SYSC 5801

# Protection and Restoration

# Introduction

- Fact: Networks fail. Types of failures:
  - ✓ Path failures
  - ✓ Link failures
  - ✓ Node failures
- Results: packet losses, waste of resources, and higher delay.
- What IGP does in the event of failures?
  - ✓ Quickly route around failures
  - ✓ Converge on the remaining topology
- What IGP doesn’t do when it comes to convergence:
  - ✓ IGP may take a few seconds (5-10 sec not uncommon) or longer.
  - ✓ A link failure can lead to congestion in some parts while leaving other parts underutilized.
  - ✓ Configuring the IGP to converge quickly can make it very sensitive to minor packet loss, causing false negatives and IGP convergence for no reason.

# How Can MPLS Help?

- Assuming IGP is used, SPF needs to be run when a link failure occurs and then again when it comes back up: time consuming and possible instability
- For MPLS, the problem is solved?
- It may be worse if a link that is part of an LSP fails.
  - ✓ The LSP is torn down.
  - ✓ The headend is notified.
  - ✓ The headend or ingress recomputes a new path (using probably CSPF) based on the topology information obtained from SPF.
  - ✓ Signal a new LSP through RSVP and run SPF for destinations that need to be routed over the tunnel.
  - ✓ This is called headend LSP reroute or headend reroute or path protection.
  - ✓ A few seconds may be acceptable in general for data traffic, but not for real-time applications like voice, video.
- Could be faster if a backup path has been **pre-established** at the headend. But …
  - ✓ What is the other performance bottleneck?

# Fast Reroute or Protection

- So, what is the benefit and how can it help?
  - ✓ Use MPLS-TE **Fast Reroute (FRR)**
- Mechanisms to address how do minimize loss as much as possible is known as FRR or simply protection.
- Practically, it means SONET-like recovery times (50ms or less) to a few hundred milliseconds of loss before FRR is effective.
- Protected resources could be physical resources (links or nodes) or logical resources (LSPs).
- Protection really means, in this context, the protection of logical resources (LSPs) from physical resources (links or nodes).
- For MPLS effectively to support failure handling,
  - ✓ Backup resources are **pre-established** and are not signaled after a failure has occurred. This is different from headend reroute.
  - ✓ Performance bottleneck is minimized: **short notification time** – local protection/repair.
- The pre-established LSPs are called *backup tunnel* or *protection tunnel.*

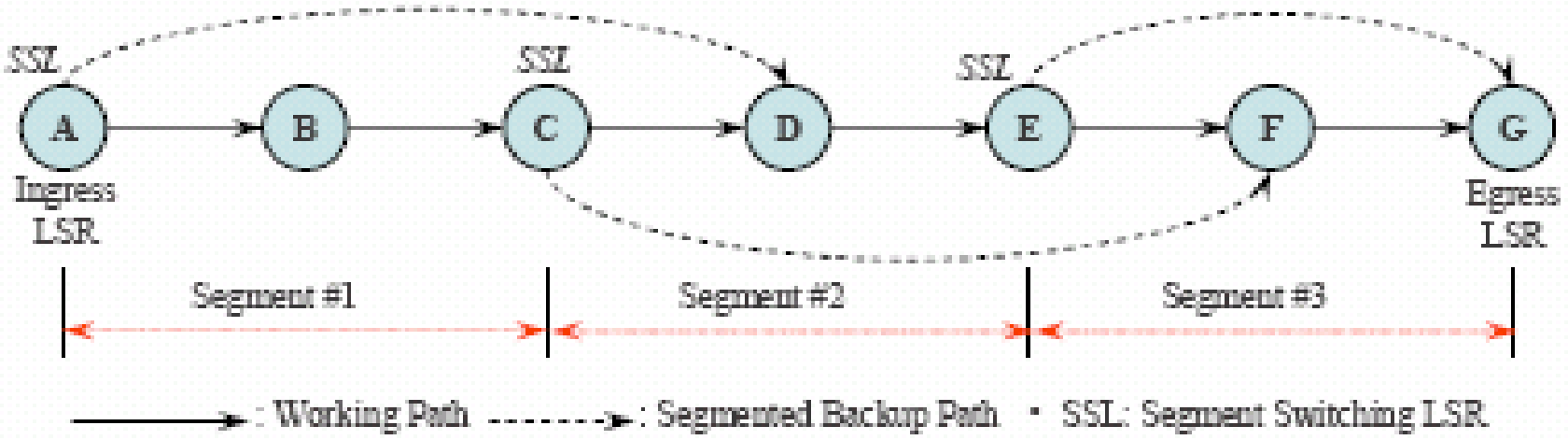
# Types of Protection

- There are different types of protection schemes:
  - ✓ Path protection
    - ➢ End-to-end protection
      - Dynamic creation of the backup path
      - Pre-established diverse LSP(s) for load balancing and TE in normal operation, and switchover in failure
    - ➢ Segment path protection
      - Designated segment heads
  - ✓ Local protection
    - ➢ Link protection
    - ➢ Node protection

# Path Protection (E2E)

- Basically, it means the establishment of one (or more) additional LSP(s) in parallel with an existing LSP.
  - ✓ **1+1**: fully protected, but less scalable and underutilized
  - ✓ **1:1**: the backup tunnel could be used for low priority traffic before switchover
  - ✓ **1:N**: what if multiple failures happen?
  - ✓ **M:N**: Multiple recovery paths are used to protect multiple working paths
- Additional LSPs can be used for backup (called backup, secondary, or standby LSPs) which means they don't carry traffic until a failure happens or they can carry less traffic or lower-priority traffic.
- What are some of the features that a backup LSP needs to consider?
  - ✓ Build along paths that are **as diverse as possible** from the primary LSP may not be easy for some networks. Also**, layer 1 and layer 3 may have different topologies.**
  - ✓ Both the primary and backup LSPs are configured at the headend and are signaled ahead of time.
  - ✓ Usually have the same constraints (i.e., bandwidth)
  - ✓ A primary LSP may require multiple backup LSPs
- **Less scalable** if every path needs to be protected.
- **Long(er) notification delay**: May take some time to notify the headend.

# Path Protection (Segment)




When a fault is detected, the fault notification needs to propagate to the Segment Switching LSR (SSL) of that domain instead of the ingress LSR

**Advantage:** Segment protection is faster than path protection because recovery can be initiated closer to the fault

**Disadvantage**: ?

# Local Protection

- The protection tunnel is built to cover **only a segment** of the primary LSP.
- Again, it requires the **pre-establishment** of the backup LSP. Reason?
- Backup LSP is **routed around a failed link or node**.
- Relationship between the primary and backup LSPs?
  - ✓ The primary LSPs that would have gone through that failed link or node are instead **encapsulated** in the backup LSP (using **label stacking**).
- What is label stacking? What feature does label stacking support?
- Better than 1+1 path protection in terms of resource utilization and scalability, i.e., a single backup LSP can protect N primary LSPs.
- Some terms for local protection:
  - ✓ PLR: Point of Local Repair
  - ✓ MP: Merge Point
  - ✓ NHop: Next-hop router
  - ✓ NNHop: Next-next hop router
  - ✓ Example

# Factors to Consider for Local Protection

- Need for **label stacking**
  - ✓ Example
  - ✓ **Global label space** instead of per-interface. Why? What if not global?
- Some traffic flows are important; some not so important.
  - ✓ Important flows: time-sensitive data requiring real-time response. Those important flows can be translated to important LSPs.
  - ✓ Important LSPs could be protected while ignoring less-important LSPs.
- **Link Protection vs. Node Protection**
  - ✓ Link protection: assume that although a protected link has gone down, the router at the other end is still up. Use **NHop** backup tunnels.
  - ✓ Node protection: protect against the failure of a downstream node (including the downstream link as well). Use **NNHop** backup tunnel.
  - ✓ Both need Label stacking.
    - ➢ Link protection: PLR knows what label the MP expects
    - ➢ Node protection: the label that MP wants is never signalled through RSVP to the PLR. Need other mechanism.

# Link Protection

- Link protection can be divided into four steps:
  - ✓ Pre-failure configuration
  - ✓ Failure detection
  - ✓ Connectivity restoration
  - ✓ Post-failure signalling

# Pre-failure Configuration

- Link protection is unidirectional. The backup tunnel does not have to carry any traffic until failure is detected on the protected link.
- Two places need to be configured:
  - ✓ At the **ingress/headend** of the tunnel interface
    - ➢ TE tunnels don't request protection by default. Why?
    - ➢ Need **explicit configuration for protection** (e.g. fast-reroute). The command will set **SESSION_ATTRIBUTE** flag 0x01 ("local protection desired) in the **PATH message** for that tunnel.
  - ✓ At the **PLR** (point of local repair)
    - ➢ Creating a backup tunnel to the NHop
      - Explicit routed path: either manually configured or CSPF calculated
      - Use the ***exclude*** option to avoid the protected link for CSPF
    - ➢ Configuring the protected link to use the backup tunnel upon failure
      - Just configuring the backup tunnel and calling the explicit path "backup" does not make traffic go over the tunnel when needed.
      - Need to tie them together, i.e., tell the interface to use that tunnel for protection:

        e.g., *mpls traffic-eng backup-path Tunnel1*:

        protecting the interface with Tunnel1
  - ✓ **MP** also needs to use global label space.

# Session_Attribute Class

- Format:

| 0 | 1 | 2 | 3 |
|---|---|---|---|
| Setup pri. | Holding pri. | **Flags** | Name length |
| Session name (variable length) | | | |

Flags:

**0x1: local protection desired**

0x2: label recording desired

0x4: Shared Explicit style

# Failure Detection

- Failure detection is critical. Why?
- Detection of a failed link has been used:
  - ✓ Specific to a particular physical layer, such as SONET
    - ➢ Requirement for SONET networks?
      - < 10 ms
  - ✓ For point-to-point links, PPP keepalives
  - ✓ RSVP hello extensions
    - ➢ Slower than layer 2 alarm-based detection
    - ➢ Refresh interval could be as low as 10ms (100ms for Cisco)
    - ➢ Can take several hundred milliseconds
    - ➢ Sufficient for local protection and generally faster than IP (no guarantee)

# Connectivity Restoration

- As soon as a failure is detected, the **PLR** is responsible for switching traffic to the backup tunnel.
  - ✓ Check if a pre-signalled backup LSP is in place, including the **new label** provided by a new downstream neighbor.
  - ✓ New adjacency information is computed based on the backup tunnel's outgoing interface. The information actually is pre-computed and ready to be installed in the FIB to minimize packet loss.
- For local protection mechanisms, while the protection is active and the backup tunnel is forwarding traffic, the **primary LSP continues to stay up**.
  - ✓ This is different from path protection scheme.
  - ✓ What effect will it have if the primary LSP goes down?

# Post-failure Signalling

- RSVP-based MPLS TE revolves around RSVP signalling. FRR is no exception.
- Three elements are needed for RSVP signalling that happens **after** the FRR has been effective:
  - ✓ **Upstream** signalling with a different PathErr subcode, “Tunnel locally repaired”
  - ✓ **IGP** notification
  - ✓ **Downstream** signalling

# Upstream Signalling

- Upstream signalling from PLR
  - ✓ Sends a different PathErr code (25) (with subcode (3) of "**Tunnel locally repaired**") to the ingress of the LSPs.
    - When ingress receives such a message, it **doesn't stop its primary LSP** and it knows that this LSP might be following a suboptimal path until it is rerouted.
    - How can the new subcode be used?
  - ✓ If the ingress calculates and signals a new path for that tunnel, a reservation (RESV) message will be received, the old path will then be torn down.
    - Make before break
    - If the ingress can't find a new path, it ingress remains on the protected path.
  - ✓ The PLR also sends Path messages downstream so that the **MP doesn't time out the protected tunnel**. The Path/Resv messages are modified (SENDER_TEMPLATE object) so that PLR becomes the sender. The tail knows this is from a new sender, but same session.

# IGP Notification

- IGP notification
  - ✓ Generally, RSVP messages reach ingress or egress before IGP, but this is not guaranteed. What if IGP info declares a link down before the RSVP?
    - ➢ If no FRR, the ingress tears down the tunnel when it receives a link-down message.
    - ➢ If FRR configured, the ingress then only tears down a protected LSP based on RSVP message and ignores IGP's error message.

# Downstream Signalling

- If no local protection, the MP sends a PathTear message to its downstream node to tear down the path.
- With FRR, the *PathTear* message is suppressed for primary LSPs that have the "Local Protection Desired" flag on.
- As long as the MP receives **Path messages** belonging to the **original RSVP session on any interface,** it does not time out.
- How does the tail know that the protected tunnel has failed?
  - ✓ It does not receive an RSVP refresh message (Path)
    - ➢ It has a grace period (usually 4 keepalive periods).
  - ✓ It receives an IGP update about the link failure.
    - ➢ With FRR, it does not take any action from the MPLS TE perspective.
  - ✓ It receives a PathTear message.
- To keep the tunnel alive, need to make sure that the tail continues to receive the RSVP refresh messages even if one of the links belonging to the primary LSP is down. How to support it?
  - ✓ Make sure that the MP continues to receive PATH messages for the primary LSP over the backup tunnel.

# Node Protection

- Similarities between LP and NP:
  - ✓ Enabling FRR at the primary tunnel headend
  - ✓ Tying the protected link to the backup tunnel
  - ✓ Failure detection
  - ✓ Connectivity restoration
  - ✓ Post-failure signalling
- Differences between LP and NP:
  - ✓ **NNHop** backup tunnel is configured at the PLR.
  - ✓ **Label recording** is required for the NNHop backup tunnel.
  - ✓ NNHop tunnel handles link and node failures.

# NNHop

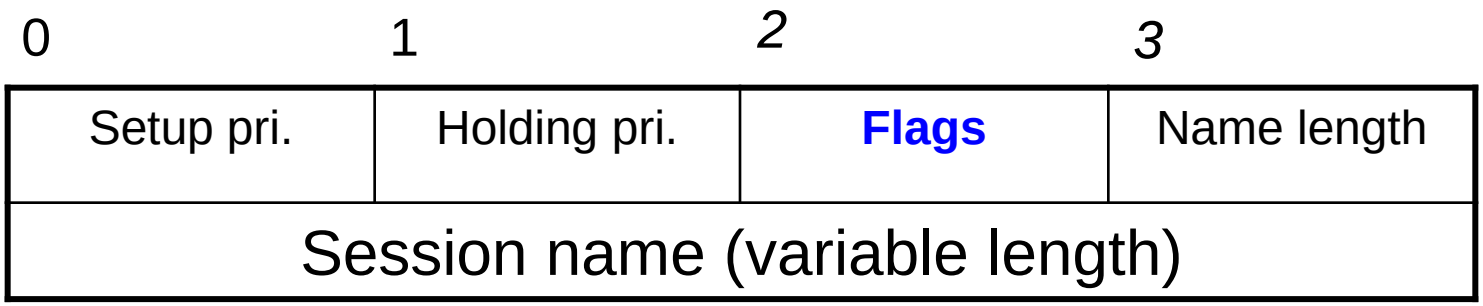
- NNHop backup tunnel is configured at the PLR
  - ✓ Similar to configuration of a backup tunnel to the NHop, except that the NNHop tunnel terminates on the NNHop instead of NHop.
- NNHop tunnel handles link and node failures
  - ✓ For PLR, there is no way to tell the difference between a link failure or a node failure.
  - ✓ With node protection, the traffic will be rerouted around both the protected link and the node.
  - ✓ Node protection is both link and node protection in one.

# Label Recording for the NNHop Backup Tunnel

- Label recording is required for node protection.
- What is label recording?
  - ✓ Records the incoming label (in **ROUTE_RECORD** object) used at each hop for an LSP
- What does label recording help?
  - ✓ For NHop, the PLR is the upstream node, so it knows what label the downstream neighbor is expecting.
  - ✓ For NNHop, it does not know what label the node expects for a LSP
  - ✓ So that a PLR doing node protection knows what label to use on a protected LSP when a switchover happens.
- How to enable label recording?
  - ✓ Label Recording flag is turned on whenever the headend configures fast reroute (including link protection). The headend doesn't know if the network supports link or node protection, so it turned on label recording. If no node protection is supported, it simply doesn't use it.

# Label Recording in Session_Attribute Class

- Format:




Flags:

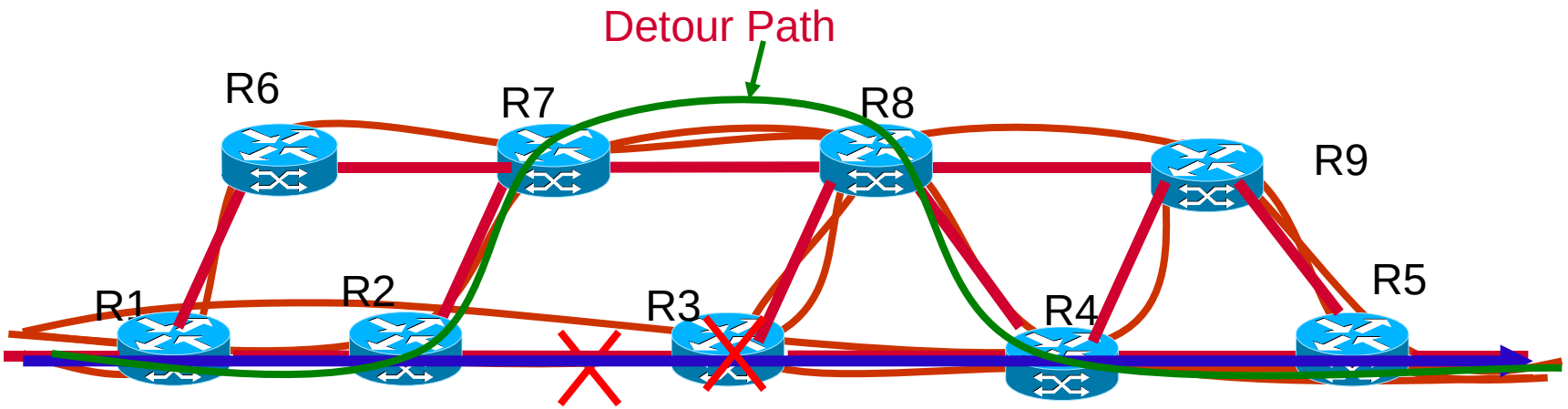
0x1: local protection desired

**0x2: label recording desired**

0x4: Shared Explicit style

# Potential Complexity for Multiple Protections

Either link or node protection has similar complexity




Protected LSP: R1-> R2 ->R3->R4->R5

R1's Backup: R1-> R6 ->R7->R8->R3

R2's Backup: R2-> R7 ->R8->R4

R3's Backup: R3-> R8 ->R9->R5

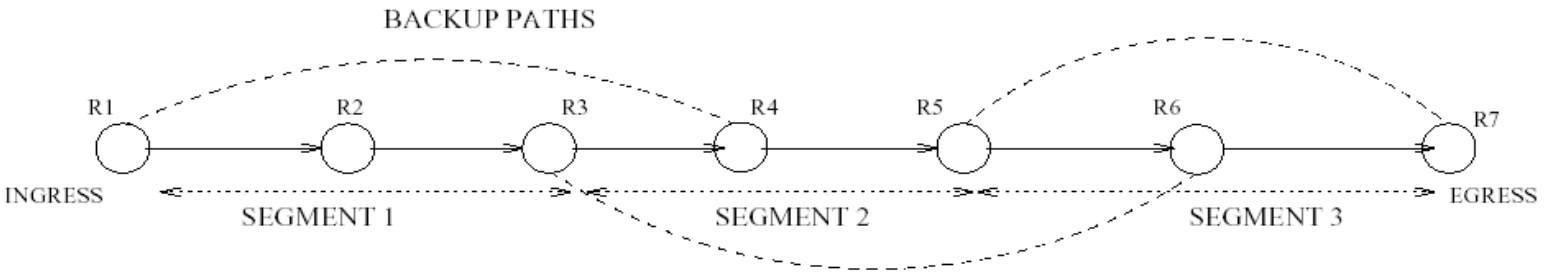
R4's Backup: R4-> R9 ->R5

# Advanced Protection Issues

- Some advanced topics:
  - ✓ Multiple backup tunnels to the same MP
  - ✓ Backup bandwidth reservation or bandwidth protection
  - ✓ Backup tunnel selection
  - ✓ Promotion
  - ✓ Combination of local protection and path protection
  - ✓ Segment protection
  - ✓ p-cycle
  - ✓ Network coding
  - ✓ Protection for multicast

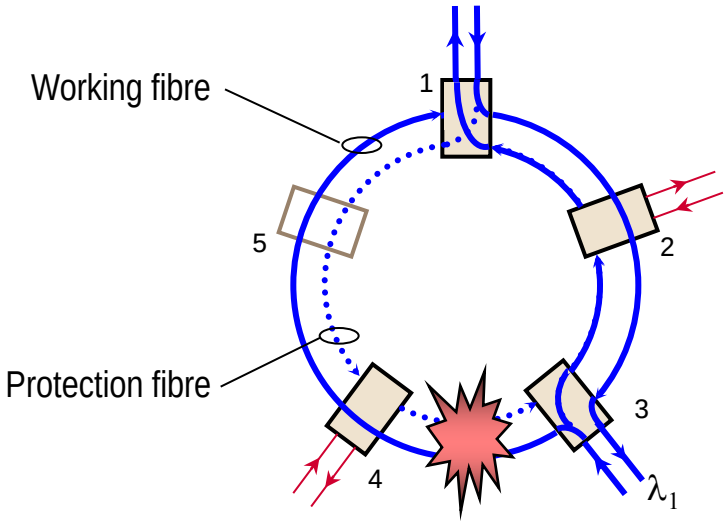
# Segment Protection

- A primary LSP is divided into segments.
- Create a backup LSP for each segment
- Less backup LSPs – bandwidth saving
- Meet QoS constrains – bounded switchover time



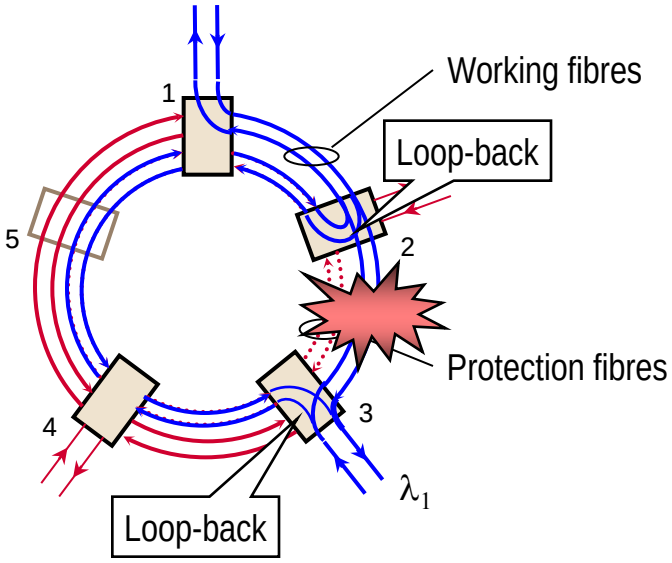
# Traditional protection and restoration

- In ring network
  - ✓ Self-healing ring
    - ➢ UPSR (Unidirectional path-switched ring)



# Traditional protection and restoration

- In ring network
  - ✓ Self-healing ring
    - ➢ UPSR (Unidirectional path-switched ring)
    - ➢ BLSR (Bidirectional line-switched ring)



# Traditional protection and restoration

- In ring network
  - ✓ Self-healing ring
    - ➢ UPSR (Unidirectional path-switched ring)
    - ➢ BLSR (Bidirectional line-switched ring)
- In mesh network
  - ✓ Link protection
  - ✓ Node protection
  - ✓ Path protection
  - ✓ 1+1,1:1,1:N,M:N

# Background and motivation

"Ring"

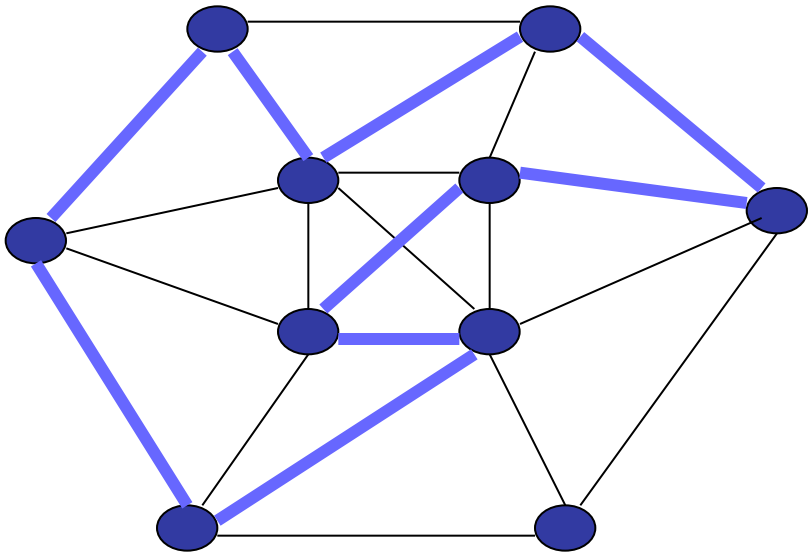
- 50msec restoration times
- Need at least 100% redundancy
- Planning of multi-ring network is complex
- Hard to accommodate multiple service classes
- Inefficient and inflexible
- Ring-constraint routing

"Mesh"

- Up to 1.5sec restoration time
- Need only 50-70% redundancy
- Simple, exact capacity planning solutions
- Easy and efficient design for multiple service classes
- Efficient and flexible
- Shortest-path routing

# P-Cycle: ring-speed and mesh-efficiency

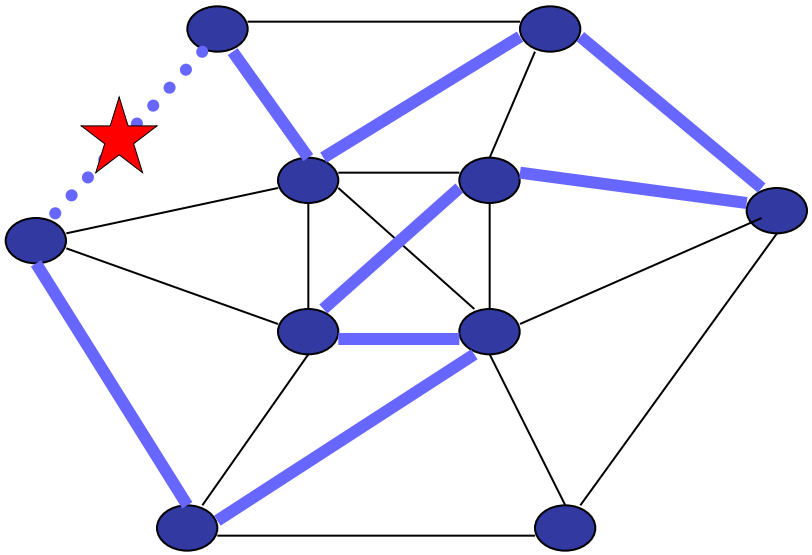
A single p-cycle in a network:



- **p**re-configured **p**rotection **cycle**s→ p-cycle

# P-Cycle: ring-speed and mesh-efficiency

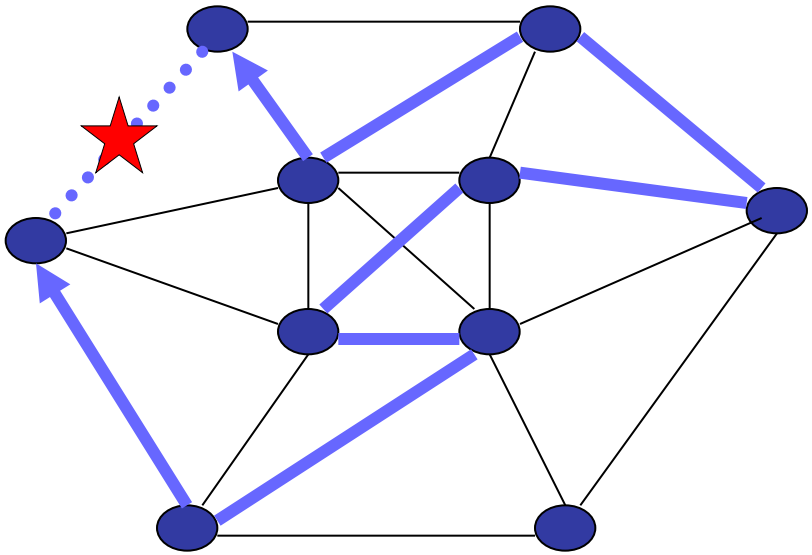
*A single p-cycle in a network:*



*On-cycle link failure*

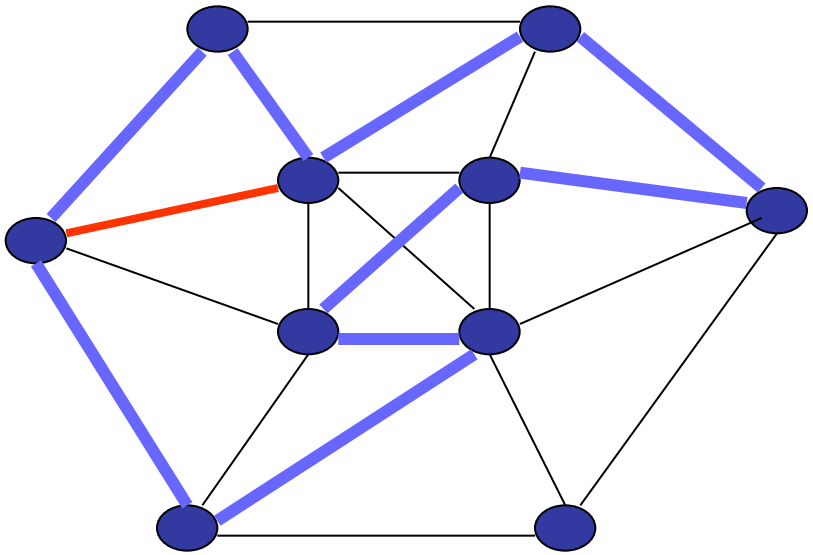
# P-Cycle: ring-speed and mesh-efficiency

*A single p-cycle in a network:*



*On-cycle link failure: 1 protection path (BLSR-like)*

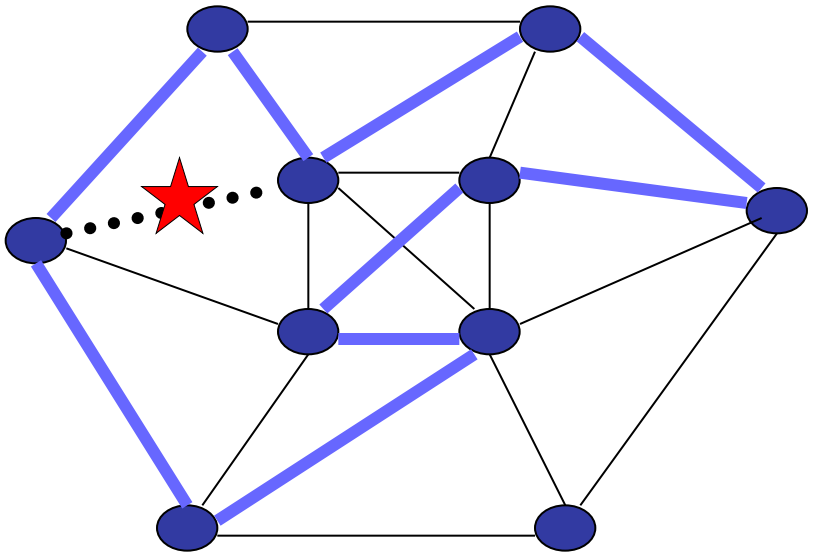
# P-Cycle: ring-speed and mesh-efficiency



***Straddling link:***

*An off cycle span having p-cycle nodes as endpoints*

# P-Cycle: ring-speed and mesh-efficiency



***Straddling link:***

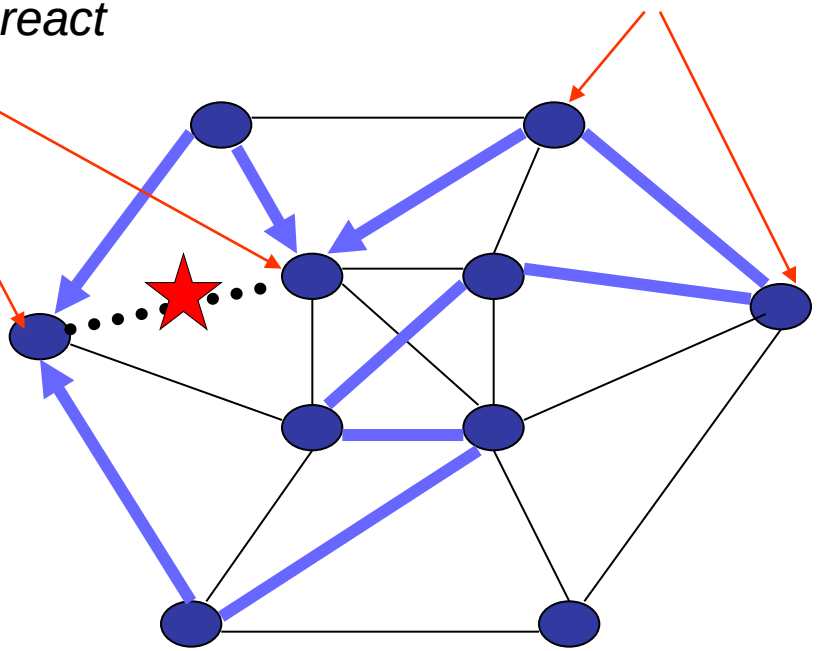
*An off cycle span having p-cycle nodes as endpoints*

# P-Cycle: ring-speed and mesh-efficiency

*No protection path signaling*

*Only nodes neighboring the failure react*

*(pre-configuration)*



***Straddling link:***

*An off cycle span having p-cycle nodes as endpoints*

*On-cycle link failure: 1 protection path*

*Straddle link failure: 2 protection path*