

# MPLS TE Fast Reroute



**Fernando Garcia**  
**fegarcia@cisco.com**

## TE - FRR Motivation

- A lot of SP are moving Voice Traffic (wireless and wireline) to IP Backbone
- IP Backbone must provide, at least, the same level of availability of PSTN
  - In general, sub-second for restoration after network failure
- TE-FRR provides this capability

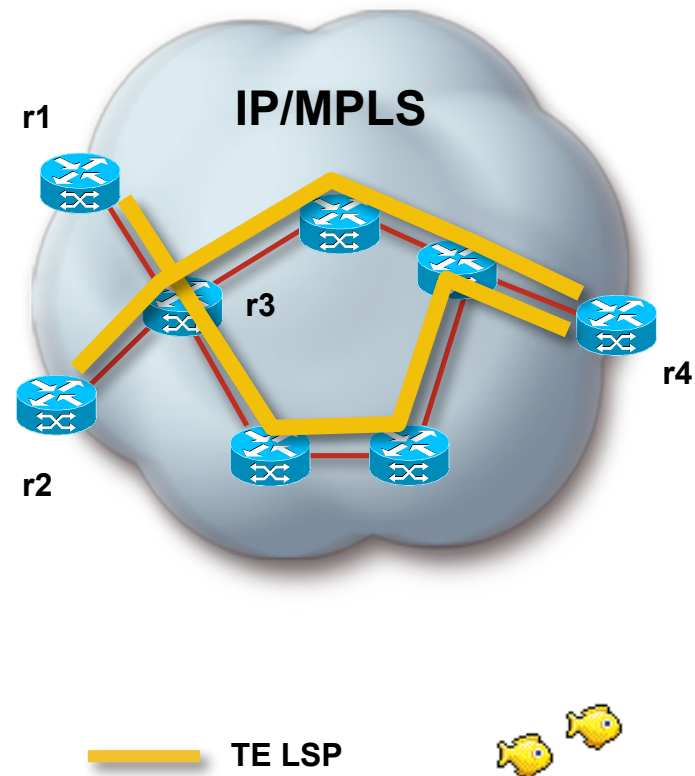
---

# Agenda

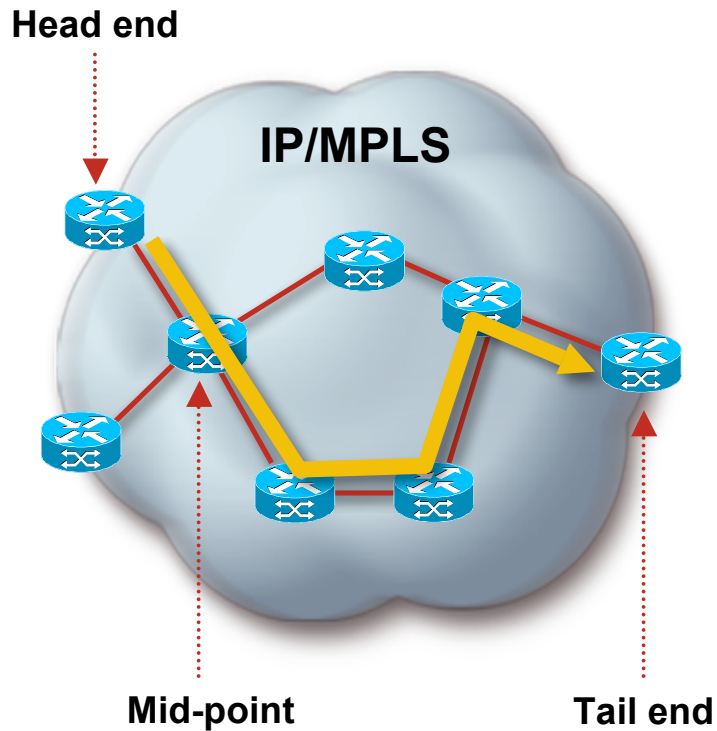
- MPLS TE Overview
  - MPLS TE Fast Re-route
-

# MPLS TE Overview

- Tunnel Based Technology - forwarding based on MPLS Label (LSP)
- Used basically to:
  - Bandwidth Optimization
  - Protection - FRR
- Supports **constrained-based routing**
- Introduces **explicit routing**
- Supports **admission control**
- Uses **ISIS and OSPF extensions** to advertise link attributes
- Uses **RSVP-TE** to establish LSPs



# How MPLS TE Works

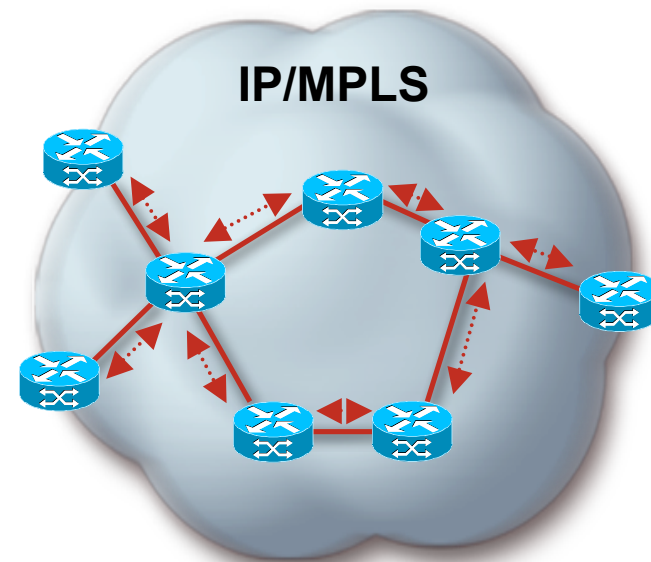


Steps necessary for MPLS TE:

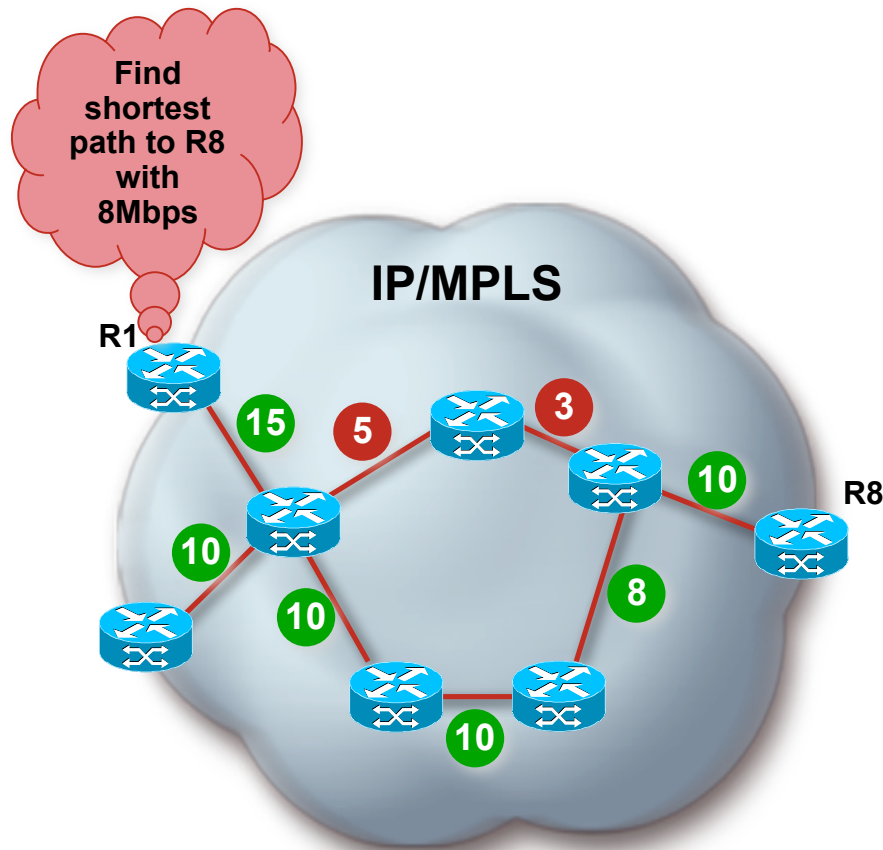
- Link information Distribution
  - ISIS-TE
  - OSPF-TE
- Path Calculation (CSPF)
- Path Setup (RSVP-TE)
- Traffic Selection



# Link Information Distribution - ISIS/OSPF

- IS-IS or OSPF extension flood link information – LSP or LSA
- Additional link characteristics
  - Physical bandwidth
  - Maximum reservable bandwidth
  - Unreserved bandwidth (at eight priorities)
  - TE metric
  - Administrative group (attribute flags)
- TE nodes build a topology database



# Path Calculation

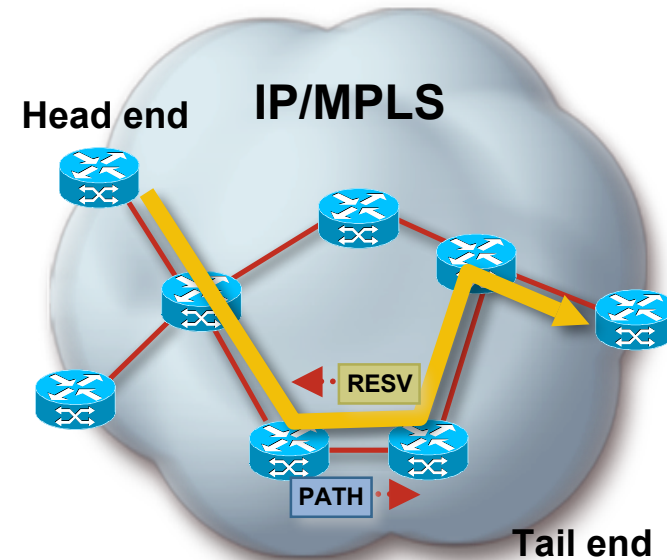


-  Link with insufficient bandwidth
-  Link with sufficient bandwidth

- Topology database as input to path computation
- TE nodes can perform constraint-based routing
- Shortest-path-first algorithm ignores links not meeting constraints
- Tunnel can be signaled once a path is found

# TE LSP Signaling

- Tunnel signaled with RSVP-TE
- Two message type:
  - PATH – From Head to Tail (downstream)
  - RESV – From Tail to Head (upstream)
  
- New RSVP objects



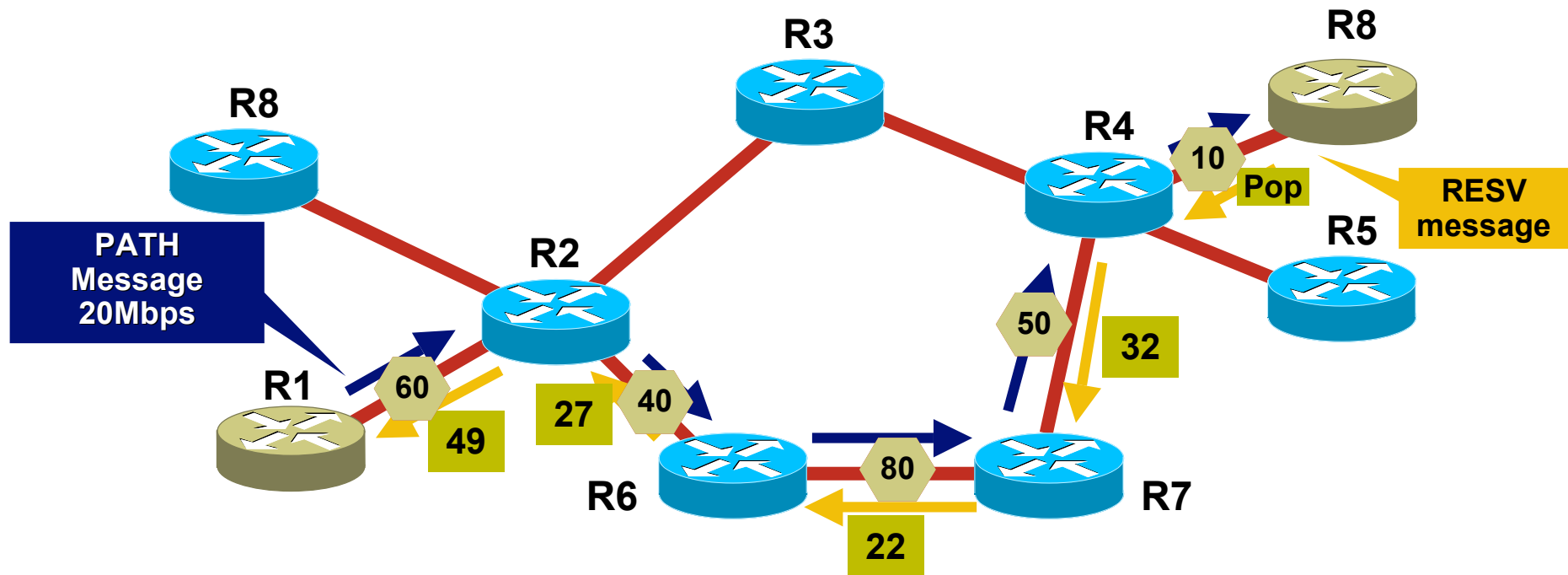
RSVP Object	RSVP Message	Description
LABEL_REQUEST	Path	Label request to downstream neighbor
LABEL	Resv	MPLS label allocated by downstream neighbor
EXPLICIT_ROUTE	Path	Hop list defining the course of the TE LSP
RECORD_ROUTE	Path, Resv	Hop/label list recorded during TE LSP setup
SESSION_ATTRIBUTE	Path	Requested LSP attributes (priority, protection, affinities)







# Trunk Admission Control

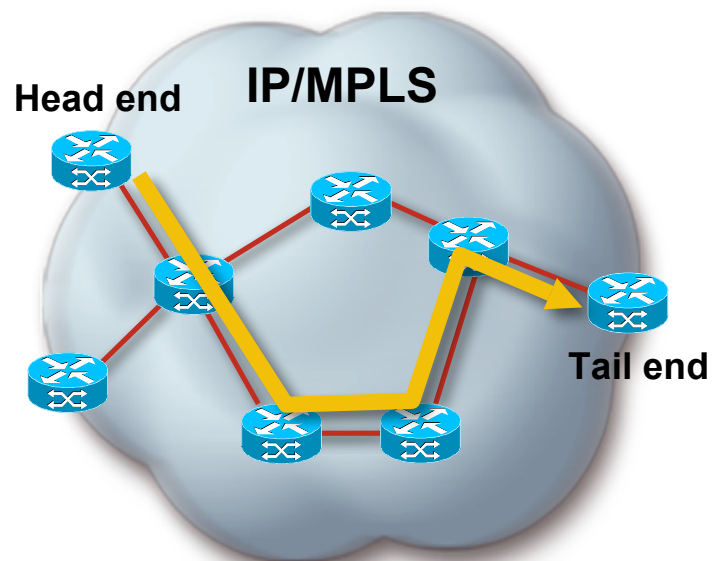
- On receipt of PATH message
  - Router will check if there is bandwidth available to honour the reservation
  - If bandwidth available then RSVP accepted
  - PATH message is sent to next hop (downstream)
- On receipt of a RESV message
  - Router actually reserves the bandwidth for the TE LSP
  - Label allocated
  - If pre-emption is required lower priority LSP are torn down
- OSPF/ISIS updates are triggered

# Path Setup Example



-  RSVP PATH: R1 → R2 → R6 → R7 → R4 → R8
-  RSVP RESV: Returns labels and reserves bandwidth on each link
-  Bandwidth available
-  Returned label via RESV message

# Traffic Selection

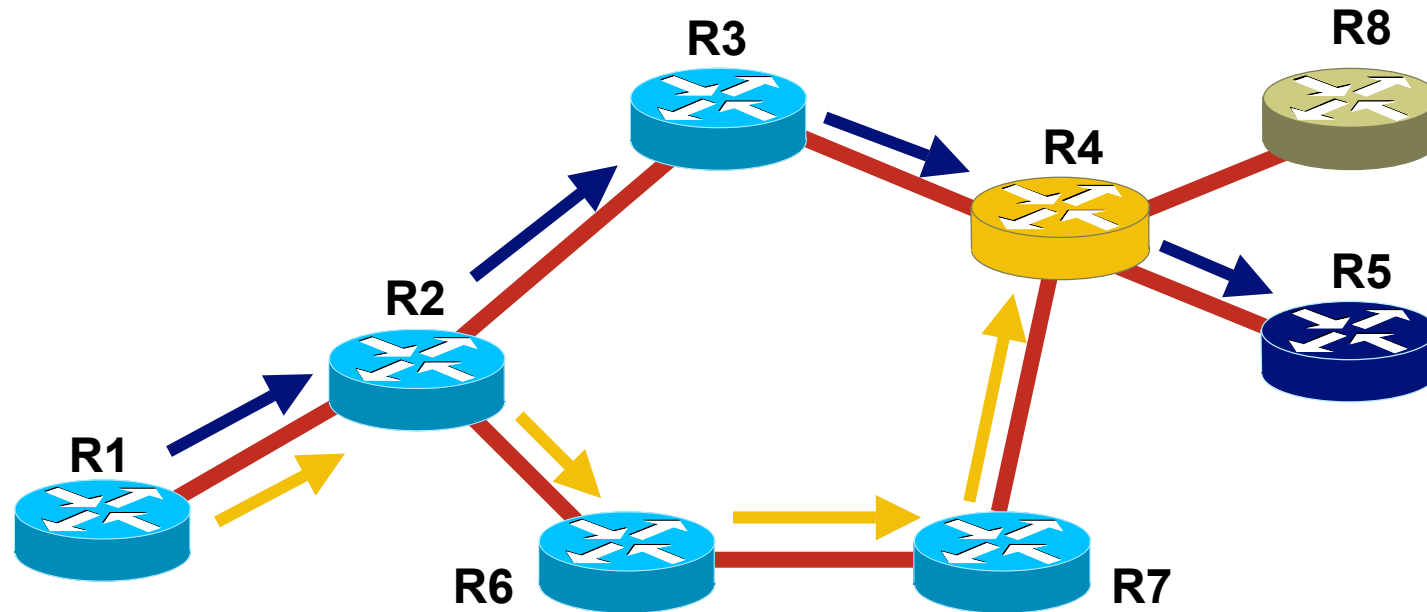


- Multiple traffic selection options
  - Auto-route
  - Static routes
  - Policy Based Routing
  - Forward Adjacency
  - Pseudowire Tunnel Selection
  - Class Based Tunnel Selection
- Tunnel path computation independent of routing decision injecting traffic into tunnel
- Traffic enters the tunnel at the head end

# Autoroute

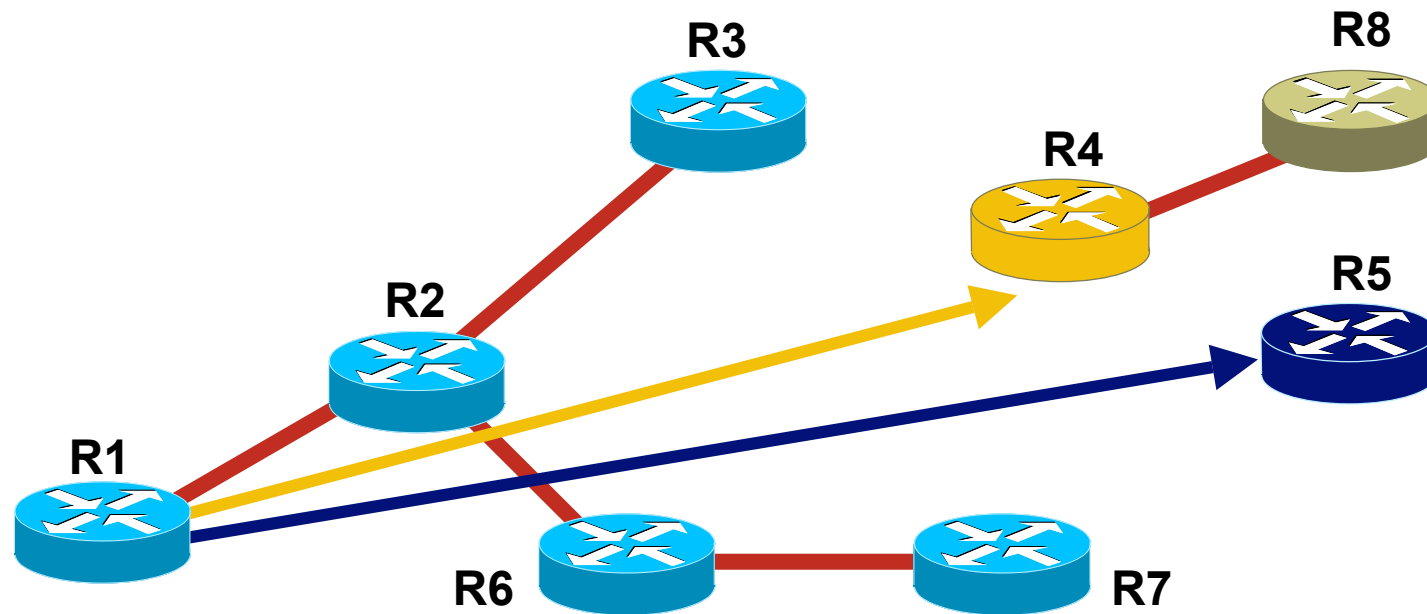
- Simplest manner to inject traffic into the tunnel
- Tunnel TE is a logical interface
- Used to include TE LSP(Logical Interface) in SPF calculations
- IGP adjacency is **NOT** run over the tunnel!
- Tunnel is treated as a directly connected link to the tail

# Autoroute Topology (OSPF and ISIS)



➡ Tunnel1: R1 → R2 → R3 → R4 → R5  
➡ Tunnel2: R1 → R2 → R6 → R7 → R4

# Autoroute Topology (OSPF and ISIS)



From R1 Router Perspective:



Next hop to R4 and R8 is Tunnel1



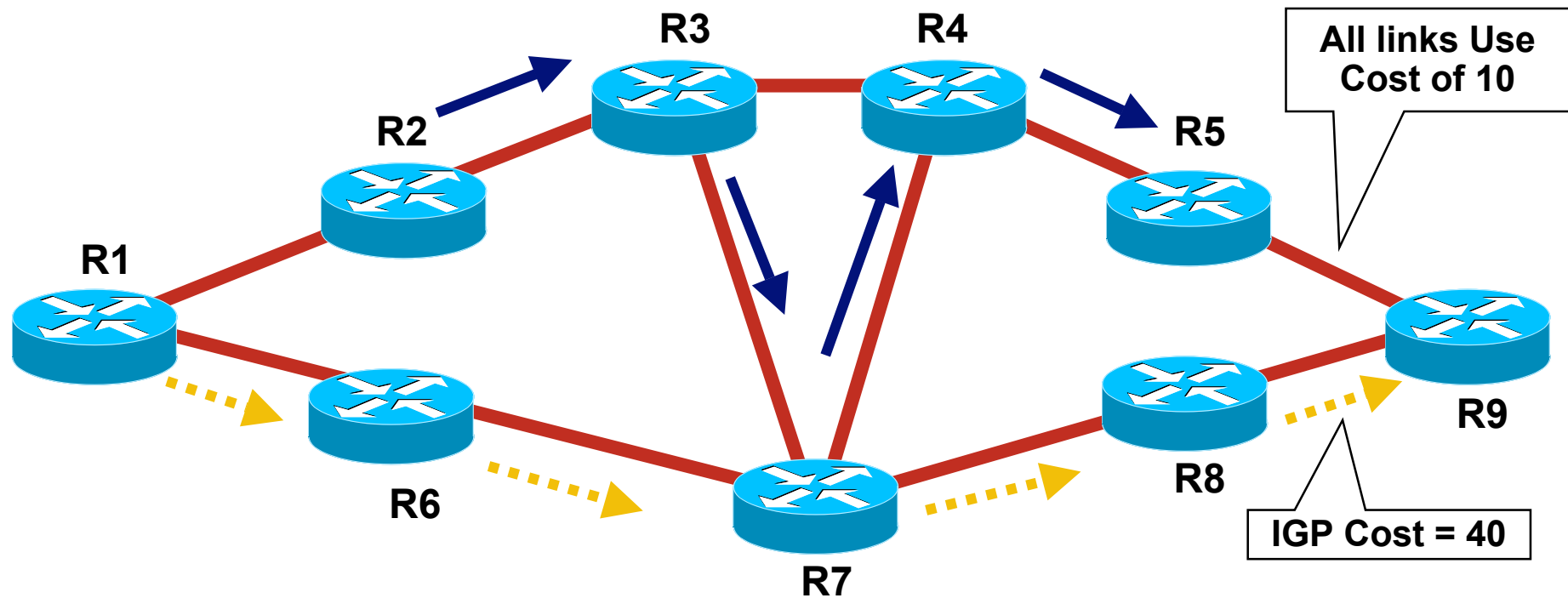
Next hop to R5 is Tunnel2

All nodes behind tunnel routed via tunnel

# Forwarding Adjacency

- Autoroute does not advertise the LSP (tunnel interface) into the IGP - Routers behind Head End maybe doesn't use TE
- FA advertises the existence of TE tunnels (new logical interfaces) into IGP
- Can get suboptimal forwarding (**NOT** loops) if you're not careful

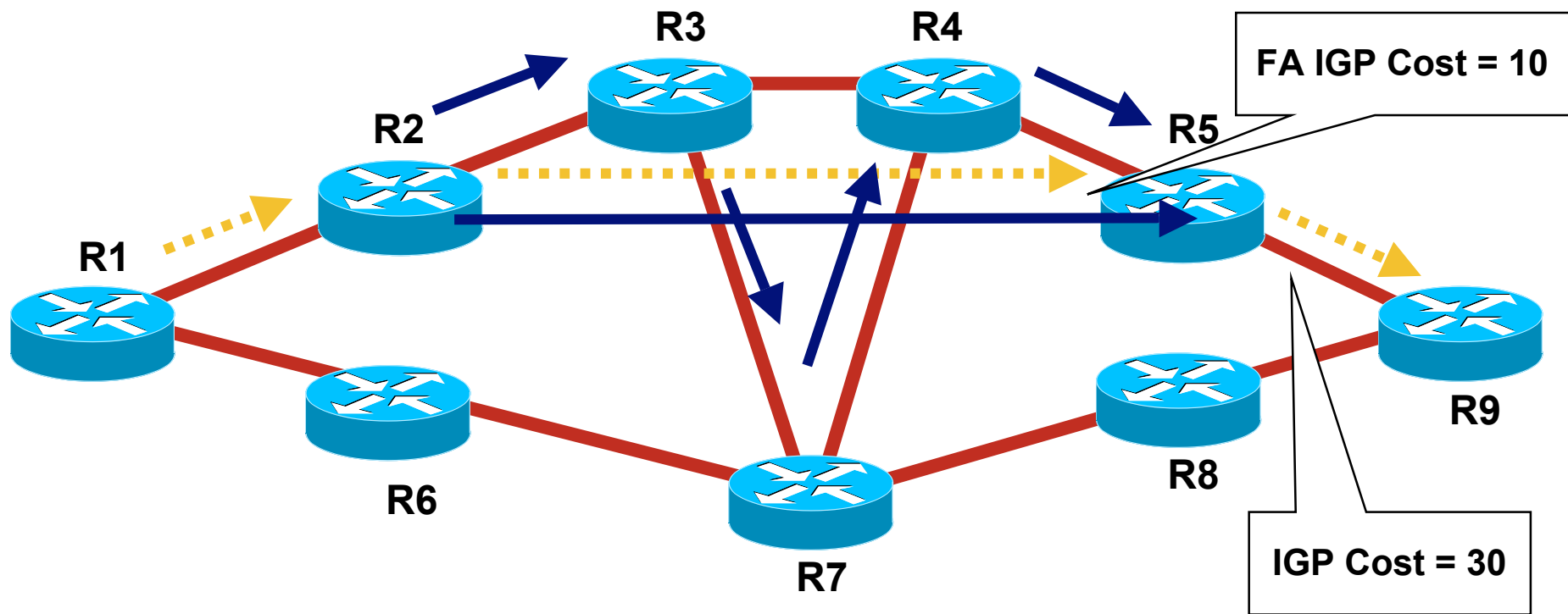
# Forwarding Adjacency



**—————▶** Tunnel: R2 → R3 → R7 → R4 → R5  
**.....▶** R1 shortest path to R9 via IGP  
Tunnel at R2 is never used as R1 can't see it

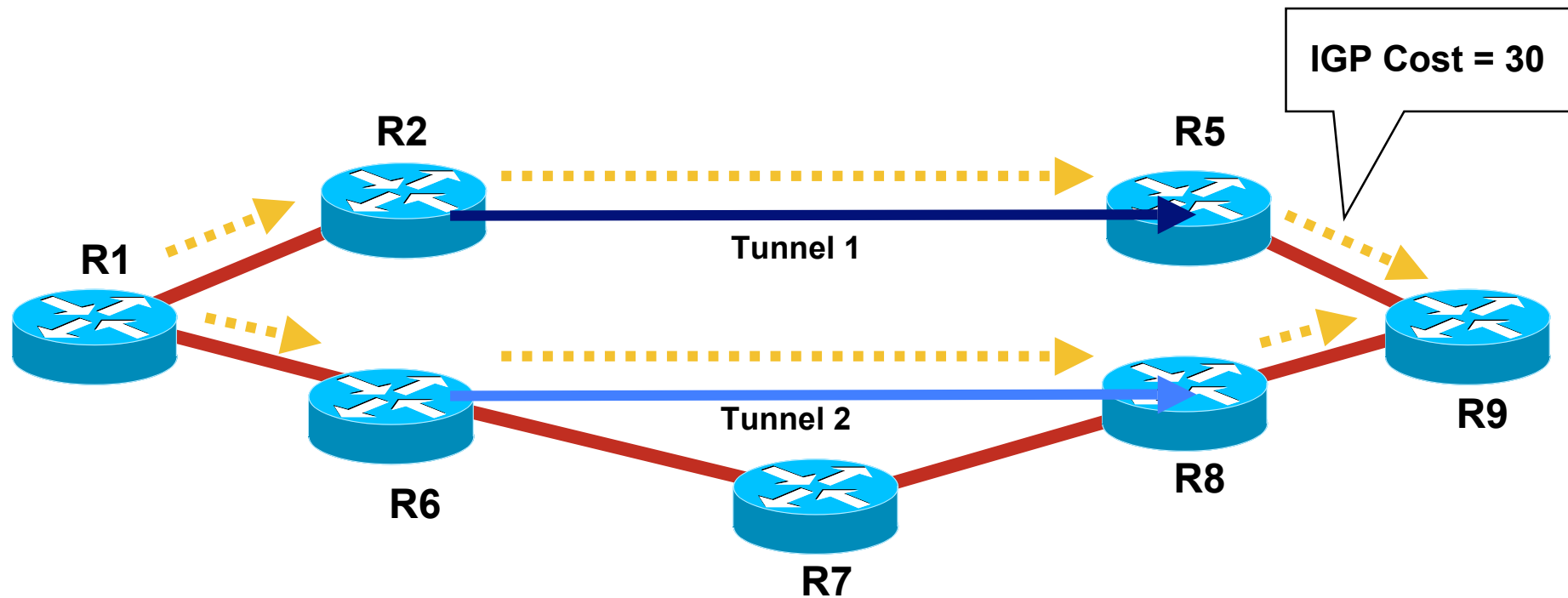



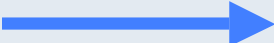

# Advertise TE Links into IGP



**→ Tunnel: R2 → R3 → R7 → R4 → R5**  
**→ R1 now uses R2 as NH. Traffic From R1 to R9 is now tunneled on R2**

# Load Balancing Across FA



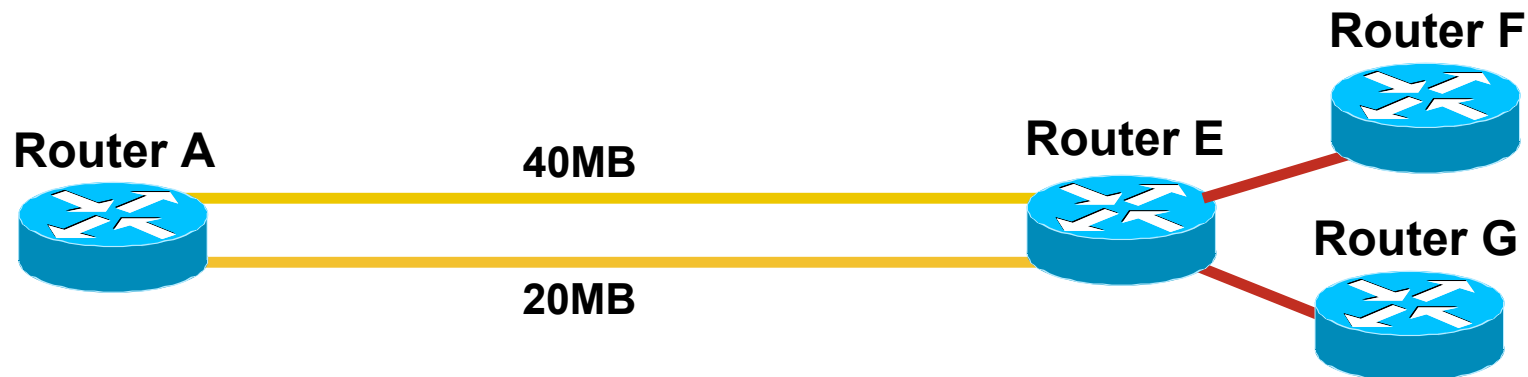
-  Tunnel1: R2 → R3 → R7 → R4 → R5
-  Tunnel2: R6 → R7 → R8
-  R1 shortest path to R9

# TE and Unequal Cost Load Balancing

- IP routing has equal-cost load balancing, but not unequal cost\*
- Unequal cost load balancing difficult to do while guaranteeing a loop-free topology
- Since MPLS doesn't forward based on IP header, permanent routing loops don't happen
- 16 hash buckets for next-hop, shared in **rough** proportion to configured tunnel bandwidth or load-share value

\*EIGRP Has 'Variance', but That's Not as Flexible

# Unequal Cost: Example 1



```
gsr1#show ip route 192.168.1.8
Routing entry for 192.168.1.8/32
  Known via "isis", distance 115, metric 83, type level-2
  Redistributing via isis
  Last update from 192.168.1.8 on Tunnel0, 00:00:21 ago
  Routing Descriptor Blocks:
  * 192.168.1.8, from 192.168.1.8, via Tunnel0
    Route metric is 83, traffic share count is 2
  192.168.1.8, from 192.168.1.8, via Tunnel1
    Route metric is 83, traffic share count is 1
```

# Unequal Cost: Example 1



```
gsr1#sh ip cef 192.168.1.8 internal
```

```
.....
```

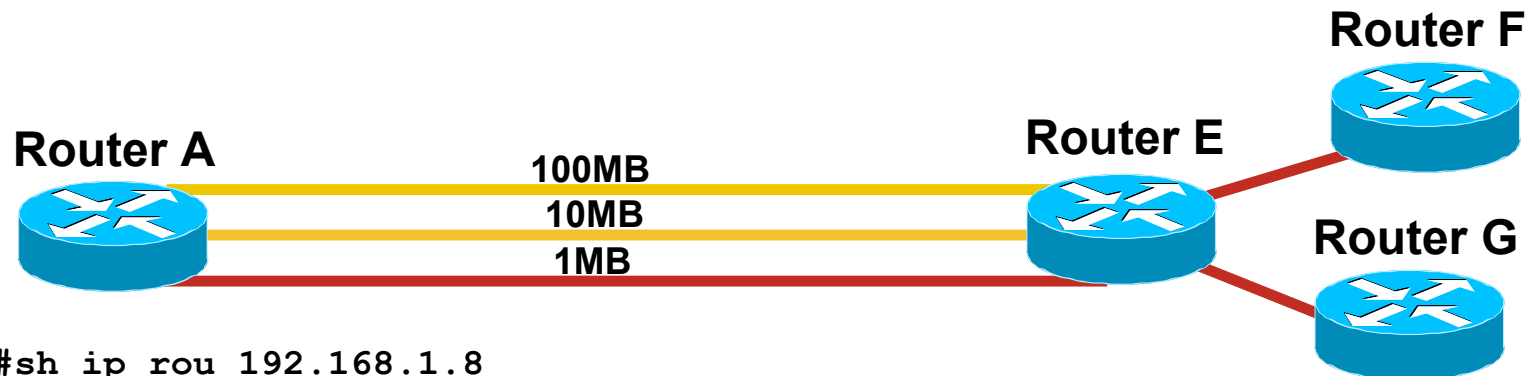
```
Load distribution: 0 1 0 1 0 1 0 1 0 1 0 0 0 0 0 0 (refcount 1)
```

Hash	OK	Interface	Address	Packets	Tags imposed
1	Y	Tunnel0	point2point	0	{23}
2	Y	Tunnel1	point2point	0	{34}
3	Y	Tunnel0	point2point	0	{23}

```
.....
```

**Note That the Load Distribution Is 11:5—Very Close to 2:1, but Not Quite!**

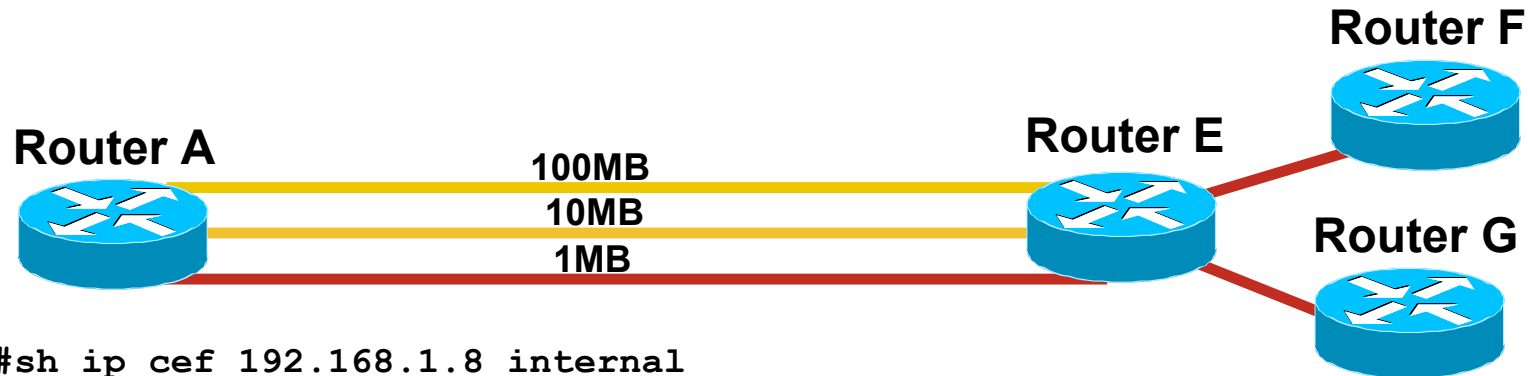
## Unequal Cost: Example 2



```
gsr1#sh ip rou 192.168.1.8
Routing entry for 192.168.1.8/32
  Known via "isis", distance 115, metric 83, type level-2
  Redistributing via isis
  Last update from 192.168.1.8 on Tunnel2, 00:00:08 ago
  Routing Descriptor Blocks:
  * 192.168.1.8, from 192.168.1.8, via Tunnel0
    Route metric is 83, traffic share count is 100
  192.168.1.8, from 192.168.1.8, via Tunnel1
    Route metric is 83, traffic share count is 10
  192.168.1.8, from 192.168.1.8, via Tunnel2
    Route metric is 83, traffic share count is 1
```

**Q: How Does 100:10:1 Fit Into a 16-Deep Hash?**

## Unequal Cost: Example 2



```
gsr1#sh ip cef 192.168.1.8 internal
```

```
.....
```

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

Hash	OK	Interface	Address	Packets	Tags imposed
1	Y	Tunnel0	point2point	0	{36}
2	Y	Tunnel1	point2point	0	{37}

```
.....
```

**A: Any Way It Wants to! 15:1, 14:2, 13:2:1, it depends on the order the tunnels come up**

**Deployment Guideline: Don't use tunnel metrics that don't reduce to 16 buckets!**

# Path Maintenance

- Path re-optimization

Process where some traffic trunks are rerouted to new paths so as to improve the overall efficiency in bandwidth utilization

For example, traffic may be moved to secondary path during failure; when primary path is restored traffic moved back

- Path restoration

Comprised of two techniques; local protection (link and node) and path protection

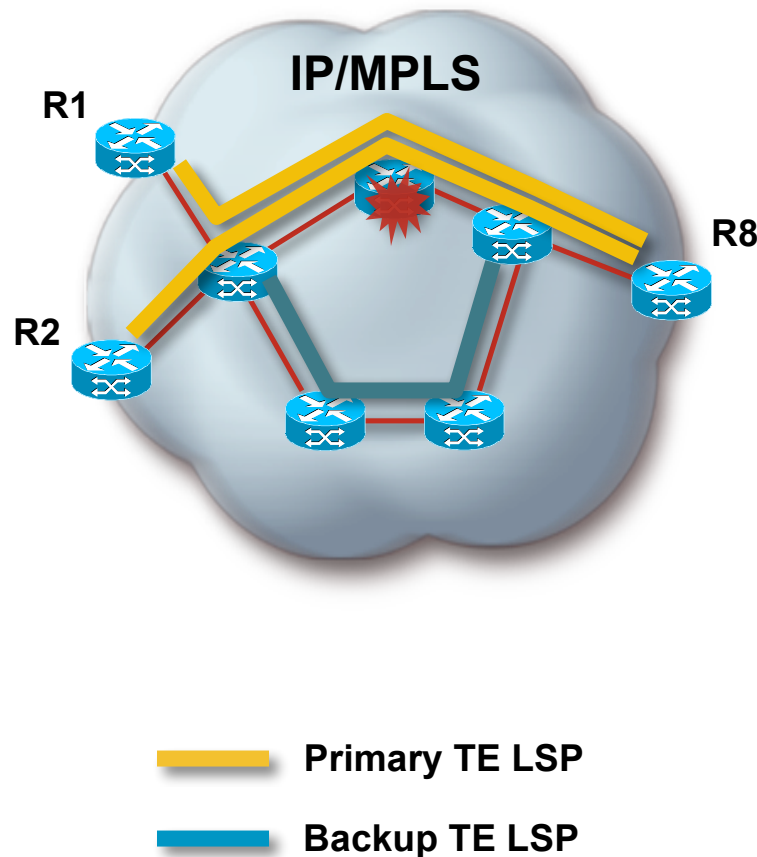
Discussed later



## Fast Reroute



# MPLS TE Fast Re-Route (FRR)



- **Subsecond recovery** against node/link failures – Fast Restoration
- Mechanism to minimize packet loss during a failure
- Scalable 1:N protection
- Cost-effective **alternative to optical protection - APS**
- **Bandwidth protection**
- A lot of SP (wireless) are implementing TE-FRR
  - T-Mobile UK, Verizon, TI, Vodafone...

# FRR - MPLS TE Protection

- MPLS TE protection also known as **FAST REROUTE (FRR)**
- Pre-provisioned protection tunnels that carry traffic when a protected link or node goes down
- FRR protects against **LINK FAILURE**
  - For example, Fibre cut, Carrier Loss, ADM failure
- FRR protects against **NODE FAILURE**
  - For example, power failure, hardware crash, maintenance

# Categories of Fast Reroute Protection

- Local protection

  - Link protection

  - Node protection

  - Protect a piece of the network (node or link)

  - 1:N scalability

  - Fast failure recovery due to local repair

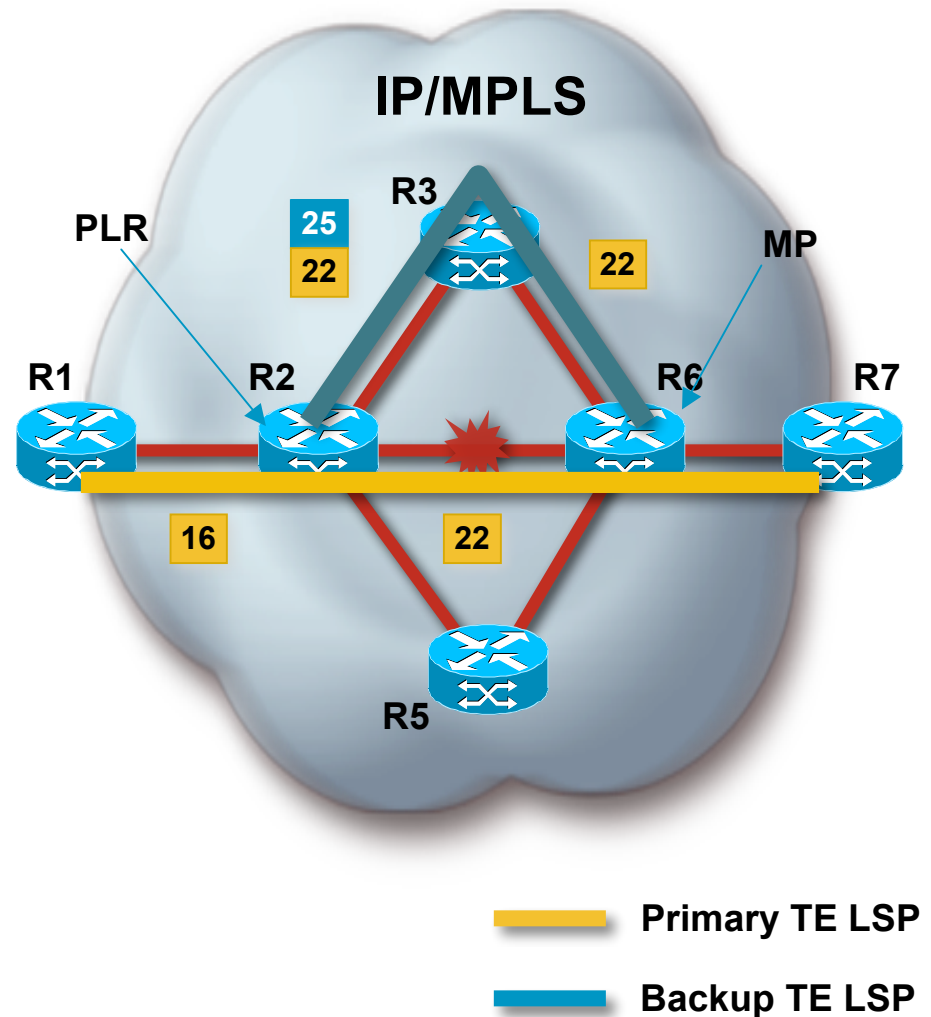
- Path protection

  - Protects individual tunnels

  - 1:1 scalability

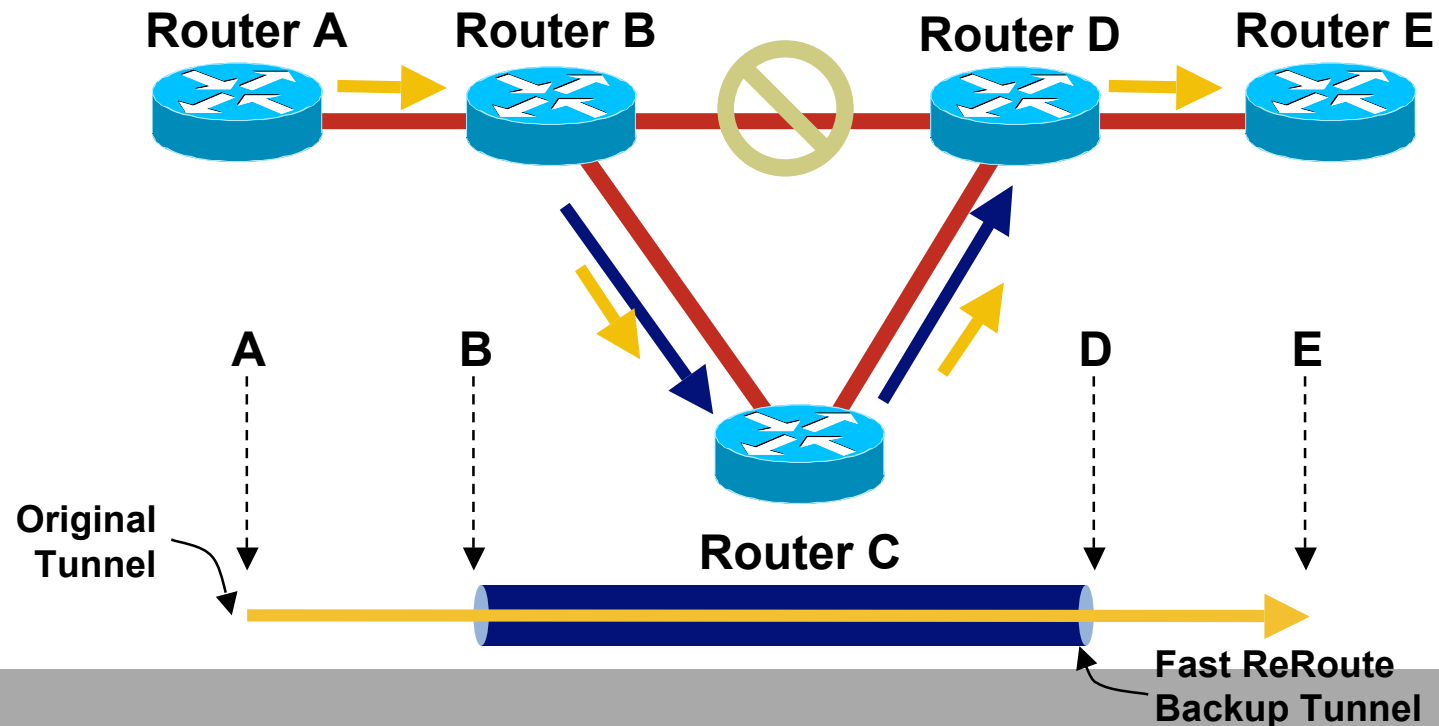
# FRR Link Protection Operation

- Requires **next-hop** (NHOP) backup tunnel
- Point of Local Repair (PLR) swaps label and pushes backup label
- Backup terminates on Merge Point (MP) where traffic rejoins primary
- Restoration time expected under ~50 ms

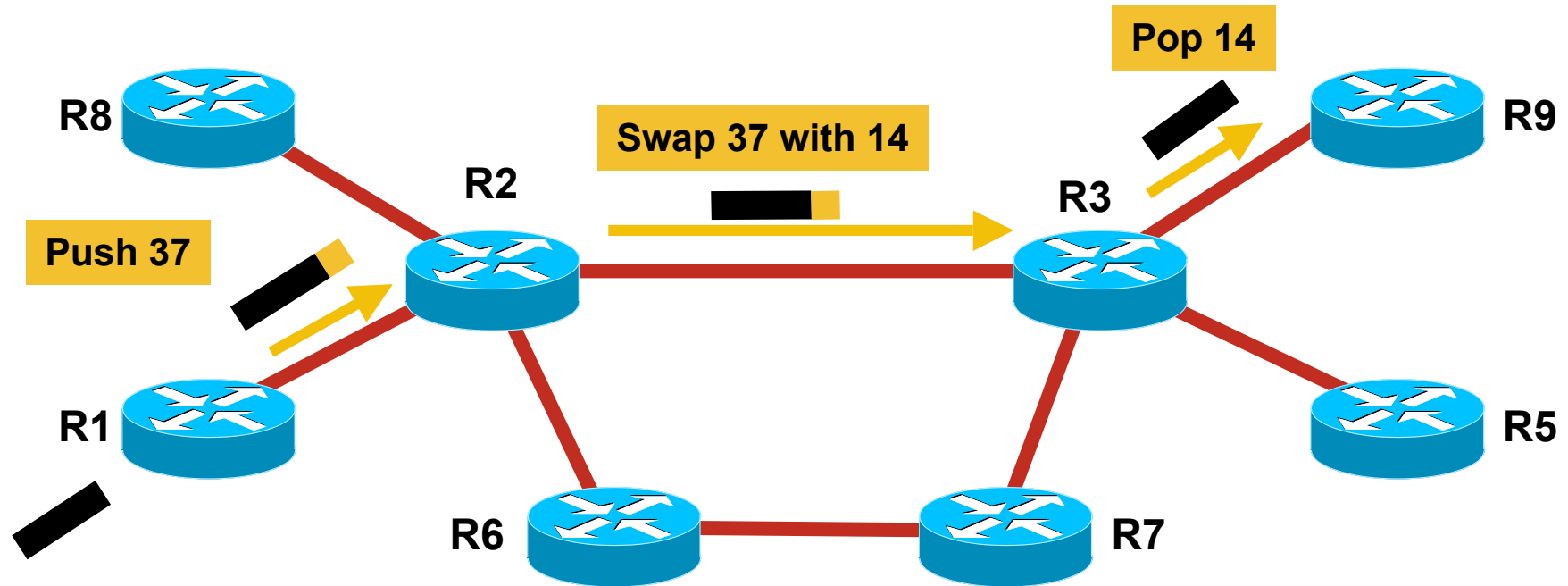


# Link Protection

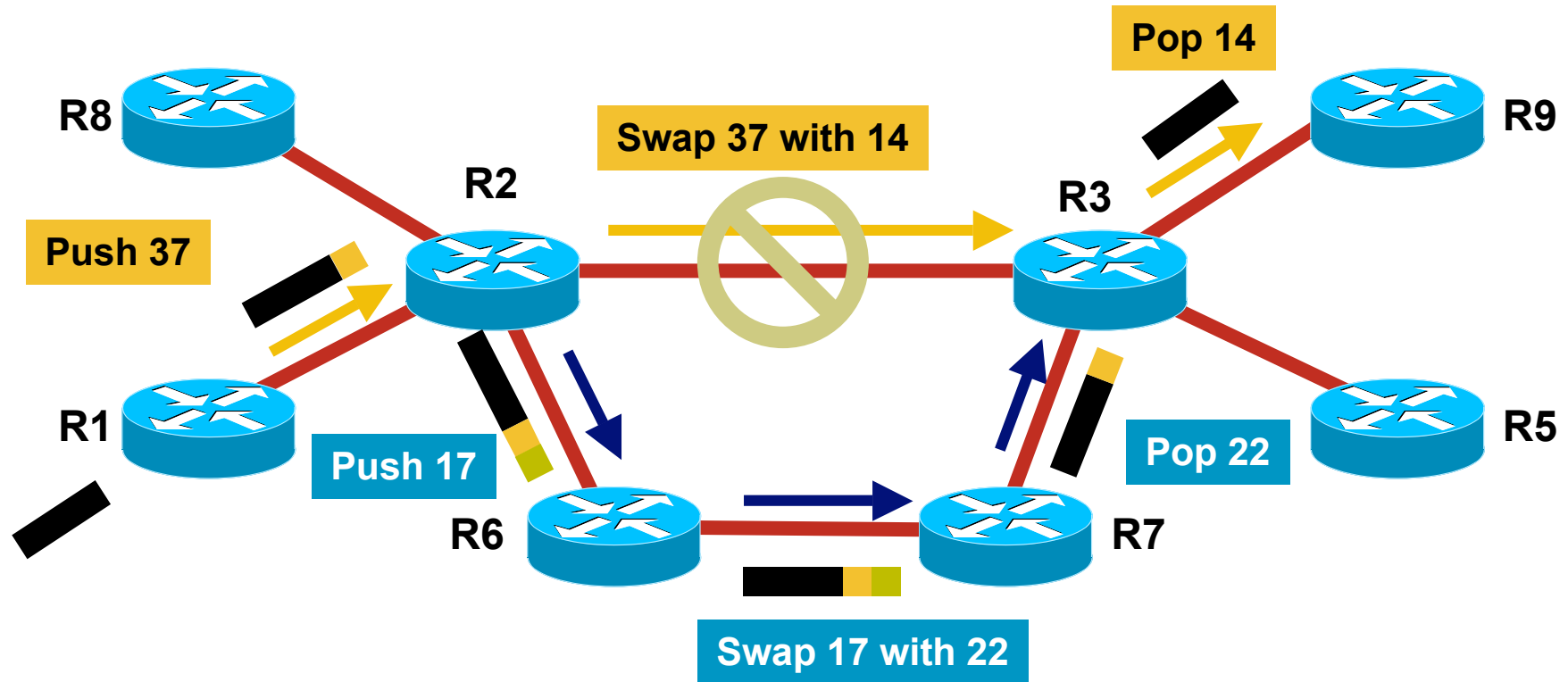
- When B → D link fails, A → E tunnel is encapsulated in B → D tunnel
- Backup tunnel is used until A can re-compute tunnel path as A → B → C → D → E (~5-15 seconds or so)



# Normal TE Operation

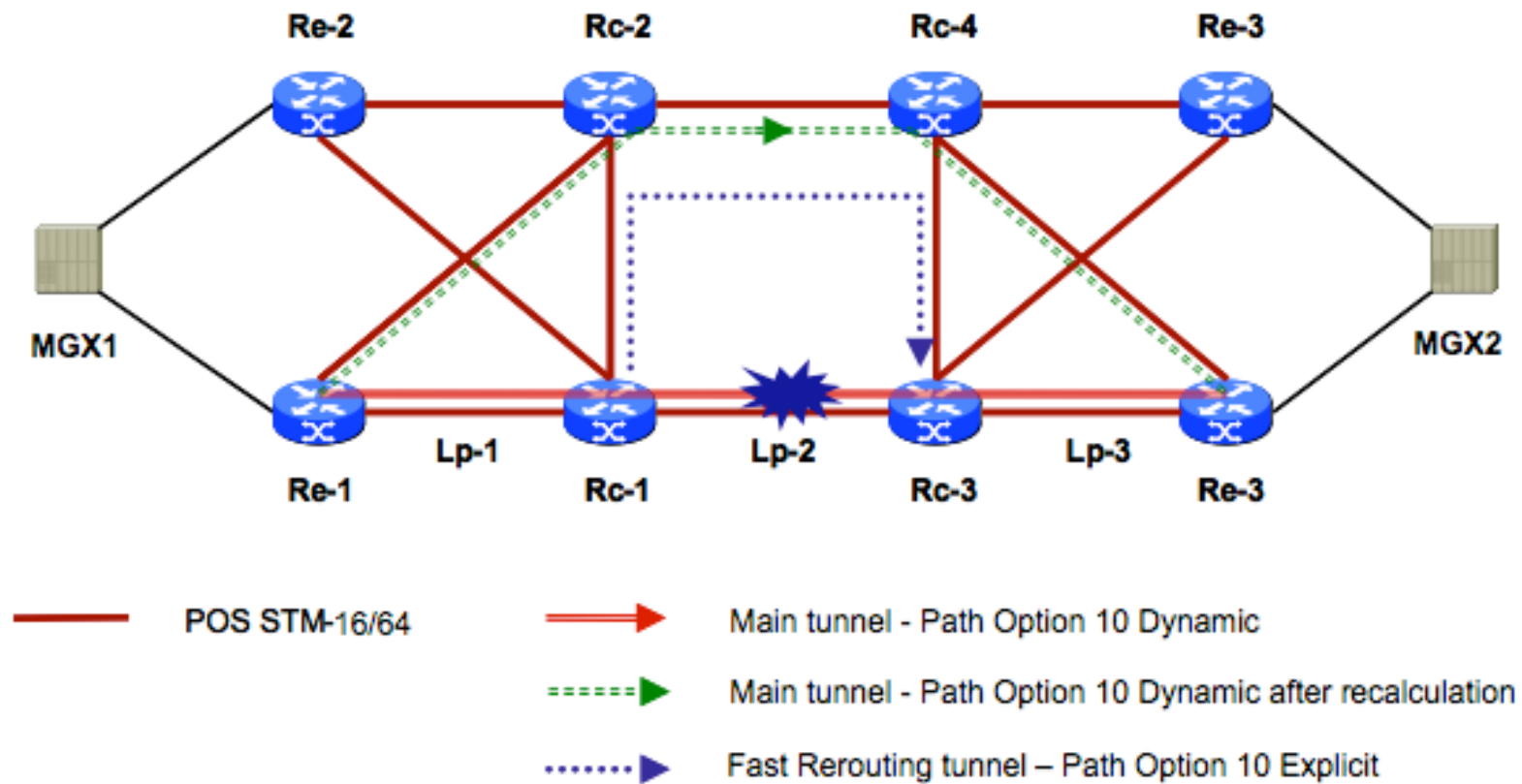


# Fast Reroute Link Failure



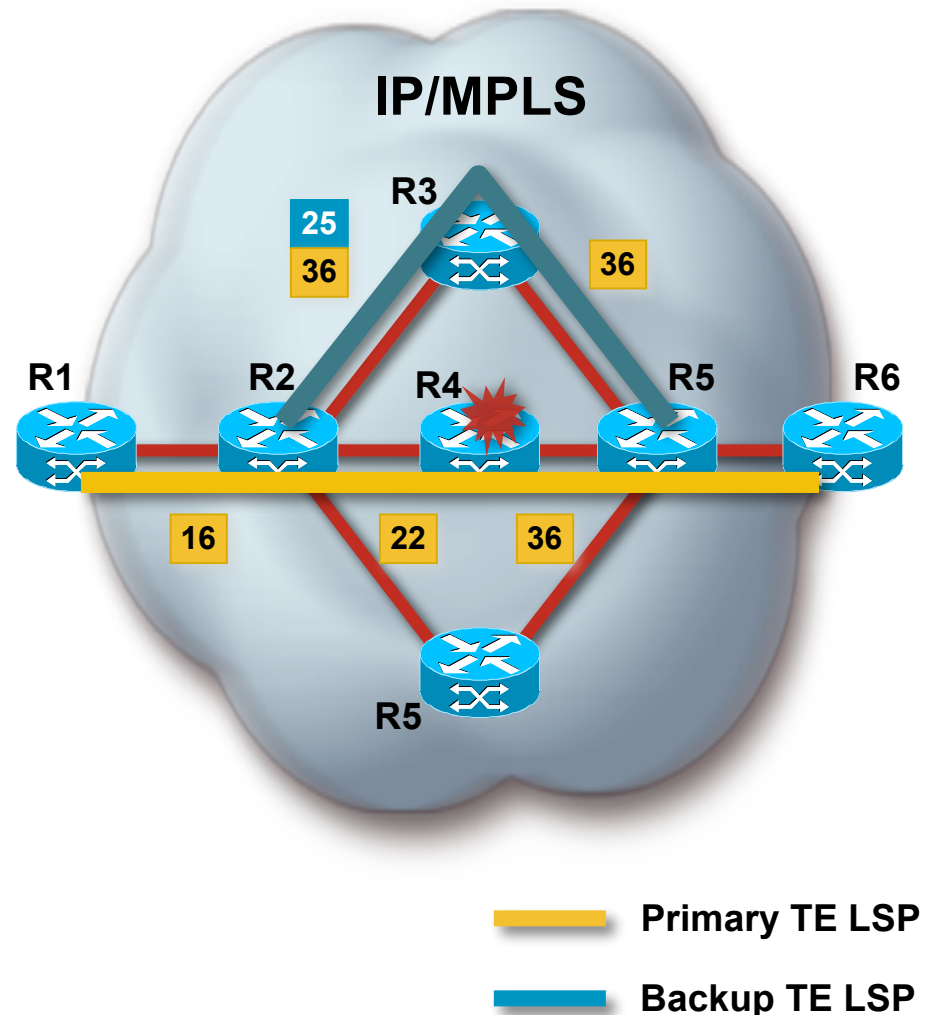


# Real Example - FRR at TI



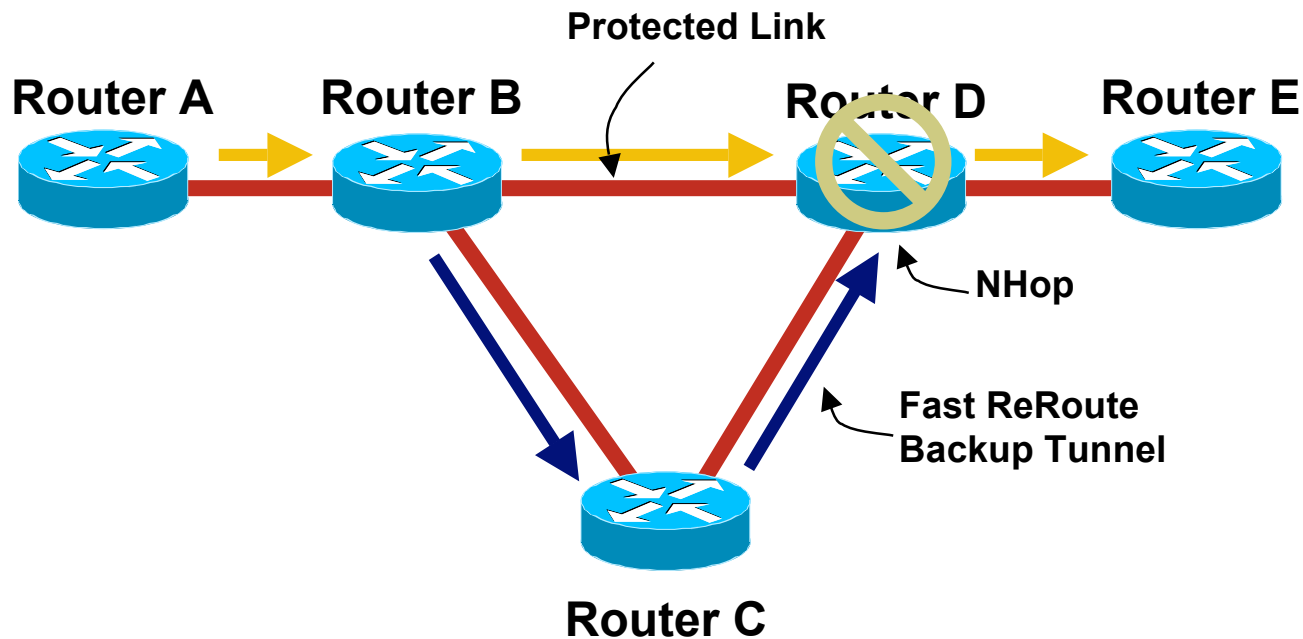
# FRR Node Protection Operation

- Requires **next-next-hop** (NNHOP) backup tunnel
- Point of Local Repair (PLR) swaps **next-hop label** and pushes backup label
- Backup terminates on Merge Point (MP) where traffic rejoins primary
- Restoration time depends on failure detection time



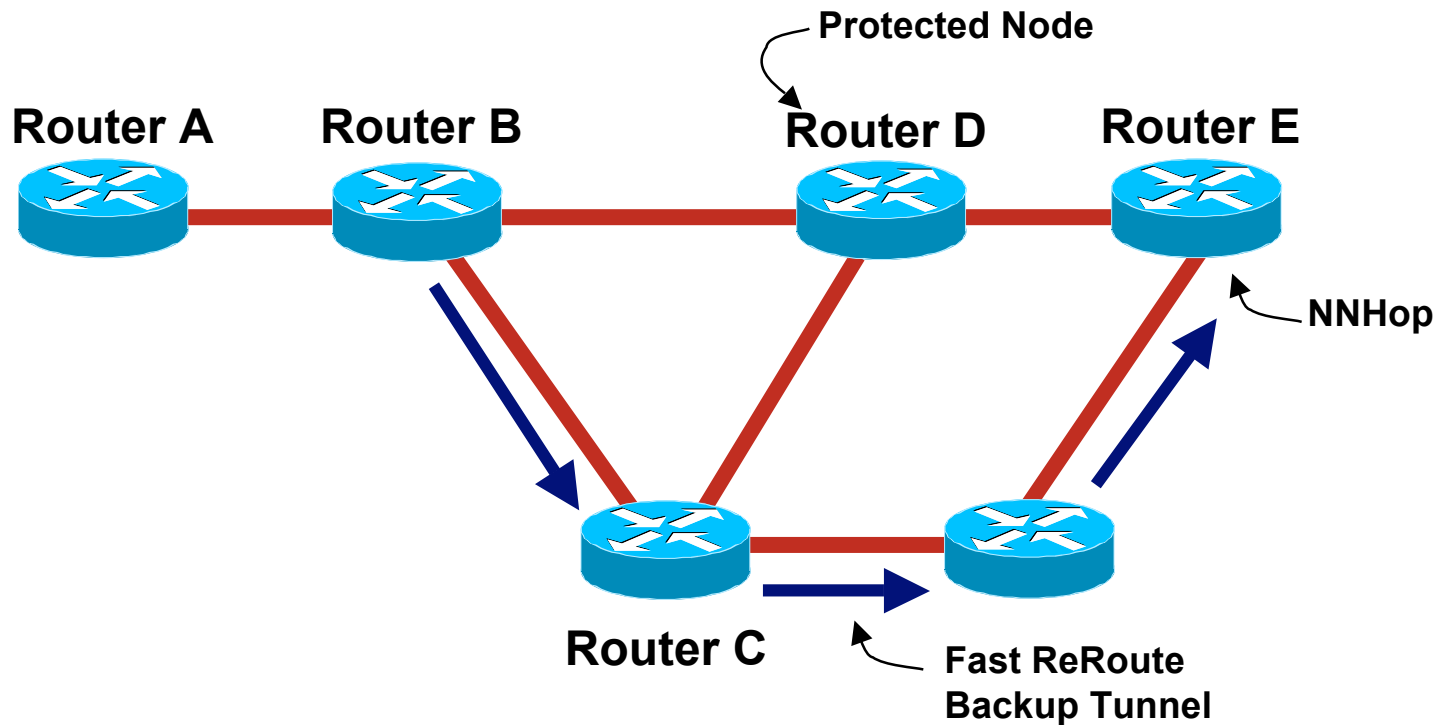
# Node Protection

- What if Router D failed?
- Link protection would not help as the backup tunnel terminates on Router D (which is the NHop of the protected link)



# Node Protection

- **SOLUTION: NODE PROTECTION** (If network topology allows)
- Protect tunnel to the next hop **PAST** the protected link (NNHop)



# Node Protection

- Node protection still has the same convergence properties as link protection
- Deciding where to place your backup tunnels is a much harder problem to solve on a large-scale
- For small-scale protection, link may be better
- Configuration is identical to link protection, except where you terminate the backup tunnel (NNHop vs. NHop)

---

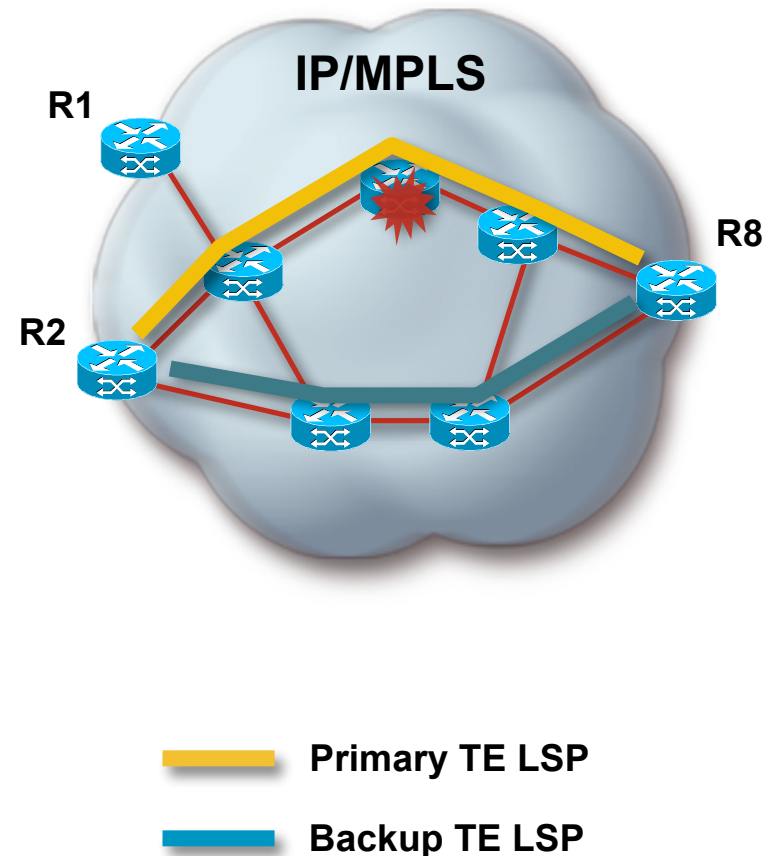
# Link and Node Protection Times

- Link and Node protection are very similar
- Protection time depends on failure detection
- SP worldwide has achieved 50~100ms



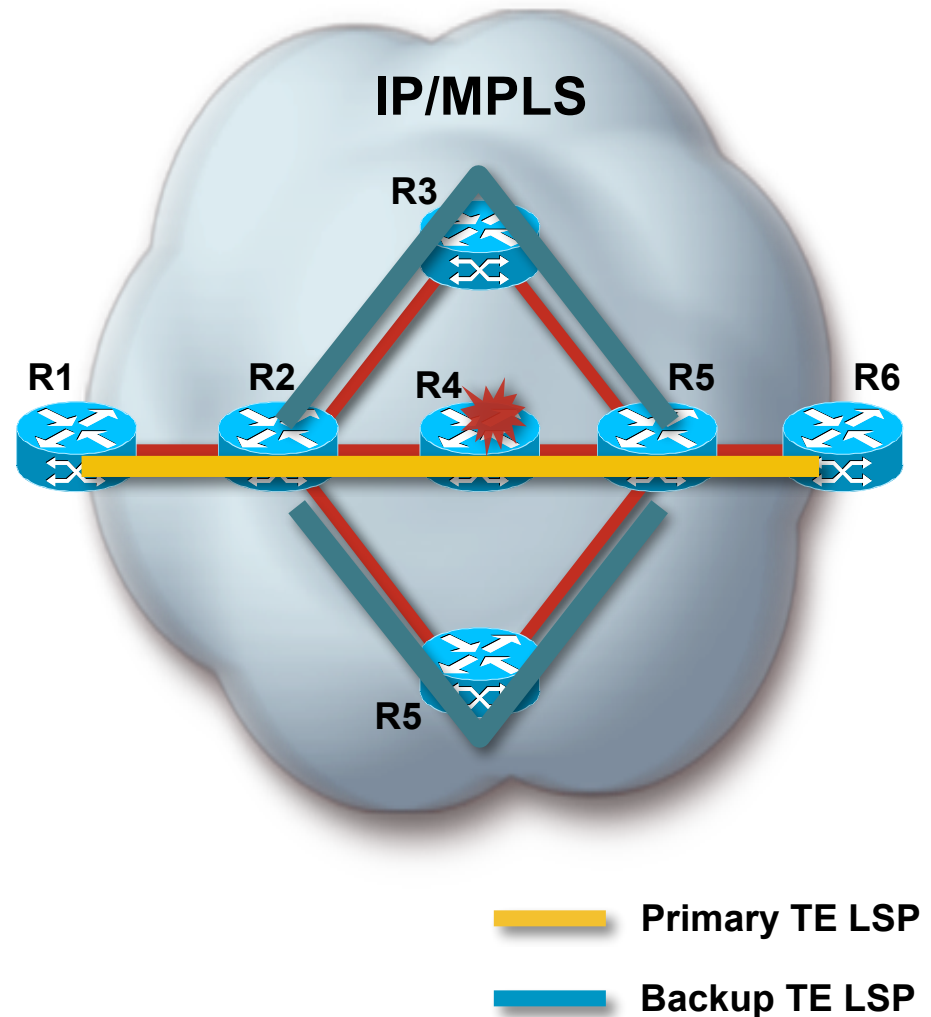
# What about Path Protection?

- Primary and backup share head and tail, but diversely routed
- Expected to result in higher restoration times compared to local protection
- Doubles number of TE LSPs (1:1 protection)



# Bandwidth Protection

- Backup tunnel with associated bandwidth capacity
- Backup tunnel may or may not actually signal bandwidth
- PLR will decide best backup to protect primary (nhop/nnhop, class-type, node-protection flag)





# Conclusion

- A lot of SP are replacing traditional voice transport network for IP network
- At least the same level of availability (5 nines) and restoration time (sub-second) needs to be achieved
- Other applications demands the same availability/restoration or even worse
  - Video Broadcast/Video ondemand (integrated at 3Play)
- TE FRR is one of the best options, at relative low cost, to provide those requirements
- Its is a mature technology - Some SP are running it successfully for more than 6 years

Obrigado !!!

