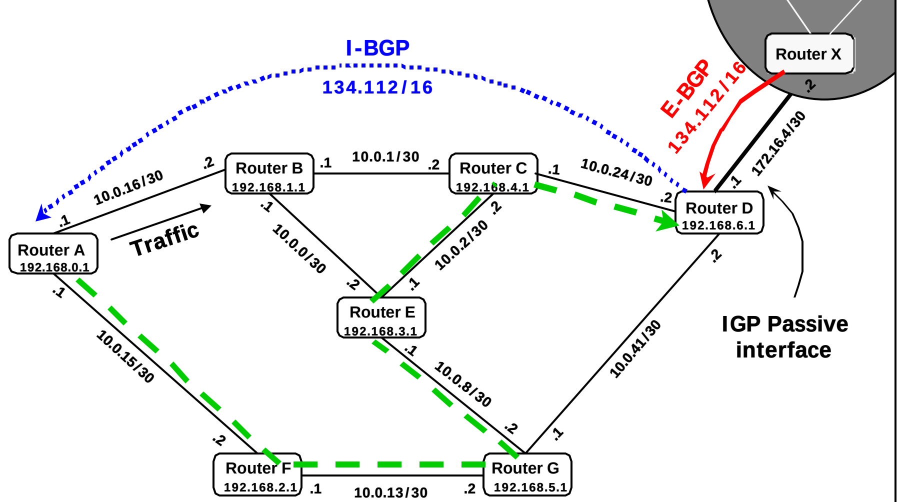
Route Resolution–Transit Traffic Example



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 136

What if BGP next hop does not align with LSP endpoint?



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 137

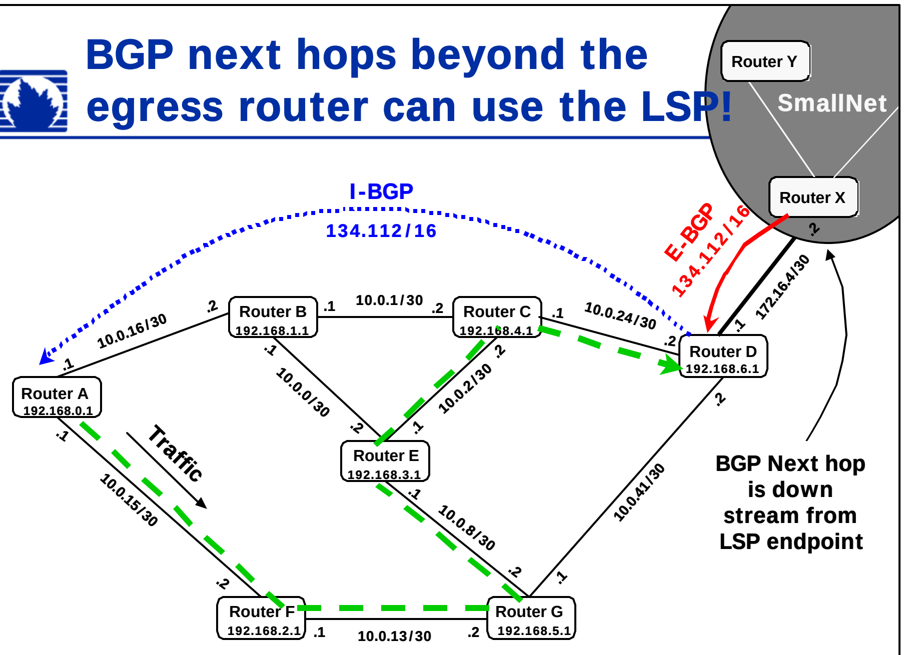
Traffic Engineering Shortcuts

- Configure TE Shortcuts on ingress router
  - Good for BGP nexthops that are not resolvable directly through an LSP
  - If LSP exists that gets you closer to BGP nexthop
  - Installs prefixes that are downstream from egress router into ingress router's inet.3 route table

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 138

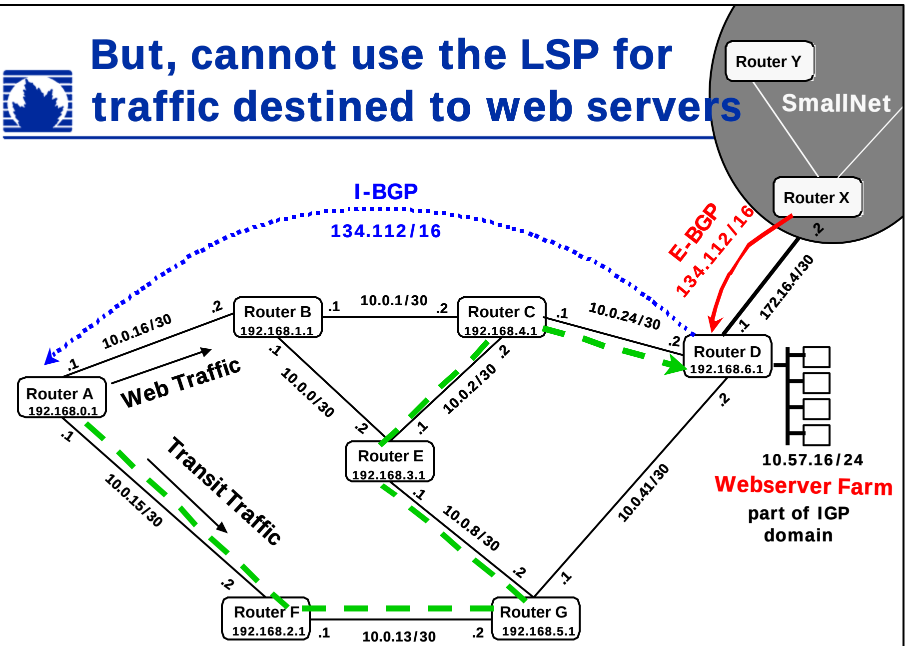
BGP next hops beyond the egress router can use the LSP!



TE Shortcuts

- By itself, still only usable by BGP
- Installs additional prefixes in ingress router's inet.3 table
- Only BGP can use routes in inet.3 for BGP recursive lookups

But, cannot use the LSP for traffic destined to web servers



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 141

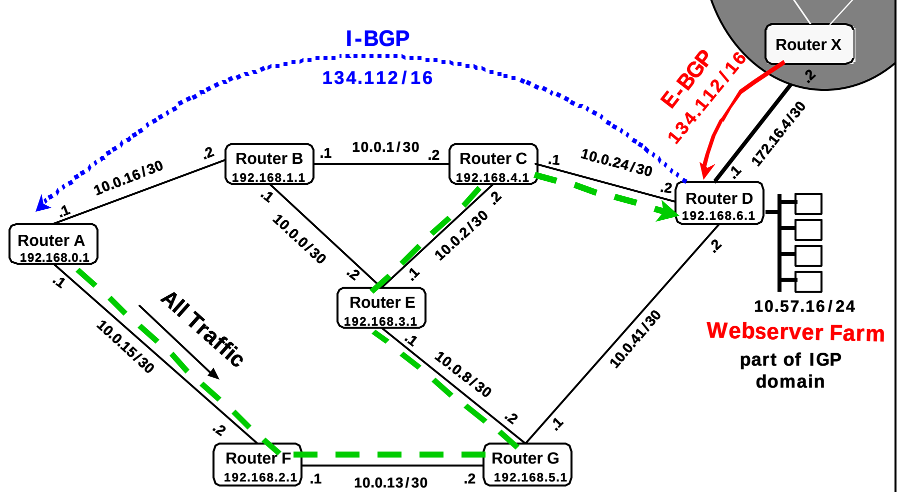
BGP-IGP knob

- **`Traffic-engineering bgp-igp` knob**
  - **Forces all MPLS prefixes into main routing table (inet.0)**
  - **All destinations can now use all LSPs**
    - **IGP and BGP prefixes**

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 142

Now all traffic destined to egress router and beyond use LSP



Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 143

TTL Decrement

- **Default is to decrement TTL on all LSR hops**
  - **Loop prevention**
  - **Topology discovery via `traceroute`**
- **Disable TTL decrement inside LSP**
  - **No topology discovery**
  - **TTL decrement at egress router only**

```
[edit protocols mpls label-switched-path lsp-path-name]
user@host# set no-decrement-ttl
```

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 144

Juniper NETWORKS

Circuit Cross Connect

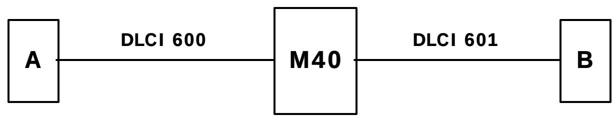
Circuit Cross-Connect (CCC)

- Transparent connection between two Layer 2 circuits
- Supports
  - PPP, Cisco HDLC, Frame Relay, ATM, MPLS
- Router looks only as far as Layer 2 circuit ID
  - Any protocol can be carried in packet payload
  - Only "like" interfaces can be connected (for example, Frame Relay to Frame Relay, or ATM to ATM)
- Three types of cross-connects
  - Layer 2 switching
  - MPLS tunneling
  - Stitching MPLS LSPs

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 146

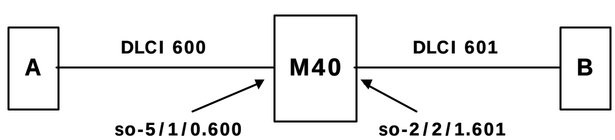
CCC Layer 2 Switching



- A and B have Frame Relay connections to M40, carrying any type of traffic
- M40 behaves as switch
- Layer 2 packets forwarded transparently from A to B without regard to content; only DLCI is changed
- CCC supports switching between PPP, Cisco HDLC, Frame Relay PVCs, or ATM PVCs
- ATM AAL5 packets are reassembled before sending

Slide 147

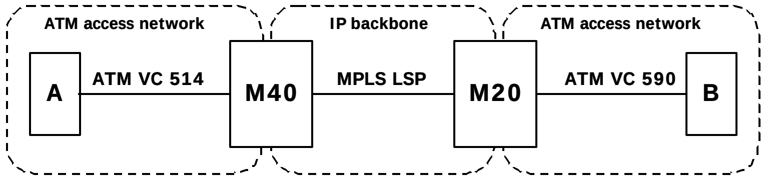
CCC Layer 2 Switching




```
[edit protocols]
user@host# show
connections {
  interface-switch connection-name {
     interface so-5/1/0.600;
     interface so-2/2/1.601;
  }
}
```


Slide 148

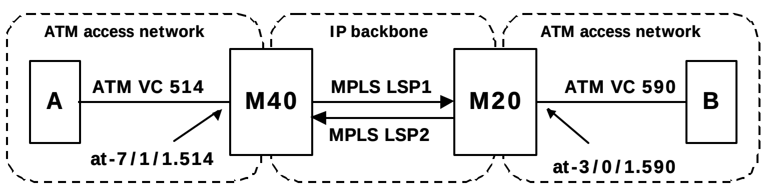
CCC
MPLS Interface Tunneling



- **Transports packets from one interface through an MPLS LSP to a remote interface**
- **Bridges Layer 2 packets from end-to-end**
- **Supports tunneling between “like” ATM, Frame Relay, PPP, and Cisco HDLC connections**

Slide 149

CCC
MPLS Interface Tunneling

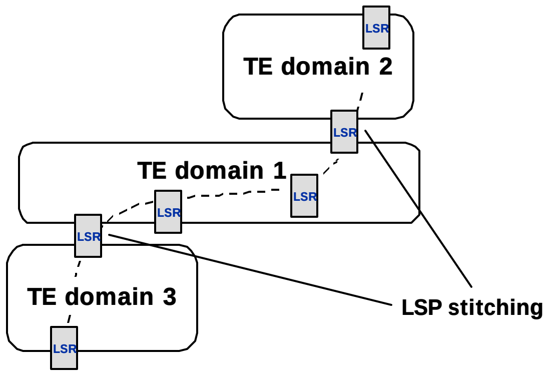



```
[edit protocols]
user@M40# show
connections {
  remote-interface-switch m40-to-m20
    interface at-7/1/1.514;
    transmit-lsp lsp1;
    receive-lsp lsp2;
  }
```

```
[edit protocols]
user@M20# show
connections {
  remote-interface-switch m20-to-m40
    interface at-3/0/1.590;
    transmit-lsp lsp2;
    receive-lsp lsp1;
  }
```


Slide 150

CCC LSP Stitching

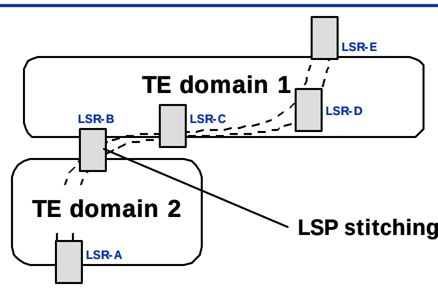


- Large networks can be separated into several traffic engineering domains (supports IS-IS area partitioning)
- CCC allows establishment of LSP across domains by “stitching” together LSPs from separate domains

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 151

CCC LSP Stitching




```
[edit protocols]
user@LSR-B# show
connections {
  lsp-switch LSR-A_to_LSR-E {
    transmit-lsp lsp2;
    receive-lsp lsp1;
  }
  lsp-switch LSR-E_to_LSR-A {
    receive-lsp lsp3;
    transmit-lsp lsp4;
  }
```

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 152

www.juniper.net

Traffice Engineering with MPLS–APRICOT 2000–22 February 2000

Slide 153