CHAPTER 5

MPLS Traffic Engineering
Recovery Mechanisms

Multi-Protocol Label Switching (MPLS) traffic engineering (TE) has encountered an ineluctable success during the past years, which led to the development of a rich set of MPLS TE recovery techniques.

This chapter starts with a refresher of the MPLS TE technology, followed by the motivation for deploying such a technology in a data network. The recovery techniques are then examined with the objective to provide a detailed description of their mode of operation and their respective pros and cons, the type of the network design they preferably apply to, and aspects of design that operators find important for deployment in their network.

Furthermore, various properties of each recovery technique are analyzed. These properties are of the utmost importance when choosing a particular recovery technique in a network: the recovery time, the impact on scalability, the ability to provide some quality-of-service (QoS) guarantees along the alternate path, and the technique efficiency with respect to the amount of bandwidth dedicated to recovery path. These are just a subset of the aspects covered for each recovery technique.

This chapter covers the default restoration mode of operation of MPLS TE, as well as the global and local protection recovery schemes. A rich set of examples are provided throughout this chapter that illustrate the mode of operation and how those various recovery techniques can be deployed in a network. An entire section is devoted to a complete set of case studies that show how an operator can use those MPLS recovery techniques to satisfy a set of recovery objectives while respecting network constraints. It is worth highlighting that most of these case studies are inspired by existing or foreseen deployment scenarios. After a summary section, this
first part of this chapter concludes with the standardization aspects of the MPLS TE recovery techniques. Then, the second part of this chapter is devoted to some advanced topics of MPLS recovery. The aim of those two sections is to cover in detail the signaling aspects of MPLS local protection (Section 5.14) and the interesting topic of the backup path computation (Section 5.15) and may be skipped by the reader without altering the good understanding of the MPLS recovery techniques. Finally this chapter concludes with a section that describes various related topics of research.

5.1 **MPLS Traffic Engineering Refresher**

In this section, we first provide a brief refresher on the notion of traffic engineering. Then the terminology specific to MPLS TE is shown through an example, and after having reviewed the main components of MPLS TE, we detail the motivation for deploying MPLS TE in a network.

5.1.1 **Traffic Engineering in Data Networks**

One of the major challenges of network design has always been traffic engineering; that is, how to route the traffic so network resources are efficiently used. The term “efficiently” requires some explanations though. An obvious objective of network design is to avoid congestion. If the network is fully congested, traffic engineering cannot really help and the network has to be upgraded (i.e., bandwidth and/or switching/routing capacities must be added). On the other hand, if some regions of the network are congested while others have spare capacity, then trying to alleviate the congestion spots by rerouting some flows along an alternate path (where capacity is available) certainly helps.

In other words, TE defines how flows should be routed to efficiently use network resources. Even in the absence of congestion, a more optimal traffic load balance may help increase the QoS. For instance, suppose that some links are used at 60% capacity (on the average), which strictly speaking cannot be considered a congested link whereas other links are loaded at 10%. It is worth noting that delay-sensitive traffic traversing a link loaded at 60% may experience some undesirable delay and jitter, especially without queuing mechanisms. Thus, achieving a better traffic load balance with the objective of minimizing the average link utilization might be another motivation for TE.

Traffic engineering is not per se specific to MPLS. Various network types have been using TE methods like public voice networks, ATM, Frame Relay, and Internet Protocol (IP).

**The Classic Fish Problem**

Let us consider the following classical fish problem to highlight how situations in which congestion appears in some parts of the network while other regions of the network have spare capacity may occur in an IP network (Figure 5.1).
5.1 MPLS Traffic Engineering Refresher

Routing Decision Based on the IP Destination Address

All Links Have a Metric = 1

Path North: R3-R4-R5-R8
Path South: R3-R4-R5-R8

Figure 5.1 The classic "fish problem."

Figure 5.1 depicts two IP routers R1 and R2 sending traffic to the router R8 (and beyond). Both R1 and R2 will compute the shortest path to reach R8 using a routing protocol like Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS). Because all the links have an equal metric of 1, the flows from R1 and R2 to R8 will both follow the same path ("north"). If the sum of their traffic exceeds the bandwidth capacity of the path "north" (R3-R4-R5-R6), this will result in some congestion, although some capacity is still available along the path "south." Changing the link metric in this case will not help because IP routing protocols base their routing decision on the IP destination address. So whether a packet whose destination is R8 is received from R1 or R2, it will be routed by R3 along the same path. Another option in this very simple case is to set up the link metric so the north and south paths have an equal cost to use load balancing, but real networks are more complicated, and if other nodes are connected to routers R4, R5, R6, and R7, load balancing becomes much more challenging to achieve.

That said, TE with IP routing is of course possible and has already been discussed in Chapter 4.

One solution to obtain better resource utilization is to use tunneling techniques between source(s) and destination(s) so intermediate nodes do not participate in the routing decision. ATM was extensively used to reach that goal; ATM permanent virtual circuits (PVCs)/switched virtual circuits (SVCs) are established between switches with characteristics based on the traffic requirements of each circuit (e.g., bandwidth and QoS). ATM PVCs/SVCs are routed based on the network resources and link costing using off-line or on-line path computation methods (e.g., Private Network–Network Interface [PNNI]). Then, once a packet (encapsulated in ATM cells) is routed onto an ATM PVC, it strictly follows the ATM PVC path.
Although relatively efficient to improve network bandwidth usage, there are several significant drawbacks with this approach:

- An additional layer (ATM) has to be managed and maintained in the network (ATM), which implies additional cost in terms of equipment and network operation.
- The number of routing adjacencies maintained by each router is potentially very high because every router has a number of routing neighbors equal to the number of routers in the mesh, which introduces some routing protocol scalability limitations. Indeed, a mesh of n routers requires for each of them to maintain n adjacencies and the route computation (shortest path first [SPF]) is also increased significantly.

This is where MPLS TE comes into play. MPLS TE is also a "tunneling" mechanism using TE Label Switch Paths (TE LSPs; the terminology TE LSP is detailed hereafter), which are established between pair of routers. Each TE LSP has its own set of constraints—like bandwidth, affinities, and rerouting constraints, to mention a few—and the network topology and resources are taken into account along with the set of constraints to compute the TE LSP path that satisfies the set of requirements. Different path computation methods can be used to achieve that objective: distributed (each router is responsible for the computation of its TE LSP path) or centralized (an off-line tool performs the path computation of all the TE LSPs in the network). Then once a TE LSP is established, IP packets are routed onto the TE LSP and strictly follow the computed path; intermediate routers do not make any routing decision.

For instance, in Figure 5.2, suppose that the sum of required bandwidths between R1 and R8 and R2 and R8 exceeds the available bandwidth on the north.

![Figure 5.2 Optimizing network resources with MPLS traffic engineering.](image)
path (R3-R4-R5). By using MPLS TE, once the TE LSP between R1 and R8 is established, R2 figures out that the bandwidth available on the north path is not sufficient to accommodate its traffic demand and selects the south path (R2-R3-R6-R7-R5-R8) to establish its TE LSP. This allows better network resource utilization and avoids traffic congestion.

Note that compared to the previous case with an ATM overlay network, just one layer is required (IP/MPLS). Moreover, routers are not required to maintain routing adjacencies over TE LSP. It is important to note that MPLS TE is a control plane reservation protocol, so this is fundamentally a Call Admission Control (CAC) mechanism. In other words, when a TE LSP is set up, no particular resources in the data plane are reserved. The purpose of MPLS TE is to ensure that a TE LSP is not routed along a path where other TE LSPs have already reserved the bandwidth. For instance, on an OC3 link, if three TE LSPs have already been reserved a total bandwidth of 120 Mbps, the remaining available bandwidth (not already reserved in the control plane) is 35 Mbps and a TE LSP requiring more than 35 Mbps will have to be routed along another path. This is in contrast to IP in which IP packets are routed along the shortest path without considering the traffic flow and available resources along this path.

5.1.2 Terminology

Because there are several terms specific to MPLS TE recovery techniques, which are used throughout this chapter, we illustrate each of them via an example (Figure 5.3).

As depicted in Figure 5.3, three TE LSPs, called T1, T2, and T3, are signaled. For instance, the TE LSP T1 starts on R1 and terminates on R8. We say that R1 is the head-end label switched router (LSR) of T1 and R8 is its head-end LSR. Any other LSR traversed by T1 is a midpoint LSR (e.g., R3, R4, and R5 are all midpoint LSRs). Note that an LSR can play the role of a head-end LSR for an LSP while being a midpoint or a tail-end LSR for other TE LSPs.

Notion of Disjoint Paths

Two TE LSPs are said to be link disjoint if they do not have any link in common (e.g., T1 and T2 in Figure 5.3 are link disjoint). The terminology link diverse is also used. On the other hand, two TE LSPs are said to be node disjoint if they do not share any TE LSR (e.g., T1 and T3 are node disjoint), except potentially their head-end and tail-end LSRs. The term node diverse is also used.

The recovery-specific terminology aspects are covered in their respective sections. For instance, several terms are specific to the local protection techniques, and these are covered in the section devoted to local protection techniques.

Shared Risk Link Group

The notion of shared risk link group (SRLG) is crucial when studying network resiliency and specifically refers to the notion of simultaneous failures of multiple
network elements that can be caused by the failure of a single element. Let us consider the network scenario in Figure 5.4.

Figure 5.4A shows a set of six optical cross-connect OXC1 through OXC6, which are interconnected by a set of fibers, which constitutes an optical layer used to interconnect the LSRs R1 through R5. More precisely, the various links are routed in the optical layer as follows:

- Link R1-R2 follows the optical path OXC1-OXC2.
- Link R1-R4 follows the optical path OXC1-OXC4-OXC5.
- Link R1-R5 follows the optical path OXC1-OXC6.
- Link R2-R3 follows the optical path OXC2-OXC3.
- Link R3-R4 follows the optical path OXC3-OXC5.
- Link R5-R4 follows the optical path OXC6-OXC4-OXC5.

In this scenario, the two optical paths followed by the links R1-R4 and R4-R5 share a common resource: the optical fiber interconnecting the OXC4 and OXC5. We say that the two links share a unique SRLG because the failure of a single resource (the optical fiber OXC4-OXC5) would provoke the simultaneous failure of the two links.

By default, the IP/MPLS layer does not have any visibility of the optical layout, which may lead to an incorrect path selection for TE a LSP. To remedy to this problem, an Internet Gateway Protocol (IGP) extension has been defined. As described in Section 5.1, the TE-related information is flooded within an OSPF area using an opaque LSA type 10 (for IS-IS the TE-related information is flooded in a specific type-length value [TLV]). This opaque LSA carries one top-level TLV, which can be one of the two following types: router address (type 1) or link (type 2). The link sub-TLV is made of several sub-TLVs. One of them is the SRLG sub-TLV (type 16); it has a variable length with 4 bytes per SRLG value.
5.1 MPLS Traffic Engineering Refresher

Figure 5.4 Shared risk link group.

Important notes:
- A link may belong to multiple SRLGs.
- The IGP extensions allow carrying the SRLG values. On the other hand, having the knowledge of the underlying optical/SONET-SDH topology is not always possible. Indeed, an operator may rely on another carrier to provide optical lambda, and in that case, the SP does not always have the knowledge of the actual physical path and the potential SRLG. Moreover, an optical path may be dynamic and so its path may change over the time. This requires updating the SRLG value each time a change occurs if the SRLG changes also.

Notion of SRLG disjoint: A TE LSP is said to be SRLG disjoint from a link L or a node R if and only if its path does not include any link or node that is part of the SRLG of that L or R. For instance, back Figure 5.4, a TE LSP T1 following the path R1-R2-R3-R4 is SRLG disjoint from the link R1-R4. Two TE LSPs are said to be SRLG disjoint if the respective set of links they traverse do not have any SRLG in common.

5.1.3 MPLS Traffic Engineering Components

The aim of this section is to review the main components of MPLS TE:

1. Configuration of TE LSP on head-end LSR: The first step consists of configuring the TE LSPs' attributes on the head-end LSR. Various attributes can be configured like the destination (address of the tail-end LSR), the required bandwidth, the required protection/restoration, the affinities, and others.
2. **Topology and resource information distribution:** To compute a path obeying the set of specified constraint(s), the head-end LSR needs to gather topology and resource information. Note that this applies only to situations in which the TE LSPs' path is dynamically computed by each LSR (also referred to as **distributed or on-line** path computation) by contrast with centralized or off-line path computation in which the LSPs' path is computed by an off-line tool. In such a case, the topology and resource information is distributed by a link state routing protocol (OSPF or IS-IS) with TE extensions that reflect links characteristics and reservation states. TE TLVs have been defined and are carried within an LSP for IS-IS and TE opaque LSA type 10 for OSPF to flood the reservation states and other parameters.

3. **TE LSP computation:** As already stated, the computation of a TE LSP path can either be performed by an off-line tool or on-line. In the former case, an external tool simultaneously computes all the TE LSPs paths according to the network resources. In the latter case, every router (LSR) uses its resource and topology database (IS-IS or OSPF), takes into account the set of requirements of the TE LSP, and computes the shortest path satisfying the set of constraints usually using a constraint shortest path first (CSPF) algorithm. Various types of CSPFs can be used.

4. **TE LSP setup:** Once the path of a TE LSP has been computed, the head-end LSR signals the TE LSP by means of the Resource Reservation Protocol (RSVP) signaling protocol with the corresponding set of extensions defined in [RSVP-TE]. For instance, in Figure 5.3, R1 computes a path for the LSP T1: R1-R3-R4-R5-R8 based on T1's attributes and the network and resources topology information disseminated by the routing protocol. Once T1's path is computed, T1 is signaled by RSVP-TE. TE LSPs are then signaled, maintained (refreshed) and potentially torn down using various RSVP messages: Path, Resv, Path Error, Path Tear, Reservation Error, Resv Confirmation, and Resv Tear. Also, various new objects have been defined in [RSVP-TE] for the purpose of MPLS TE, for example, to allocate labels to TE LSPs that will then be used in the MPLS data plane. Note that labels are assigned in the upstream direction using RSVP messages (Resv message) and intermediate LSRs are programmed accordingly. For instance, when the TE LSP T1 is signaled, labels are assigned by LSRs in the upstream direction: R8 provides a label to R5, R5 provides a label to R4, and so on.

**Note:** It is worth mentioning that RSVP has often been criticized for its scalability, in particular the number of states required in the network. As a matter of fact, currently deployed networks can handle thousands of RSVP TE reservations (TE LSPs) on a single router without any problem. Moreover, various protocol enhancements have been defined (see [REFRESH-REDUCTION]) to further increase the scalability, if needed. Finally, MPLS TE can be deployed with multiple levels of hierarchies, if required, in very large networks.
5. Packet forwarding: Once a TE LSP is set up, the head-end LSR can update its routing table and start using TE LSP to forward IP packets. A label of 32 bits is pushed onto the IP packet, which is then label switched across the network (intermediate routers do not make any routing decision).

5.1.4 Notion of Preemption in MPLS Traffic Engineering

There is one interesting property called “preemption” defined in MPLS TE, which deserves to be slightly elaborated in the chapter because upon network element failure, preemption mechanisms may be triggered. [RSVP-TE] defines the notion of preemption or priority for a TE LSP. This parameter is signaled in the SESSION-ATTRIBUTE object of the RSVP TE Path message (more precisely, the RFC defines two priorities known as the “setup” and “holding” priorities, which define the priority of a TE LSP with respect to taking and holding resources, respectively).

When a new TE LSP is signaled, an LSR considers the admission of this newly signaled TE LSP by comparing the requested bandwidth with the bandwidth available at the priority specified in the setup priority. If the requested bandwidth is available but this requires preempting other TE LSPs having a lower priority, then the newly signaled TE LSP is admitted and one or more TE LSPs with a lower priority are preempted. Note that the selection of the set of lower priority TE LSPs to be preempted is a local decision and is generally implementation specific. More details of preemption policies can be found in [PREEMPTION-POL].

The preemption process implies the set of following actions for each preempted TE LSP:

- The corresponding local RSVP states are cleared and the traffic is no longer forwarded.
- Messages are sent both upstream (RSVP Path Error message) and downstream (RSVP Resv Error) so all the states corresponding to the preempted TE LSP are cleared along its path. Then the head-LSR LSR of a preempted TE LSP initiates a TE reroute procedure as detailed earlier to reroute the TE LSP along another path.

This means that hard preemption is by nature a disruptive mode. So the concept of soft preemption has been introduced in [SOFT-PREEMPTION] and proposes a different mode of preemption. If a TE LSP must be preempted to accommodate a higher priority TE LSP requests, the preemtting LSR performs the following actions:

- The preemtting LSP signals to the respective head-end LSR the need to reroute the TE LSP in a nondisruptive fashion (so-called “make before break” procedure).
- The local states of the soft preempted TE LSP are not cleared and no RSVP Path Error/RSVP Error messages are sent.
Hence, the preempting node keeps forwarding the traffic of a soft preempted TE LSP for a certain period. This gives a chance for the soft preempted TE LSPs head-end LSR to reroute their TE LSPs along an alternate path without disrupting traffic flow.

It is worth pointing out that this implies to temporary provoke reservation overbooking on some links because until the soft preempted TE LSPs are rerouted by their respective head-end LSR, the sum of admitted bandwidth is higher than the maximum allowed. Note that some algorithms can be carefully designed to preempt hard preemtible\textsuperscript{62} TE LSPs first. Moreover, appropriate MPLS Diffserv mechanisms can be used to make sure that high-priority traffic is served adequately.

5.1.5 Motivations for Deploying MPLS Traffic Engineering

Once the concept of TE and the main components of MPLS TE have been reviewed, it is time to highlight the various motivations for deploying MPLS TE in a network.

1. Bandwidth optimization: As pointed out in Section 5.1, MPLS TE can be deployed to achieve better network resource utilization, usually referred to as bandwidth optimization.

2. Strict QoS guarantees: Another motivation for deploying MPLS TE in a network is to enforce strict QoS guarantees for various service types including sensitive traffic flows like voice, video, and circuit emulation. As already mentioned, MPLS TE acts on the control plane and as such takes care of the routing decision. For instance, consider a network with a single class of service (CoS), MPLS TE allows an operator to reduce the average and maximum link utilization. Hence, a direct implication is that the probability of traffic queuing delay is decreased, which correlates with a better QoS. Another example is the case of a network with multiple classes of service. Making sure that appropriate treatment of sensitive flows is performed in the data plane requires various mechanisms like marking, queuing, and congestion avoidance in the data plane. In such networks, MPLS TE will allow control over the proportion of high-priority traffic versus medium- and low-priority traffic on a per-link basis, which will increase the QoS.

Although this has already been highlighted, to provide QoS guarantees between two nodes, specific actions must be taken in the IP/MPLS data plane, implementing the Differentiated Services (Diffserv) model. Indeed, MPLS TE is responsible for finding a path obeying a set of constraints, but once the packets are sent onto that TE LSP, each node along the path has to serve the packet appropriately according to the required CoS.

3. Fast recovery: Several mechanisms for MPLS TE are described throughout this chapter, allowing for fast recovery along with other requirements like QoS protection during failure. Those mechanisms have been generating a

\textsuperscript{62}The hard/soft preemtible property of a TE LSP is explicitly signaled in RSVP Path message.
5.2 Analysis of the Recovery Cycle

Growing interest for MPLS TE, and the sole interest for fast convergence, even if bandwidth optimization or strict QoS guarantees are not required, may justify the deployment of MPLS TE. Several large networks have deployed MPLS TE to benefit from the set of fast recovery mechanisms.

The aim of the previous short paragraph was to introduce the motivation of deploying MPLS TE in a network: bandwidth optimization, strict QoS guarantees, and fast recovery. They are of course nonexclusive. For example, consider an IP/MPLS network where the resource utilization is not optimal and fast recovery is desired. Then MPLS TE with, for instance, any fast recovery technique described in this chapter can be deployed. Another example is an IP/MPLS network where strict QoS guarantees are required for the voice traffic, for instance, as well as fast recovery for the virtual private networking (VPN) traffic and the voice traffic. Finally, as already pointed out, MPLS TE can be deployed for the sole motivation of benefiting from fast recovery. Consider an overprovisioned network in which neither bandwidth optimization nor strict QoS guarantees are necessary (QoS guarantee is achieved by overprovisioning), but fast recovery is a must. Then MPLS TE is a good candidate for its fast recovery property.

5.2 Analysis of the Recovery Cycle

Before studying the various recovery techniques used in IP/MPLS networks, it is worth spending some time on the recovery cycle analysis introduced in Chapter 1 and depicted in Figure 5.5.

5.2.1 Fault Detection Time

As with any other recovery techniques at any layer, the fault detection time is a key component of the total recovery time and highly varies depending on the fault detection mechanism in use and the underlying layer 1 and layer 2. For instance, the
fault detection time can vary from a few tens of milliseconds when two LSRs are interconnected via a SONET/SDH VC or an optical lightpath to a few hundreds of milliseconds or seconds when hello mechanisms are required. (Section 4.3 in Chapter 4 has been entirely devoted to the important aspects of failure profile and fault detections aspects.)

5.2.2 Hold-Off Timer

A hold-off timer can be very useful if the underlying layer has a recovery scheme. Those aspects of multilayer protection/restoration strategies are covered in detail in Chapter 6. In a nutshell, consider, for instance, a multilayer network where fast recovery mechanisms are implemented both at the optical layer and at the MPLS layer. Then, when the failure occurs, one should generally avoid any racing conditions where both recovery mechanisms simultaneously try to perform a reroute along an alternate path. In that case, a bottom-up timer-based approach can be adopted, in which the MPLS layer will wait for a hold-off timer to expire before trying to perform a reroute, to give the optical layer a chance to restore the failed resources. If the optical layer does not succeed in restoring the failed resource before the hold-off timer expires, the MPLS recovery mechanism will be triggered to restore the failed resource at the MPLS layer (the interlayer recovery mechanisms are more extensively discussed in Chapter 6).

5.2.3 Fault Notification Time

To perform traffic recovery, an LSR must first be informed of the failure. As we will see in this chapter, depending on the MPLS TE recovery mechanism used, the traffic recovery may be performed on the node immediately upstream to the failure or on the head-end LSR (the LSR originating the TE LSP); we call the fault indication signal (FIS) the signal of the failure to the node in charge of performing the traffic recovery. Hence, once the fault has been detected by an LSR R, the FIS is propagated until reaching an LSR that has the ability to reroute the TE LSP affected by the failure. The fault notification time (time for the FIS to be received by the node in charge of the traffic recovery) will vary depending on whether the recovery technique is local or global, as shown in Chapter 1, Section 1.5.4.

It is usually desirable to guarantee through appropriate scheduling on the various LSRs that the FIS receives the proper QoS, to minimize and guarantee the fault notification time. For instance, as mentioned in Chapter 4, the IGP flooding should be prioritized. In addition, IGP and RSVP messages should be queued appropriately and of course should never be dropped in the case of congestion. Refer to Chapter 4, Section 4.5, for further details on QoS mechanisms.

RSVP Reliable Messaging

As we saw in the Chapter 4, IGP updates are always sent in reliable mode; this is inherent to link state routing protocols. By contrast, RSVP messages are sent by default in nonreliable mode. So a loss of a Path Error message (which is used to
report an LSP failure to upstream nodes) may significantly increase the fault
notification time, especially if the IGP has not been tuned to provide fast notifica-
tion (see Chapter 4 for details). [REFRESH-REDUCTION] proposes a mechanism
to send RSVP messages in reliable mode.

Two additional RSVP objects are defined: the MESSAGE-ID and the MES-
SAGE-ID-ACK objects. Each RSVP message sent in reliable mode contains a
unique MESSAGE-ID object and is acknowledged by a MESSAGE-ID-ACK
object (note that it may be piggybacked to any other RSVP messages or to
an RSVP acknowledgment message). The retransmission of a nonacknowledged
message for which an explicit acknowledgment had been requested is based on
an exponential back-off procedure; when an LSR has to send a message in
reliable mode, it inserts a MESSAGE-ID object in the RSVP message and sets a
particular flag in the MESSAGE-ID header called the ACK-Desired flag.
Upon receiving the RSVP message, a neighboring LSR will send back an
RSVP message containing a MESSAGE-ID-ACK object. When the message is
acknowledged, the transmission procedure is terminated. If the sending LSR does
not receive any acknowledgment before a dynamic timer has elapsed, the message
is retransmitted. The dynamic timer Tk is exponentially increased until a maximum
value is reached. Tk is first set to an initial retransmission value (generally a
short value).

For example, let us suppose that a message is sent for the first time, and Tk = T1
is set to initial timer (the recommended value is 500 ms).

- If the message is not acknowledged after T1, then it is retransmitted.
  Otherwise the procedure is stopped.
- Then Tk is set to Tk-1* (1+delta) (the recommended value for delta is 1).
- The maximum value for k is set to a fixed value (k = 3 is recommended).

In summary, the sending LSR waits 500 ms and then retransmits the message,
then waits for the 500 ms*2, then 500 ms*4 with exponential increased waiting
times. If the maximum retransmission value is set to 3, the message is no longer
retransmitted after three trials.

5.2.4 Recovery Operation Time

Any recovery technique involves a set of actions to be completed. This includes
potential synchronization between network elements to coordinate.

5.2.5 Traffic Recovery Time

The traffic recovery time represents the time between the last recovery action and
the time the traffic is completely recovered. Each component described earlier is
analyzed for the various recovery techniques described in this chapter. We just saw
a brief description of each phase of the recovery cycle.

There are multiple types of MPLS TE recovery techniques (Table 5.1):
CHAPTER 5  MPLS Traffic Engineering Recovery Mechanisms

Table 5.1 Categories of MPLS Recovery Mechanisms

<table>
<thead>
<tr>
<th>Protection</th>
<th>Restoration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local recovery</td>
<td>Local protection</td>
</tr>
<tr>
<td>(Section 5.5)</td>
<td>(Section 5.4)</td>
</tr>
<tr>
<td>Global recovery</td>
<td>Global protection</td>
</tr>
<tr>
<td>Global default</td>
<td>Global default restoration</td>
</tr>
<tr>
<td>(Section 5.3)</td>
<td>(Section 5.3)</td>
</tr>
</tbody>
</table>

- **MPLS TE global default restoration** (Section 5.3). This is the default mode of recovery of MPLS TE, whereby the failure is notified to the head-end LSR by means of RSVP and the routing protocol, which in turn recomputes a new path and finally resignals the TE LSP along that new path.

- **MPLS TE global protection** (Section 5.4): The basic principle is that two TE LSPs are set up by the head-end LSR: a primary LSP and a backup. Once the head-end LSR is notified of a failure along the LSP path, it starts using the backup LSP.

- **MPLS TE local protection** (Fast Reroute; Section 5.5) is a local repair recovery scheme in which upon failure detection the LSPs affected by the failure are locally rerouted by the node immediately upstream to the failure.

5.3  **MPLS Traffic Engineering Global Default Restoration**

MPLS TE global default restoration is the default recovery technique. Once a failure is detected by some downstream node, the head-end LSR is notified by means of RSVP and the routing protocol (FIS). Upon receiving the notification, the head-end LSR recomputes the path and signals the LSP along an alternate path.

5.3.1  **Fault Signal Indication**

It is probably worth elaborating on the nature of the FIS in the context of MPLS TE because this aspect might be a source of confusion. In the context of an IP/MPLS TE network, the FIS is either an IGP update or an RSVP Path Error message. Actually, both will be generated independently. In the case of the IGP, a node detecting a loss of routing adjacency will generate an LSA/LSP update (see Chapter 4 for a detailed description of IP routing from a recovery perspective). When a link fails between two nodes, the nodes attached to the failed link will send an IGP update. In the case of a node failure, all the neighbors of the failed node will send an IGP update. As discussed in Chapter 4, the timing sequence will highly

---

63In the rest of this chapter, the term *IGP* will be used in place of *routing protocol*. 
depend on the failure detection time and IGP parameter tuning. Moreover, every node detecting a failure will also generate an RSVP Path Error message sent to each head-end LSR having a TE LSP traversing the failed resource. For instance, in Figure 5.3, if the link R3-R4 fails, as soon as the node R3 detects the link failure, it sends a notification (RSVP Path Error message) to R1, the head-end LSR of T1 because T1 traverses the failed link. In addition, an IGP update will be sent by both the nodes R3 and R4 to reflect the new topology. Again, the timing sequence depends on the IGP tuning (see Chapter 4). Usually, the RSVP Path Error message is received by the head-end LSRs within a few tens of milliseconds so generally before the IGP update, but regardless of which FIS is first received, the head-end LSR will get notified. As pointed out in Section 5.2, the FIS delivery is of the utmost importance with MPLS global default restoration, because it triggers the rerouting of the affected LSPs by the head-end LSR.

5.3.2 Mode of Operation

When a TE LSP is configured on a head-end LSR, its set of attributes is specified: destination (IP address of the tail-end LSR), bandwidth, priority, protection/restoration requirements, and other MPLS TE parameters. As far as the recovery is concerned, an important parameter is the TE LSP path. As mentioned in Section 5.1, the path of a TE LSP can be computed in either a distributed or a centralized fashion. In the former case, the configuration does not specify any particular path and the head-end LSR dynamically computes the LSP path, taking into account the constraints and available resources in the network. In the latter case, the path for the TE LSP is statically configured on the head-end LSR. Some MPLS TE implementations allow the configuration of both options with an order of preference.

In Table 5.2, a TE LSP is defined, with its corresponding parameters/constraints: destination address (10.0.1.100), bandwidth (10000), and priority (1). In addition, the notion of path-option allows specifying in order of preference the list of paths that the LSP should follow. In this example, the preferred path is a static path (path 1) for which the set of hops is statically configured on the head-end LSR. If path 1 is not available (path broken, not all the required constraints can be satisfied along this path), path 2 is the second preferred path. Note that this corresponds to the off-line path computation method already mentioned for MPLS TE where the LSP path is computed by some other tool (not by the head-end LSR itself). Then, if none of the static paths is available, the head-end LSR will try to find a path that complies with the requested constraints using the CSPF algorithm (this is the path option 3). Note that in addition, it might be possible to have different sets of constraints for different path options. For example, suppose that no path satisfying the bandwidth constraint (10000) can be found. Then one solution could be to try a lower value. Of course, that example of configuration shows a combination of static and dynamic paths for the sake of illustration. Just one dynamic path could have been configured or one or more static paths.
Table 5.2 An example of MPLS Traffic Engineering TE LSP Configuration

interface Tunnel1
  ip unnumbered Loopback0
  no ip directed-broadcast
  tunnel destination 10.0.1.100
  tunnel mode mpls traffic-eng
  tunnel mpls traffic-eng priority 1 1
  tunnel mpls traffic-eng bandwidth 10000
  tunnel mpls traffic-eng record-route
  tunnel mpls traffic-eng path-option 1 explicit name path1
  tunnel mpls traffic-eng path-option 2 explicit name path2
  tunnel mpls traffic-eng path-option 3 dynamic

Path1 = {192.170.14.2, 192.170.10.1, 192.170.4.5}
Path2 = {192.170.13.2, 192.170.17.1, 192.170.20.5}

Recovery Cycle with Global Default Restoration

The mode of operation of global default restoration is relatively simple: When the head-end LSR is informed of the link/node failure, if an alternate path is specified, the head-end LSR will check to see whether the configured path satisfies the constraints for the TE LSP. If so, the TE LSP is reestablished along that path. If no preconfigured path is specified on the head-end router and if configured as such, then it triggers a new path computation for the set of affected TE LSPs, calling the CSPF process (this exactly corresponds to the example in Table 5.2: If a notification is received reporting that path 1 is unavailable, the head-end LSR tries to determine whether it can use path 2, and if path 2 is not valid for some reason, it tries to compute a path itself).

Note 1: Various existing MPLS TE implementations allow relaxing constraint(s) upon failure, which might sometimes be necessary. A slightly more complicated example could be given in which for each path option, a set of different constraints is specified. For instance, consider a network with relatively high link utilization in terms of bandwidth reservation; a major node failure may cause the inability for several TE LSPs to find an alternative path. In this case, one of the options is to relax some constraints, like the bandwidth constraint so the TE LSP can be routed. There is one undesirable side effect though: Allowing a TE LSP to be rerouted as a 0 bandwidth TE LSP implies that traffic will flow over this tunnel without any CAC. Thus, no bandwidth can be guaranteed in this case. There are also various constraints a TE LSP can be configured to support. Bandwidth is just one of them. Another example is affinities. This allows, for instance, to ensure some TE LSPs will avoid particular network resources, using some bit masks. This can be seen as color. As an
example, some network links might be colored in red (with red meaning “high propagation delay” or “poor quality”). This affinity link property is propagated through IGP TE extensions (see [OSPF-TE] and [IS-IS-TE]). This way, a TE LSP carrying very sensitive traffic like voice-over-IP (VoIP) will be configured so red links are excluded from the path selection. In such a case, a major network failure may imply for the affected TE LSP to be non-reroutable without crossing one or several red links. In this case, it might be desirable to relax the affinity constraint.

Note 2: A large proportion of deployed MPLS TE networks rely on distributed computation in which no static path is configured; in this case, just a dynamic path is configured and the head-end just recomputes a new path based on the LSP constraints and its knowledge of the network and resource topology information provided by the IGP.

A usual question is: What is the CSPF duration time? And the systematic answer is: That depends. Indeed, the CSPF duration time is a function of the network size and the CSPF algorithm in use. Finding the shortest constraint path in a very large network obviously requires more time than in a small network. Furthermore, the CSPF complexity may be variable depending on the algorithm in use. Finally, the router CPU should also be taken into account. That said, in an order of magnitude, an average CSPF computation time using a classic CSPF algorithm on a network with hundreds of nodes rarely exceeds a few milliseconds. It is worth noting that one CSPF must be triggered per affected TE LSP. Indeed, if N LSPs starting on a head-end LSR R1 traverse a failed link, R1 will have to compute a new path for each of them.

Once a new path has been found and computed, the TE LSP is signaled along the new path. The final operation before any traffic can be routed over the newly signaled TE LSP consists of updating the routing table for the destinations that can be reached via the TE LSP.

5.3.3 Recovery Time

Providing hard numbers is not a realistic exercise because a significant number of factors influence the rerouting time, but we describe the different components of the recovery cycle with global default restoration through an example. Figure 5.6 shows the different steps of the recovery cycle with MPLS TE global default restoration.

Step 1: The link R3-R4 fails, and an FIS (RSVP and IGP update) is sent to the head-end LSR. As already pointed out, the sequence timing of IGP update and the RSVP Path Error depends of many factors. The receipt of one of them is sufficient for the head-end LSR to be notified of the failure.

Step 2: The FIS is sent to the head-end LSR. Note that the propagation delay might be nonnegligible and is made up of two components: the propagation delay (on wide area networks; this can be on the order of tens of milliseconds and can become as large as 100 ms between two continents where the optical path can be very long) and the queuing and processing delays for the FIS to
reach the head-end router. As mentioned in Section 5.2, an appropriate marking and scheduling in the forwarding path is highly recommended to ensure that the queuing and processing delays are both minimized.

**Step 3:** Upon receiving the failure notification, the head-end LSR (R1 in this example) tries to find an alternate path satisfying the set of constraints for each TE LSP affected by the failure.

**Step 4:** The TE LSP is signaled along the new path. The RSVP signaling set up time is also made of several components: the propagation delay along the path (round trip) and the queuing and processing delays at each hop in both directions (upstream and downstream).

**Step 5:** The routing table of R1 is updated to use the newly signaled LSP.

In conclusion, because the different components of the recovery time are highly dependent of the network characteristics, the resulting recovery time may vary from a few milliseconds to hundreds of milliseconds, sometimes a few seconds. Testing MPLS traffic reroute in a lab made of a few routers will probably result in a very short convergence time (a few milliseconds); indeed, the propagation delays are negligible, as is the FIS processing delay. The CSPF computation is also very short because the network size is limited, and finally the set up time will also be negligible. In contrast, a network with 1000 nodes, links with high propagation delays, and hundreds of TE LSPs to reroute will require a much more significant amount of time to converge.

### 5.4 MPLS Traffic Engineering Global Path Protection

MPLS TE global path protection (also usually referred to as path protection) is a global 1:1 protection recovery mechanism. As defined in Chapter 1, Section 1.5.4, this implies that the head-end LSR performs the rerouting (global recovery) and a presignaled backup LSP is used (protection) if the protected LSP fails.
5.4.1 Mode of Operation

Figure 5.7 describes the mode of operation of global path protection. In this figure, there are two primary TE LSPs, T1 (which follows the path R2-R3-R4-R5-R6) and T2 (which follows the path R7-R8-R9-R6). For each primary TE LSP, a dedicated backup LSP is set up, before any failure occurs. It is worth noting that a backup TE LSP (also called secondary TE LSP) must be link diverse or node diverse from the primary TE LSP. In this example, the backup LSP of T1 follows the path R2-R1-R10-R11-R5-R12-R6, which is link diverse \(^{64}\) from T1. By contrast, the backup LSP of T2 follows the path R7-R2-R3-R4-R5-R6 and is node diverse from T2. The aspects related to the backup path computation are covered in Section 5.15.

A backup (secondary) TE LSP is a regular TE LSP; that is, as far as RSVP signaling is concerned, a backup TE LSP is signaled as any other TE LSP and the backup TE LSP can be configured with either the same attributes as the primary TE LSP (in this case, the backup TE LSP satisfies the same set of constraints as the primary TE LSP) or with different constraints (e.g., no affinities, less bandwidth [say, 50% of the primary TE LSP]). For instance, if the backup TE LSP is configured with 50% of the primary TE LSP bandwidth, when used, the traffic will be forwarded along a path where 50% of the bandwidth has been reserved. This does not mean that the traffic will suffer from QoS degradation, depending on the actual use of the other LSPs sharing the same network resources along its backup path.

\(^{64}\)The terms disjoint and diverse are used interchangeably.

---

**Figure 5.7** MPLS traffic engineering path protection.
CHAPTER 5  MPLS Traffic Engineering Recovery Mechanisms

The mode of operation is quite straightforward: Once the failure is detected by some downstream node, an FIS is sent to the head-end LSR of each affected LSP (by affected LSP, we mean each LSP traversing the failed resource).

Note that all the aspects related to the FIS delivery described in Section 5.3 identically apply here because both the global default restoration and the global path protection rely on the FIS delivery to trigger an LSP recovery.

Then upon receiving the FIS, the head-end LSR immediately switches the traffic onto the backup TE LSP and updates its routing table accordingly.

5.4.2 Recovery Time

Compared to global default restoration, no routing computation has to be done “on the fly” to find an alternate route for the failed TE LSP. Moreover, with global path protection, the backup tunnel is already signaled, so no signaling round is required to set up the backup TE LSP. It is important to note that the saving in convergence time is predominately provided by the presignaling of the TE LSP.

5.5 MPLS Traffic Engineering Local Protection

After a brief section introduction to the specific terminology used for MPLS TE local protection, we describe the principle and mode of operation of two local protection techniques called MPLS TE Fast Reroute. The last section describes two deployment strategies of local protection recovery techniques. Note that the terms MPLS TE local protection and Fast Reroute are used interchangeably throughout this chapter.

5.5.1 Terminology

We begin this section by defining the terminology specific to MPLS TE Fast Reroute through an example (Figure 5.8).

As shown in Figure 5.8, an LSP T1 is signaled that follows the path R1-R2-R3-R4-R5. T1 is said to be “fast reroutable” if it is signaled with a specific attribute set in the RSVP Path message that indicates its desire to benefit from local recovery in the case of a failure.65 As shown in further subsections, Fast Reroute is a local protection recovery scheme; hence, the LSPs affected by a failure are locally rerouted by the node immediately upstream to the failure. This node is called the point of local repair (PLR). For instance, the node R2 is a PLR if the link R2-R3 or the node R3 fails. Fast Reroute uses backup tunnels to reroute affected LSPs. When a backup tunnel terminates to PLR’s next hop (direct adjacent neighbor), it is an NHOP backup tunnel. When the backup tunnel terminates on the neighbor of the

65See Section 5.14 for the details on RSVP signaling for Fast Reroute.
PLR's neighbor, the backup tunnel is an *NNHOP backup tunnel*. Back to our example, B1 is an NHOP backup tunnel of the PLR R2 and B2 is an NNHOP backup tunnel of R2. The node where the backup tunnel terminates is called the *merge point* (MP); hence, R4 is the MP of B2. Finally, a fast-reroutable LSP is said to be protected at a node R if there exists a backup tunnel that can be used in the case of a failure. T1 is protected at R2 by B1 and B2.

The terminology of detour merge point used in one Fast Reroute technique (one-to-one protection) is discussed in Section 5.14.

### 5.5.2 Principles of Local Protection Recovery Techniques

We use the generic term *MPLS TE Fast Reroute* or *Fast Reroute* to describe local protection techniques. There are two techniques of Fast Reroute (both are local protection techniques) that are described in this chapter:

- *Facility backup* (also referred to as *bypass*)
- *One-to-one backup* (also referred to as *detour*)

Although the terminology might appear difficult to understand, the terminology used in this section is in line with the corresponding standardized documents.

Both methods described are local repair techniques using *local protection*:

- *Local*: In the case of a link or node failure, a TE LSP is rerouted by the node that is immediately upstream to the failed link or node. Compared to the global default restoration and global path protection where the TE LSP is rerouted by the head-end LSR, in the case of local protection, the protected LSP is rerouted at the closest location upstream to the failure. This presents the very significant advantage of eliminating the need for the FIS to be received by the head-end LSR to reroute the affected TE LSP along an alternate path.
CHAPTER 5  MPLS Traffic Engineering Recovery Mechanisms

- **Protection:** As seen in the Chapter 1, with protection recovery mechanisms, a backup resource is preallocated and signaled before the failure. With both local protection recovery methods (facility backup and one-to-one backup), the backup LSPs are established before the failure occurs. When a failure occurs and is detected, every protected TE LSP traversing the failed resource (usually referred to as affected TE LSP) is rerouted over a backup TE LSP without having to compute a backup path “on the fly.”

Although both methods are local repair techniques, they significantly differ in terms of backup LSPs. With facility backup, a single (or a very limited number of) backup LSP(s) is used to protect all the fast-reroutable TE LSPs from the failure of a link or node, which is a major benefit of the MPLS label stacking property. By contrast, the one-to-one backup creates a separate backup LSP for each protected TE LSP at each hop. More details about their respective scalability are provided in Section 5.5.8.

To ease the understanding on each local protection technique, the following approach is followed: First, a quick overview of each local protection method is provided via an example. Then each method is described in detail in subsequent subsections.

5.5.3  **Local Protection: One-to-One Backup**

As depicted in Figure 5.9, with one-to-one backup, at each hop, one backup LSP (called a Detour LSP) is created for each fast-reroutable TE LSP. So, for instance, at the node R3, to protect the set of fast-reroutable TE LSPs T1, T2, and T3, the following set of backup TE LSPs are set up:

- One Detour LSP D1 for the protected TE LSP T1, following the path R3-R10-R11-R5-R6

![Figure 5.9 Illustration of the Detour LSP with one-to-one backup.](image)
- One Detour LSP D2 for the protected TE LSP T2, following the path R3-R8-R5-R9
- One Detour LSP D3 for the protected TE LSP T3, following the path R3-R10-R11-R12

Note that this only protects the fast-reroutable TE LSPs T1, T2, and T3 against a failure of the link R3-R4 and the node R4. Similarly, each node along the fast-reroutable TE LSP paths will perform the same operation.

At each PLR along the fast-reroutable TE LSP path, a local backup tunnel called Detour LSP that avoids the protected resource and terminates on the tail-end LSR for the fast-reroutable TE LSP is set up. In the previous example, for the fast-reroutable TE LSP T1, R3 sets up a Detour LSP D1 originated at R3 and terminated at R6 that avoids both the link R3-R4 and the node R4.

Figure 5.10 shows the label allocation for both the primary TE LSP T1 and the Detour TE LSP D1 protecting T1 against a failure of either the link R3-R4 or the node R4. The respective labels of the protected TE LSP T1 and the Detour LSP D1 originated on R3 are shown in Figure 5.10. For example, when a failure of the node R3 occurs, as soon as the PLR R3 detects the failure, the fast-reroutable TE LSP T1 is locally rerouted by the PLR to follow the Detour LSP, as shown in Figure 5.11. It is worth noting the label swapping change here: Once R3 detects the R4 node failure, the label 1 is no longer swapped from 1 to 2 and forwarded to the R3-R4 interface but is now swapped from 1 to 10 and is sent to the outgoing interface R3-R10.

_Detour LSP merging:_ Various merging rules allow for the reduction of the number of Detour LSPs and are described in Section 5.14.

---

**Figure 5.10** Mode of operation of one-to-one backup.
5.5.4 Local Protection: “Facility Backup”

By contrast with one-to-one backup, with facility backup, just one backup tunnel per NHOP is required to protect against a link failure and one NNHOP backup tunnel is required to protect against a node failure. Of course, an NNHOP protects against not only a node failure (the bypassed node) but also the link between the immediately upstream node and the bypassed node. As discussed later, there are some benefits in setting up both NHOP and NNHOP backup tunnels. More accurately, a small set of backup tunnels may be required if bandwidth protection must be guaranteed (see Section 5.15 for more details on bandwidth protection), but the key point is that the number of required backup tunnels is not a function of the number of TE LSPs in the MPLS network, which is a crucial property to preserve scalability.

In Figure 5.12, a single NNHOP backup tunnel (bypass) is configured on R3 (PLR) to protect any fast reroutable TE LSP traversing the node R3 and following the R3-R4-R5 path against a failure of the link R3-R4 or the node R4 (indeed, the same NNHOP backup tunnel can be used in both failure scenarios). R5 is the merge point. Hence, for instance, the two fast-reroutable TE LSPs T1 and T2 are protected by the NNHOP bypass tunnel B1 that follows the path R3-R10-R11-R5.

Let us now consider a fast-reroutable TE LSP T1 that follows the path R2-R3-R4-R5-R6. As shown in Figure 5.12, the corresponding labels are distributed in RSVP Resv messages (R5 distributes the label “3” to R4, R4 distributes the label “2” to R3, R3 distributes the label “1” to R2). In this example, a bypass tunnel B1 starting at the PLR R3 is also set up to protect against a link failure of the link R3-R4 and a node failure of R4. The corresponding labels are depicted in Figure 5.12.

Note: In the case of an NHOP backup tunnel, this is often referred to as MPLS TE Fast Reroute link protection. When the backup tunnel is an NNHOP backup tunnel, this is usually called MPLS TE Fast Reroute node protection.

![Figure 5.11 One-to-one backup: Example of the mode of operation when the node R4 fails and the protected TE LSP T1 is locally rerouted by the PLR R3 onto its Detour LSP D1.](image-url)
Figure 5.12 Facility backup operation.

Figure 5.13 Facility backup: Example of the mode of operation when the node R4 fails and the protected TE LSP T1 is locally rerouted by the PLR R3 onto the NNHOP backup tunnel B1.

A PLR can have NHOP and NNHOP backup tunnels. Furthermore, a PLR can have multiple NHOP backup tunnels and multiple NNHOP backup tunnels between a pair of LSRs to guarantee the bandwidth to the protected LSPs. This is discussed in detail in Section 5.15.

Let us now consider a node failure and see the mode of operation of facility backup (Figure 5.13). As shown in Figure 5.13, in the case of a node failure of R4, as soon as the failure is detected by the PLR (R3), each protected TE LSP following
the path R3-R4-R5 will be rerouted onto the bypass tunnel B1. The rerouting operation consists of swapping the incoming label to the appropriate outgoing label, pushing an additional label corresponding to the backup tunnel label, and redirecting the traffic onto the outgoing interface of the backup tunnel. The "appropriate" label is the label expected at the MP for the protected TE LSP.

It is worth elaborating on what the expected label is. So let us consider the two following situations:

**Situation 1:** The backup tunnel is an NHOP backup tunnel, in which case, the MP is also the PLR's NHOP for the protected LSP before failure occurs. Upon link failure, the PLR must perform a similar swap (no label change) as before the failure occurs; then the MP will receive the same label as before the failure but from a different interface.

This is illustrated in Figure 5.14.

In Figure 5.14, an NHOP backup tunnel B1 is set up from R3 to R4, which follows the path R3-R8-R4, protecting against a failure of the link R3-R4. The backup label distributed by R8 to R3 is 10 and a PHP (penultimate hop popping [PHP]) operation is performed between R8 and R4. Once the link failure is detected by the PLR (R3 in this example), for all the protected TE LSPs traversing the link R3-R4, the PLR R3 performs the following operations:

- Label swap of the protected TE LSP using the same label as before the failure
- Push of the label corresponding to the NHOP backup tunnel
- Redirect the traffic onto the backup tunnel outgoing interface

Figure 5.15 shows the situation after the link R3-R4 has failed and the PLR R3 has locally rerouted the protected TE LSP T1 onto the NHOP backup tunnel B1.

![Figure 5.14 Facility backup: Example of the mode of operation when the node R4 fails and the protected TE LSP T1 is locally rerouted by the PLR R3 onto the NHOP backup tunnel B1.](image-url)
The PLR R3 performs the following operations to locally reroute the protected TE LSP T1 onto the NHOP backup tunnel B1: R3 swaps 1 to 2 (as before), pushes the label 10 and redirects the traffic onto B1's outgoing interface (R3-R8). R4 (the MP) receives a label-switched packet containing the same label as before the failure but from a different interface.

Situation 2: With an NNHOP backup tunnel, the MP is now the PLR's next-next hop of the protected LSP before the failure. So the PLR must perform a swap so the MP receives a label switched packet with the expected label (but from a different interface).

To highlight this mechanism, let us consider the example depicted in Figure 5.16. Remember, at steady state (without any failure) the label swapping operation performed by R3 for the fast-reroutable TE LSP T1 is 1 to 2. In the case depicted in Figure 5.16, the MP R5 expects to receive a label 3 (label distributed by R5 to R4 for T1). So when the failure of the link R3-R4 or the node R4 occurs, R3 must swap 1 to 3 (instead of 2 before the failure), push the label 10, and redirect the traffic onto B1's outgoing interface (R3-R10). This way, R5 (the MP) receives an identical packet as before the failure but from a different interface. By default, the PLR does not have the knowledge of the label used between its NHOP LSR and NNHOP LSR; it just learns from its direct downstream neighbor the label it must use for the TE LSP. An extension to an existing RSVP object (RRO object) is used to learn the label used between the NHOP and the NNHOP LSR (that signaling extension is described in Section 5.14).
CHAPTER 5  MPLS Traffic Engineering Recovery Mechanisms

Figure 5.16 Situation after the failure of the link R3-R4 and the PLR R3 has locally rerouted the protected TE LSP T1 onto the NNHOP backup tunnel B1.

Important notes:

Note 1: An identical operation is performed for every protected LSP rerouted onto the same backup tunnel; indeed, with facility backup, the same backup LSP is used for all the rerouted TE LSPs that intersect the backup tunnel on both the PLR and the MP.

This is illustrated in Figure 5.17. This figure shows two primary tunnels T1 and T2 that used to follow the paths R1-R3-R4-R5-R6 and R7-R3-R4-R5-R6 before the failure. The labels in use are 100 (between R1 and R3), 101 (between R3 and R4), 102 (between R4 and R5) and PHP (between R5 and R6) for T1 and 110 (between R7 and R3), 111 (between R3 and R4), 112 (between R4 and R5) and PHP between R5 and R6. Because both T1 and T2 intersect at R3 and R5, the same NNHOP backup tunnel B1 can be used in the case of failure of the link R3-R4 or node R4. This is of course a very important scaling property of facility backup that uses MPLS stacking. Note also that the same property applies to NHOP backup tunnels.

Note 2: In both cases (NHOP and NHOP bypass tunnels), no additional RSVP states are created along the backup paths for the rerouted TE LSPs. In other words, the LSRs along the backup path do not “see” the rerouted TE LSPs as far as the control plane is concerned. This is also a crucial property for the scalability properties of this solution.
5.5 Properties of a Traffic Engineering LSP

When using MPLS TE local protection, there are three properties a TE LSP can have:

1. Fast Reroute desired
2. Bandwidth protection desired
3. Node protection desired

**Fast Reroute desired TE LSP:** Fast Reroute is a technology that can be used for some TE LSPs only (as already stated, such TE LSPs are called fast-reroutable TE LSPs), so if a backup tunnel has been configured on a PLR, just the TE LSP signaled as “fast reroutable” will be fast rerouted in the case of a failure. Typically, this provides fast recovery using local protection to a subset of TE LSPs having stringent recovery requirements (e.g., the TE LSPs carrying sensitive traffic like VoIP or ATM-over-MPLS), whereas other TE LSPs carrying less sensitive traffic (e.g., Internet traffic) will be rerouted using TE LSP reroute. This obviously requires the ability to explicitly signal this fast-reroutable property of a TE LSP. The details of the signaling aspects are covered in Section 5.15.

**Bandwidth protection desired:** The notion of bandwidth protection is extensively covered in Section 5.15, but here is a high-level description of this important notion. The previous section described the mode of operation of Fast Reroute for both the facility backup and the one-to-one backup method. When a TE LSP is signaled, one of the TE LSP attributes of the TE LSP is the bandwidth. A TE LSP is said to be bandwidth protected at a node R only if it can be fast rerouted and the selected backup tunnel offers an equivalent bandwidth as the primary TE LSP used to receive along the primary path (before the failure). In other words, the TE LSP does not suffer any QoS degradation along the alternate path. Note that the QoS may be a function not just of the bandwidth.
but also of the propagation delay or jitter. Section 5.15 details how backup paths can be computed to provide such guarantees. When signaled, a protected TE LSP can explicitly request bandwidth protection.

Node protection desired: In some cases, also further discussed in Section 5.15, it might not be possible for a PLR to find both an NHOP and an NNHOP backup tunnel offering full bandwidth protection. For example, let us consider the simple case of three routers R1, R2, and R3 connected in a row, and the R1-R2 link bandwidth is 20 Mbps and the R2-R3 link is 10 Mbps. Then the PLR may try to find an NHOP backup tunnel with 20 Mbps worth of bandwidth and an NNHOP backup tunnel with \( \min(20, 10) = 10 \) Mbps worth of bandwidth. Suppose that no such NNHOP backup tunnel can be found but just an NNHOP backup tunnel of 5 Mbps. Then as new TE LSPs requesting for bandwidth protection are signaled, it may happen that no NNHOP backup tunnel offering bandwidth protection can be found. In this case, having an additional signaled parameter explicitly requesting node protection is desirable and can be used as a tie break. So if the PLR has two requests for bandwidth protection and cannot select an NNHOP backup tunnel for both of them because of insufficient bandwidth on the NNHOP backup tunnel, it can preferably select the NNHOP backup tunnel for the TE LSP having expressed a desire to get node protection in addition to bandwidth protection. Such a parameter has been standardized in [FAST-REROUTE] and is described in Section 5.15.

**Notion of Class of Recovery**

The various TE LSP recovery requirements mentioned earlier allow an operator to define multiple CoRs and assign a different CoR to each TE LSP according to its recovery requirements. For instance, very sensitive traffic like voice-over-IP/MPLS or ATM-over-MPLS could be routed over protected TE LSPs with bandwidth and node protection. In the case of a link or node failure, those TE LSPs would be very quickly rerouted, while maintaining an equivalent QoS. On the other hand, MPLS VPNs traffic could be routed onto protected TE LSPs without bandwidth protection. Finally the less sensitive traffic could be routed over non-protected TE LSPs.

Defining multiple classes of recovery provides the two following benefits:

- The set of rerouting operations can be prioritized. Indeed, every LSR will preferably start to recover the TE LSPs that belong to the highest CoR.
- When bandwidth protection is required, this implies reserving some backup capacity in the network. With multiple CoRs, the amount of backup capacity is limited to the set of TE LSPs that belong to the CoR for which bandwidth protection is required. This allows to significantly optimize the required backup capacity.
5.5.6 Notification of Tunnel Locally Repaired

As described earlier, upon detection of a link/node failure, the PLR immediately starts rerouting the set of protected TE LSPs over their respective backup tunnels (bypass tunnels or Detour LSPs). This may result in following a suboptimal end-to-end path. Consequently, in addition to performing the local reroute, the PLR sends a specific RSVP Path Error message for each rerouted TE LSP to their respective head-end LSR to indicate that a local reroute has occurred. This type of RSVP Path Error is sometimes qualified as nondisruptive because no RSVP states are cleared; it serves as a pure indication to the head-end LSR. The receipt of such a message will then trigger a reoptimization on the head-end LSR for the affected TE LSP. Indeed, as previously mentioned MPLS TE Fast Reroute is a temporary network recovery mechanism; the protected TE LSPs are quickly and locally rerouted onto backup tunnels using a local protection technique, but the path followed by the rerouted flows might no longer be optimal. This is illustrated in Figure 5.18.

In Figure 5.18, a protected TE LSP, T1, following the path R0-R1-R2-R8 is set up. At router R1 (PLR), T1 is protected by an NHOP backup tunnel B1 against a failure of the link R1-R2 (B1 follows the path R1-R3-R4-R5-R2). When the link R1-R2 fails, upon detecting the link failure, the PLR (R1) reroutes the LSP T1 onto B1 and sends a Path Error “tunnel locally repaired” to T1’s head-end LSR (R0). As you can see in Figure 5.18, the path followed by T1 is not optimal (R0-R1-R3-R4-R5-R2-R8). The receipt of the Path Error triggers a reoptimization on R0, which in turn reroutes the TE LSP T1 along the path R0-R3-R4-R5-R2-R8, which is more optimal than the path followed by the rerouted flows during failure (R0-R1-R3-R4-R5-R2-R8). In this example, we assume that all the links have the

![Figure 5.18](image-url)
same metric. Of course, the TE LSP reoptimization should always be performed using the "make before break" procedure, avoiding any traffic disruption.

The head-end will also be informed of the link failure via the receipt of an IGP update from one of the routers adjacent to the failed link. Either upon the receipt of an RSVP Path Error notify message "tunnel locally repaired" or an IGP update, the head-end triggers a TE LSP reoptimization.

Case of a Multiarea (OSPF) or Multilevel (IS-IS) Network

In the case of a multiarea (OSPF), multilevel (IS-IS), or multiautonomous systems network, if the failure does not occur in the head-end LSR area/level, no IGP notification will be received by the head-end LSR. This means that the head-end LSR exclusively relies on the receipt of the RSVP Path Error message to be informed that a local repair has been performed on a downstream node. Consider the network depicted in Figure 5.19.

In Figure 5.19, a fast-reroutable interarea TE LSP (T1) is routed from R0 to R4 and spans multiple areas. On R2, a NHOP backup tunnel that follows the path R2-R5-R6-R7-R3 protects any fast-reroutable TE LSPs traversing the link R2-R3 from a failure. When the link R2-R3 fails, the TE LSP T1 is rerouted onto the backup tunnel B1, but in this case the head-end LSR R0 does not receive any IGP update. Indeed, the failure occurred in the backbone area, so R0 does not have any visibility of the backbone area topology. A failure in the backbone area is invisible to R0 (R2 might send a new summary LSA if some addresses are no longer reachable, but generally the address aggregation scheme will be such that no summary LSA will be flooded into the area 0.0.0.1). Because the RSVP Path Error notify message is the only mechanism allowing the head-end LSR to be informed of a local repair that occurred on a downstream node that does not reside in the head-end area, a best common practice consists of sending the RSVP Path Error message in reliable mode.

Figure 5.19 Notification of local repair followed by head-end reoptimization in a multiarea routing domain.
5.5.7 Signaling Extensions for MPLS Traffic Engineering Local Protection

By contrast with MPLS global default protection and MPLS TE global protection, which do not require any signaling protocol extensions beyond those of RSVP TE defined in [RSVP-TE] for the signaling of MPLS TE LSP, MPLS TE local protection (Fast Reroute) requires several signaling extensions. Although they are undoubtedly important, their detailed understanding is not a prerequisite to grasp how local protection works. Consequently, the signaling aspects of Fast Reroute are covered in detail in Section 5.14.

5.5.8 Two Strategies for Deploying MPLS Traffic Engineering for Fast Recovery

As mentioned in Section 5.1, there might be several motivations for deploying MPLS TE:

- **Bandwidth optimization**: So that the network resources are used in a more efficient way. This also helps in providing better QoS.
- **Providing strict QoS guarantees** to some specific traffic flows.
- **Fast recovery**.

In some networks, there might be an interest in MPLS TE for its fast recovery property only. In other words, bandwidth optimization and/or strict QoS guarantees are not required, but the operator would like to benefit from the fast recovery property of Fast Reroute without tuning its IGP parameters as described in Chapter 4. This section proposes two strategies for deploying MPLS TE when the only objective is to get fast recovery by using Fast Reroute.

For instance, consider an underutilized (or overprovisioned) network. Such a network does not require any bandwidth optimization because it is not congested. Also, depending on the network load, QoS guarantees could rely on the simple assumption that no link is congested and the link loads are very low. In such a situation, MPLS TE is not required, and paths computed by the routing protocol are perfectly satisfactory. However, such a network may require fast recovery of link or node failures, making Fast Reroute a good candidate. Because Fast Reroute requires TE LSPs, the solution includes deploying TE LSPs but in a quite specific way, which we describe in this section.

There are two strategies for deploying MPLS TE when the sole objective of the operator is to use Fast Reroute:

1. With a full mesh of unconstrained TE LSPs
2. With one-hop unconstrained TE LSPs

**Network Design with a Full Mesh of Unconstrained TE LSPs**

A simple and efficient strategy is to deploy a full mesh of unconstrained TE LSPs. An unconstrained TE LSP is an LSP without any constraint. For instance,
the required bandwidth is 0, and no affinities are defined. The only property of such a TE LSP is to be fast reroutable. Indeed, the objective is not to use the traffic engineering property of MPLS TE (in the sense of “traffic engineer” the flows across the network). So the available bandwidth and other TE link–related information are still flooded by the IGP TE extensions but will never change. When a head-end LSR computes a path for an unconstrained TE LSP, the same CSPF algorithm is used as with any other TE LSP, but the obvious outcome is that the TE LSP will follow the IGP shortest path. In other words, the traffic routed onto unconstrained TE LSPs will follow the same paths as IP routed traffic, but in the case of link and/or node failures, fast-reroutable TE LSPs will be rerouted by MPLS TE Fast Reroute, which was the initial objective.

**Network Design with Unconstrained One-Hop TE LSPs**

If the requirement is to use Fast Reroute for link protection only, then exactly one primary unconstrained TE LSP plus one single NHOP backup tunnel are required for every link to protect.

The idea is to set up a one-hop tunnel following the same path as the link to protect. One way of achieving this is to set up an unconstrained TE LSP. This way the CSPF algorithm will just follow the most direct path between the head-end LSR and the tail-end LSR (the next hop of the head-end LSR in this case). Note that in this case the PLR node is also the head-end LSR. Then the one hop primary TE LSP must be configured so that all the traffic follows the TE LSP.

It is important to note that because the TE LSP is a one-hop LSP, if PHP is used, no label is added once the traffic is routed over the primary TE LSP. Such a strategy is depicted in Figure 5.20.

In the example shown in Figure 5.20, the objective is to protect the link R2-R3. So a single-hop tunnel (T1) is configured from R2 to R3 and all the traffic is routed onto this one-hop primary TE LSP through this link. T1 has no constraint, so this TE LSP follows the path R2-R3. An NHOP backup tunnel B1 is configured between R2-R3 with the constraint of being diversely routed from the protected link and follows the path R2-R8-R3. As discussed in Section 5.15, additional constraints may be added to also provide bandwidth protection. In the case of failure of the link R2-R3, the PLR (R2) will trigger Fast Reroute and all the traffic that used to be routed over the link R2-R3 will be rerouted over B1, following the path R2-R8-R3. Then the primary TE LSP T1 will be rerouted (reoptimized) and will follow the new shortest path between R2 and R3. Finally, the routing protocol will be informed of the link failure and will recompute a new path, which may or not follow B1’s path.

The same configuration has to be repeated for each link to protect using Fast Reroute. Note that if the link R2-R3 is protected using a SONET/SDH protected VCs, Fast Reroute may also be used to protect against a router interface failure on the R2 or R3 side. In that case, one must ensure that both mechanisms are not
simultaneously triggered. This aspect is covered in Chapter 6. Existing implementations support mechanisms to automate the creation of both the primary and the backup TE LSP, because in this case their set of attributes is known in advance to alleviate the configuration burden. The only constraint of the backup tunnel is to be diversely routed from the link to protect (some implementations support the computation of SRLG-diverse paths).

Protection against link and node failures: To guard against both link and node failures, a similar approach is followed, with the only difference that at each hop, both one unconstrained TE LSP and one NNHOP backup tunnel per next-next hop must be configured.

Why are one primary and one NNHOP backup tunnel required per NNHOP?

Let us consider the example in Figure 5.21. As shown in Figure 5.21, in the case of a node failure of R3, all the traffic traversing the protected LSR needs to be rerouted onto some appropriate backup tunnels. That requires setting up one primary TE LSP for each possible traffic path traversing the protected node. In Figure 5.21, there are three paths leaving R2 that traverse the node R3 to consider: R2-R3-R4, R2-R3-R7 and R2-R3-R8. So three unconstrained TE LSPs are configured and set up on R2: T1, T2, and T3. Because each of these tunnels needs to be rerouted over a diversely routed backup tunnel, three NNHOP backup tunnels are configured: B1 protecting the traffic following the path R2-R3-R4 and routed onto the tunnel T1, B2 protecting T2, and finally B3 protecting T3. As in the case of link protection, the protected TE LSPs are unconstrained and follow the shortest IGP path.

This explains the requirement for one unconstrained TE LSP and one backup tunnel per NNHOP. In the previous example, the number of NNHOPs of R2 is equal to 3: R7, R3, and R8.
Comparison of Both Approaches

Both the “unconstrained full mesh TE LSPs” and the “unconstrained approach” can be used and have their respective pros and cons. Indeed, the unconstrained approach clearly has the advantage to require the configuration and set up of a very limited number of TE LSPs. If just link protection is required, for every link to protect with Fast Reroute, just two TE LSPs are required: the primary one-hop TE LSP and an NHOP backup tunnel diversely routed from the link to protect. If node protection is required, one pair of TE LSPs (primary and backup) is needed for every next-next hop, as described earlier, which is still a very reasonable number. Note that at the time of publication, commercial implementations support only the 1-hop unconstrained approach. Moreover, some implementations ease the configuration process with the use of very few commands to automate the configuration of such primary and backup TE LSPs.

On the other hand, the unconstrained full mesh TE LSPs approach also offers a very easy migration path to the use of MPLS TE for other purposes like bandwidth optimization and strict QoS guarantees. Indeed, if at some point, one of those requirements appears, the operator will just need to set constraint(s) on the TE LSPs. For instance, bandwidth can be configured and then the TE LSPs will start using alternate path(s), if required.

In terms of existing implementations, some solutions are available that automate the configuration process when setting up a full mesh of TE LSPs. In a nutshell, those solutions rely on several components:

- A discovery process is in charge of discovering the members of a mesh. In some MPLS TE networks, there might be multiple TE LSP meshes: one mesh of TE LSPs between LSRs acting as VoIP gateways, for instance, and...
another full mesh of TE LSPs between routers carrying the Internet traffic. Each TE mesh has its own set of characteristics in terms of bandwidth, priority, and protection/restoration, to mention just a few requirements. Then each router uses an IGP extension (OSPF or IS-IS) to advertise that it is a member of one or multiple TE meshes. This mechanism allows every router to discover all the other routers that belong to the same TE mesh.

- Then, once a router has discovered all the routers that belong to the same mesh, it can use a "template" (where the constraints specific to that particular mesh are locally specified) to set up the mesh of TE LSPs. Note that in this particular context of using MPLS TE for fast recovery only, the template is very restricted because the primary TE LSPs are unconstrained.

In terms of IGP, both methods are equivalent. The TE-related information is flooded by the IGP but will never changed because the TE LSPs are unconstrained and never reserve bandwidth.

5.6 Another MPLS Traffic Engineering Recovery Alternative

Another MPLS TE recovery alternative has been proposed but never got any traction in the industry because of severe limitations: 1+1 packet protection whose principle is to permanently bridge the IP/MPLS traffic over two diversely routed TE LSPs. The traffic bridging is made on the head-end LSR, and the decision to switch the traffic is performed by the tail-end LSR, which permanently compares the two identical received flows from the primary and secondary TE LSPs. When a failure occurs in the network, the traffic received from one of the TE LSPs is affected. Once the tail-end LSR detects the failure, it switches to the secondary TE LSP. Note that such a mechanism is also called a single-ended protocol because the switching decision process is made by a single entity (the tail-end LSR in this case) without requiring any signaling exchange between the nodes. A failure may be a traffic interruption, an unacceptable error rate, or any other kind of defects. Once the tail-end LSR has performed the switch, it can either decide to stay indefinitely on this TE LSP and start using the other TE LSP (once restored) in the case of failure of the currently selected TE LSP or decide to switch back to the original TE LSP, once restored.

Although this mechanism is simple and efficient in terms of recovery time, it has two major drawbacks that drastically limit its applicability:

- The amount of traffic forwarded in the network is doubled for each TE LSP protected with this 1+1 mechanism. This is a serious issue because it basically implies at least^66 a bandwidth wastage of 50%.

---

^66This technique implies at least a bandwidth wastage of 50% because one of the constraints of the backup TE LSP is to be disjoint from the protected TE LSP, which usually means that it will follow a longer path.
The failure discovery at the tail-end LSR usually requires some hardware changes and thus equipment replacement, which can also be expensive. For those reasons, such a mechanism has never been implemented or deployed but is just mentioned here for the sake of completeness in describing MPLS TE recovery techniques.

5.7 Load Balancing

Load balancing is a technique to forward the traffic from a source to a destination across multiple paths. With equal load balancing, the traffic is balanced across multiple equal-cost paths. IGP, like OSPF or IS-IS, performs equal load balancing. This can be done on a per-packet basis (packets are sent along N equal-cost paths using a round-robin algorithm) or via some more sophisticated techniques avoiding packet reordering described in Chapter 4. With MPLS TE, both equal and unequal load balancing are supported. For instance, if there are two TE LSPs, T1 and T2 between two LSRs, LSR1 and LSR2, with respective bandwidth Bw1 and Bw2, then LSR1 can decide to balance the traffic whose destination is LSR2 (or beyond) in proportion to the respective bandwidths Bw1 and Bw2. Usually load balancing in MPLS TE-enabled networks is used when a single path obeying the set of constraints cannot be found between two LSRs. For instance, a TE LSP of B Mbps is required and no path with the required amount of bandwidth is available. Then the solution is to set up N LSPs so the sum of their bandwidth is equal to B. Another constraint can be added when the path computation of the set of N LSPs is performed like path diversity (the set of network elements traversed by the TE LSPs are disjoint).

Strictly speaking, load balancing is not an MPLS TE recovery technique, so why dedicate a section to it?

Because a positive side effect of load balancing is that when the flow between two points is balanced across multiple paths, the probability of simultaneous failures of all those paths is reduced compared to a single path, especially if those paths are explicitly diversely routed. Hence, the overall availability is increased. This property has been used by some operators to reduce the impact of network failure on specific flows.

Let us illustrate that statement through the example in Figure 5.22. In this case, strictly speaking, the network availability is not increased but the impact of a network element failure on the traffic flows between two points is reduced.

In Figure 5.22, the two POPs of Sevilla and Barcelona are made of two VoIP gateways and one LSR connected to the core of the network. The VoIP traffic is carried onto TE LSPs. In this case, even if all the traffic between LSR1 and LSR2 could be carried onto a single TE LSP, two diversely routed TE LSPs are established between LSR1 and LSR2 (with the same bandwidth or different bandwidths) and the traffic is balanced onto those two TE LSPs. In the case of a network failure,
just a proportion of the traffic between LSR1 and LSR2 is affected (the traffic
carried onto T1 in the previous example).

That said, we must admit that such a design choice has the two following
drawbacks:

- The number of states in the network is nonnegligibly increased: indeed, at
  least two TE LSPs are required between two LSRs.
- The constraint of computing diverse paths may result in computing non-
  optimal paths compared to a single TE LSP.

But on the other hand the impact of a single element failure on the voice traffic
between the two POPs is reduced.

**Note:** One can, for example, increase the capacity of each TE LSP to be able to
absorb the excess traffic resulting from the failure of one TE LSP. For instance, if N
TE LSPs of B Mbps are set up between two routers R1 and R2 (let us call it a
bundle of N TE LSPs), by allocating B * N/(N-1) Mbps to each of them instead of B
Mbps; this allows the survival from the failure of one of them. In this case, the
backup capacity reserved in the network is strictly equal to the capacity of one TE
LSP in the bundle.
5.8 **Comparison of Global and Local Protection**

As previously described in Chapter 1, the evaluation of a recovery mechanism requires the consideration of several parameters: scope of recovery (link, node, SRLG), recovery time, guaranteed bandwidth, backup capacity requirements, state overhead, scalability, reordering, additive latency and jitter, signaling requirements, stability, and others. Throughout this chapter, we saw several MPLS TE recovery techniques, so the natural question that comes to mind is, which one to use. Although there is no unique answer because each network has its own constraint and requirements, the aim of this section is to provide a comparison of the global protection and local protection techniques with a particular focus on three key performance aspects:

- The recovery time
- The state overhead, which is directly correlated to the scalability
- The ability to perform bandwidth sharing when bandwidth protection is required

### 5.8.1 Recovery Time

With global protection, rerouting is performed by the head-end LSR, which means that this requires for the head-end LSR to receive the failure notification to reroute the affected traffic onto their respective backup paths (whose paths have been precomputed and signaled). So in terms of recovery time, the delta between global and local protection is the failure indication signal propagation time to the head-end LSR. How large this delta is highly depends on the network characteristics. Thus, for instance, a network confined to a small country generally implies short propagation delays (less than 10 ms); on the other hand, an international network may easily experience much longer propagation delays... up to a few hundreds of milliseconds. In that case, convergence of a few tens of milliseconds requires the use of local protection techniques. Furthermore, queuing delays to process the control plane notification (RSVP and/or IGP) messages can be reduced via the use of QoS mechanisms.

Note that in terms of recovery time, the two local protection schemes described earlier (i.e., "one-to-one" and "facility backup") are equivalent; they both rely on local protection where fast-reroutable TE LSPs are locally rerouted on presignaled backup tunnels and then reoptimized by their respective head-end LSR.

In summary, as far as the recovery time is concerned, the key difference between local and global protection is in the failure propagation notification time to the head-end LSR which, in the case of global protection is made of incompressible propagation delays and queuing delays that can be reduced by means of QoS mechanisms.

### 5.8.2 Scalability

Scalability is undoubtedly one of the major aspects to consider when evaluating a recovery mechanism, and to that respect, global path protection, one-to-one backup, and facility backup local protection differ very significantly.
Scalability is a relatively generic term that requires clarification in this context. Protection mechanisms require setting up backup tunnels before any failure to provide fast convergence (by contrast with global default restoration, the backup path is already computed and signaled). The configuration of backup tunnels can be facilitated via an automatic process, but setting up backup tunnels in a network is not entirely cost free. Although the scalability of RSVP is very high, in large networks, the number of backup tunnels can be significant as shown below, which requires to potentially handle a large number of states on routers. Moreover, the troubleshooting task is even more complicated for the team in charge of operating the network. So scalability is considered in terms of number of required backup tunnels in this context.

Let us evaluate the number of required backup tunnels with global path protection, Fast Reroute facility backup, and one-to-one, based on the following assumptions:

- $D$: network diameter (average number of hops between a head-end LSR and a tail-end LSR)
- $C$: degree of connectivity (average number of neighbors)
- $L$: total number of links to be protected with Fast Reroute\(^{67}\)
- $N$: total number of nodes (LSRs)
- $T$: total number of protected TE LSPs in the MPLS network
- $Bu$: number of backup tunnels required
- $K$: number of class of recovery (as mentioned in Section 5.5.5, there might be several classes of TE LSPs, each requiring different CoR. In this case, each CoR has a dedicated set of backup tunnels)
- $S$: average number of splits (as discussed in Section 5.15, in some cases where bandwidth protection is required and backup bandwidth is a very scarce resource, more than one backup tunnel per protected link/node may be required if a single backup tunnel with enough bandwidth cannot be found)

**Note:** Realistic assumptions for $S$ and $K$ are as follows:

- $S < 4$: Generally $S$ will very rarely exceed 3. In a network where bandwidth protection is required but backup capacity is not a very scarce resource $S = 1$. If bandwidth protection is not required, then $S = 1$.
- Also $K < 3$.

- $M$: number of meshes in the network (e.g., there may be multiple meshes of TE LSPs in a network serving different purposes: one mesh for the voice traffic and one mesh for the data traffic).

It follows that

- $L < N \cdot C$ (because some links may not be protected by Fast Reroute)
- $T = M \cdot N \cdot (N-1)$ (assuming a full mesh TE deployment)

\(^{67}\)Some links may be protected via other means like SONET/SDH and optical protection/restoration.