MPLS and GMPLS:

Principles, Implementation, and Advanced Concepts

Adrian Farrel, Old Dog Consulting
Zafar Ali, Cisco Systems, Inc.
Mallik Tatipamula, Cisco Systems, Inc.

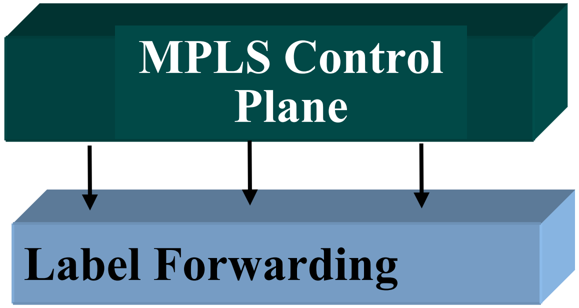
MPLS
2004

Cisco Systems

Outline

- Principles of MPLS-TE
- Extending the Concepts to GMPLS
- Fundamental Concepts
- Implementing and Deploying MPLS-TE and GMPLS
- Inter-Domain Traffic Engineering
- Components of MPLS/ GMPLS High Availability
- MPLS O&M
- Future Work

MPLS Architectural Principles



Separation of Forwarding and Control plane

- Forwarding Plane
  - Simple “label swapping” mechanism to forward packets along a “Label Switched Path” (LSP)
  - Map traffic to LSP based on “Forwarding Equivalence Class” (FEC).
- Control Plane: various application dependent mechanisms for exchanging labels
- MPLS Applications can be broadly classified as MPLS Traffic Engineering and MPLS VPN.

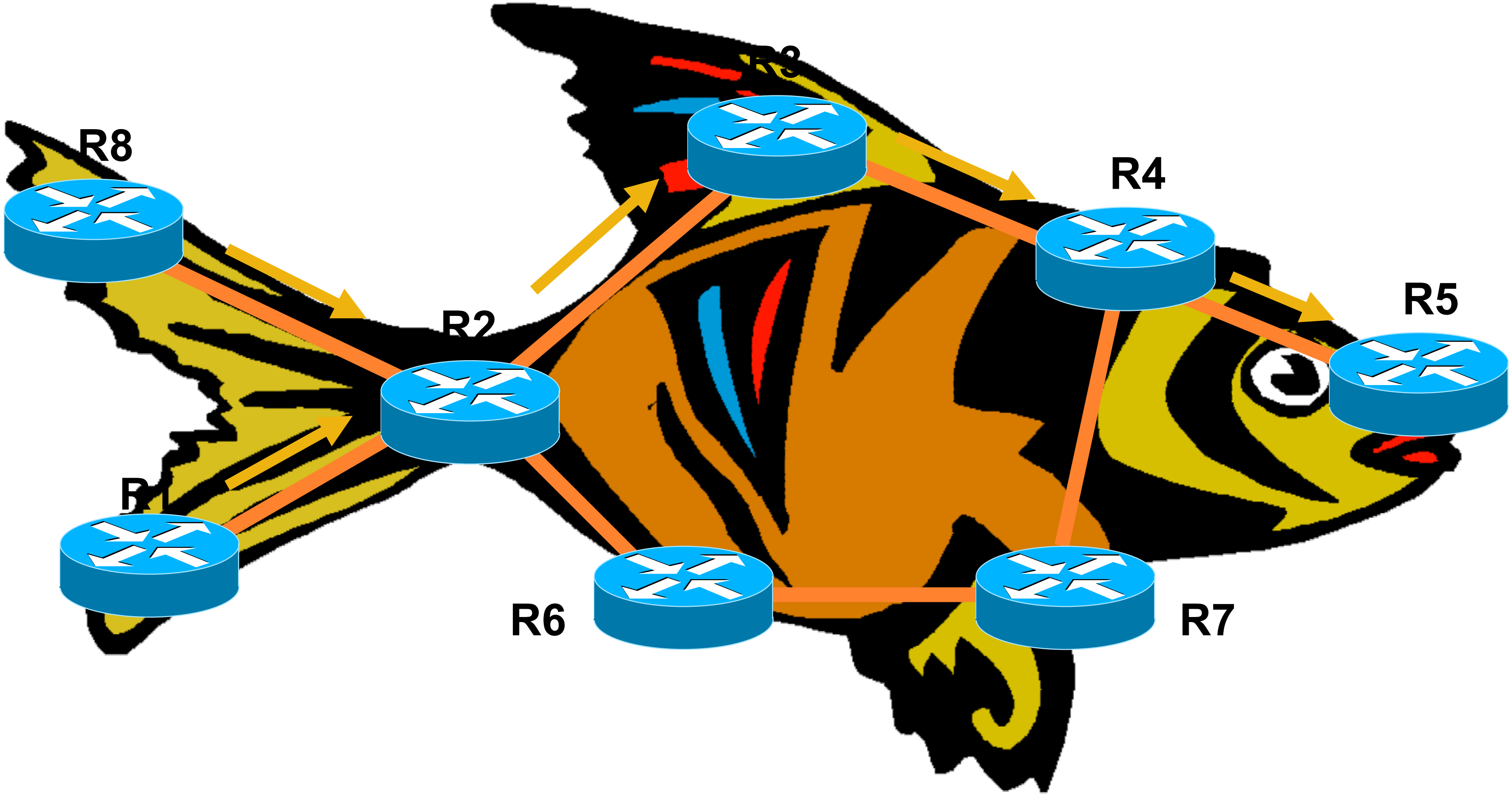
Motivation for MPLS Traffic Engineering

- Reduce the overall cost of operations by more efficient use of bandwidth resources
- Ensures the most desirable/appropriate path for certain traffic types based on certain policies
- MPLS FRR
- The ultimate goal is COST SAVING

MPLS 2004

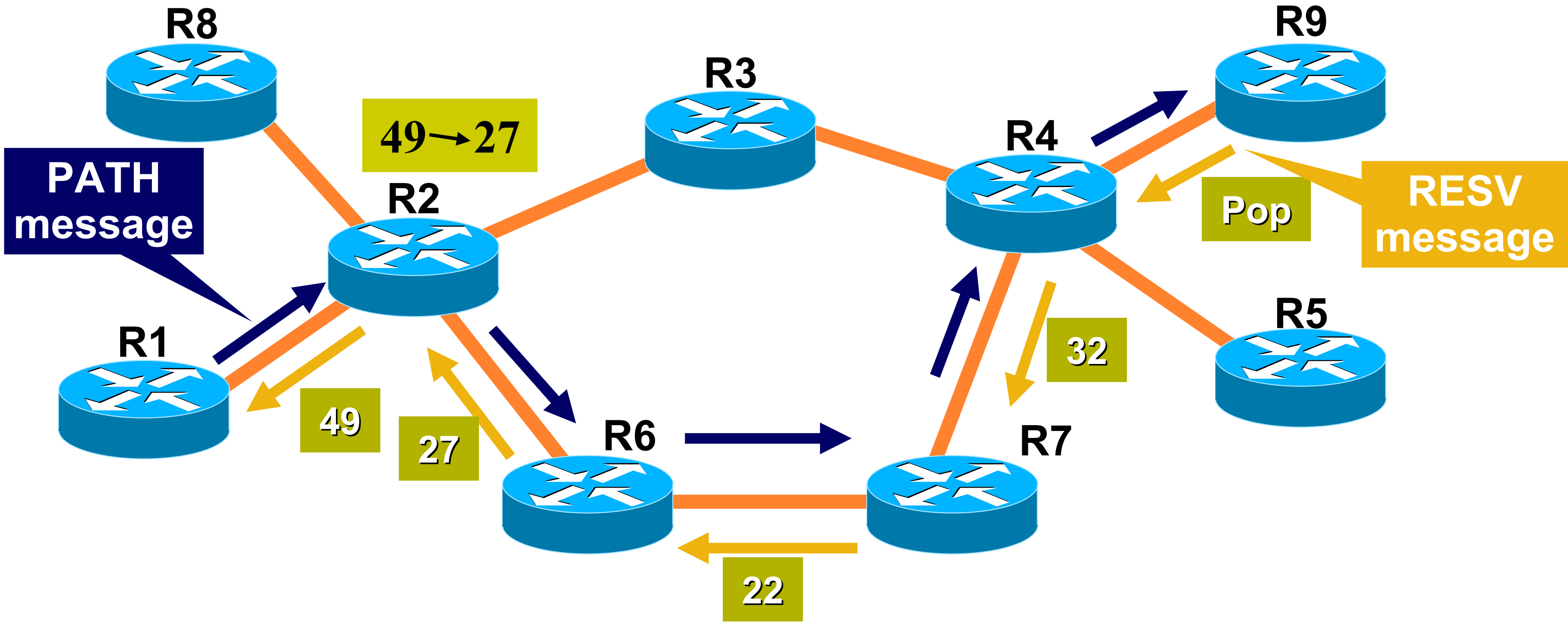
CISCO SYSTEMS

The “Fish” Problem (Shortest Path)



- IP uses shortest path destination-based routing
- Shortest path may not be the only path
- Alternate paths may be under-utilized
- Whilst the shortest path Is over-utilized

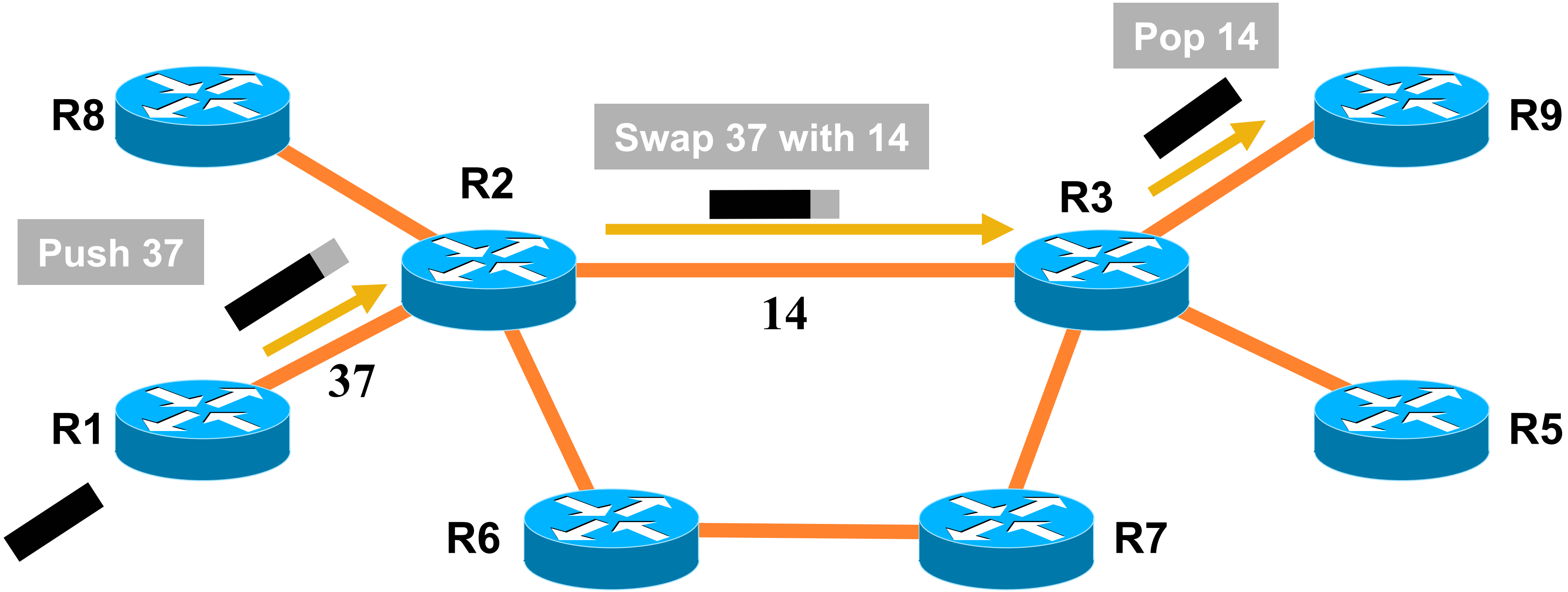
Traffic Engineering Tunnel Creation



⟶ RSVP PATH: R8 ➔ R2 ➔ R3 ➔ R4 ➔ R5

⟶ RSVP RESV: RSVP communicates labels and and reserves bandwidth on each link

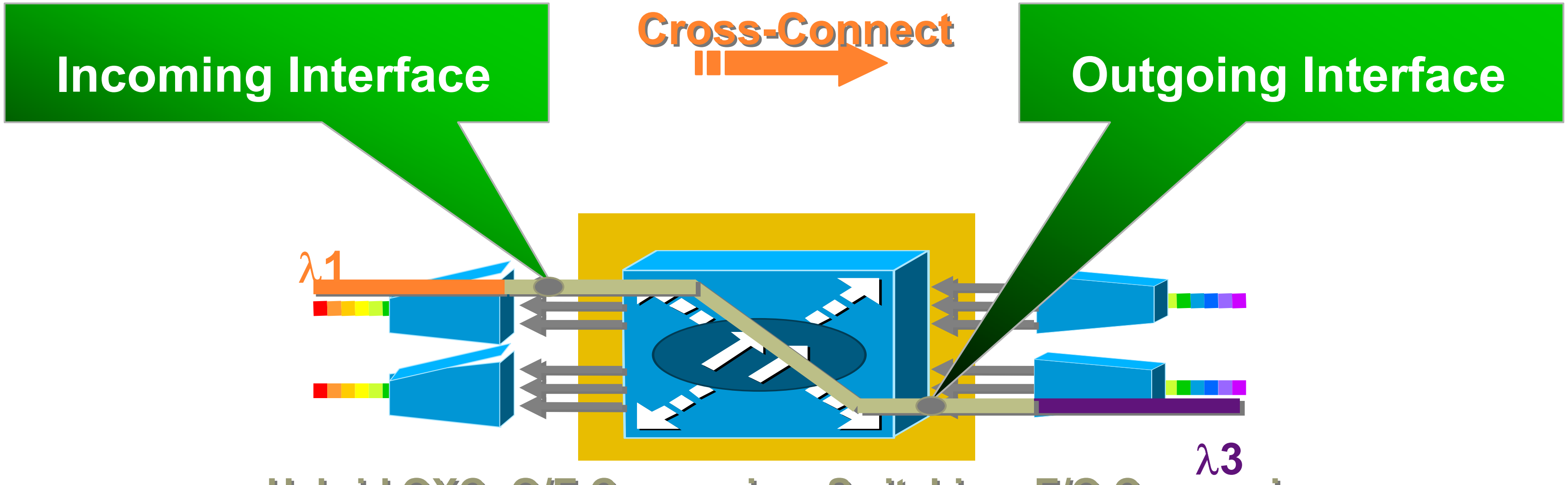
Traffic Forwarding over an MPLS LSP



Outline

- Principles of MPLS-TE
- Extending the Concepts to GMPLS
- Fundamental Concepts
- Implementing and Deploying MPLS-TE and GMPLS
- Inter-Domain Traffic Engineering
- Components of MPLS/ GMPLS High Availability
- MPLS O&M
- Future Work

LSR and OXC Similarities: Birth of G-MPLS

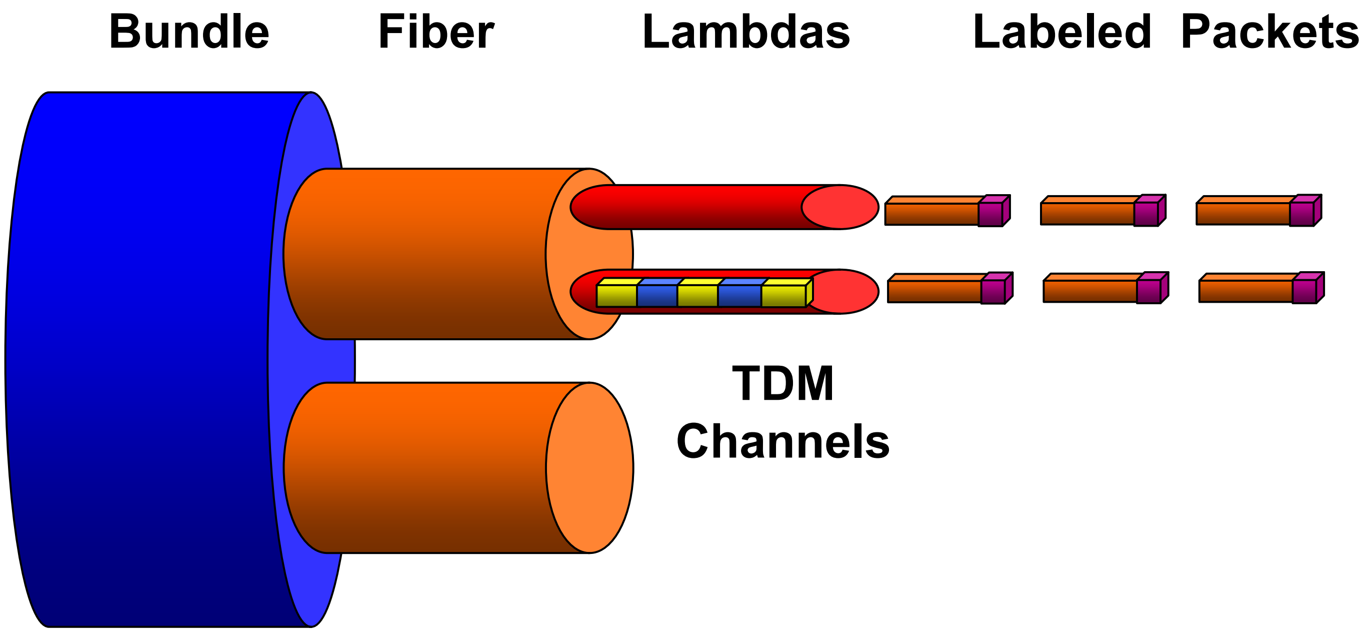


Hybrid OXC: O/E Conversion, Switching, E/O Conversion
Pure OXC: Lambda Conversion

- Data plane driven by a switching matrix
  - LSR: (i_if, ingress label) => (o_if, egress label)
  - OXC: (i_if, ingress λ) => (o_if, egress λ)

A Label is a label

GMPLS Label Hierarchy

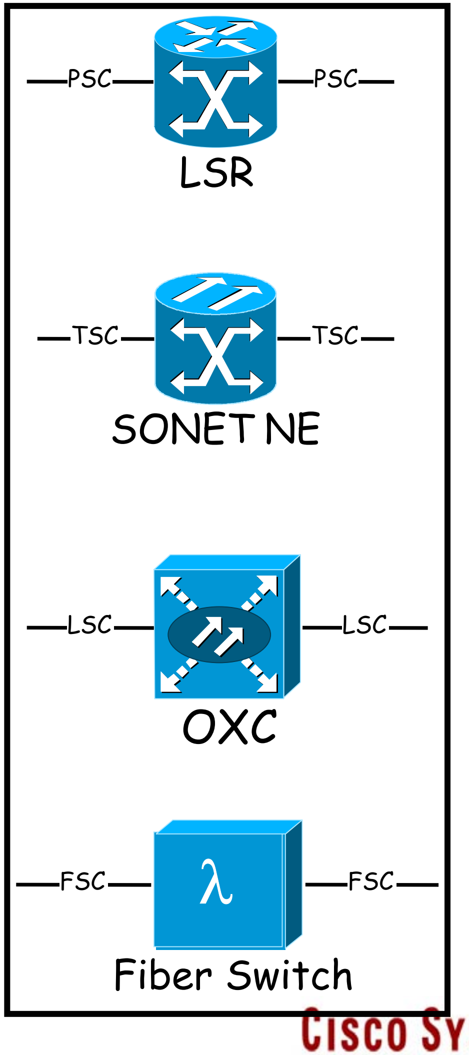


Generalized MPLS (GMPLS)

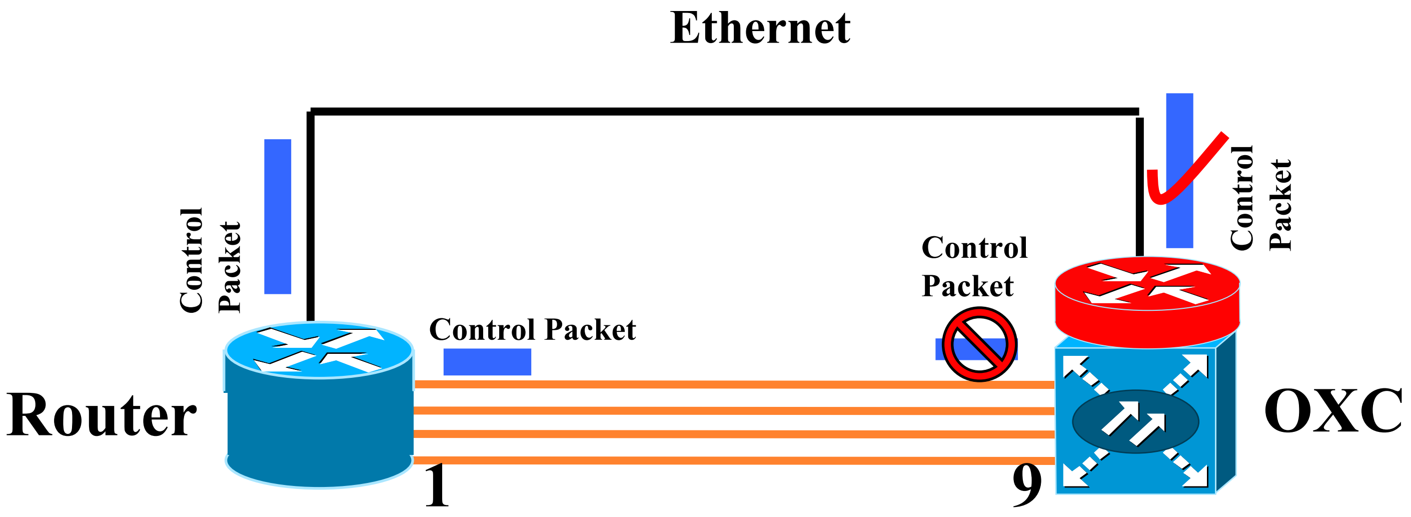
- **MPLS control plan extended for circuits, lambdas, fiber and ports.**

Support for New Type of Devices

- Need support for devices that make forwarding decision on other than packet/cell boundaries
- Unified Control Plane for the following type of devices/ interfaces:
  - packet-switch capable (PSC)
  - Layer2-switch capable (L2SC)
  - TDM switch capable (TSC)
  - Lambda switch capable (LSC)
  - Fiber switch capable (FSC)

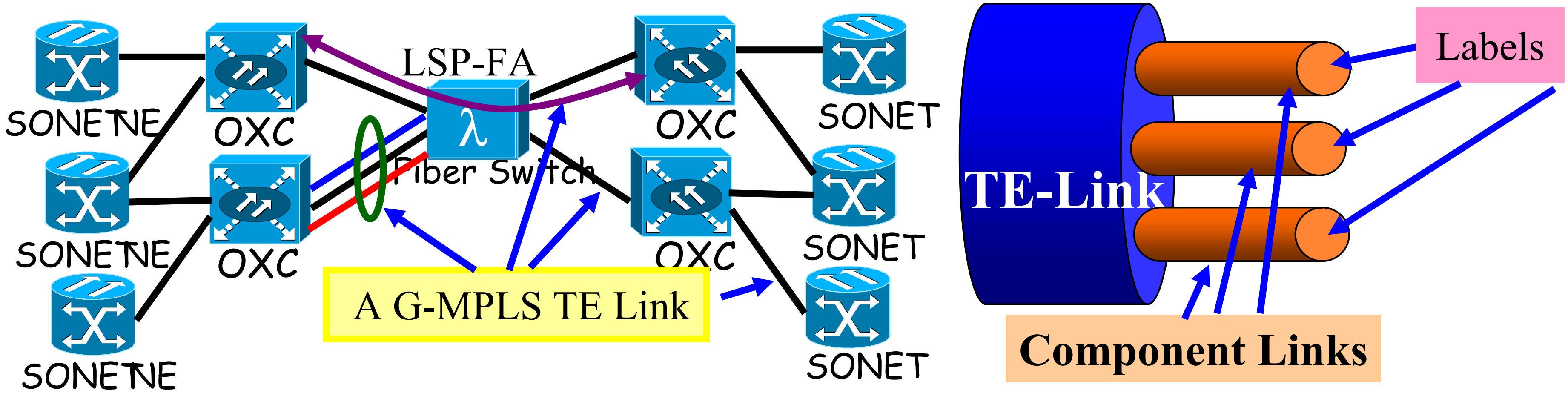


Need for Separation of Control and Data Planes?



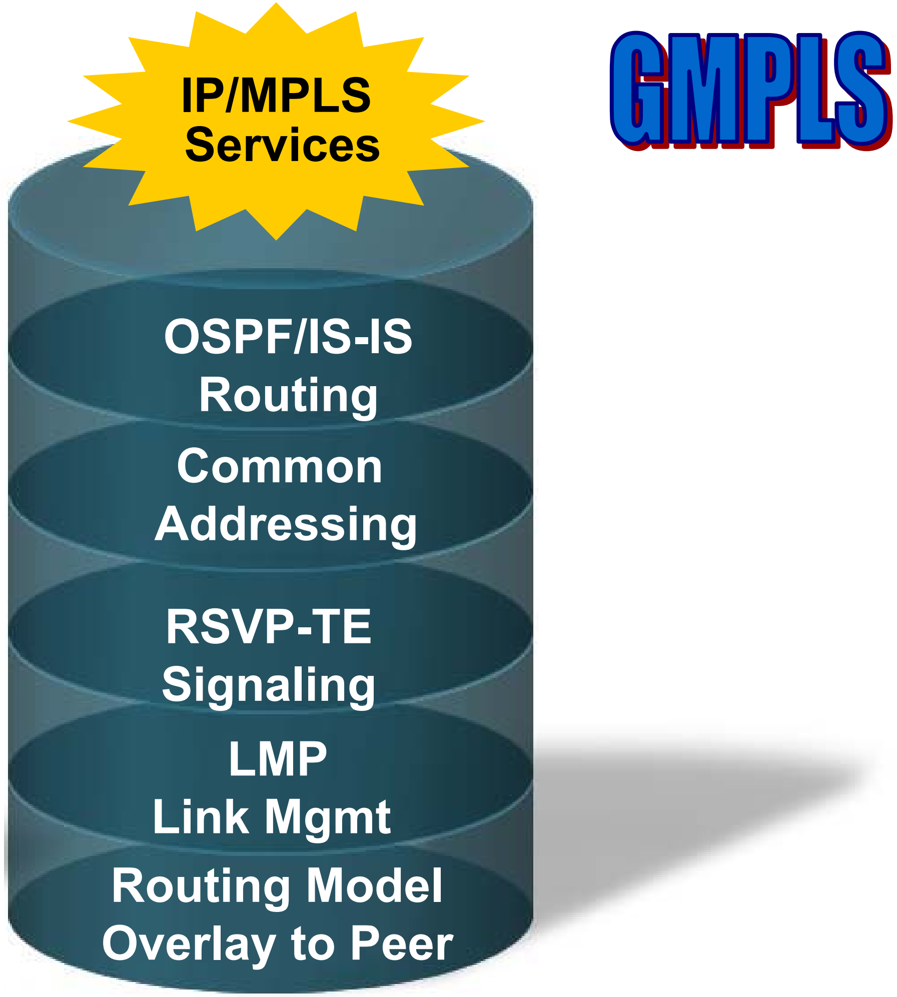
- Inability of optical devices to terminate data links.
- Solution: Add a router blade to the optical devices.

GMPLS TE Links Vs. MPLS TE Links



- **A routing adjacency cannot be brought up on optical (non-packet) links**
  - **A TE link between a pair of LSRs doesn't imply the existence of a routing adjacency (e.g., when TE link is a Forwarding Adjacency).**

GMPLS Building Blocks



IP/MPLS Services

OSPF/IS-IS Routing

Common Addressing

RSVP-TE Signaling

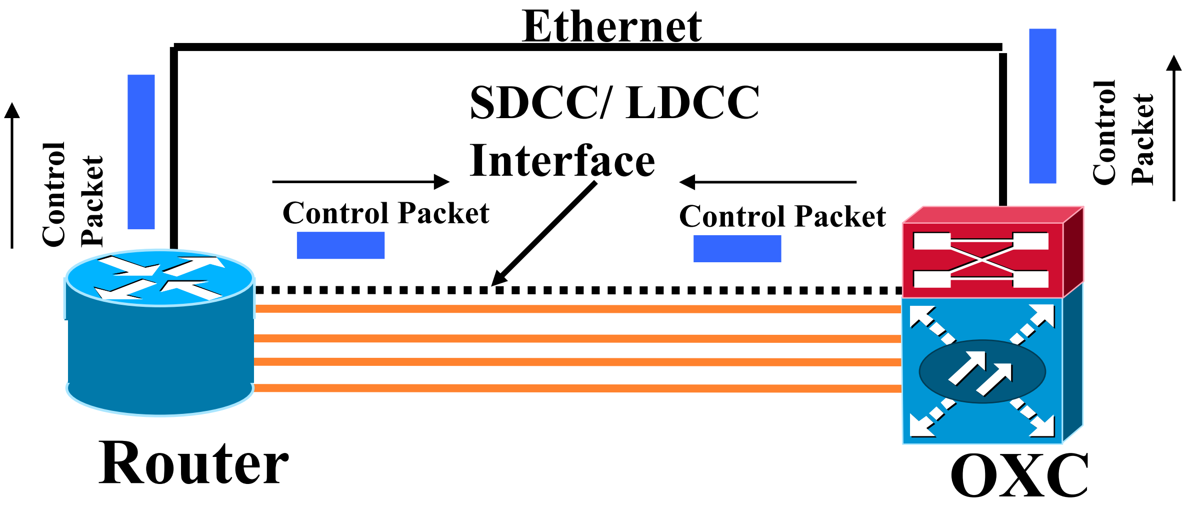
LMP Link Mgmt

Routing Model Overlay to Peer

GMPLS

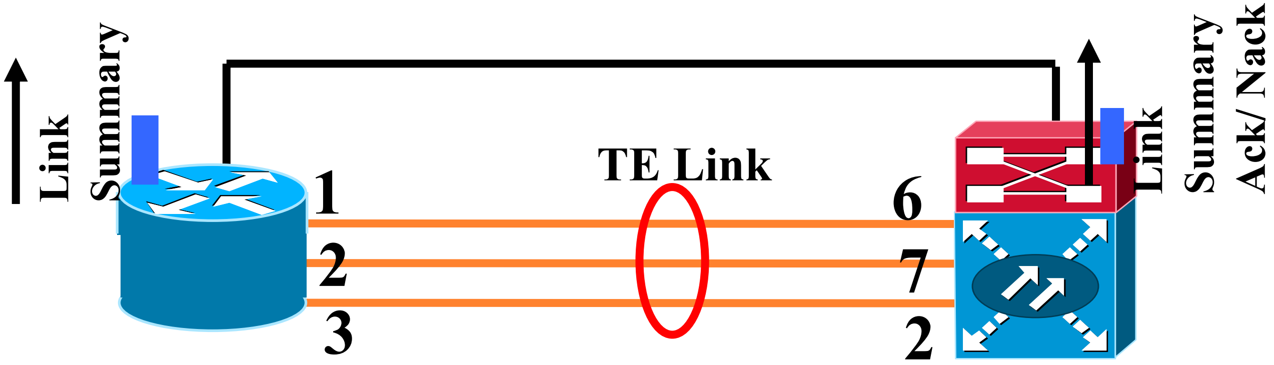
MPLS 2004

LMP Functionality:
Control Channel Management



- Bi-directional control channel(s) between two neighbors.
- Control Channel implementation is unrestricted.
  - OF/ OB (e.g., Ethernet, POS, etc.).
  - IF/ IB (e.g., SONET SDCC/ LDCC).
  - IP-in-IP, GRE tunnels, etc.

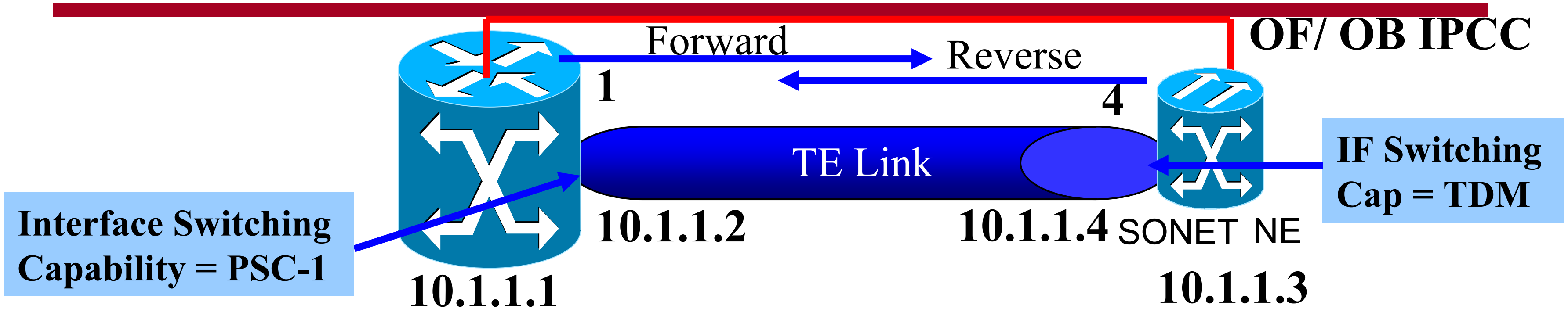
LMP Functionality:
TE Link Property Correlation



- Non-applicability of IGP to exchange TE Link properties, e.g., local and remote interface addresses, etc.
- New TE Link attributes to be exchanged.
- Configuration or LMP.

- **Synchronizes link state:**
  - **TE Link ID**
  - **Component Interface Id mapping.**
  - **Link properties, e.g., link mux type, encoding, protection.**

New IGP Parameters of TE Link in G-MPLS



Link Protection Type

Used to compute paths with the desired protection type.

Extra Traffic, Unprotected, Shared, Dedicated 1:1, 1+1, Enhanced

Source: LMP/ configuration.

Shared Risk Link Group (SRLG),

Physical fiber diversity - e.g. two fibers with same SRLG are in the same conduit

Source: Configuration.

Link Descriptor

Link encoding type and bandwidth granularity. Source: LMP

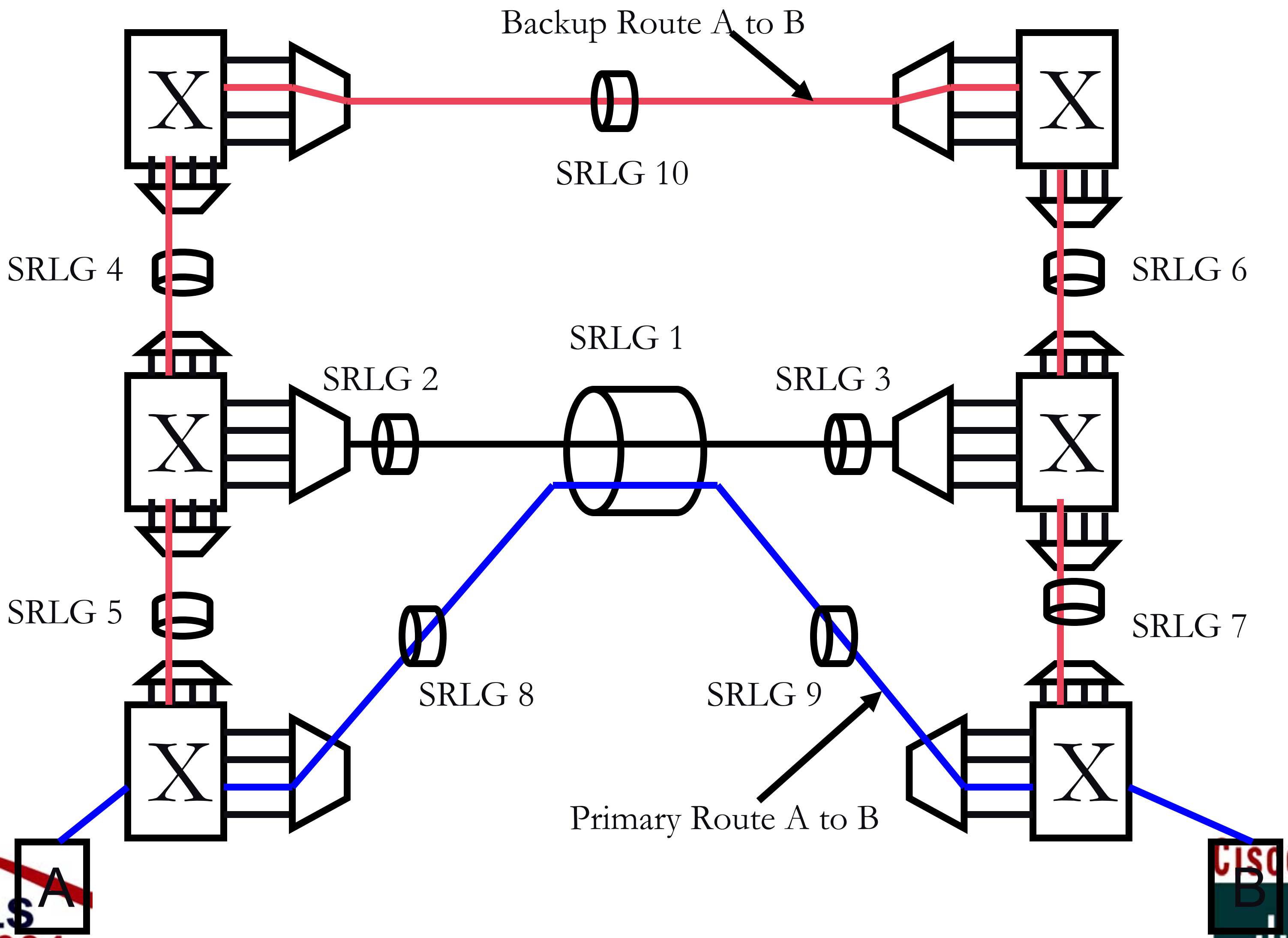
Interface Switching Capability

Defines the receiving nodes ability to demultiplex data based on packets, TDM timeslots, lambdas or fiber.

Source: LMP (for optical links).

draft-ietf-ccamp-gmpls-routing-09.txt
draft-ietf-ccamp-ospf-gmpls-extensions-12.txt
draft-ietf-isis-gmpls-extensions-19.txt

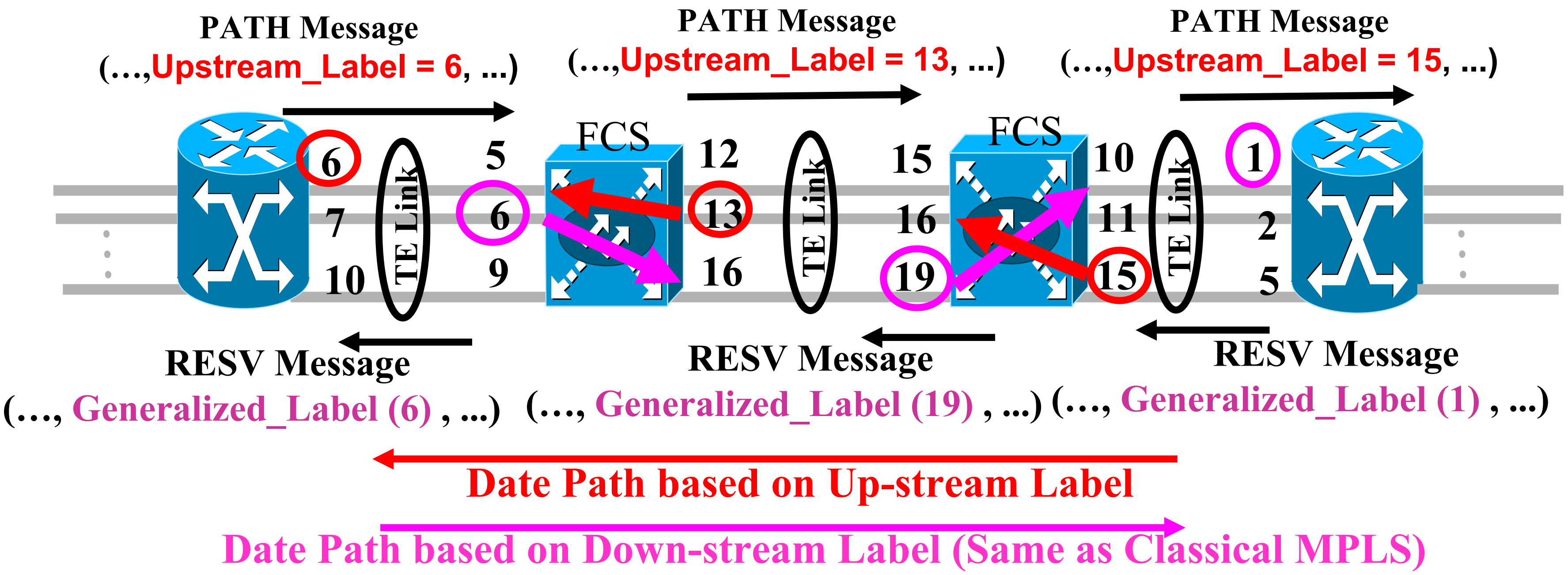
Example of SRLG disjoint Paths



Generalized MPLS Signaling

- RFC 3471 and 3473 (RSVP-TE).
- Extended label semantics for TDM, Lambda, Waveband and Fiber Labels.
- Extend RSVP-TE to support new label objects over explicit/non explicit path.
- Bidirectional LSP setup.
- Signaling with desired protection attributes.
- Label Set - limits choice of labels that downstream LSR can choose from.
  - If no wavelength conversion available then same lambdas must be used, etc.
- Reducing set-up latency: Suggested Label.
- RSVP error notification.
- Egress Control.

Bi-directional LSP Setup


PATH Message
(…,Upstream_Label = 6, ...)
PATH Message
(…,Upstream_Label = 13, ...)
PATH Message
(…,Upstream_Label = 15, ...)
6
7
10
TE Link
5
6
9
FCS
12
13
16
TE Link
15
16
19
FCS
10
11
15
TE Link
1
2
5
RESV Message
(…, Generalized_Label (6) , ...)
RESV Message
(…, Generalized_Label (19) , ...)
RESV Message
(…, Generalized_Label (1) , ...)
Date Path based on Up-stream Label
Date Path based on Down-stream Label (Same as Classical MPLS)

- Upstream and downstream labels (generalized label object) are mandatory for bidirectional LSP setup.

MPLS 2004

Cisco Systems

TE Link and GMPLS Protocols

TE link Resources/ Protocols	TE Link ID	Component Link ID	Labels
IGP (OSPF/ ISIS)	X		
LMP	X	X	
RSVP-TE	X	X	X

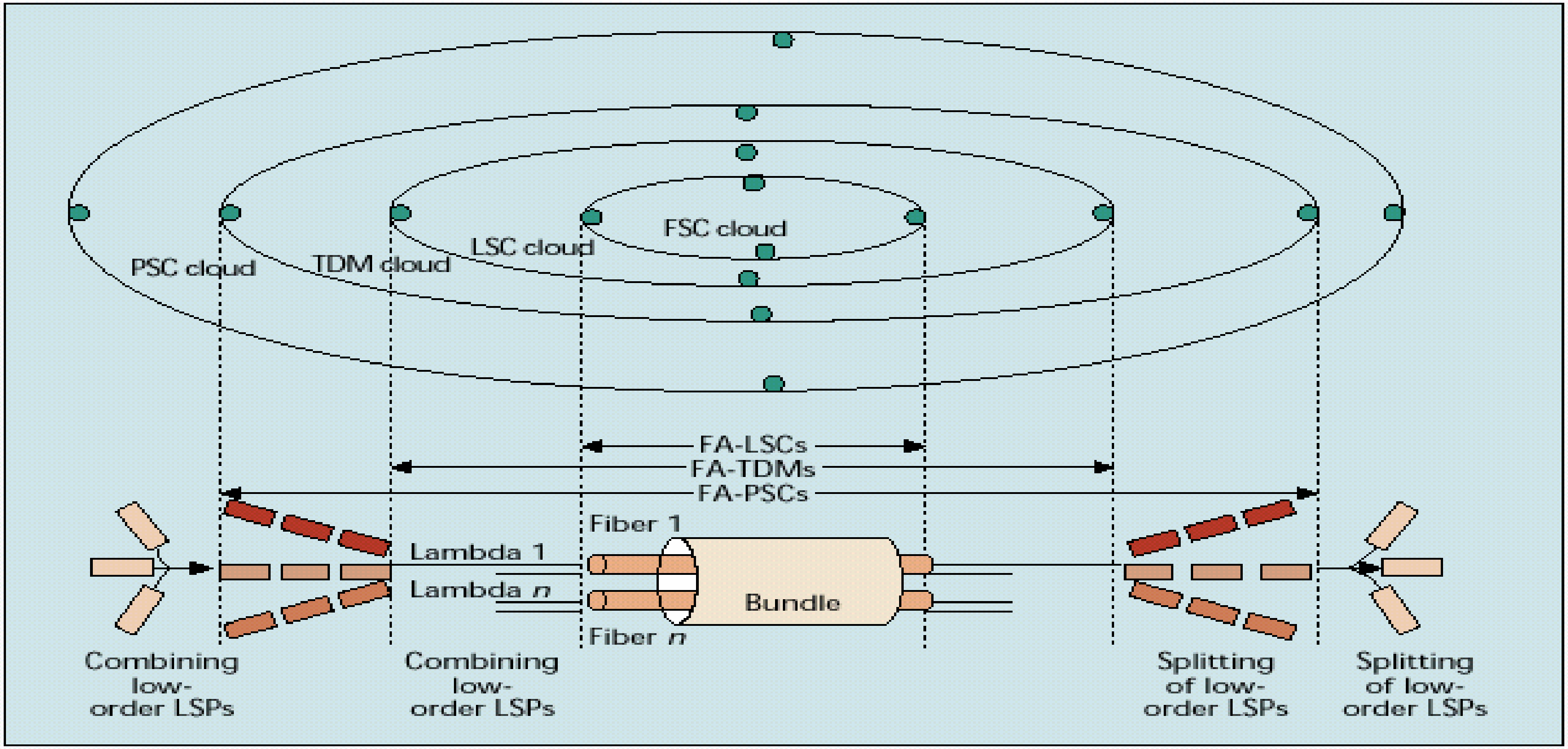
Summary for whereabouts of TE Link Attributes

Attribute	*LMP*	*IGP*	*RSVP*
Mux Type	*Link Mux Type*	Interface Switching Capability	LSP Switching Type (PSC-1, TDM, LSC, etc.)
	Link Summary Message	Link TLVs	Generalized Label Request Object
Encoding	*TE Link Type*	Link Encoding Type	LSP Encoding Type (SONET ANSI T1.105, Ethernet 802.3, etc.)
	Link Summary Message	Link Descriptor TLVs	Generalized Label Request Object
G-PID	*Not a TE Parameter*	*Not a TE Parameter*	Generalized Protocol ID (G-PID), which is concerned with the End Points.
	Not a TE Parameter	*Not a TE Parameter*	Generalized Label Request Object.
Protection	*Protection Type*	*Link Protection*	LSP Protection
	Link Summary Message	Link TLV	Protection Object
SRLG	*Local value only*	*SRLG*	Only important during ERO computation.
	Local value only	Link TLV	Not signaled.

Outline

- Principles of MPLS-TE
- Extending the Concepts to GMPLS
- Fundamental Concepts
- Implementing and Deploying MPLS-TE and GMPLS
- Inter-Domain Traffic Engineering
- Components of MPLS/ GMPLS High Availability
- MPLS O&M
- Future Work

LSP Hierarchy

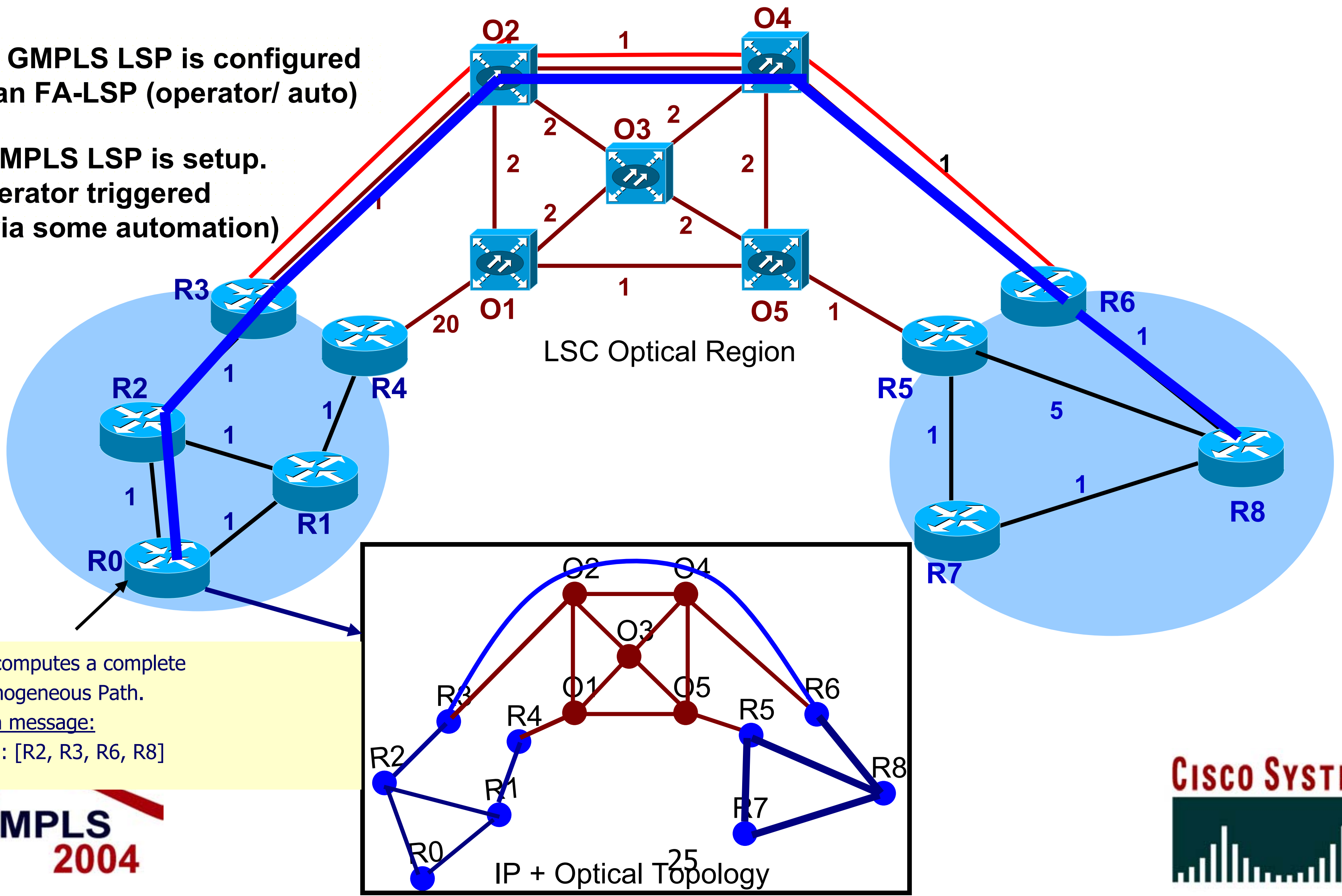


- Enables aggregation of GMPLS LSP tunnels
  - **Fewer high-order labels (e.g.lambdas) consumed**
  - **Nested LSPs can be of non-discrete bandwidth**
  - **FA-LSP can "hide" topology**
- draft-ietf-mpls-lsp-hierarchy-08.txt

Notion of Forwarding Adjacency

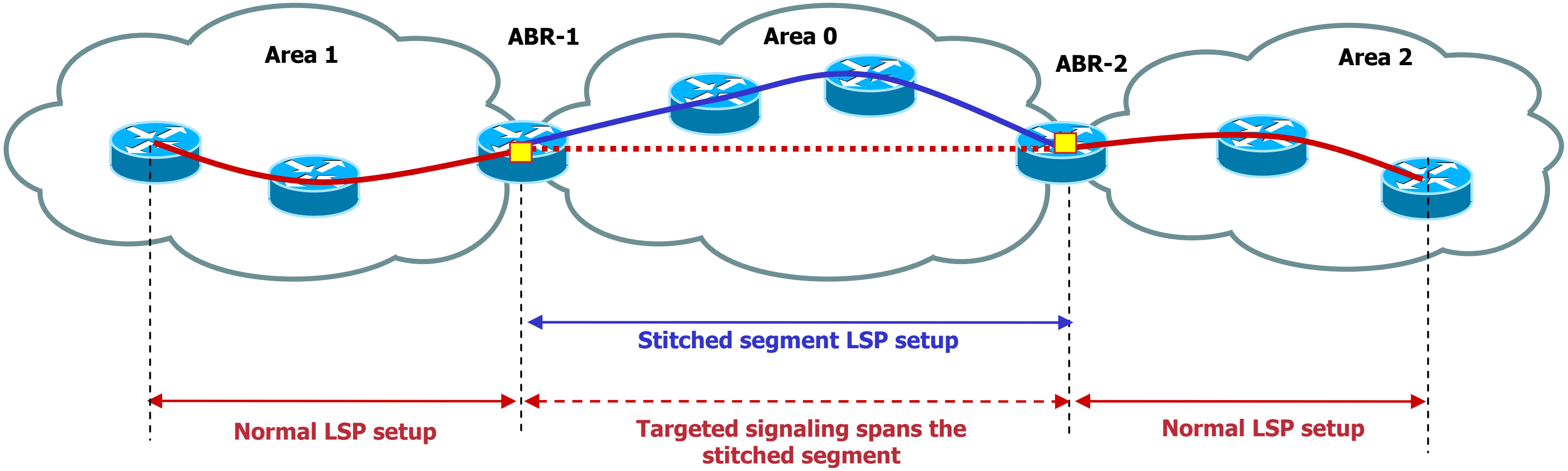
The GMPLS LSP is configured
As an FA-LSP (operator/ auto)

A GMPLS LSP is setup.
(Operator triggered
or via some automation)



R0 computes a complete
Homogeneous Path.
Path message:
ERO: [R2, R3, R6, R8]

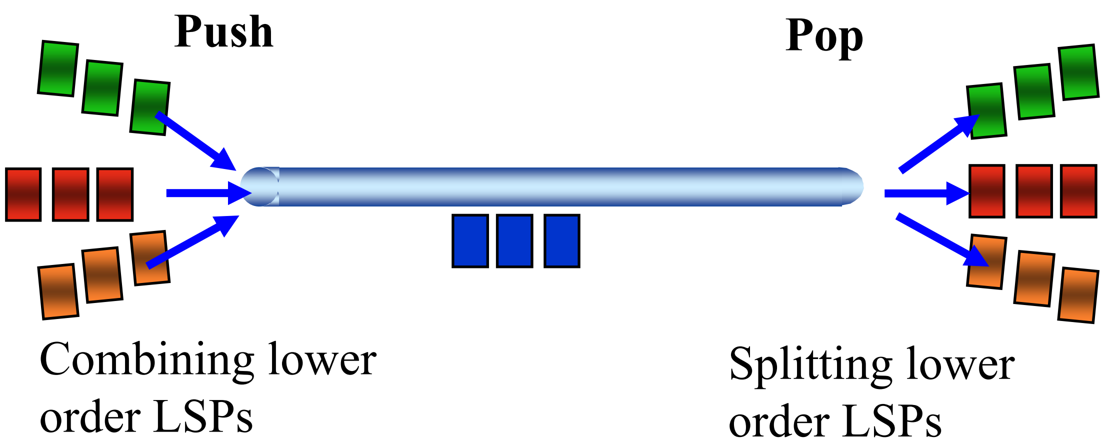
LSP Stitching



Area 1
ABR-1
Area 0
ABR-2
Area 2
Stitched segment LSP setup
Normal LSP setup
Targeted signaling spans the stitched segment
Normal LSP setup

- **LSP Stitching is where two or more contiguous LSP segments are joined together to form an end-to-end LSP.**
- **In this example, the ABRs are responsible for "cross-connecting" the LSP segments.**
- **The central segment is like an FA-LSP but:**
  - Label stacking is not used.
  - There is a one-to-one correspondence between LSP segments.
- **Signaling for the remote segment is carried using targeted signaling just as for hierarchies.**

LSP Hierarchy Vs. Stitching



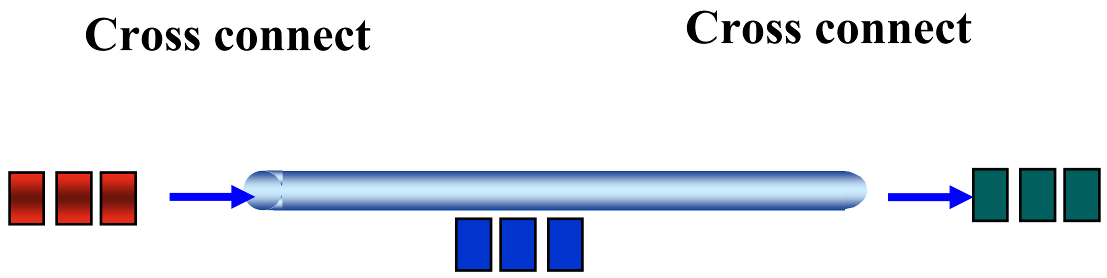
Push

Pop

Combining lower order LSPs

Splitting lower order LSPs

LSP Hierarchy

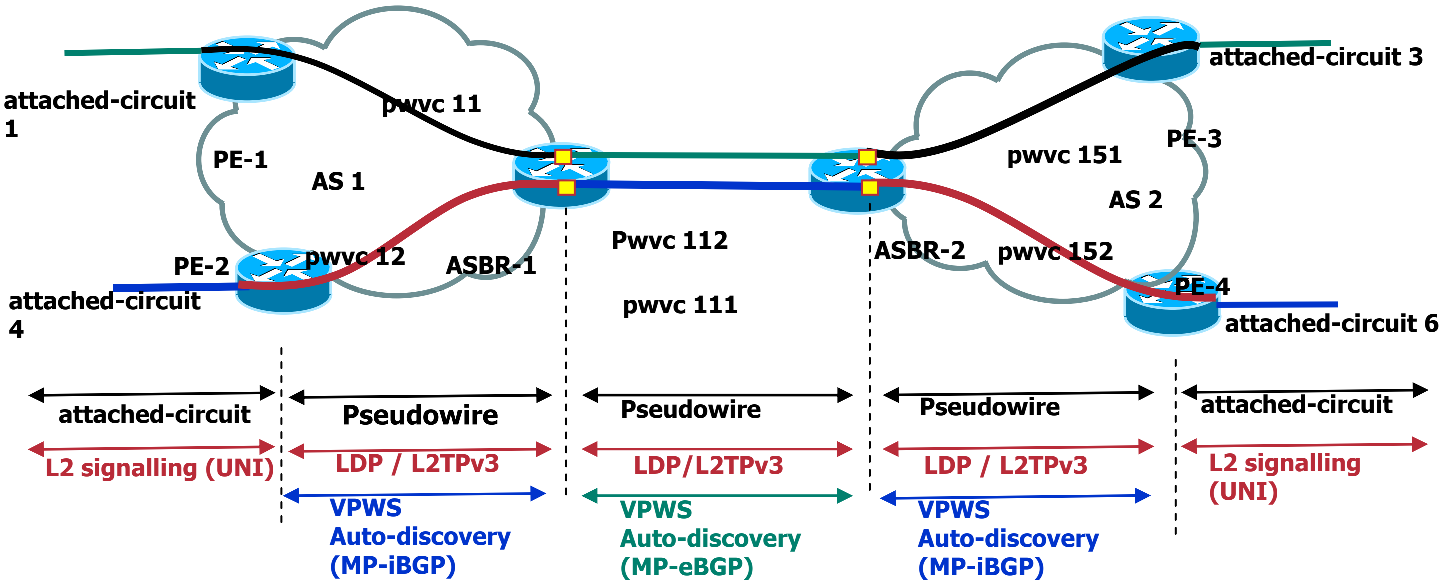


Cross connect

Cross connect

LSP Stitching

Pseudowire Stitching



- The Pseudowire Stitching Model allows service provider(s) to extend an existing pseudowire with another pseudowire.
- Each pseudowire segment can independently employ draft-martini or L2TPv3 signaling and encapsulations.
- The ASBRs are responsible for "cross-connecting" the pseudowire control channels and pseudowire data planes.

Outline

- Principles of MPLS and MPLS-TE
- Extending the Concepts to GMPLS
- Fundamental Concepts
- Implementing and Deploying MPLS-TE and GMPLS
- Inter-Domain Traffic Engineering
- Components of MPLS/ GMPLS High Availability
- MPLS O&M
- Future Work

Implementing and Deploying MPLS-TE and GMPLS

- Software Components
  - Thoughts about how to construct a system
- Common Implementation Issues
  - Random thoughts about various issues
- Deployment Scenarios
  - Some general ways that MPLS-TE and GMPLS might be deployed

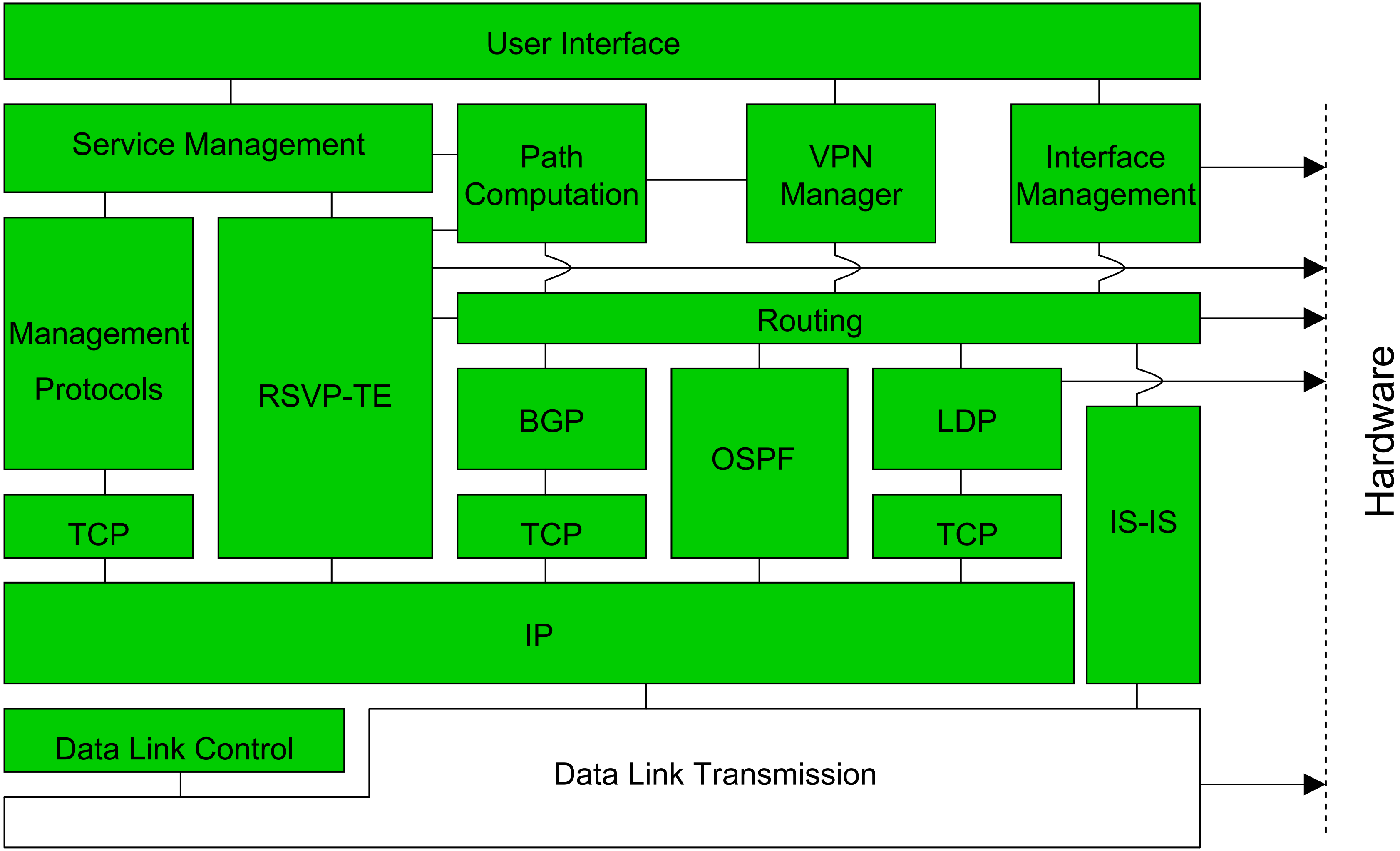
Implementation Homilies

- Modular code is easier to implement and test
  - Flexible addition or replacement of function
    - Swap between OSPF-TE and ISIS-TE
    - Add BGP
  - Maintenance, upgrades and bug fixes
  - Modules may map to:
    - Individual protocols (RSVP-TE, OSPF, etc.)
    - Logical units (protocols on interfaces, TED maintenance)
    - Specific functions (neighbor keepalive, label space management)
- Long code paths can block high priority work
- Multi-threading modules can cause:
  - OS churn
  - Additional code to protect data structures
  - Lock contention

Relationships Between Components

- Components need to exchange information
- Function calls (APIs) are fine, but:
  - They can lead long, synchronous code paths
  - They can make components re-entrant
- APIs with locks can work, but:
  - Still have long code paths
  - Enormous risks of lock contention on long code paths
- Message-passing through message queues works well
  - It does increase the code path slightly
  - It can be common code hidden by an API (as in most OSs)
  - It makes for easy shuffling of modules (including distribution across different hardware components)
  - Improved work granularity and prioritization
- Simplicity is usually a benefit

A Possible Decomposition



Implementation Issues: Where is My Label Space?

- Or maybe the question is: what is my label space?
- Packet Switch Capable (PSC)
  - MPLS or GMPLS
  - Label is simply an identifier and can be global or pre-interface
    - Even per-interface labels can be managed from a global pool
  - If the label is 'borrowed' from the transport technology (e.g. ATM or Frame Relay) it will be a per-interface label
- In other technologies a label ***is*** a resource.
  - Means that labels must be pre-interface
  - But port/interface labels are clearly global
  - Label is usually scoped in the downstream node's context
    - Even for bi-directional LSPs
    - TDM labels have clearer definitions

Implementation Issues: Managing the Product

- The finished product must be manageable!
- Usually through CLI
    - Also SNMP, CORBA, XML, TL1, etc., etc.
    - Many data schemas are ‘standardized’
- Useful to design with a common management interface
    - All protocols and user interfaces map to this
    - Important to get the managed objects right
        - Look at existing MIBs and schemas
        - Work out the CLI commands in advance
- Must have relatively rapid response to users
    - Must not block protocol operations
        - Some ‘display’ commands require a lot of processing
- Management and Control planes should be mutually survivable

Implementation Issues: When is Routing not Routing?

- The Traffic Engineering Database (TED) is not a routing table
- Computations
  - SPF computations are performed periodically to build a routing table
  - CSPF computations are done on demand and are resource-intensive
- The TED is built from topology information
  - Assumed to be distributed by extensions to the routing protocols
  - Much more detailed than simple link state database
- Path computation should be achieved using a separate module

Implementation Issues: You Want *How Many* LSPs?

- Scalability must be designed into the implementation
- Scaling issues:
  - Data occupancy per LSP
  - Speed of searching data structures
  - Basic background processing
    - Per neighbor Hello processing (every three seconds?)
    - Per LSP soft state processing
      - Refresh Reduction
      - Checksum applied to state

Implementation Issues: Achieving Scalability

- Prioritizing work
  - Small granularity work items
  - Preemptable tasks
- Load sharing/distribution
  - Multi-CPU cards or multiple CPUs (CPUs on line cards?)
  - What work should I off-load?
- Cost-benefit analysis
  - What scaling do I need to achieve?
  - Will distribution really help?

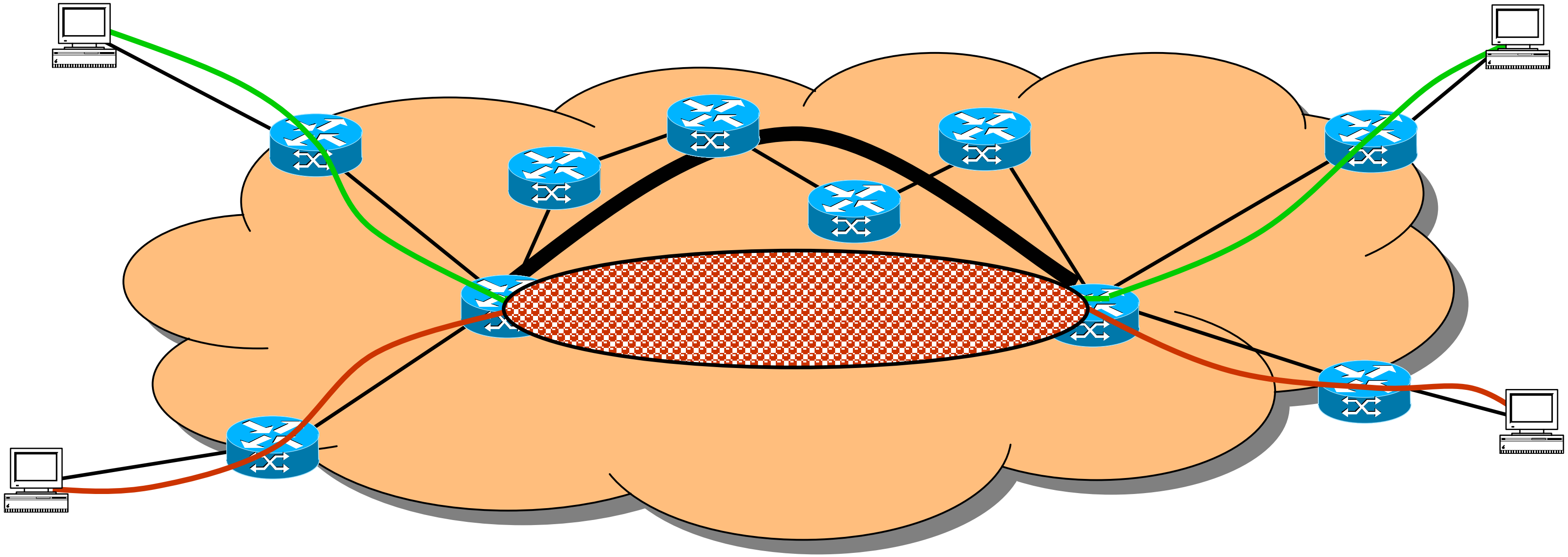
MPLS 2004

Implementation Issues: Fault Tolerance

- What are you trying to achieve?
  - “Carrier class” and “five-nines” are sometimes over-used terms
  - Routers have a poor reputation for software quality and customers need comforting
- Many options
  - Data plane survives control plane failures
  - Control plane recovers and resynchs with data plane
  - Control plane switches seamlessly to back-up instance
  - Data plane switches seamlessly to back-up
  - Service-level protection and restoration
- Cost-benefit analysis
  - Carrier class control plane is very expensive to develop and nearly impossible to test
  - Full data plane redundancy is very expensive
  - Packet and transport networks have different requirements
- Many systems choose
  - Separation of control/data plane
  - Rapid recovery and resynch of control plane
  - Service-level protection

Deployment: Basic Traffic Engineering

- Well-understood technique to improve network efficiency, increase traffic performance, reduce costs, and increase profitability.
- Increasingly achieved through MPLS

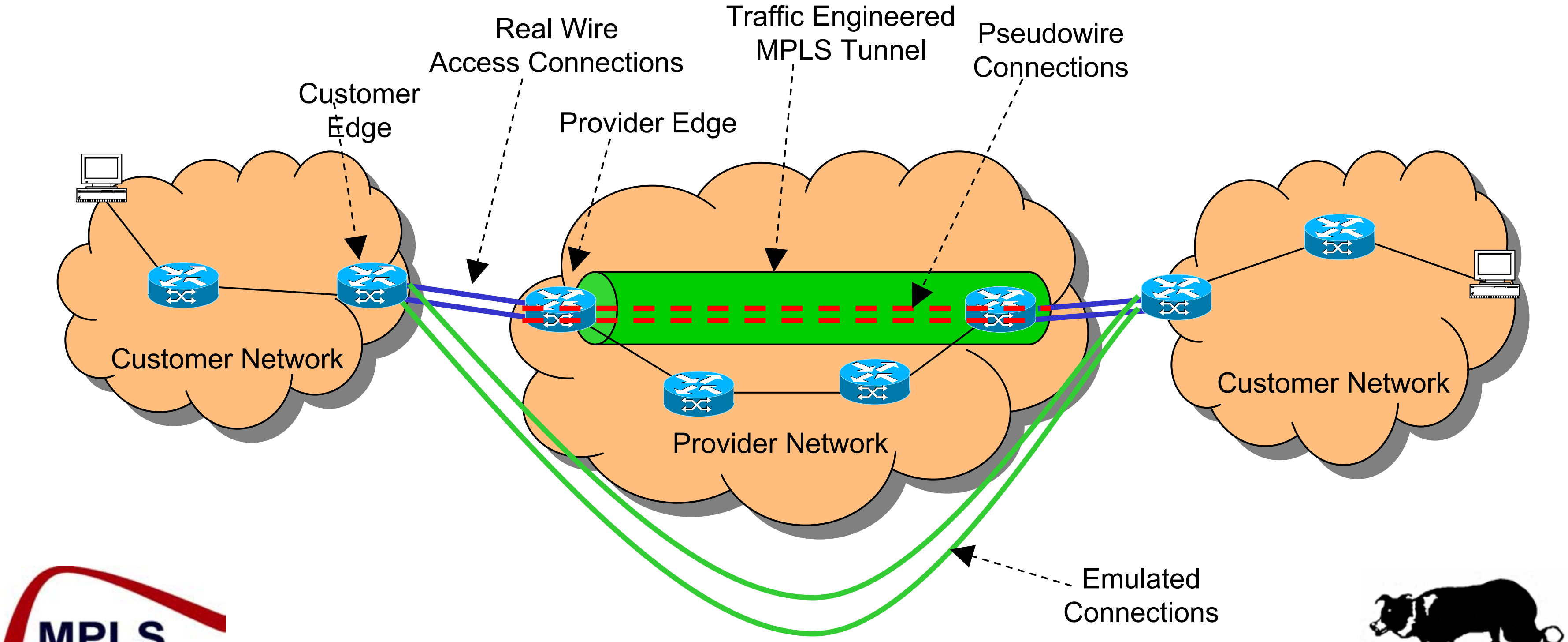


Traffic Engineering Deployment Options

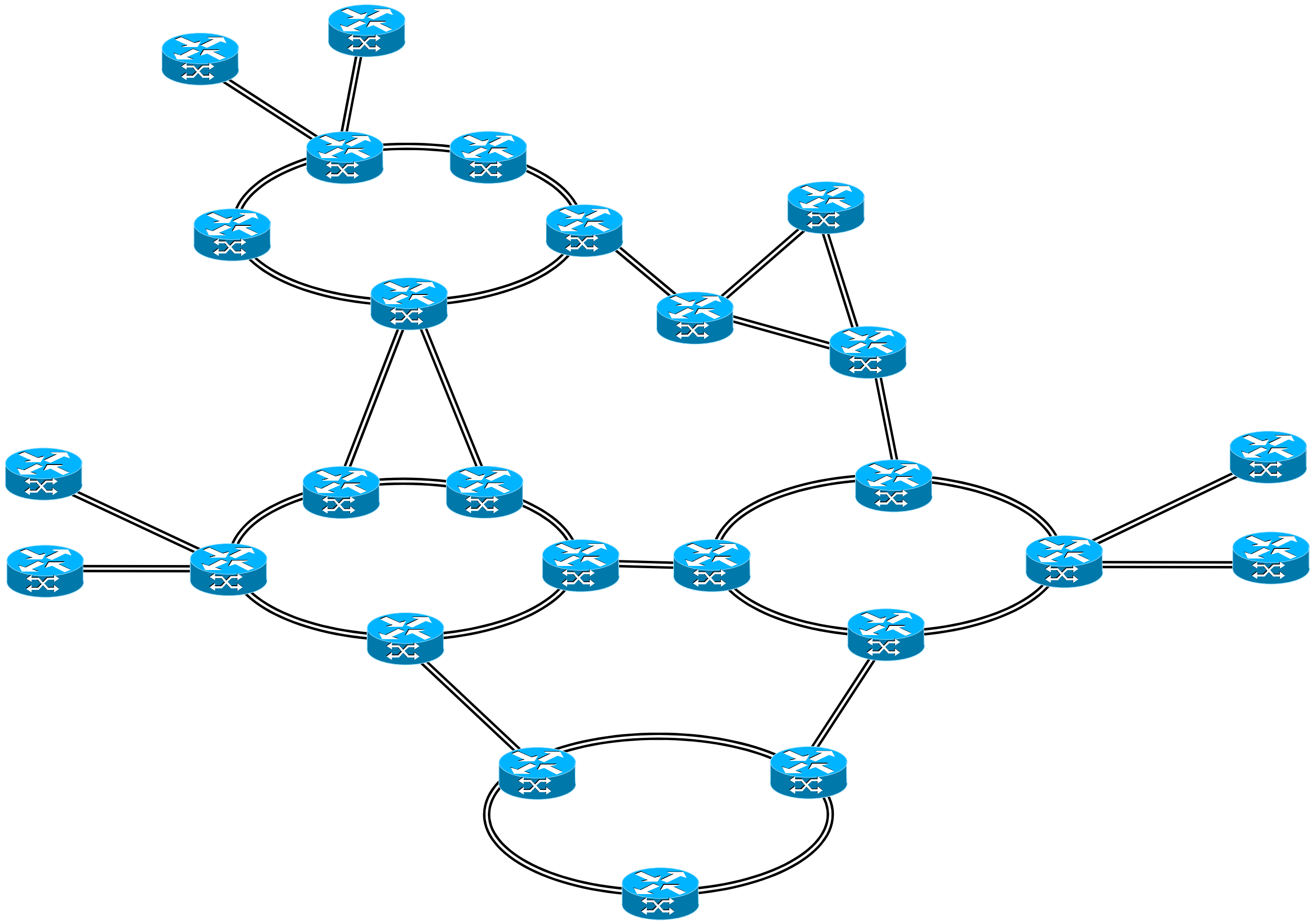
- Placement of tunnels can be automatic or under operator control
- Full mesh
  - Connect all edge nodes
  - A management nightmare
- Partial mesh
  - How do you decide which edge nodes to interconnect?
- Mesh groups
  - Tell an edge node that it is in a group and let it get on with it
- Management control
  - Careful placement of selected tunnels according to
    - Current hot spots
    - Known current and future traffic flows
- On demand
  - Automatically triggered by network congestion

Deployment: Pseudowire and Private Wire

- Pseudowire emulates a data service over MPLS packet switching
- Pseudowires are often (usually) carried across the network using MPLS-TE tunnels



Deployment: Optical Rings



Outline

- Principles of MPLS and MPLS-TE
- Extending the Concepts to GMPLS
- Fundamental Concepts
- Implementing and Deploying MPLS-TE and GMPLS
- Inter-Domain Traffic Engineering
- Components of MPLS/ GMPLS High Availability
- MPLS O&M
- Future Work

Inter-Domain Traffic Engineering

- What is a Domain?
- GMPLS Switching Regions
- Using Forwarding Adjacencies
  - Hierarchies and Stitching
  - The Virtual Network Topology
- Migrating MPLS to Packet-Switched GMPLS
- Multi-Layered Networks
- Path Computation Elements

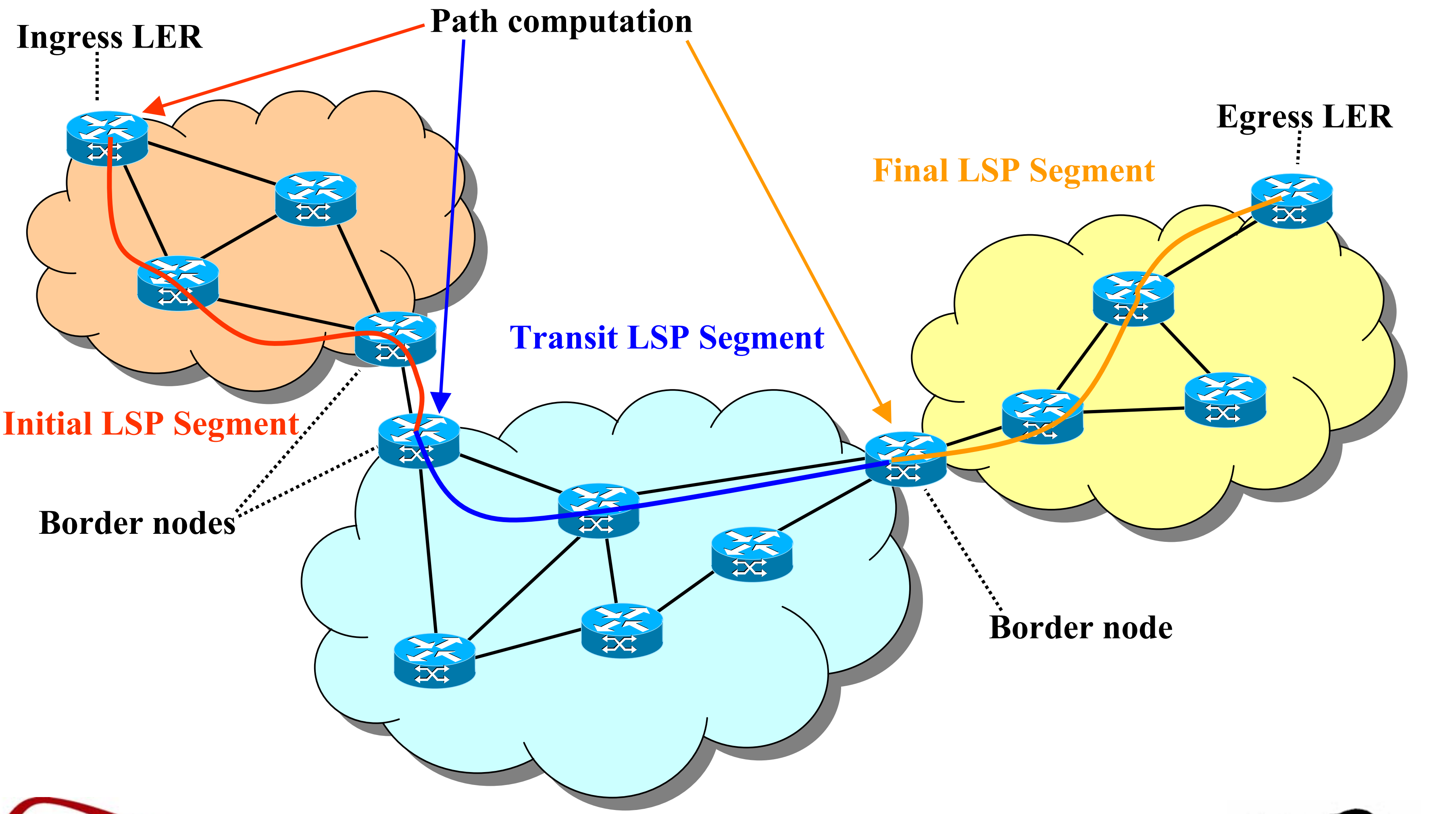
The Domain

- Defined by the IETF’s CCAMP working group in draft-ietf-ccamp-inter-domain-framework as…

  > *Any collection of network elements within a common sphere of address management or path computational responsibility.*

- Classic examples…
  - IGP Areas
  - Autonomous Systems
- More complex examples…
  - Administrative sub-domains of areas or ASs with limited “view” into other domains, or limited responsibility for path computation.

Arbitrary Domain Representation



Path computation

Ingress LER

Egress LER

Final LSP Segment

Transit LSP Segment

Initial LSP Segment

Border nodes

Border node

GMPLS Switching Regions

- An *LSP region* is defined as a collection of LSRs that can support an LSP of a homogenous switching type
  - Example: a SONET ring
  - Example: a mesh of lambda routers
  - Example: a packet switching network
- Modern networks are often *multi-region networks* (MRN)
  - Example: packet network connected by optical network
  - Example: optical network built from SONET and lambda
- Each region is a domain
- Computational visibility is usually limited to the region

Using Forwarding Adjacencies

- Span a domain as a single hop
  - An FA may provide a tunnel to carry multiple LSPs across a domain
    - The FA is advertised as a TE link
    - May enable full path computation
    - Particularly useful in MRN
      - Switching granularity of transit region is coarser
- Use similar techniques to stitch LSPs at domain boundaries

The Virtual Network Topology

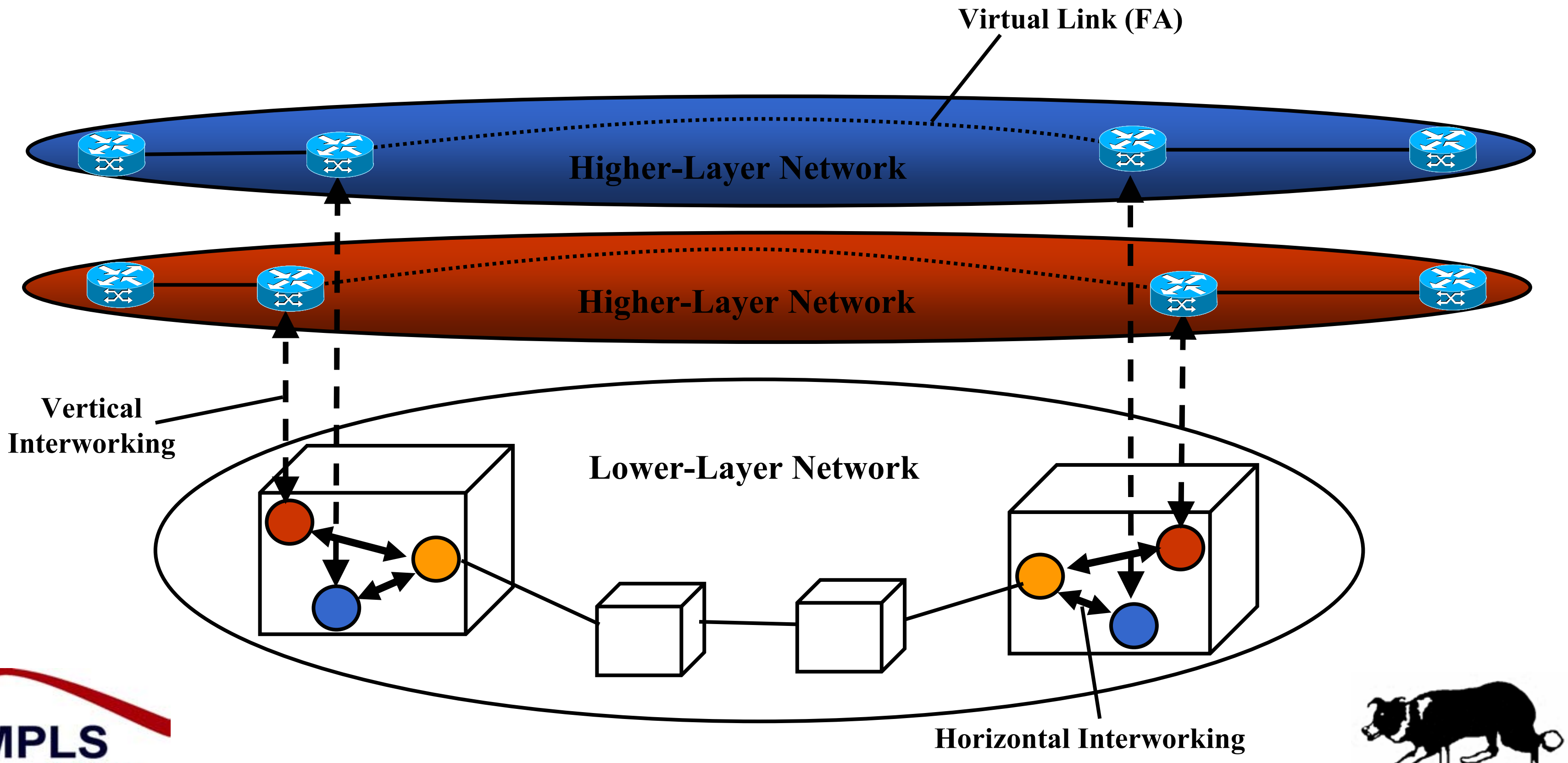
- Forwarding Adjacencies are “virtual” links
- An FA ties up core network resources even when it is not in use
- Wouldn’t it be nice to have on-demand FAs?
  - Could leave it to the border nodes
    - How does the ingress pick the border node?
  - Can advertise the FA TE links but only signal them on demand
    - Core resources used more flexibly
    - Ingress can still do full computation
  - Addition/removal of signaled links under management control

Migrating from MPLS to Packet-Switched GMPLS

- Packet networks are increasingly requiring features of GMPLS signaling:
  - Bi-directional LSPs
  - Extended Hello processing
  - Diverse control and data plane paths
  - Hierarchies and bundles
  - etc.
- Inevitable migration since GMPLS is packet-switch-capable
  - Leakage of features into MPLS routers?
  - New networks deployed as GMPLS?
  - Dual-capable network nodes?
  - Distinct domains of MPLS and GMPLS capability?

Multi-Layered Networks

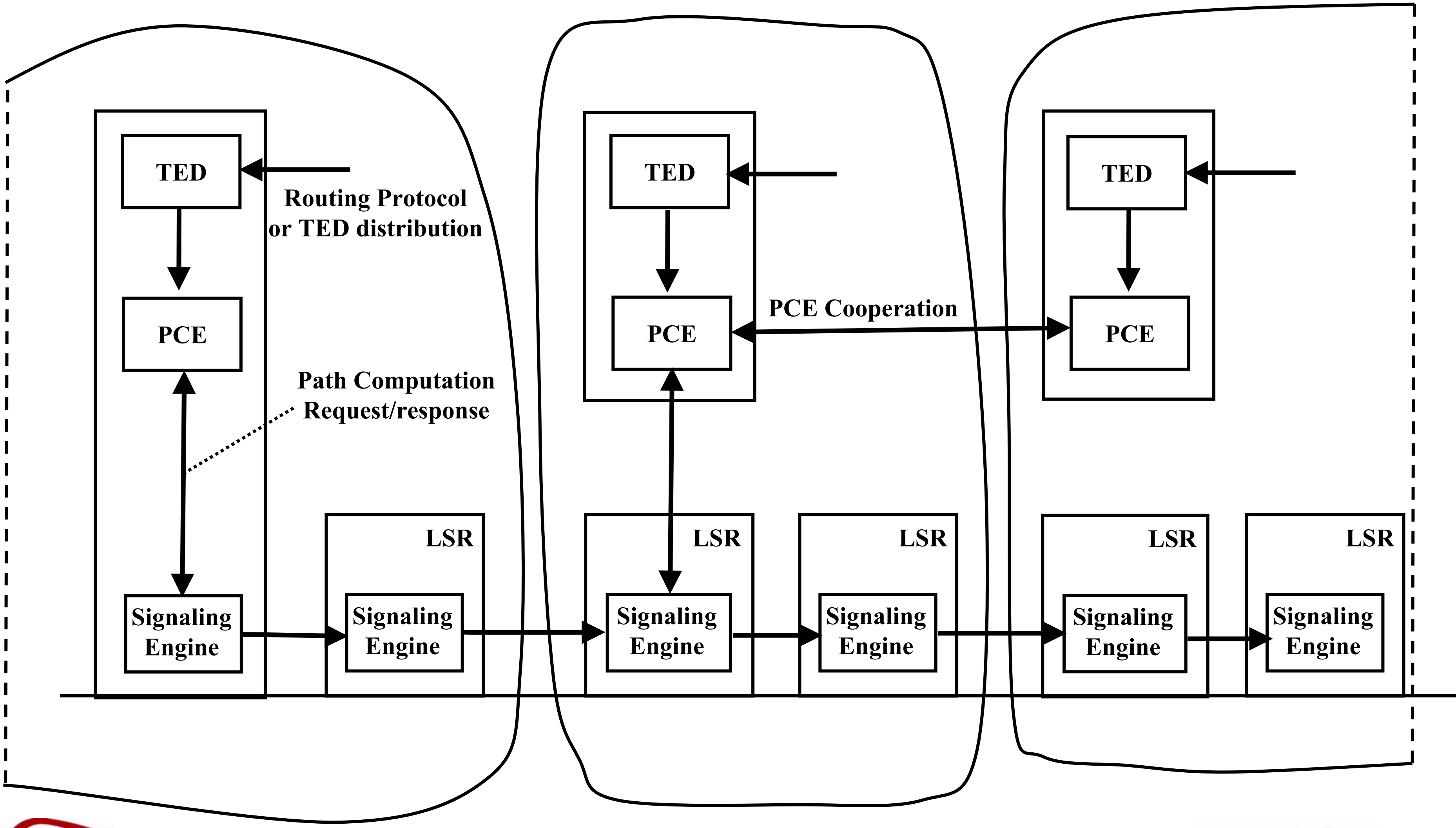
- The Multi-layered Network is a broader architectural concept
  - Switching regions define a layering of technologies
  - So do signaling capabilities
  - Administrative and operational boundaries create client-server relationships between networks



Path Computation Elements (PCE)

- Any entity that is capable of performing path computation
  - In most networks, this is the ingress LSR
  - Sometimes an off-line tool supplies the path to the ingress
  - When loose paths are used, transit LSRs must do computation
  - Border nodes do it in multi-domain networks
- Many issues concerned with LSR-based computation
  - Path computation is often resource intensive
  - There are optimality issues with incomplete path computation
  - Confidentiality and policy mean that full TED isn’t circulated
  - May want to use additional information (traffic flow, existing LSPs, etc.)
- Define one or more special nodes (PCEs) to perform path computation
- Particularly valuable in multi-domain networks
  - A domain is a zone of path computational responsiblity

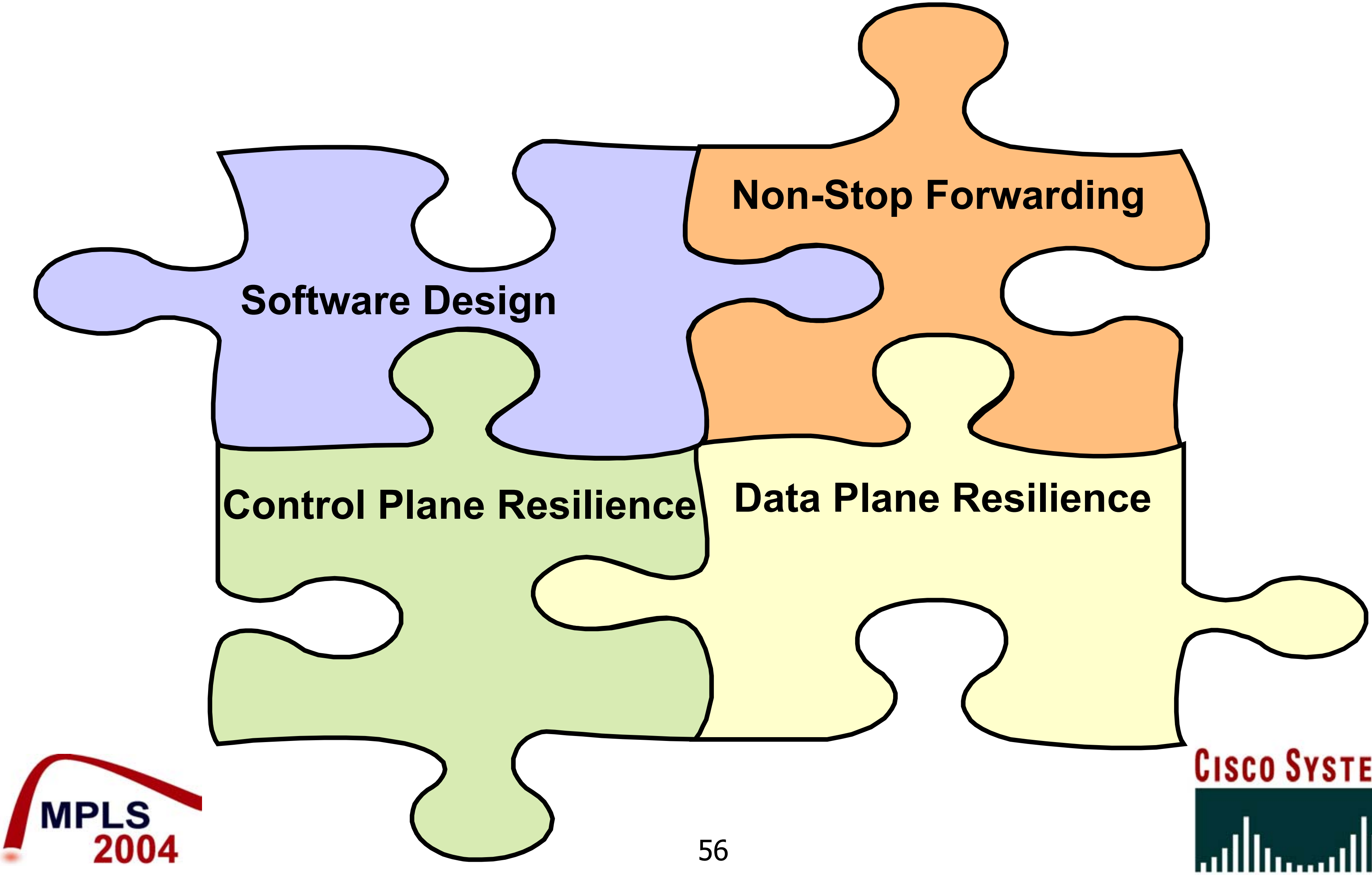
PCE Model


TED
PCE
Signaling Engine
Routing Protocol or TED distribution
Path Computation Request/response
LSR
Signaling Engine
TED
PCE
PCE Cooperation
LSR
Signaling Engine
LSR
Signaling Engine
TED
PCE
LSR
Signaling Engine
LSR
Signaling Engine

Outline

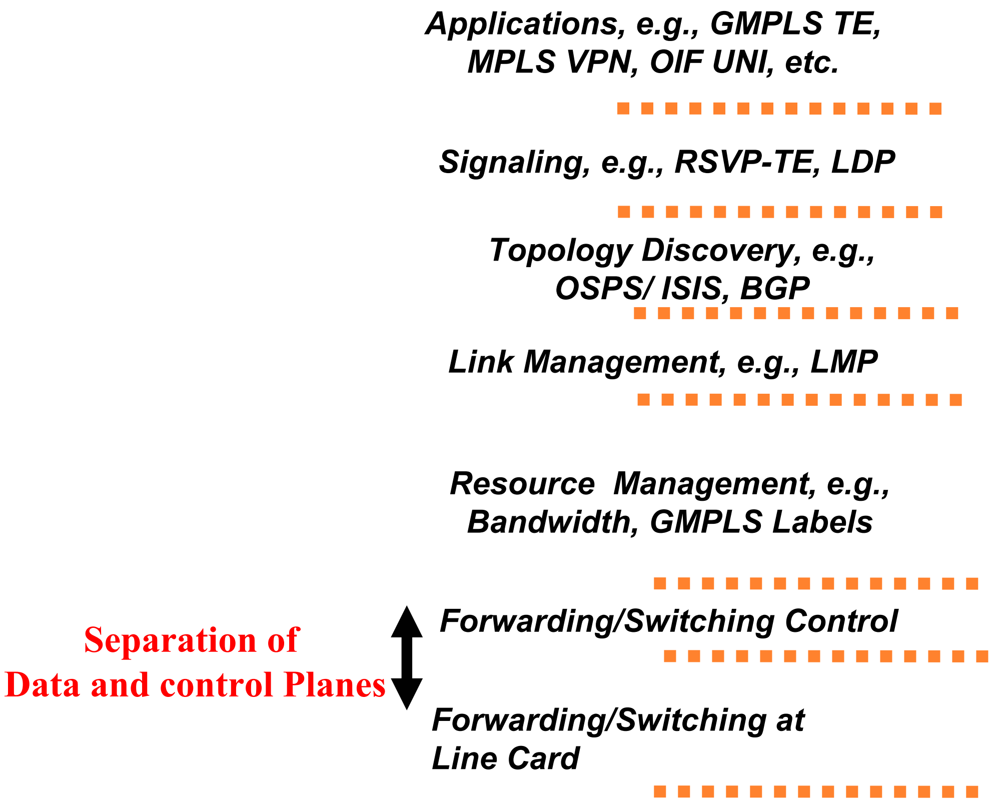
- Principles of MPLS-TE
- Extending the Concepts to GMPLS
- Fundamental Concepts
- Implementing and Deploying MPLS-TE and GMPLS
- Inter-Domain Traffic Engineering
- Components of MPLS/ GMPLS High Availability
- MPLS O&M
- Future Work

Components of High Availability: Outline



GMPLS High Availability Goals

- Availability of 99.999% or 5.25 minutes of downtime/year
- Ability to perform hitless software upgrades
- Control and Data Planes separation
- Modular Approach



Applications, e.g., GMPLS TE, MPLS VPN, OIF UNI, etc.

Signaling, e.g., RSVP-TE, LDP

Topology Discovery, e.g., OSPS/ ISIS, BGP

Link Management, e.g., LMP

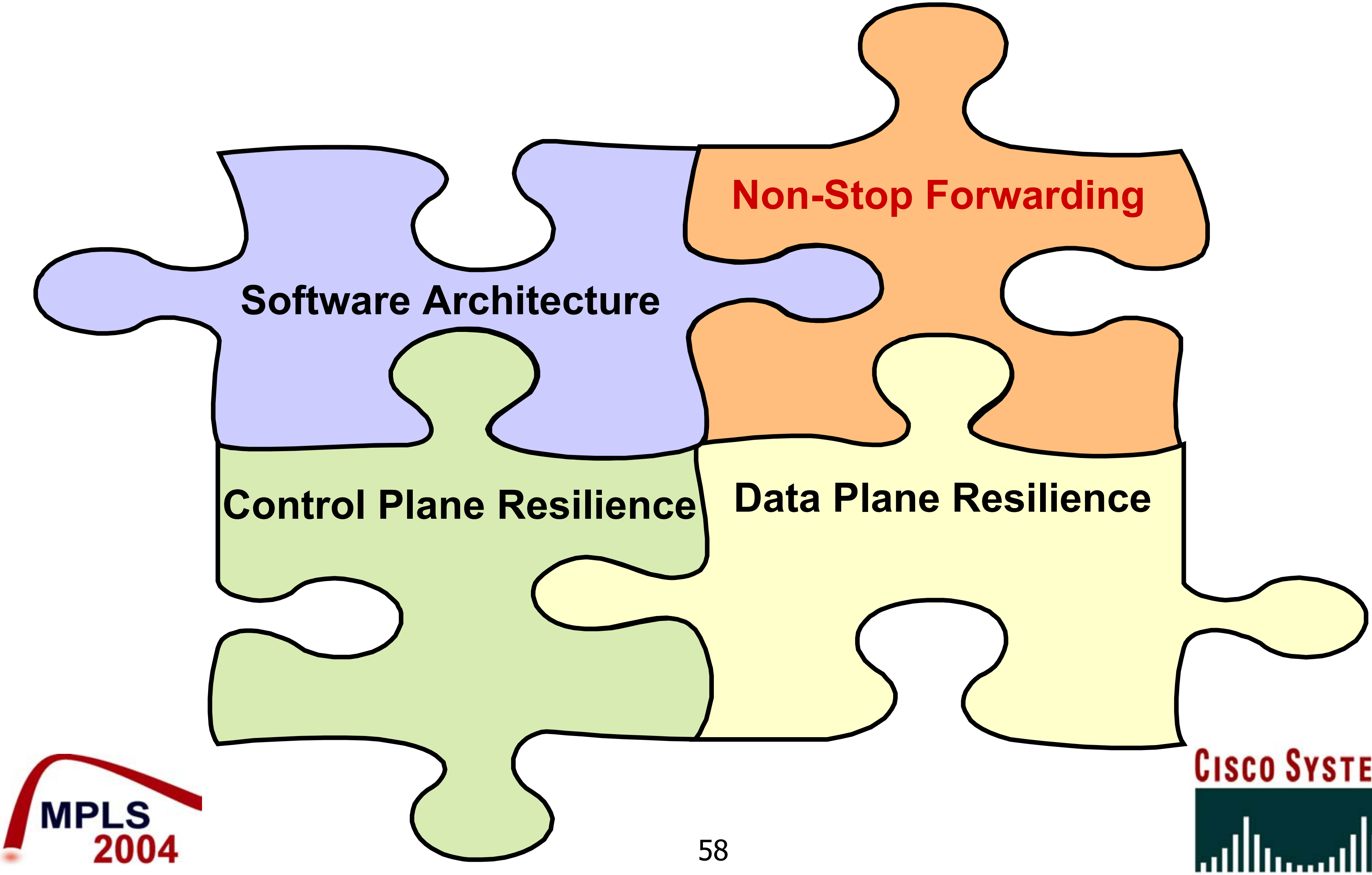
Resource Management, e.g., Bandwidth, GMPLS Labels

Forwarding/Switching Control

Forwarding/Switching at Line Card

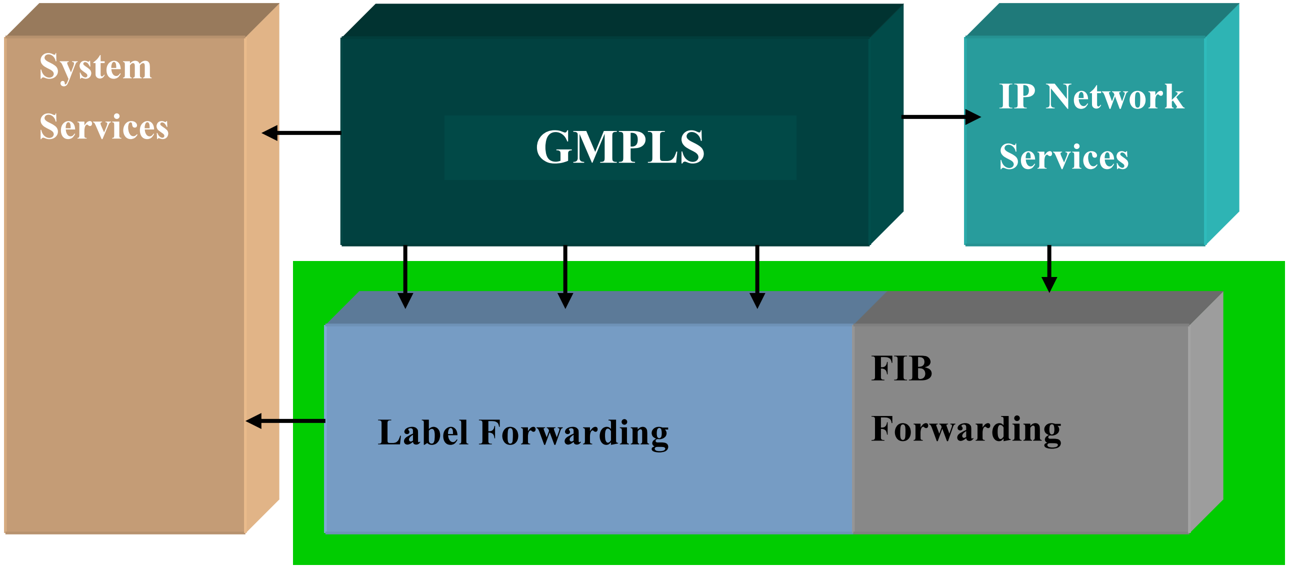
Separation of Data and control Planes

Components of High Availability



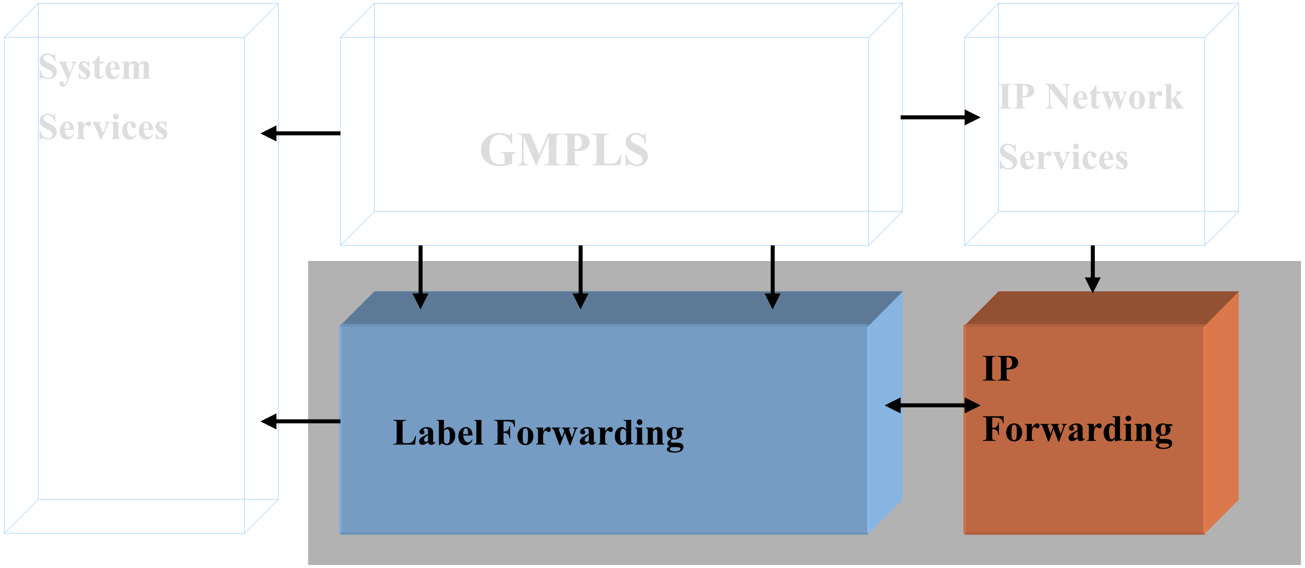
GMPLS HA - What is NSF?

- Forwarding can survive failure of entire or parts of control plane.
  - No impact on forwarding/ switching when control plane fails.
- Separation of Forwarding and Control Planes.

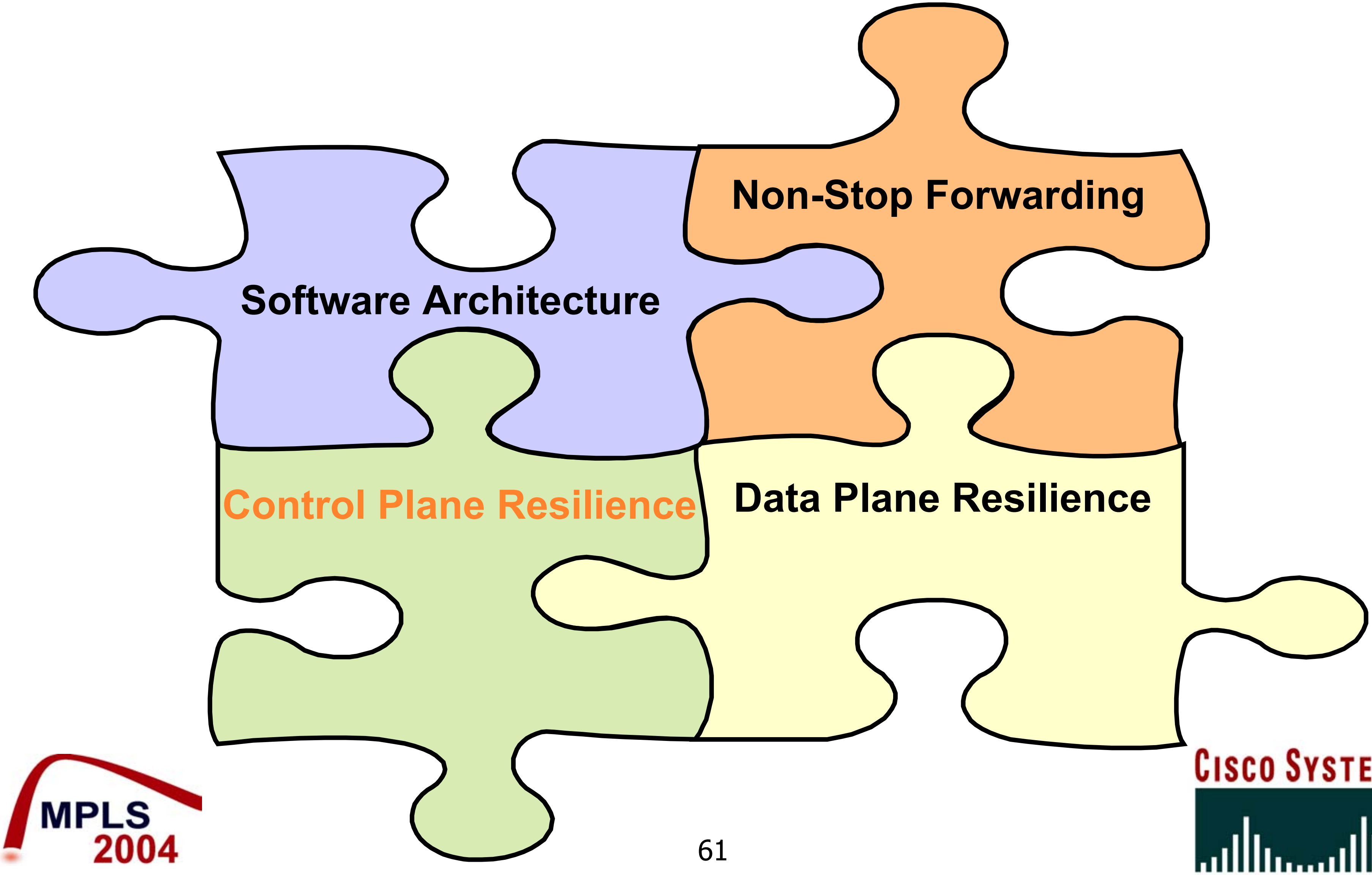


NSF as Headless Forwarding

- If the control plane fails, the forwarding plane can continue to send traffic. Headless forwarding.
- Minimize the time forwarding remains headless.



Components of High Availability: Outline



Non-Stop Forwarding

Software Architecture

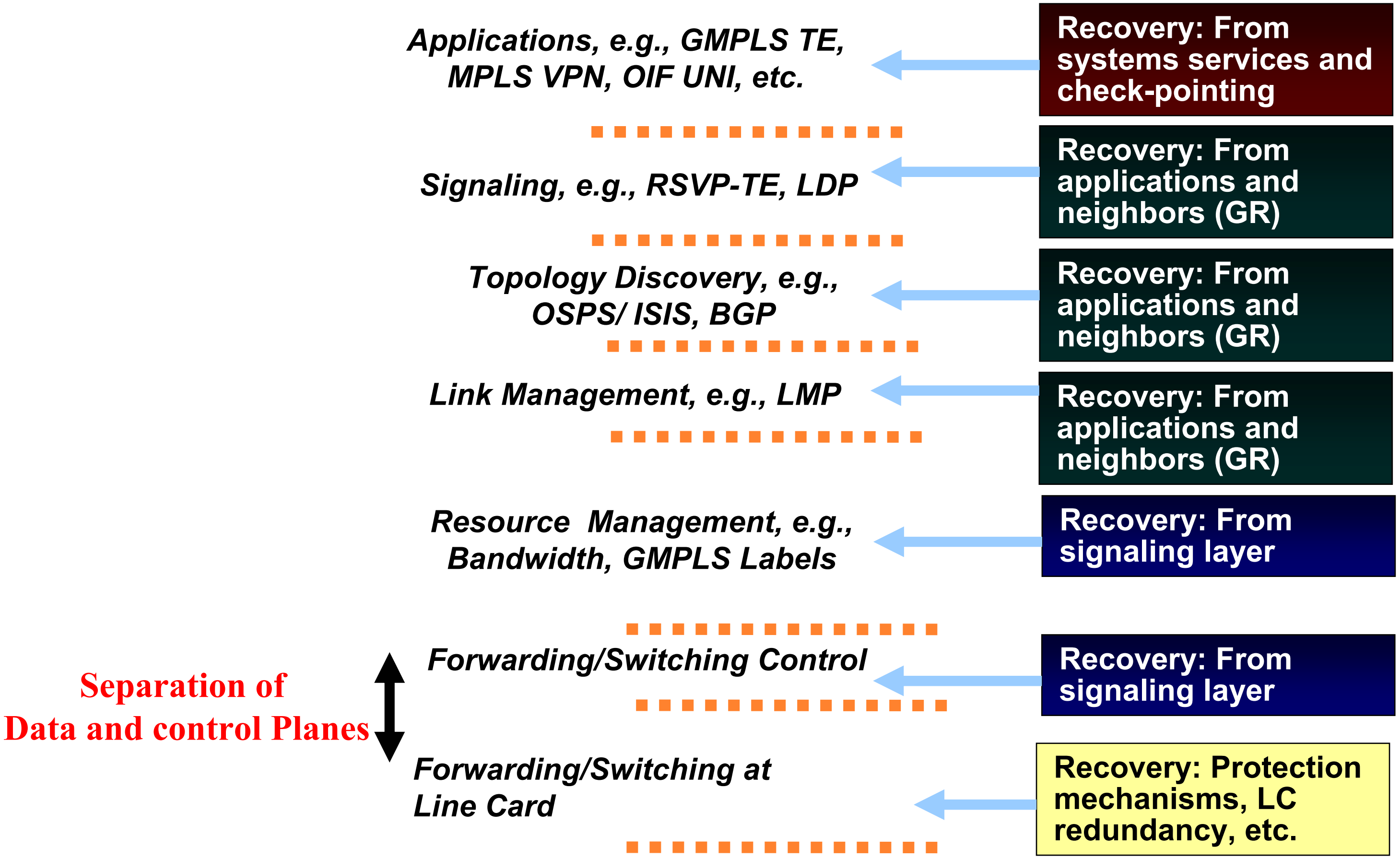
Control Plane Resilience

Data Plane Resilience

Control Plane Resilience: General Design Goals

- Recovery based on information from collaborators, checkpointing and network control messages.
  - Ability to perform dynamic state recovery using the standard's based procedures, e.g., RSVP GR, IGP GR, BGP GR, LDP GR, LMP GR.
  - Check-pointing for data that cannot be recovered dynamically.

MPLS/ GMPLS Dynamic State Recovery



Applications, e.g., GMPLS TE, MPLS VPN, OIF UNI, etc.

Recovery: From systems services and check-pointing

Signaling, e.g., RSVP-TE, LDP

Recovery: From applications and neighbors (GR)

Topology Discovery, e.g., OSPS/ ISIS, BGP

Recovery: From applications and neighbors (GR)

Link Management, e.g., LMP

Recovery: From applications and neighbors (GR)

Resource Management, e.g., Bandwidth, GMPLS Labels

Recovery: From signaling layer

Forwarding/Switching Control

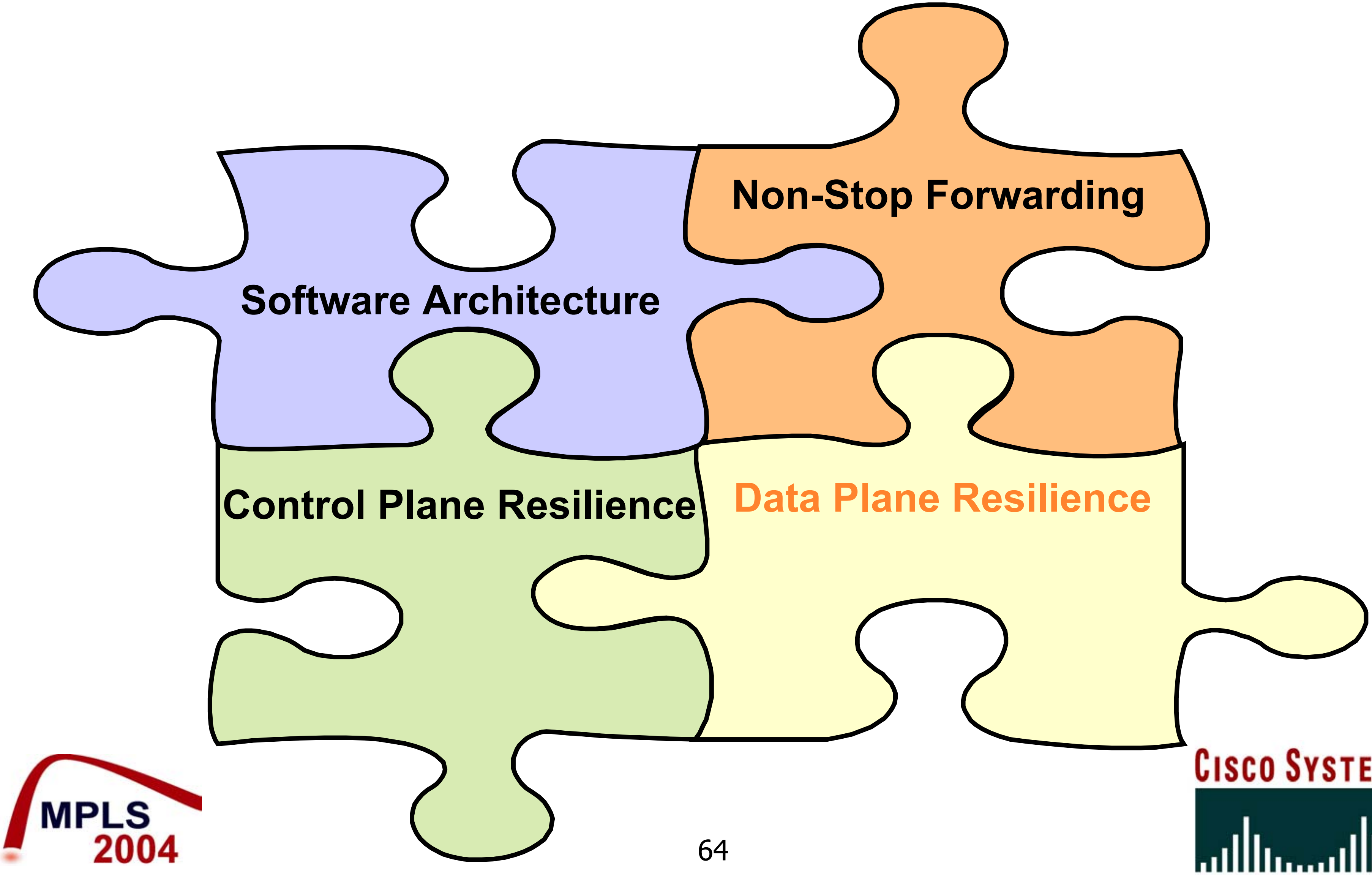
Recovery: From signaling layer

Separation of Data and control Planes

Forwarding/Switching at Line Card

Recovery: Protection mechanisms, LC redundancy, etc.

Components of High Availability: Outline



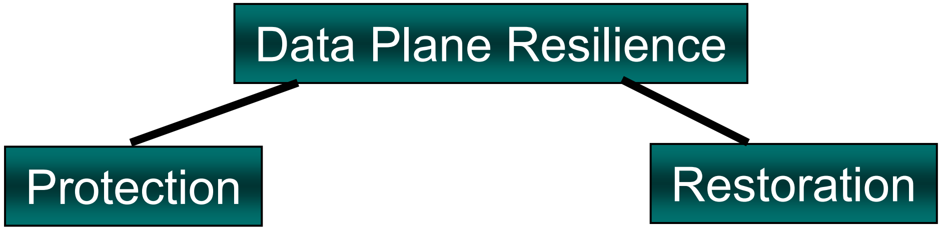
Non-Stop Forwarding

Software Architecture

Control Plane Resilience

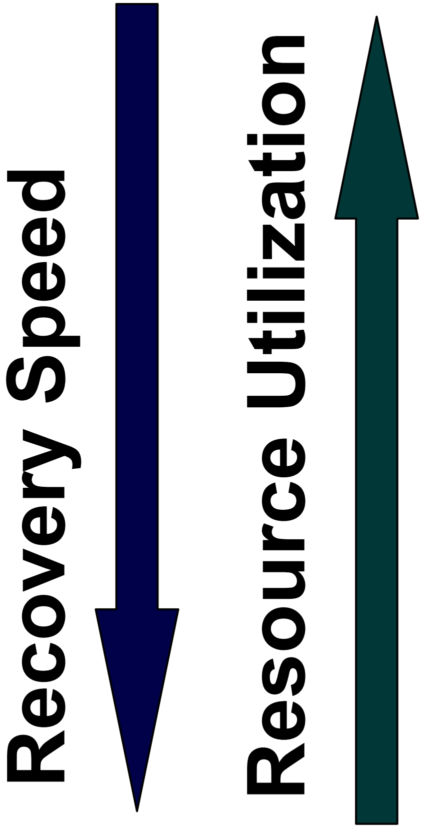
Data Plane Resilience

Data Plane Resilience: Protection Vs. Restoration



- Protection (capacity is pre-assigned to ensure survivability, protection path is known BEFORE the failure, generally faster)
  - 1+1 SONET APS (at L2)
  - Bundled Interfaces (at L2)
  - Load Balancing (at L3)
  - MPLS/IP Fast Reroute (at L2/3)
- Restoration (traffic is rerouted using a path discovered AFTER the failure, using the available capacity, slower)
  - Dynamically signaled GMPLS LSPs
  - L3 re-route
- Protection + Restoration = Data Plane Resilience

Protection and Restoration Tradeoffs

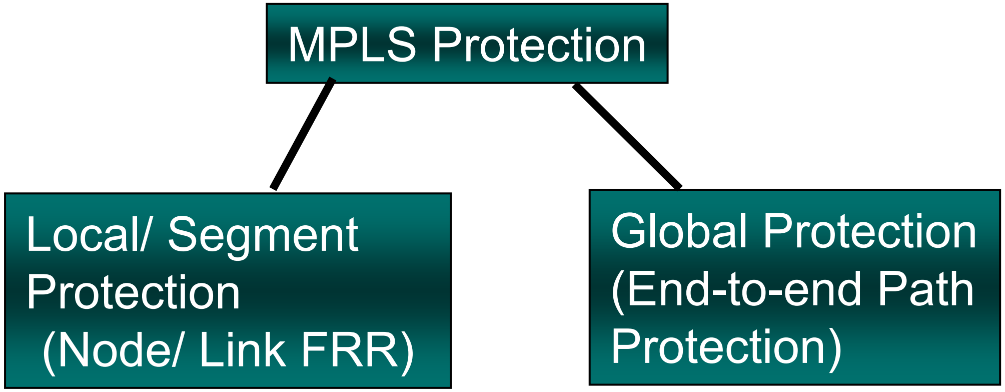


Recovery Speed

Resource Utilization

	Pre-computed	Pre-signaled	Resource Reservation
Dynamic/ On-demand	NO	NO	NO
Pre-computed Path	Yes	NO	NO
Pre-signaled Path	Yes	Yes	NO
Protection	Yes	Yes	Yes

MPLS Protection

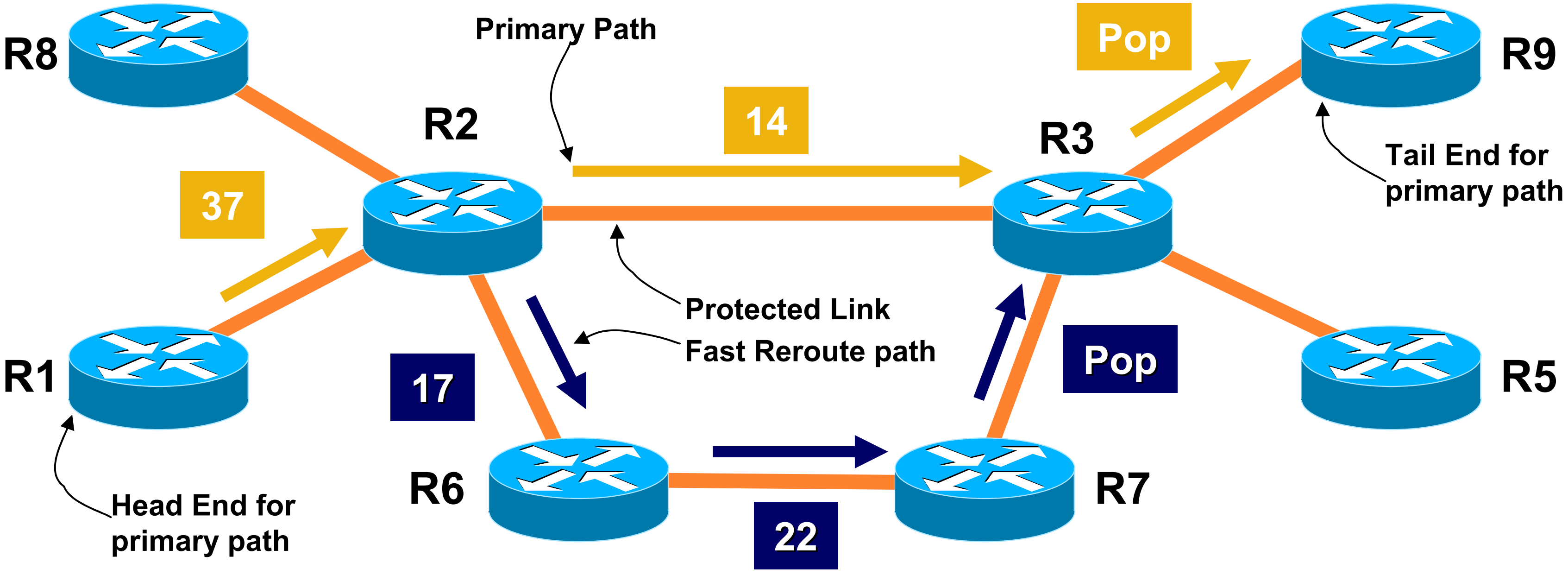


MPLS Protection

Local/ Segment Protection (Node/ Link FRR)

Global Protection (End-to-end Path Protection)

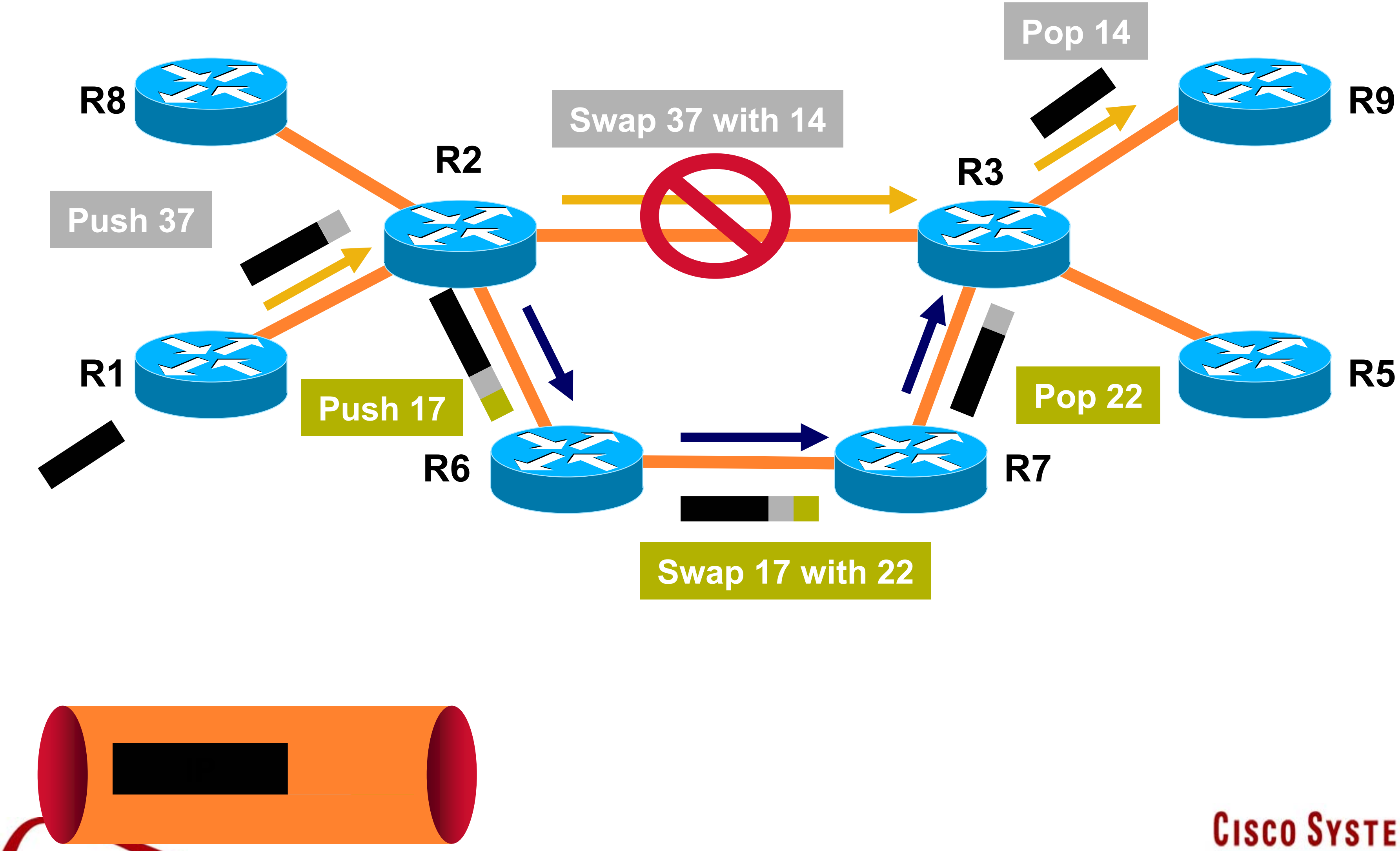
FRR Link Protection Example



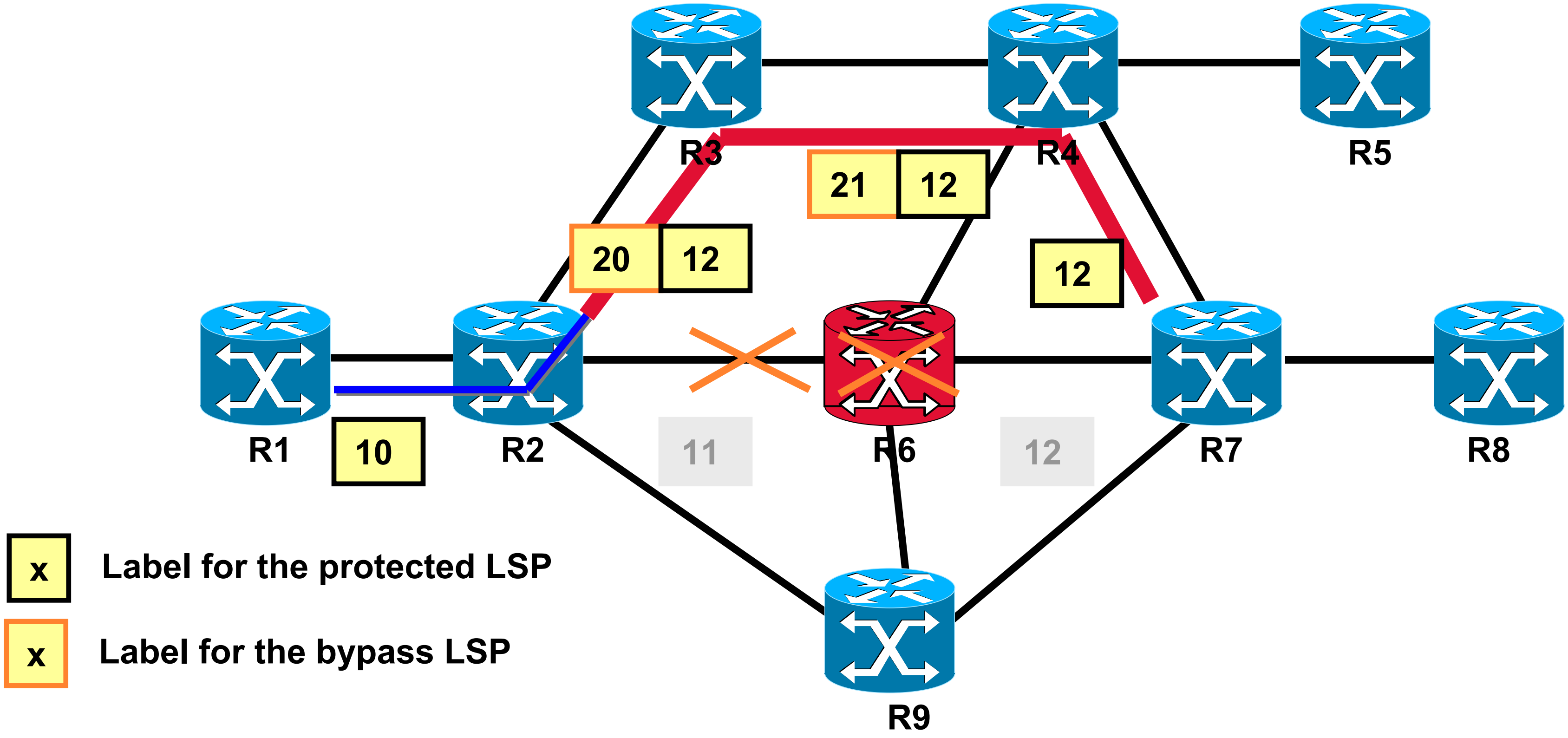
Primary path: R1 → R2 → R3 → R9

Fast Reroute path: R2 → R6 → R7 → R3

Fast ReRoute Link Failure

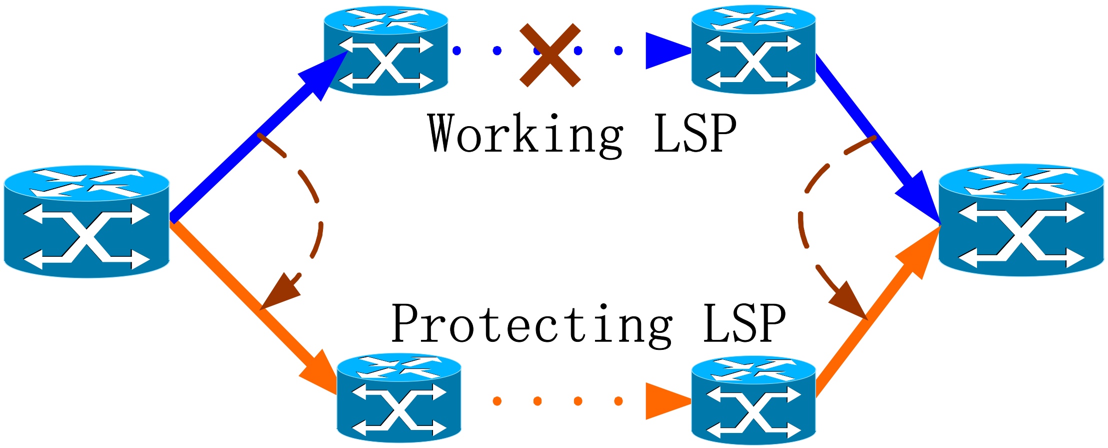


MPLS TE FRR – Node Protection



- **The PLR learns the label to use from the RRO object carried in the Resv message when the reroutable LSP is first established – With global label space allocation on the MP**

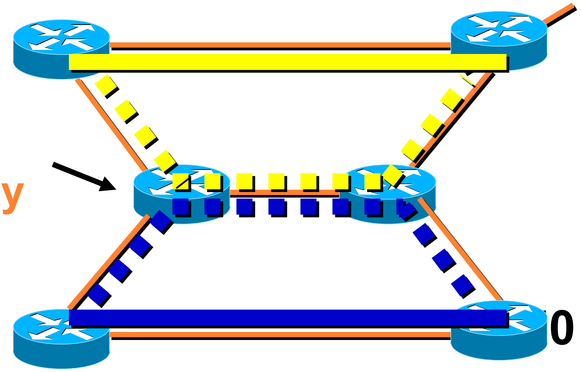
MPLS TE Path Protection



- MPLS TE Path Protection is a global repair mechanism using protection switching
- No path computation and signalling of the new LSP once the failure has been detected and propagated to the head-end (compared to LSP reroute)
- Diversely routed paths are calculated by the CSPF on the head-end (they may be link, node or SRLG diverse)

MPLS TE Path Protection

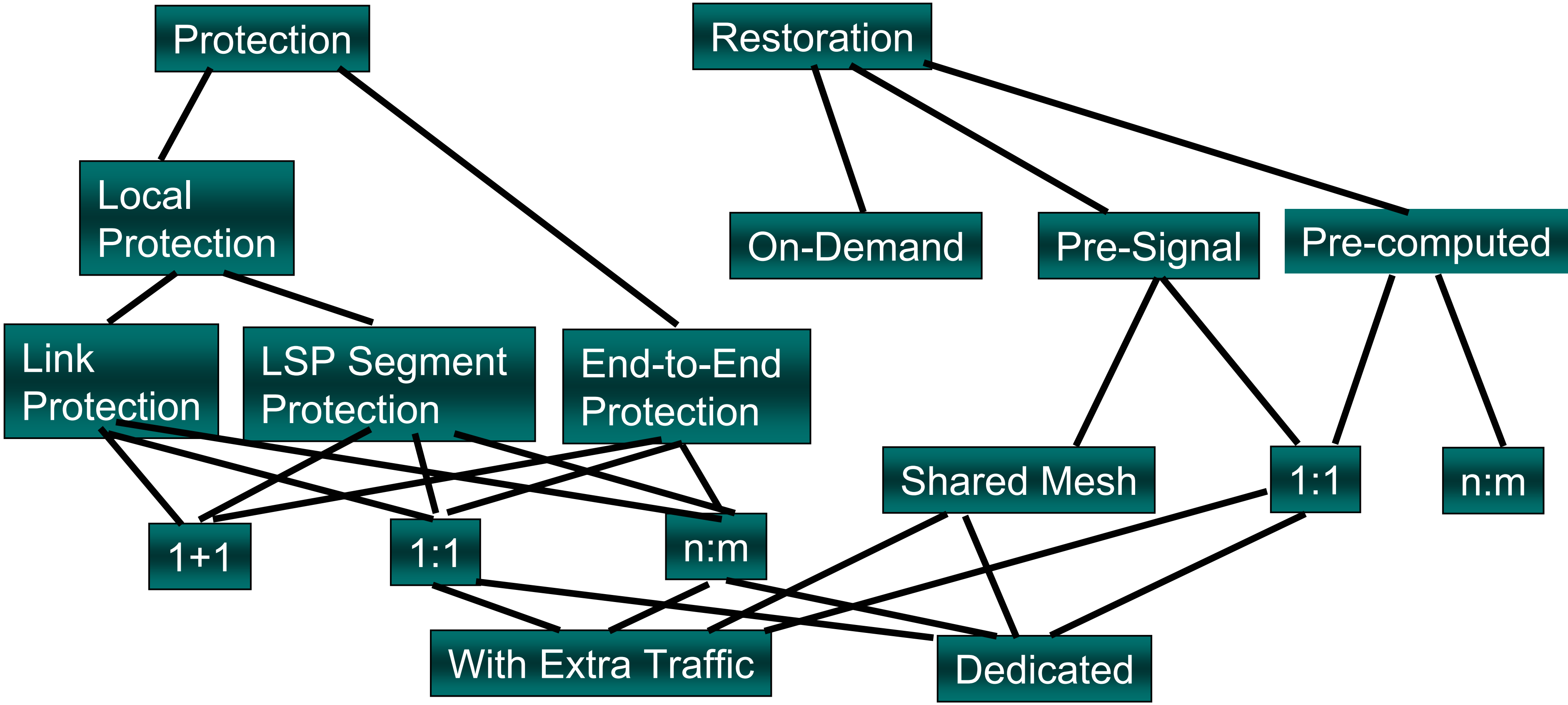
- Limitation of MPLS TE Path protection
  - The FIS propagation may be unacceptable especially for very sensitive traffic,
  - The number of states in the network is doubled !!
  - CSPF is likely to be highly inefficient in term of bandwidth usage.
- Path protection may be an attractive solution if and only if:
  - Just a few LSPs require protection
  - A few hundreds of msecs convergence time is acceptable



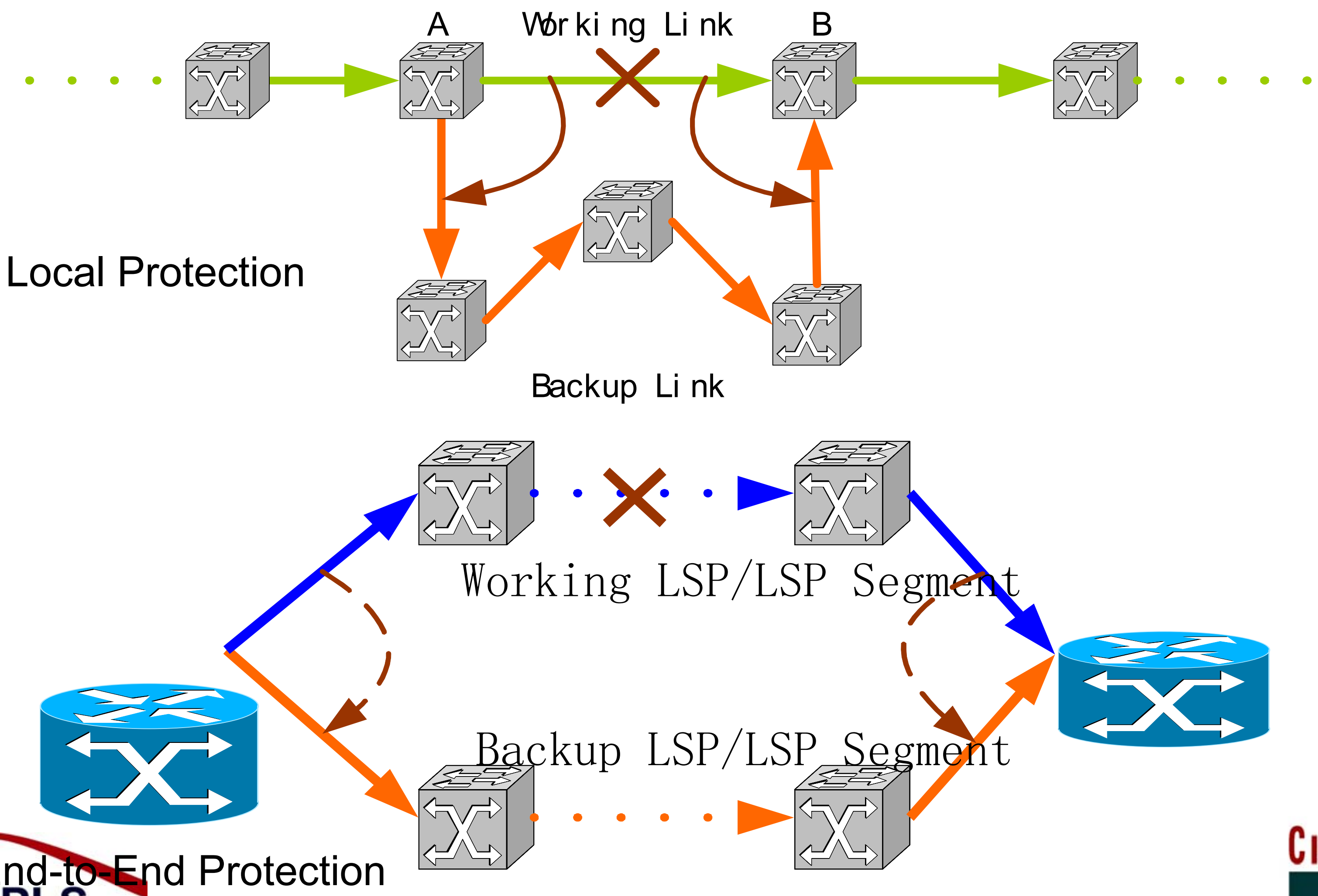
Shared capacity

primary diversely routed paths may share backup bandwidth
(under the assumption of single network element failure)

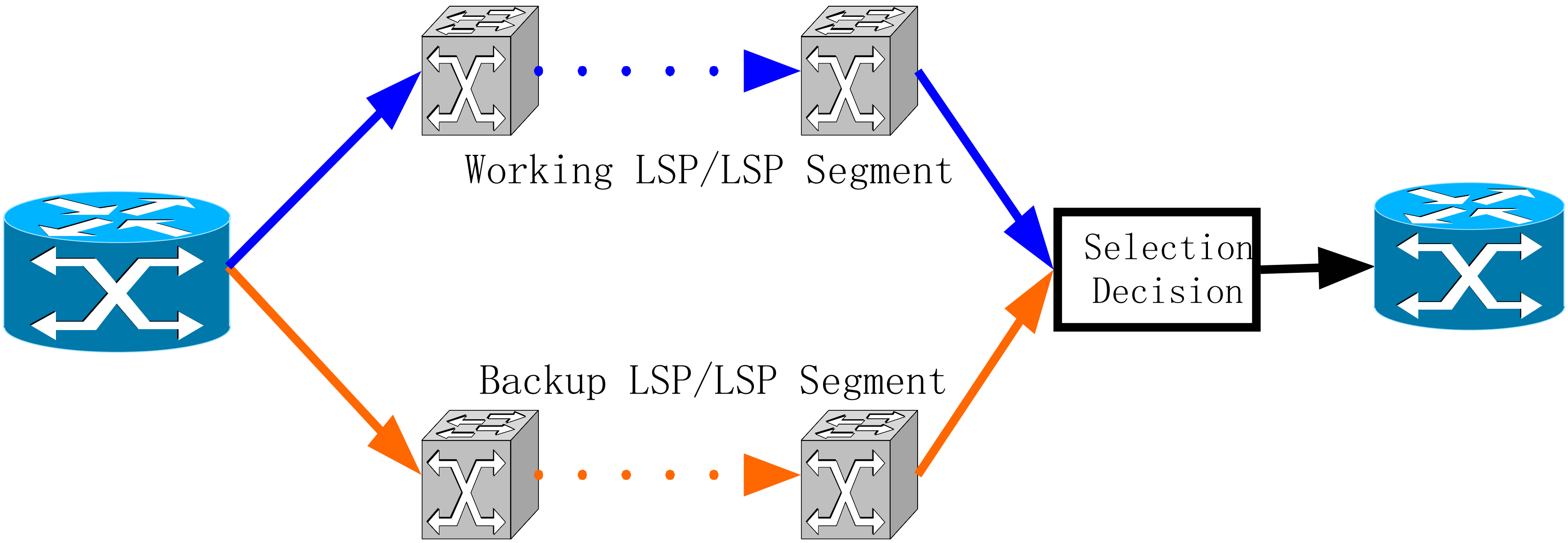
Taxonomy of Protection & Restoration in GMPLS Network



Recovery Scope: Local Vs. end-to-end

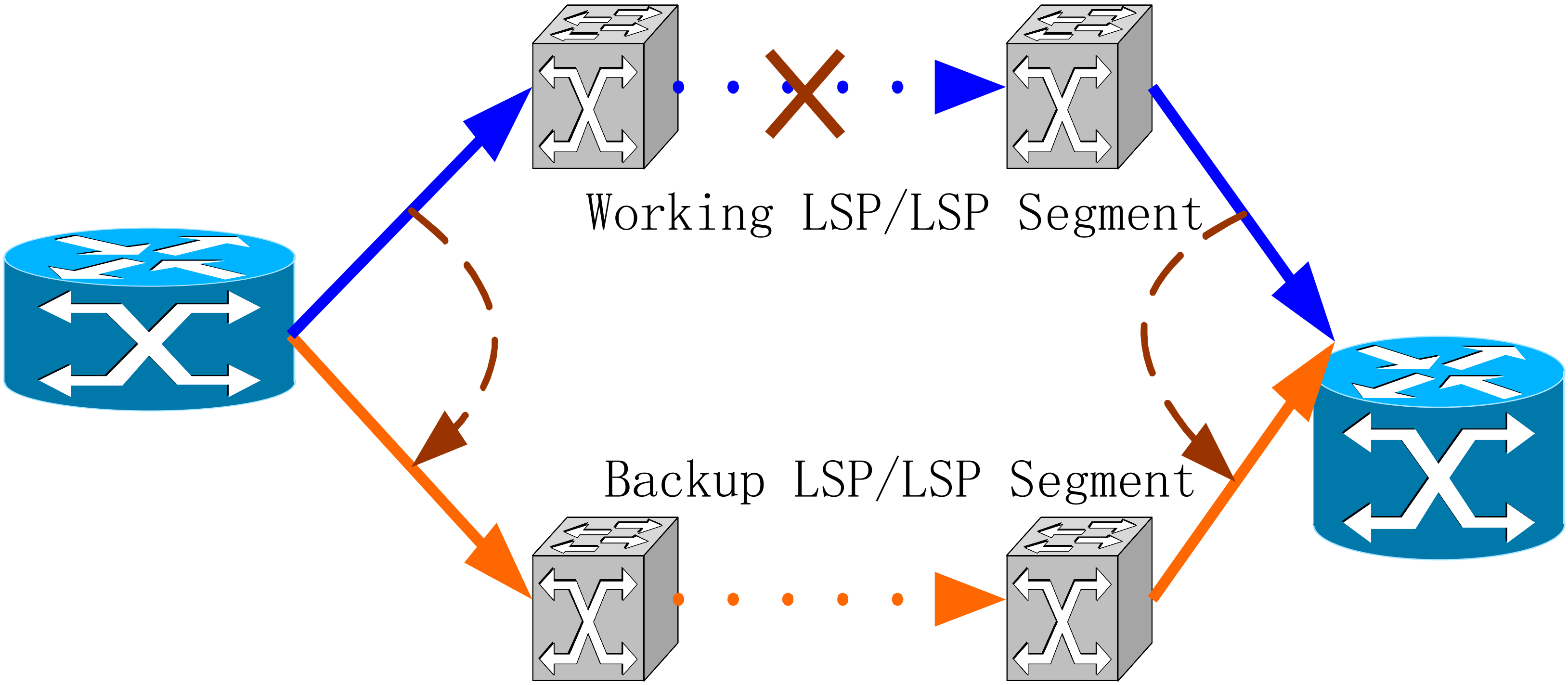


1+1 Protection



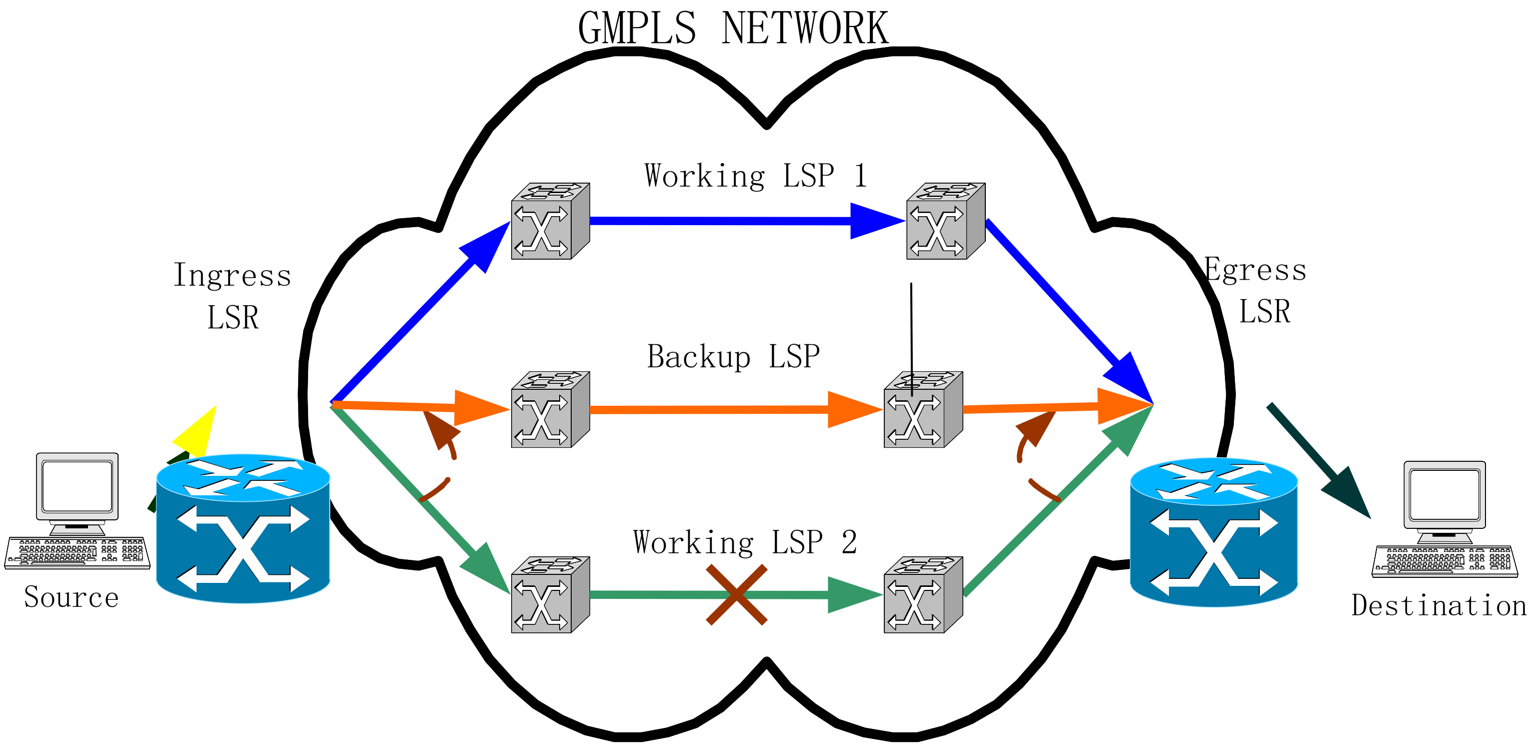
- Traffic is forwarded to both Primary and backup LSPs.
- Backup capacity cannot be used.

1:1 Protection



- All required resources are pre-allocated in backup LSP before the working LSP fails
- Backup capacity may be used for lower priority traffic.

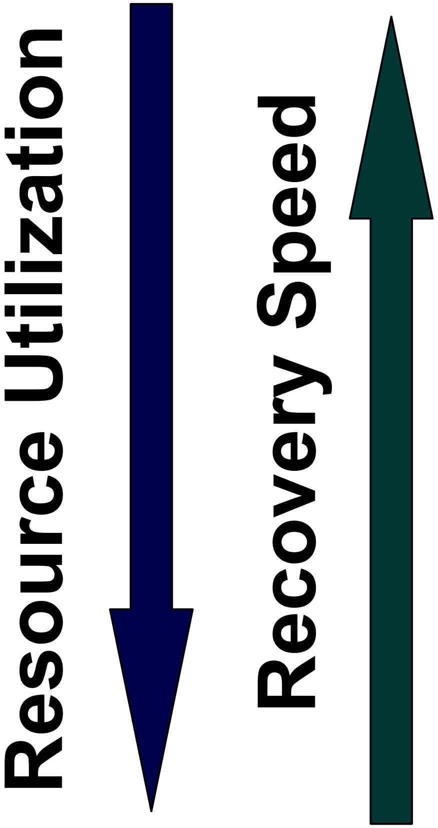
n:m Protection/ Shared Mesh Restoration



- Backup resources are shared.
- All required resources are pre-allocated in backup LSP before the working LSP fails
- Backup capacity may be used for lower priority traffic.

Resource Sharing Tradeoffs

- 1+1 Protection
- 1:1Protection (Dedicated).
- 1:1 with extra Traffic
- n:m Protection (Dedicated)
- n:m (Extra Traffic)

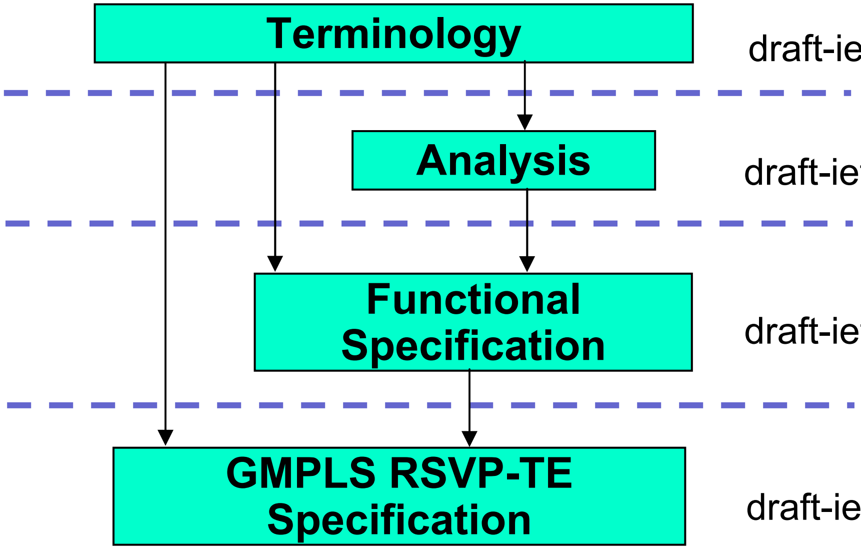


Resource Utilization

Recovery Speed

- 1:1 Restoration (pre-signaled)
- Shared Mesh Restoration (Pre-signaled)
- Pre-computed Restoration
- On-Demand Restoration

IETF Drafts on GMPLS Based Recovery



draft-ietf-ccamp-gmpls-recovery-terminology-xx.txt

draft-ietf-ccamp-gmpls-recovery-analysis-xx.txt

draft-ietf-ccamp-gmpls-recovery-functional-xx.txt

draft-ietf-ccamp-gmpls-recovery-e2e-signaling-xx.txt

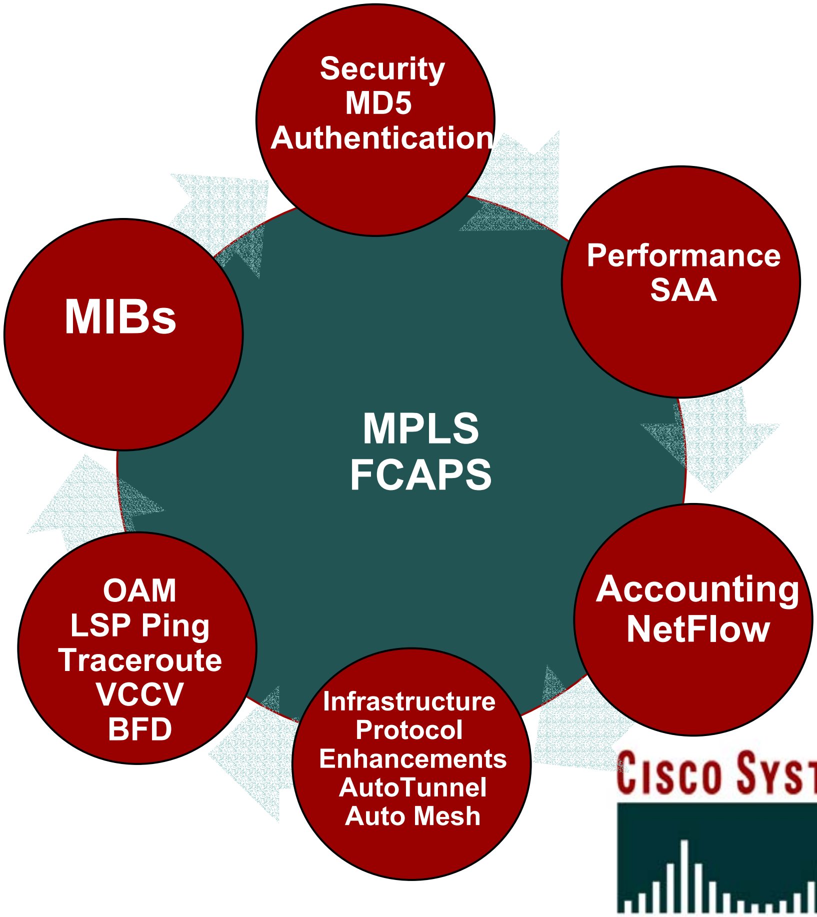
Outline

- Principles of MPLS-TE
- Extending the Concepts to GMPLS
- Fundamental Concepts
- Implementing and Deploying MPLS-TE and GMPLS
- Inter-Domain Traffic Engineering
- Components of MPLS/ GMPLS High Availability
- MPLS O&M
- Future Work

What is MPLS OAM?

MPLS Operations Administration & Management (O&M) are tools and techniques needed to address FCAPS in deploying and operating an MPLS network successfully

Fault-management
Configuration
Accounting
Performance
Security



MPLS FCAPS
Security MD5 Authentication
Performance SAA
Accounting NetFlow
Infrastructure Protocol Enhancements AutoTunnel Auto Mesh
OAM LSP Ping Traceroute VCCV BFD
MIBs

MPLS 2004

Cisco Systems

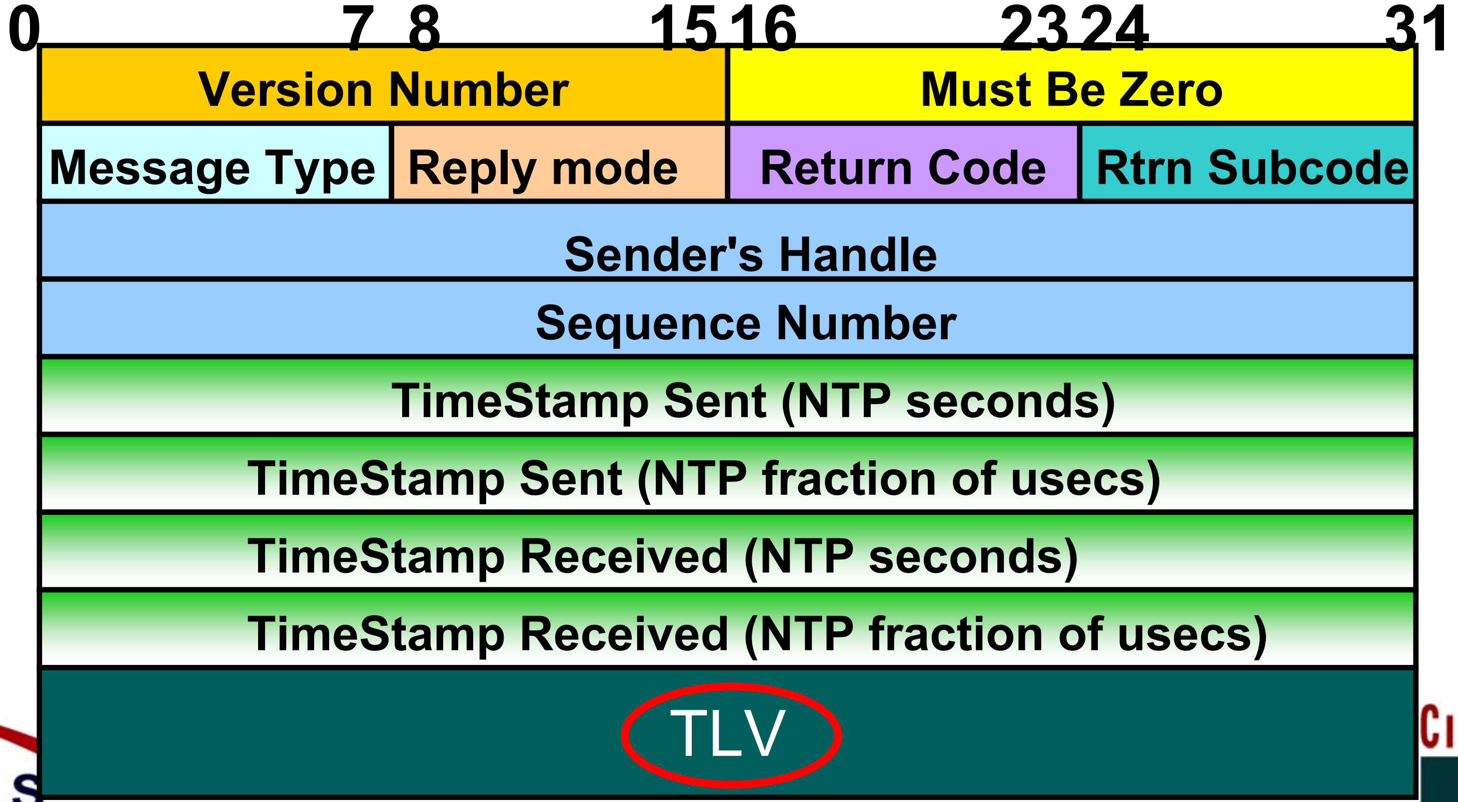
MPLS Embedded Management and FCAPS

Fault Management	MPLS Ping/Traceroute, VCCV, Mib, Auto SAA (Service Assurance Agent)
Configuration	MPLS TE Auto Tunnel, Auto Tunnel Mesh Groups, Auto SAA
Accounting	NetFlow, MIB
Performance	SAA, Auto SAA, NetFlow, Mib
Security	RSVP Message Authentication LDP Message Authentication MD5 Authentication for Routing Protocol: BGP, OSPF

Packet format of an MPLS LSP Echo

LSP Ping, like the traditional IP Ping, is based on echo request and echo reply.

MPLS LSP echo request and reply are UDP packets with following format.



0 – 7	8 – 15	16 – 23	24 – 31
Version Number		Must Be Zero	
Message Type	Reply mode	Return Code	Rtrn Subcode
Sender's Handle			
Sequence Number			
TimeStamp Sent (NTP seconds)			
TimeStamp Sent (NTP fraction of usecs)			
TimeStamp Received (NTP seconds)			
TimeStamp Received (NTP fraction of usecs)			
TLV			

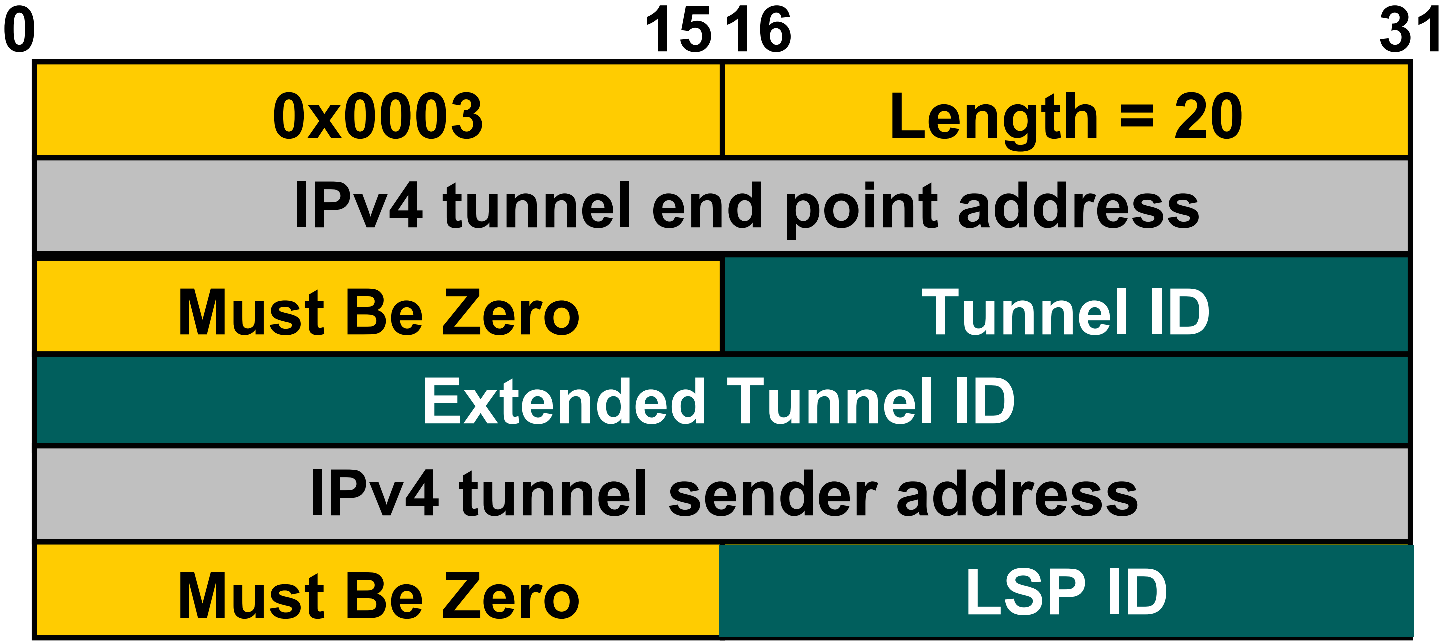
LSP Ping TLVs

Value	Meaning
1	Target FEC Stack
2	Downstream Mapping
3	Pad
4	Error Code
5	Vendor Enterprise Code

- An MPLS echo request MUST have a Target FEC Stack that describes the FEC stack being tested.
  - The LSP to be tested is identified by the "FEC Stack".
- Examples of Target FEC Stack
  - LDP IPv4/ v6 prefix
  - RSVP IPv4/ v6 session query
  - VPN IPv4/ v6 prefix
  - BGP labeled IPv4 prefix, etc.

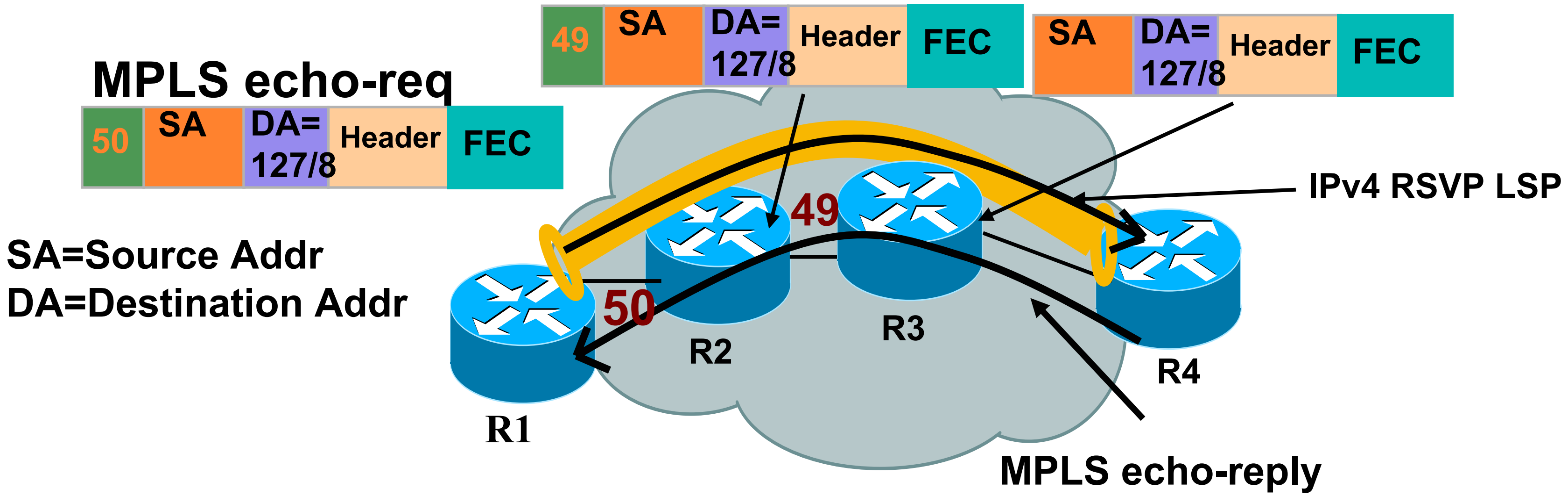
IPv4 RSVP Session Query TLV

- The LSP to be tested is identified by the "FEC Stack".
- E.g., IPv4 RSVP Session Query TLV contains enough information to uniquely identify an RSVP signaled LSP.



0 15	16 31
0x0003	Length = 20
IPv4 tunnel end point address	
Must Be Zero	Tunnel ID
Extended Tunnel ID	
IPv4 tunnel sender address	
Must Be Zero	LSP ID

LSP Ping: Theory of Operation



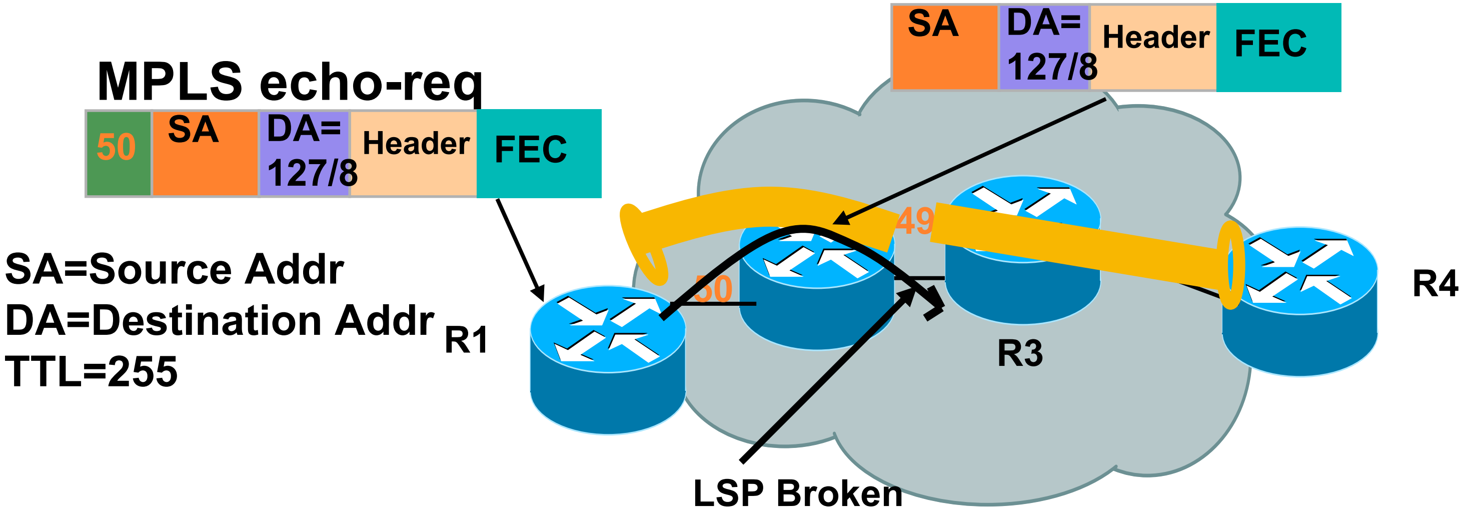
SA=Source Addr
DA=Destination Addr

Use the same label stack as used by the LSP to make the echo to be switched in band of LSP under test.

The IP header destination address field of the echo request is a 127/8 address.

An Echo reply, which may or not be labeled, has outgoing interface IP address as the source. Destination IP address/port are copied from the echo-request's source address/port

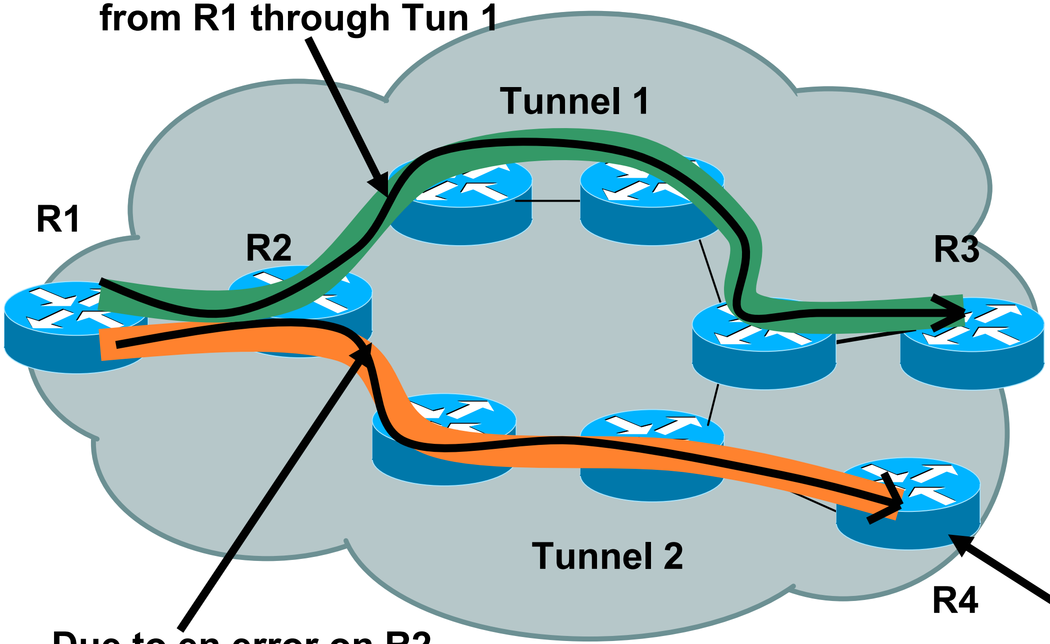
LSP Ping: Theory of Operation (Broken LSP)



- Presence of the 127/8 address in the IP header destination address field causes the packet to be consumed by any routers trying to forward the packet using the IP header.
- In this case R3 would not forward the echo-req to R4 but rather consumes the packet and reply to it accordingly.

LSP Ping: Theory of Operation (Misrouted LSP)

LSP Ping is initiated
from R1 through Tun 1

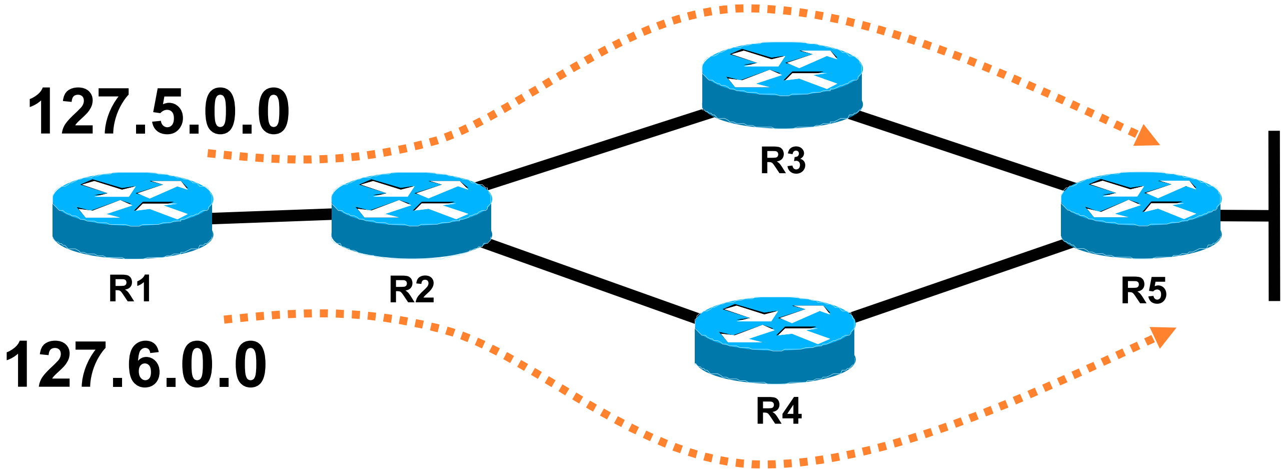


Due to en error on R2
the LSP Ping is
switched into Tun 2

R4 would recognize
that dest addr, LSP id
and Tu id are different
and would reply with a
return code 4

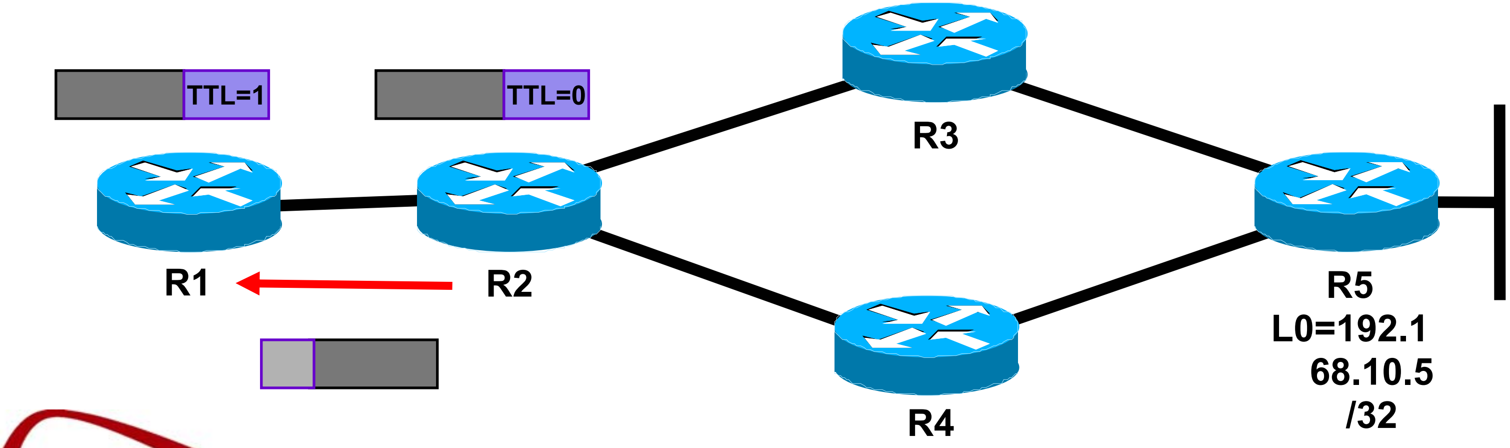
MPLS Ping: Handling Equal Cost Multi-Paths

- Packet needs to follow data path
- Not trivial when Multiple Paths available
- No standard ECMP algorithm
- Use Destination address in the 127/8 Range

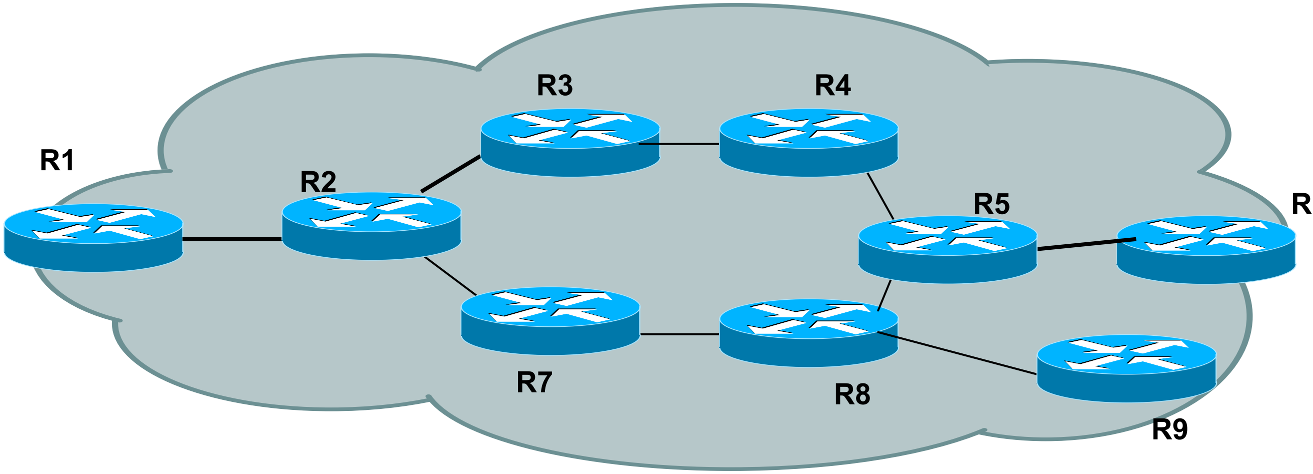


MPLS LSP Traceroute, Packet Flow

- Traceroute is used for hop-by-hop fault localization as well as path tracing.
- MPLS Ping Packets are sent with increasing TTL to “probe” the ECMP tree from downstream LSRs.
- Label switched if TTL > 1, Processed by control plane where TTL expires.
- Reply contains downstream mapping TLV (i.e. the label, interface for reaching the downstream router).

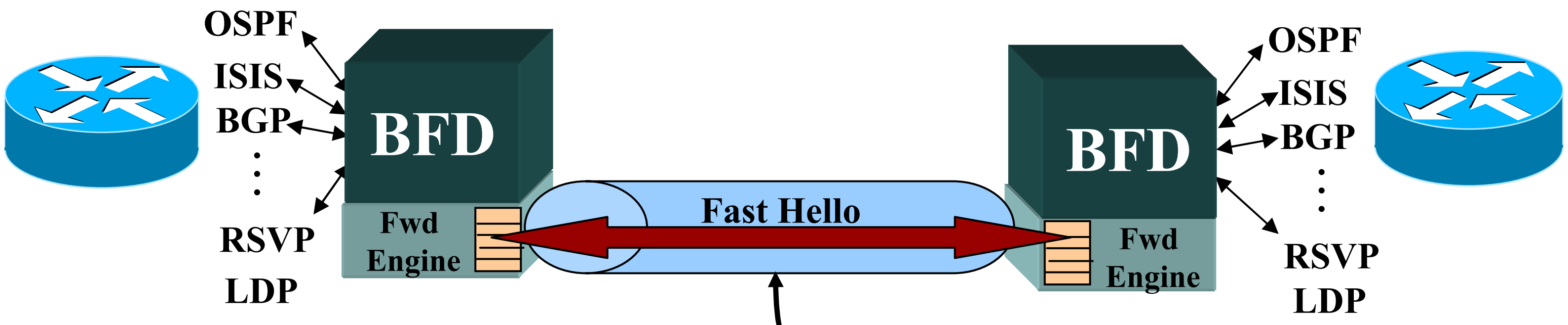


LSP Trace: Path/Tree Trace



- Path trace would give us information of only one path out of all the possible ECMP paths
- In the above example a path trace from R1 to R6 may only report Path: R1-R2-R3-R4-R5-R6
- Tree trace returns ALL of the possible paths between one source and destination
- So in the above case the LSP (tree) trace would give us information about both the paths R1-R2-R3-R4-R5-R6 and R1-R2-R7-R8-R5-R6

Attributes of BFD



- Direct physical links
- Multi-hop routed paths
- Virtual circuits, Tunnels
- MPLS LSPs
- Bi/uni-directional links

- Simple Hello Protocol
- Protocol Independence
- Media Independence
- Fast failure detection
  - Light Weight, Fixed Length; simple to parse
- Forwarding plane liveliness
  - E.g., Link may be up but forwarding engine may be down or an entry may be incorrectly programmed.
- **No discovery mechanism in BFD**
  - Applications bootstrap a BFD session

MPLS BFD Vs. LSP Ping

Method	Data Plane Failure Detection	Control Plane Consistency	Protocol Overhead
LSP Ping	**YES**	**YES**	**Higher than BFD**
MPLS-BFD	**YES**	**NO**	**Low**

MPLS-BFD can complement LSP Ping to detect a data plane failure in the forwarding path of a MPLS LSP

Supported FECs:
RSVP IPv4/IPv6 Session, LDP IPv4/IPv6 prefix
VPN IPv4/IPv6 prefix, Layer 2 VPN, Layer 2 Circuit ID

Outline

- Principles of MPLS and MPLS-TE
- Extending the Concepts to GMPLS
- Fundamental Concepts
- Implementing and Deploying MPLS-TE and GMPLS
- Inter-Domain Traffic Engineering
- Components of MPLS/ GMPLS High Availability
- MPLS O&M
- Future Work

Future Developments

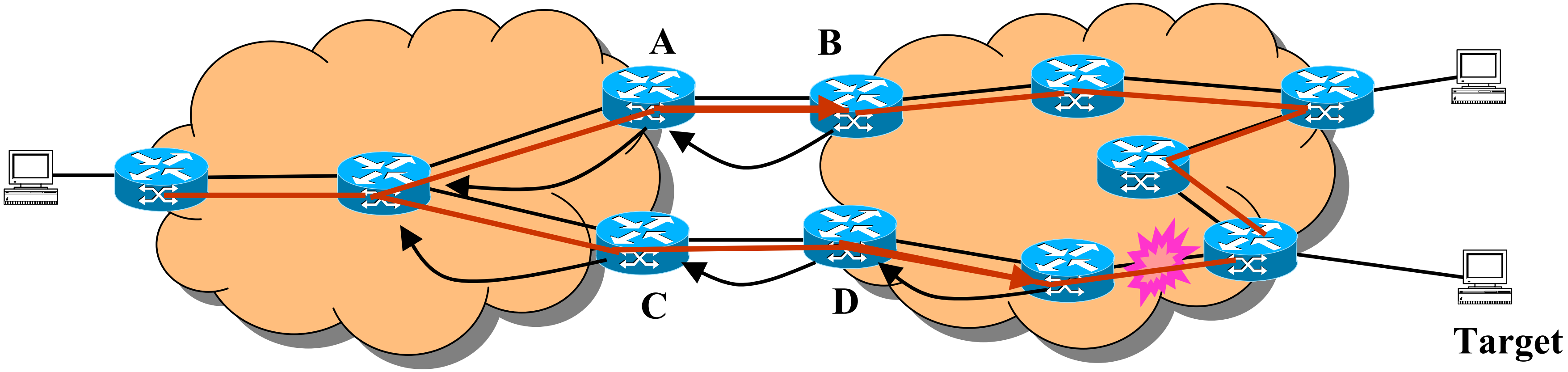
- Alarm and Error Reporting
  - Signaling Alarm Information
  - Enhancing Error Reporting for Crankback
- O&M for GMPLS
- Layer One Virtual Private Networks (L1VPN)
- The ITU-T's ASON Architecture
  - Reference Points
  - Calls and Connections
  - Integration with GMPLS
- Point-to-Multipoint Traffic Engineering

Signaling Alarm Information

- Not alarm reporting
- Dissemination of alarms so that all LSRs on an LSP have the same view of the alarms
- New Alarm Spec object modeled on Error Spec
  - Error Spec TLVs indicate interface etc.
  - Error Value carries Alarm code
  - Other TLVs carry additional information
    - Severity
    - Timestamp
    - Count
    - Text string
- Carried on Path and Resv message
  - Forward all alarms received, and add local alarms
- Enabled/disabled using GMPLS Administrative Status object
- All alarm correlation, soaking, reporting etc. is unchanged
- Alarm signaling is open to local policy
- Applicable to all networks; focused on optical networks

Crankback

- Enhanced error reporting to indicate LSP setup blockages
- Can report link and node failures
- Also:
  - Resources (labels)
  - Component links
  - Areas
  - Autonomous Systems
- Aggregation of errors
- Control of retry attempts



GMPLS O&M

- Aim is to build on existing MPLS Techniques
- Issues:
  - Additional features and functions
    - Bid-directional LSPs
    - Control of labels in EROs
    - More extensive definition of labels
    - Extra parameters such as switching and encoding types
  - Technology is not necessarily packet-based
    - Makes some statistics harder to capture
    - Means that data plane traceroute techniques don’t work

Additions to the MPLS MIB Modules

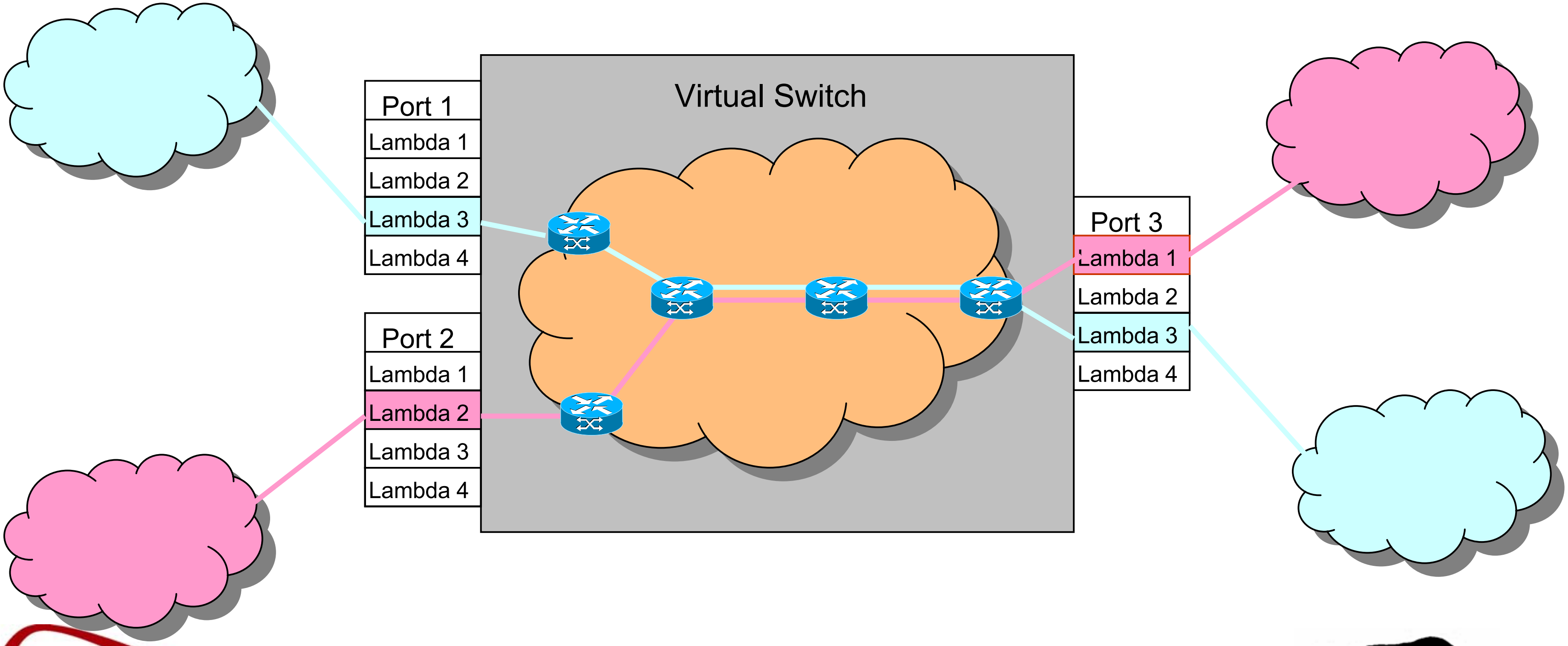
- LSR MIB module
  - Allow configuration of Hello period per interface
  - Mark segments according to their direction
    - Forwards or backwards
  - Show amount to decrement TTL for out-segments
- New Label MIB Module
  - One table – gmplsLabelTable
    - Rows pointed to from other tables
    - Allows context-sensitive encoding of labels
    - Allows label concatenation
- TE MIB Module
  - Request additional parameters
    - Label recording, LSP encoding and switching types, Link protection, G-Pid
    - Protection and directionality indicators
    - Notify recipients and Administrative Status
  - Add forward and reverse label control to EROs (hop tables)
  - New table to track errors
    - RSVP error codes and time stamps
  - Add error information to SNMP Notifications

Tunnel Trace

- RFC3609 sets out requirements
  - Build on existing techniques
  - Add security features to tracing
  - Trace through hierarchies and reveal the path of the outer tunnels
  - Work in non-packet technologies
- Generic Tunnel Trace Protocol (GTTP) is a ‘work in progress’
  - Like LSP Ping, but:
    - Cannot use TTL in the data plane
    - Uses new control plane messages in UDP packets
    - Must ‘digress’ to trace the tunnel at each layer of hierarchy

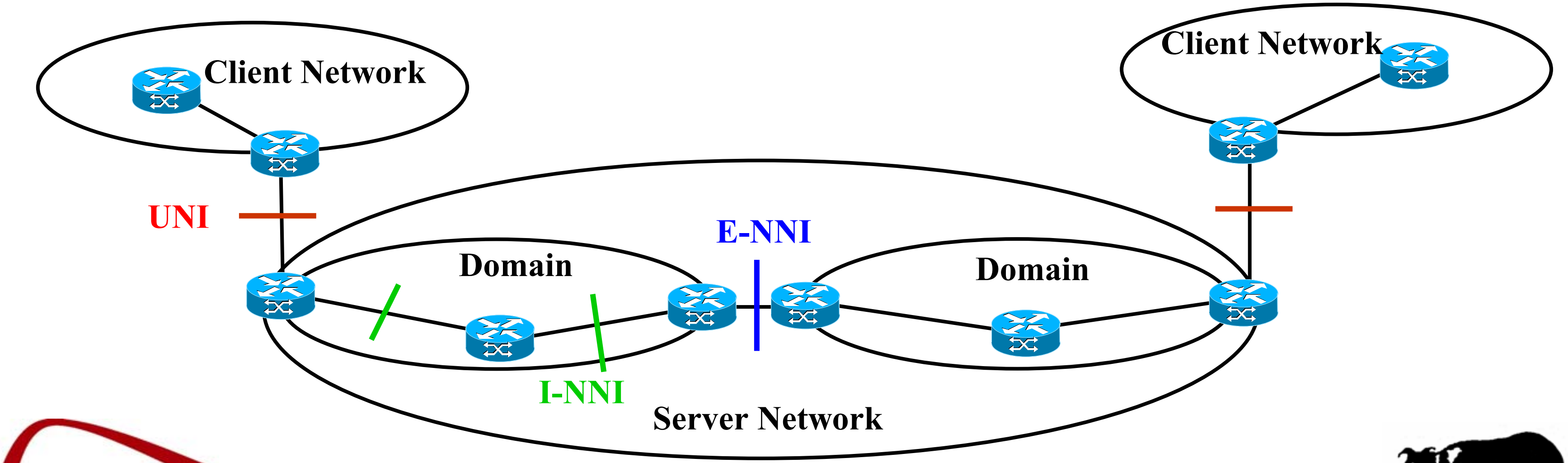
Layer One Virtual Private Networks

- VPN concepts can be extended to transport networks
- The whole network is presented as a single switch
  - Cross-connects are installed between labels (lambdas) on virtual ports
- BGP/GMPLS is the favoured technique



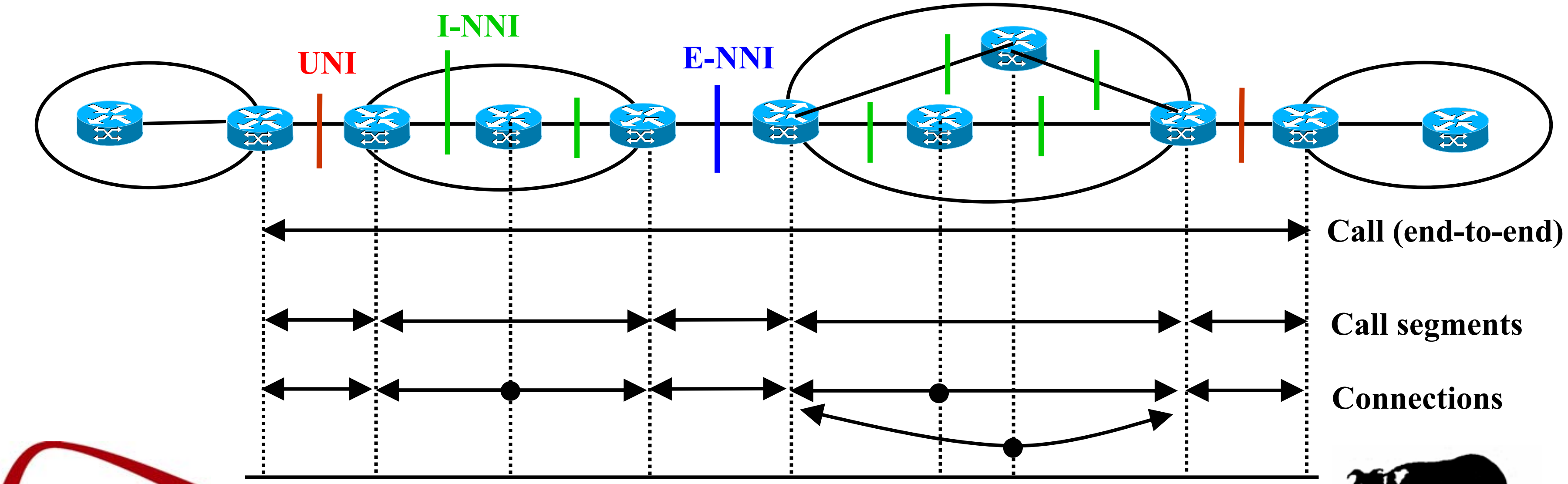
Reference Points in the ITU-T's ASON Architecture

- Reference points are *abstract functional interfaces*
  - They may lie between or within network nodes
- Client/Server split depends on data plane technology
- Domains allow for:
  - Differences in technology implementation
  - Administrative or operational splits
  - Different protection or computation policies
  - Different signaling capabilities (different protocols or just management)



ASON Calls and Connections

- A *connection* is part of the realization of an end-to-end service between two nodes or across a subnetwork (domain)
  - Initiated at UNI and E-NNI reference points
  - Processed at all reference points (including I-NNI)
- A *call* is used to coordinate the connections and to enable the service in an en-to-end manner.
  - Initiated at UNI reference points
  - Processed at UNI and E-NNI reference points (***NOT*** at I-NNI)
  - Signaling of call setup does not have to follow same path as connection setup

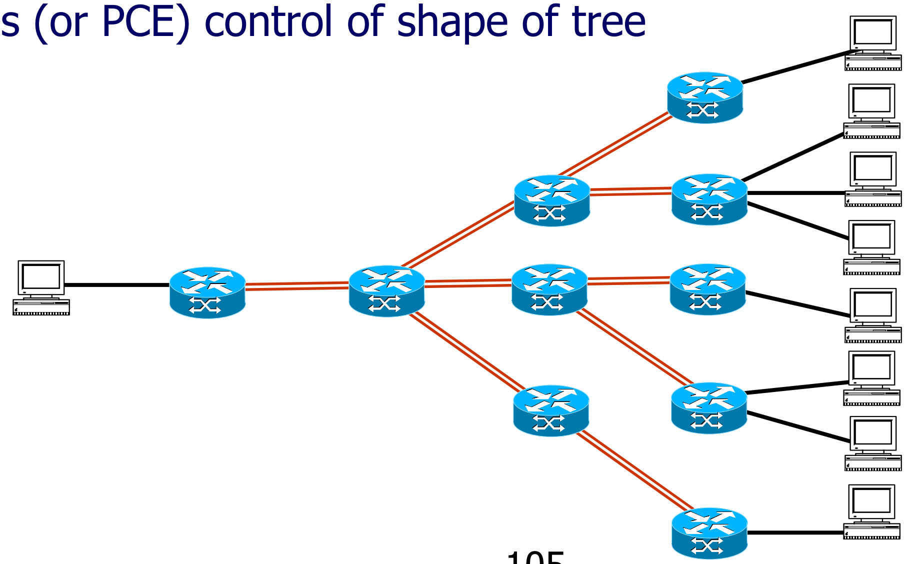


Integrating ASON with GMPLS

- GMPLS as specified is based on services and connections
  - Existing GMPLS mechanisms can be used to meet functional requirements at UNI, I-NNI and E-NNI
- The IETF has a strong end-to-end philosophy
  - Service state should not be held at transit nodes
- ITU-T and OIF have made additions to GMPLS
  - Signaling protocol for UNI and E-NNI reference points
  - Assumes that Call and Connection are established at same time
- IETF is working on extensions to GMPLS to add support for Calls
  - Connections must be identified with Calls
  - Call setup with/without Connections is required

Point-to-Multipoint Traffic Engineering

- Traffic engineering is not IP multicast!
- Applications
  - Content distribution
  - TE support of IP multicast
  - Multicast VPNs
  - MPLS-TE and all forms of GMPLS
- Significant functions
  - Graft/prune
  - Re-optimize
- Major requirements
  - Single copy of data on common paths
  - Ingress (or PCE) control of shape of tree



Questions?

Adrian Farrel (adrian@olddog.co.uk)
Zafar Ali (zali@cisco.com)