Chapter 5
Resilience Schemes for Fast Recovery in Packet-Switched Communication Systems

Packet-switched networks, invented independently by Paul Baran and Donald Davies during the 1960s, have been playing a key role worldwide in delivering communication services in numerous deployment scenarios, including the Internet, data center networks, or enterprise networks [7]. In *packet switching*, data is organized into packets of a limited length consisting of the *packet header* and the *packet payload*. Packet headers include data utilized by the network nodes to deliver the packets to destination nodes. Packet load, in turn, denotes data used by higher layer protocols and applications. Concerning the TCP/IP protocol family, major forms of packets include Layer-2 Ethernet frames and IP Layer-3 datagrams.

The uninterrupted availability of packet-switched networks has become crucial for the operation of many classes of applications, e.g., related to business or health. However, in failure scenarios, it is often common that the response of the conventional resilience mechanisms deployed in the control plane is not efficient enough to provide a fast recovery of the affected communication paths. Indeed, the time needed for conventional control plane mechanisms to recompute communication paths can be high and even involve tens of seconds [12].

In this chapter, we discuss the properties of mechanisms extending the operation of conventional Layer-2 and Layer-3 route calculation schemes, which are necessary to reduce the noticeably long convergence time, i.e., the time needed for network nodes to obtain a new joint view of the network state and the set of updated transmission paths that are valid after a failure. In the remaining part of this chapter, we first discuss in Sect. 5.1 the properties of Layer-2 message dissemination schemes, namely the spanning tree protocol (STP) characteristic of Ethernet networks (being the most common IP Layer-2 technology), and further explain the major variants of STP aimed at ensuring fast recovery of affected spanning trees. Next, in Sect. 5.2, we discuss the properties of the selected IP Layer-3 fast recovery mechanisms, while

J. Rak, *Resilient Routing in Communication Networks*, Computer Communications
and Networks, https://doi.org/10.1007/978-3-031-64657-7_5

in Sect. 5.3, we highlight the mechanisms of fast recovery in IP-MPLS networks. Section 5.4 concludes the chapter.

5.1 Link-Layer Recovery Mechanisms in Packet-Switched Networks

This section aims to discuss the mechanisms of resilient transmission for Ethernet networks being the most common IP Layer-2 technology [7]. In general, assuring resilience is much more challenging in Ethernet networks since, contrary to IP Layer-3 multi-hop transmission, Layer-2 frames do not include fields similar to the Layer-3 Time-to-Live (TTL) to prevent forwarding loops in failure scenarios.

To avoid forwarding loops while restoring the Layer-2 communication paths affected by failures, solutions based on the concept of the *spanning tree* were proposed. In this context, the first notable scheme is the IEEE 802.1D Spanning Tree Protocol (shortly, *STP*) standardized as IEEE 802.1D [15] using a single spanning tree, i.e., a tree connecting all the nodes in the network. In this way, any pair of network nodes remains connected by a single path being part of that tree. In the event of a failure, the spanning tree is reconfigured in a way that provides transmission opportunities for any pair of nodes surviving the failure.

However, the procedure for reconfiguring a spanning tree in STP is relatively slow and often unacceptable for many applications. Indeed, following [11], the recovery of an affected spanning tree can even take tens of seconds, depending on the network size. Therefore, this section, apart from discussing the properties of STP, will also review the characteristics of two other representative approaches aimed at reducing the time needed for the recovery of the affected spanning tree, namely the Rapid Spanning Tree Protocol (RSTP) referred to as IEEE 802.1w [18] and a scheme using multiple spanning trees (IEEE 802.1s standard [17]), both later on included in the IEEE 802.1Q-2014 standard [16].

5.1.1 Spanning Tree Protocol

As already mentioned in this section, the purpose of the *Spanning Tree Protocol* (*STP*) proposed by Radia Perlman is to establish and maintain a tree topology connecting all nodes of an Ethernet network. In a tree topology, for every pair of network nodes, there is exactly one path in that tree connecting them (i.e., there are no loops).

Prevention of loops is indeed one of the major objectives of STP since, as already mentioned in this chapter, Ethernet frames do not provide a field similar to the Layer-3 Time-to-live (TTL) field to avoid the endless forwarding of frames likely to occur in mesh topologies. For this purpose, STP disables network links not

Table 5.1 Link cost vs. link bandwidth in STP (IEEE 801.1D-1998)

Link bandwidth	4 Mbps	10 Mbps	16 Mbps	100 Mbps	1 Gbps	2 Gbps	4 Gbps
STP link cost	250	100	62	19	4	3	2

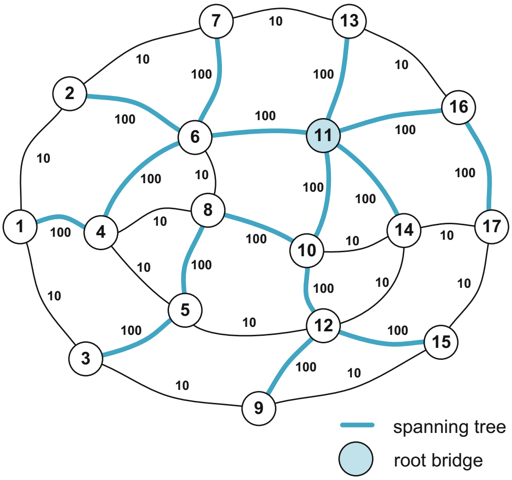


Fig. 5.1 An example spanning tree determined by STP for a 17-node network (the values next to links denote the nominal link capacity in Mbps)

belonging to the spanning tree and, therefore, maintains only a single path between each pair of network nodes.

In STP, one switch in the network is elected as a *root bridge*. This election takes place based on the lowest value of bridge priorities configured for each switch manually. In the case of several equal lowest values of bridge priority configured for several switches in the network, a switch with the lowest MAC address among these switches becomes the root bridge. After that, each non-root switch determines the best communication path (i.e., of the lowest cost) between itself and the root bridge. This path will next become part of the tree. Table 5.1 illustrates the reference costs of links in STP in relation to link bandwidth based on IEEE 802.1D-1998, while Fig. 5.1 gives an example spanning tree for a 17-node topology with node 11 elected the root bridge. In general, as the costs of links are inversely proportional to their bandwidth, links of higher capacity are preferred in path computations.

During path calculations, STP switches exchange information using *bridge protocol data units* (*BPDUs*). After all paths between switches and the root bridge are determined, each switch configures one of its ports as a root port, which connects it with the root bridge. Links not present in any path between switches and the root bridge are thus excluded from the tree (i.e., blocked).

Upon a change of the network topology (as a result of, e.g., adding a new node or following a failure of a given network element), topology change notification (TCN) BPDUs are sent by the respective non-root node (i.e., the switch at which the change was detected on one of its ports) toward the root bridge. Upon receiving the TCN BPDU, the root bridge initiates the topology update procedure by setting the related “topology change” flag in exchanged BPDUs. Setting this flag triggers the spanning tree update by forcing the non-root nodes to recalculate their best paths to the root bridge.

Since the exchange of BPDUs in STP is periodic (typically once every 2 seconds), the reaction of STP to failures leading to a reconfiguration of the spanning tree, measured even in tens of seconds, is indeed slow. For this reason, as well as owing to the preference for high-capacity links when forming the spanning tree, remarkable data losses may occur in STP in failure scenarios. Therefore, the focus of mechanisms discussed in the remaining part of this section is on the fast recovery of spanning trees.

5.1.2 Rapid Spanning Tree Protocol

The main reason behind the introduction of the *Rapid Spanning Tree Protocol* (*RSTP*) was to reduce the long time of the STP algorithm convergence, e.g., in post-failure periods. Compared to STP, which typically requires from 30 to 50 seconds to re-establish the spanning tree, the time needed to finalize a new configuration of a spanning tree in RSTP is significantly improved. As verified in [22], RSTP is able to converge even within milliseconds.

RSTP is similar to STP concerning the rules for electing the root bridge, root ports, designated ports (i.e., ports leading to certain segments of a network), and in terms of blocking certain ports to avoid loops.

Compared to STP, apart from the root ports configured at each switch, RSTP introduces additional roles illustrated in Fig. 5.2 that can be assigned to ports of switches to improve the algorithm convergence time, namely:

- *Alternate port*: a port providing the alternate path from a given switch to the root bridge (i.e., a path that is different from the main one via the root port).
- *Backup port*: a port being a backup port for a given root port providing a backup path from the root bridge to a given network segment.

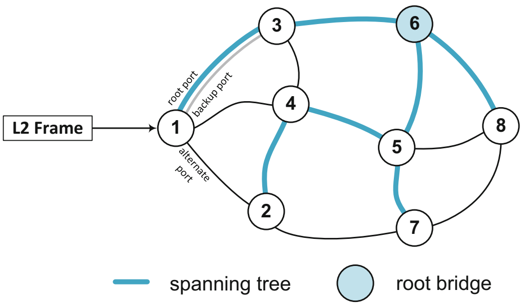


Fig. 5.2 An example configuration of a spanning tree including information on roles of selected ports of node 1 specified in RSTP

In RSTP, these ports can immediately enter the forwarding state instead of waiting for the final result of the algorithm convergence (as in STP) due to the ability of neighboring switches (i.e., connected by a point-to-point link) to acknowledge messages indicating that a given port asks to enter the forwarding mode.

As RSTP continuously monitors the network to detect any changes in network configuration (as in a link-state algorithm), it can detect changes in the network topology in a fast way. Also, unlike in STP, in RSTP, any switch can respond to the BPDUs received from the direction of a root bridge. This, in turn, enables switches to propose a spanning tree by sending the details of the suggested tree via their designated ports. Such a strategy of a rapid transition to the proposed variant of a spanning tree can visibly accelerate the entire convergence procedure.

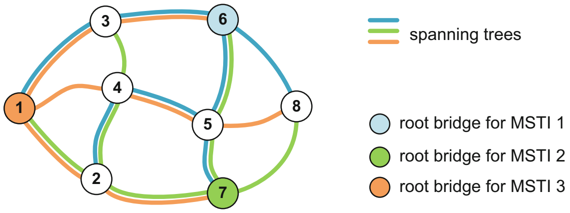
Among several modifications of the RSTP protocol available in the literature, it is worth mentioning the following schemes:

– The strategy from [25] of Fast Spanning Tree Reconfiguration (FSTR) by means of executing an offline ILP program to identify for a set of predefined failure scenarios the best sets of links that could be added to the spanning tree (called reconnect links). As the preconfiguration of these reconnect links is done in advance (prior to failures), recovery time can be visibly reduced.
– The scheme for a spanning tree recovery after a simultaneous failure of two links from [26] utilizing a similar idea of adding links to the spanning tree as in the FSTR scheme, however, here aimed at avoiding loops in scenarios of failures of two links.
– An extension of the FSTR scheme provided in [28] that assumes protection of only those flows that require protection by triggering the recovery of a tree only with respect to failures of a certain subset of links. This is indeed a reasonable assumption since not all flows require full protection.
– The update of the spanning tree by reusing parts of the former spanning tree not affected by the failure [19]. In the event of a link failure, if that failed link belongs to the spanning tree, the technique from [19] would replace the failed link with a non-tree link that remains operational. This scheme involves three phases: fault detection, failure propagation (for broadcasting the information about the failure), and reconfiguration. Due to the reactive nature of this mechanism, maintenance of multiple structures of spanning trees (valid in certain failure scenarios) can be avoided.

5.1.3 Multiple Spanning Trees

The *Multiple Spanning Tree Protocol* (*MSTP*) originally proposed in the IEEE 802.1s standard provides an extension/evolution of STP and RSTP protocols. It is particularly useful for *virtual local area networks* (*VLANs*), i.e., isolated broadcast Layer-2 domains. It allows for a parallel operation of multiple instances of spanning trees within the network (also called Multiple Spanning Tree Instances—MSTI) as illustrated in Fig. 5.3.

Fig. 5.3 An example configuration of multiple spanning trees



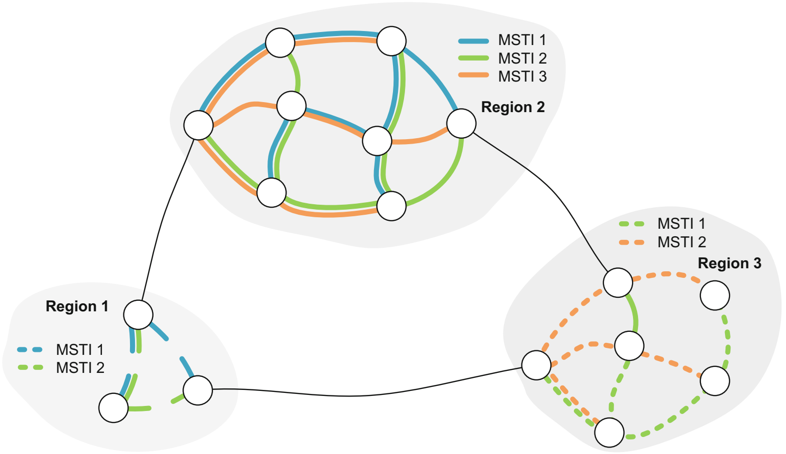


Fig. 5.4 An example configuration of multiple regions of MSTP operation

In MSTP, each spanning tree is assigned a unique VLAN number, which is included in the header of Layer-2 frames. Therefore, in MSTP, frames can be forwarded by switches within a given spanning tree if the VLAN identifier included in the header of the Layer-2 frame matches the VLAN number of a given spanning tree. The existence of multiple spanning trees thus allows Layer-2 frames to follow different paths depending on the value of the VLAN identifier stored in the frame header.

The possibility to set up multiple spanning trees within a network also allows for the configuration of multiple regions, where each region can be served by its own subset of spanning trees, as given in Fig. 5.4. These regions (together with other switches and local area networks) can be, in turn, connected by a single *Common Spanning Tree* (*CST*) and *Common and Internal Spanning Tree* (*CIST*) for connectivity among MST regions and other STP and RSTP LANs in a way to avoid forwarding loops on a global scale (i.e., beyond the reach of particular segments).

Similar to RSTP, MSTP also uses the concept of alternate ports and backup ports for fast restoration of end-to-end connectivity of network nodes in the case of failures affecting spanning trees. As discussed in [7], the fastest recovery can be achieved by substituting the root ports with the respective alternate ports. The explanation for this is that in failure scenarios, switches located farther away from the failed element will not experience a network topology change. Otherwise, if the alternate port is not activated on time, the MSTP would trigger a conventional procedure for re-establishing spanning trees.

Among several alternatives/extensions to MSTP available in the literature, as noticed in [7], the following ones are worth mentioning:

- The Viking scheme from [29], which, contrary to MSTP, allows spanning trees to cover the network topology instead of being confined to particular network segments.
- A scheme involving the deployment of alternate trees configured before a failure occurrence from [24]. In the event of a failure, data transmission is switched onto a backup tree at a local node located upstream of the failed element. Two respective variants of this scheme were proposed in [24], namely the connection-based (where switching the traffic onto a given backup spanning tree depends on the source node, destination node, and the original VLAN ID of frames) and the destination-based (where the backup spanning tree for given frames is determined based on only the destination node, and the original VLAN ID of these frames, i.e., regardless of their source node).

5.2 Mechanisms of Fast Recovery in IP Networks

In this section, we discuss mechanisms of recovery designed for the Layer 3 (the network layer) of the Internet protocol stack. The concepts covered here are thus suitable for operation in IPv4 and IPv6 environments. However, as Layer 3 offers connectionless best-effort data transmission services, the implementation of fast recovery mechanisms based on preplanned protection (with backup paths established before failure) is hardly possible. In fact, ensuring a certain level of service quality in IP networks is already difficult in the normal (operational) scenario and becomes even more challenging in scenarios of failures since the connectionless behavior of the IP network layer does not allow for the association of packets with certain alternate paths before the failure occurrence. Therefore, without additional mechanisms of resilience deployed, it is common for IP datagrams to be served in a best-effort manner by means of backup detours determined reactively [7].

Despite these difficulties, there are several data plane mechanisms available for IP networks, which are designed to make the best use of the properties of link-state routing algorithms to recover the affected traffic as fast as possible in the IP domain by focusing on the adoption of preplanned local detours. These mechanisms of fast recovery in IP networks, often called *IP Fast-Reroute* (shortly, *IPFRR*), described in this section are designed to operate on top of unicast connectionless IP data plane service and typically require at most minimal updates (extensions) of the original IP specification [7, 8]. Notable examples include approaches based on shortest path rerouting, such as Loop-Free Alternates, Remote Loop-Free Alternates, Not-Via addresses, or Failure Insensitive Routing.

The shortest path rerouting schemes discussed here extend the operation of common link-state routing algorithms (e.g., Open Shortest Path First—OSPF [21] or Intermediate System-to-Intermediate System —IS-IS [6]). Link-state routing is, by default, based on a flooding mechanism used to periodically disseminate the actual information on the network topology to all routers in the network and, based on that, to recalculate the related primary paths by all routers. The shortest path rerouting schemes extend these schemes by also calculating one or more backup paths configured before the failure event at routers as the secondary next hops. Therefore, when a failure occurs, backup paths are already available and can be immediately used as bypasses for the affected flows.

Loop-Free Alternates (LFA)
LFA [1] is one of the simplest techniques focused on the deployment of repair paths (i.e., backup paths providing local detours over the failed element). The alternate paths are computed in a way to avoid loops (i.e., scenarios when the secondary hops, being not aware of the failure, are looping back packets to the router that initiated the switchover).

In order to ensure that the computed routes are loop-free, LFA verifies the fulfillment of a set of conditions given by formulas (5.1)-(5.4). In particular, for a given node s and a next hop e of node s on the shortest path toward t, assuming that $\text{dist}(i, j)$ is the shortest path distance between i and j, any node $n \neq e$ is classified as:

– An *ECMP alternate* if

$$\text{dist}(s, n) + \text{dist}(n, d) = \text{dist}(s, d) \tag{5.1}$$

– A *downstream neighbor LFA* if

$$\text{dist}(n, d) < \text{dist}(s, d) \tag{5.2}$$

– A *node-protecting LFA* if

$$\text{dist}(n, d) < \text{dist}(n, e) + \text{dist}(e, d) \tag{5.3}$$

– A *link-protecting LFA* if

$$\text{dist}(n, d) < \text{dist}(n, s) + \text{dist}(s, d) \tag{5.4}$$

These equations are ordered descending their coverage, i.e., any equal cost multipath (ECMP) alternate router is always a downstream neighbor LFA. Every downstream neighbor LFA is always a node-protecting LFA, and every node-protecting LFA is always a link-protecting LFA [7]. As discussed in [7], during the operation of LFA, when deciding about the alternate next hop, a stronger property is always preferred.

The major shortcomings of LFA are as follows:

1. By allowing only local detours, LFA can provide protection in the case of about 80% of single link failures and 40–50% of node failures due to topological constraints.
2. During the time of recovery, loops may be encountered when not all routers have a consistent view of the failure scenario.
3. To verify conditions given by formulas (5.1)–(5.4), additional execution of Dijkstra's algorithm is needed to determine distances dist(i, j).

Remote Loop-Free Alternates (rLFA)

The rLFA was proposed in [9] as an extension of the LFA scheme to improve the ratio of failure scenarios (over the result provided by the LFA) successfully covered by backup paths. For this purpose, compared to LFA, the scope of rLFA is extended to multi-hop backup paths [7]. In rLFA, any remote router is also allowed to become an alternate router if the three following conditions are met:

1. The originating router is able to perform packet tunneling from itself to that alternate router.
2. The shortest path between a pair of the originating router and the alternate router does not include the failed element.
3. For the remote router, there is a valid path to the destination node available in a considered failure scenario.

The failure coverage of rLFA backup paths, although higher than for LFA, may still not be able to reach 100%.

Not-Via

The *Not-Via* approach [5] was proposed to overcome the problem of limited coverage of failure scenario characteristic of LFA and rLFA schemes. In particular, in LFA and rLFA schemes, although for a given primary path node, there exists a suitable router, which could serve as a proper alternate next hop, all shortest paths to the candidate next hops may actually converge to a given failed next hop.

To avoid this scenario, Not-Via uses explicit signaling for advertising exclusions of certain failed elements when disseminating the reachability information. For example, as illustrated in Fig 5.5, a given router A having two interfaces a_1, a_2 and being aware of a failure of router C disseminates its reachability information, however, not via router C, following the recognition of a failure of node C. Any router receiving such advertisements will update all backup paths heading toward node A in a way that they omit the failed node C.

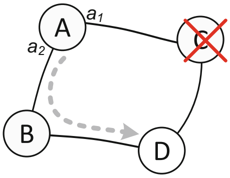


Fig. 5.5 Dissemination of explicit notifications on excluded next hops following a failure of router C

The traffic to be sent along a given backup path from a given source router toward a given destination router via router A needs to be tunneled between that source router and router A to make sure that any transit router between the source router and router A (e.g., router B in Fig. 5.5) will not forward the traffic back to the source router.

Failure Insensitive Routing (FIR)
FIR is able to provide full protection in scenarios of single link failures. Similar to Not-Via, FIR is also able to exclude failed elements from communication paths. However, contrary to Not-Via (which excludes certain elements by means of explicit Not-Via addresses explicitly communicated across the network), such exclusions are applied by routers in FIR by deducing them based on the way packets arrive at these routers. For instance, if certain packets from a given source router arrive through a nontypical interface of a given router (i.e., the one that would never be in use for that purpose in a normal scenario), a set of potentially failed links that may cause such behavior of packets can be identified. Such inferred information is next used by routers to redirect packets to other next hops.

A drawback of FIR is that full coverage in scenarios of all single link failures requires updates of the conventional IP data plane. This is because decisions on packet forwarding are based not only on the destination addresses but also on the interface of a given intermediate router they arrived at.

5.3 IP-MPLS Mechanisms for Fast Recovery

The architecture of *Multiprotocol Label Switching* (*MPLS*) [27] was introduced to assure a certain level of QoS in IP networks by default offering the best-effort services only. In MPLS, packets are forwarded across the network based on 20-bit *labels* contained in the MPLS packet header between the headers of Layer 2 and Layer 3 as given in Fig. 5.6.

IP-MPLS networks are formed by *label switch routers* (*LSRs*), a subset of which localized at the border of the system is referred to as *label edge routers* (*LERs*) [31]. Contrary to conventional IP networks, packet processing at transit nodes is not based on the longest prefix matching but is solely determined by the values of the mentioned labels. These labels are assigned to packets by edge routers (i.e., when entering the IP-MPLS network) based on several parameters related to the IP destination address, QoS requirements, VPN identifiers, etc. and can be updated later on by transit LSRs.

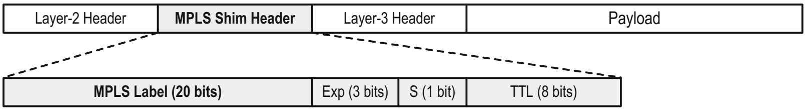


Fig. 5.6 The structure of an MPLS header

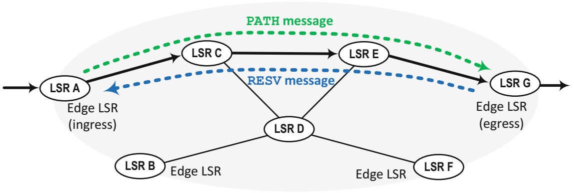


Fig. 5.7 Illustration of a procedure of setting up the LSP

MPLS labels, in fact, assign packets to certain *Forwarding Equivalence Classes* (*FECs*) defined as the groups of packets forwarded by several consecutive LSRs in a consistent manner [7], i.e., following the same path. Indeed, each LSR determines the next hop for a given packet solely based on the packet label following the respective entry from the label forwarding table of that LSR. The utilization of MPLS labels thus makes IP-based systems behave in a way that is closer to connection-oriented systems. Dissemination of information on the association of certain labels with FECs among the LSRs is provided by *Label Distribution Protocol* (*LDP*).

As illustrated in Fig. 5.7, before packets are sent along a given *label switched path* (*LSP*), the path needs to be established between a given pair of ingress and egress LSRs. For this purpose, the respective `PATH` message is first sent from the ingress LSR toward the egress LSR. It is important to note here that the demanded path can also be explicitly included in that `PATH` message. Otherwise, the installation of a path is determined by the `RESV` message sent back from the egress LSR to the ingress LSR via the sequence of transit LSRs in reverse order to the one for the `PATH` message. The `RESV` message also includes the label assigned to that path by the egress LSR. While forwarding the `RESV` message, the transit LSRs also reserve the necessary resources for the path. The reception of the `RESV` message by the ingress node completes the procedure of setting up the LSP.

Failures of LSRs or MPLS links may undoubtedly affect the label switched paths. Among various resilience schemes, we can distinguish the proactive ones using preestablished dedicated or shared backup LSPs, as well as reactive approaches where backup LSPs are determined only after the occurrence of a failure. Selected techniques belonging to these two classes are discussed in the remaining part of this section. However, as noted in [7], only proactive schemes are able to ensure fast recovery of the affected working LSPs.

5.3.1 Proactive Schemes of Resilient Routing in MPLS Networks

Mechanisms of fast recovery in MPLS networks typically involve local protection schemes, where backup LSPs provide local detours over the failed transit links or nodes of a working LSP. Such local protection techniques are commonly called **Fast**

Reroute schemes. As discussed in [7], they can be classified into *one-to-one backup* and *facility backup* schemes illustrated in Fig. 5.8.

In one-to-one backup schemes, a given backup LSP is designed to protect only a given working LSP. The facility backup approach, in turn, allows a single backup LSP to protect a set of working LSPs traversing the same sequence of MPLS links.

As both classes, in fact, imply local recovery operations, in the event of a failure, one of the end nodes of the backup LSP located closest to the failed element (called *Point of Local Repair—PLR*) redirects the traffic from the affected working LSP onto the backup LSP. Both types of schemes are considered by fast recovery procedures of the RSVP-TE (Resource Reservation Protocol-Traffic Engineering) solution from [23] commonly used in practice in MPLS networks.

Resilience schemes in IP-MPLS networks are undoubtedly resource demanding due to the need for reservation of link capacity also for backup LSPs. However, as these backup LSPs often remain unused in normal (i.e., non-failure) periods, techniques of **backup LSP sharing** can help lower the total cost of backup LSP installation. As discussed in several papers on fast reroute covering this aspect (see, e.g., [4, 30]), a set of several backup LSPs can share resources at a given link as long as the corresponding working LSPs are guaranteed not to fail at the same time. This, in turn, is assured by a mutual disjointness of these working LSPs as illustrated in Fig. 5.9.

Apart from solutions based on local detours, fast redirection onto the backup LSPs can be achieved by some of the global protection schemes with the redirection of the affected flow made by the LSP located close to the failed network element. Such an idea of **local redirection** is utilized, e.g., in the *local-to-egress protection* from [13, 14] involving a backup LSP configured in the reverse direction from the last-hop working LSP node toward the working LSP source node and next back to the destination node of a working LSP via a path being node-disjoint with the related working LSP, as shown in Fig. 5.10.

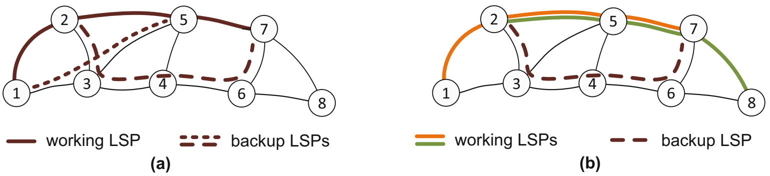


Fig. 5.8 Illustration of (**a**) one-to-one local protection method where each node of the primary LSP is protected by its own backup LSP and (**b**) facility backup scheme involving the use of one backup LSP to protect a certain joint segment of several working LSPs

Fig. 5.9 Illustration of a possibility for sharing the resources of backup LSPs at a link between nodes 4 and 5

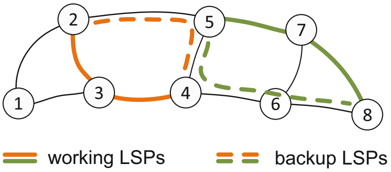


Fig. 5.10 Illustration of the local-to-egress configuration of a backup LSP

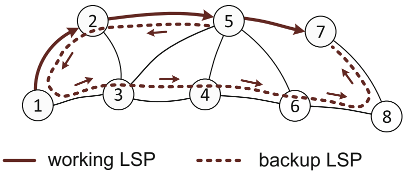
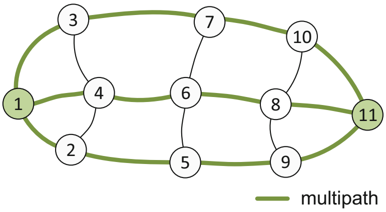


Fig. 5.11 An example configuration of a self-protecting multipath (SPM) between LER 1 and 11



In this scheme, after a network node/link traversed by the working LSP fails, switching the traffic onto the backup LSP is done at the working LSP node adjacent upstream to the failed element. As this operation does not involve any multi-hop recovery signaling, it is, therefore, fast. Also, only a single backup LSP needs to be set up for a given working LSP.

As soon as the upstream (source) LSR recognizes the backward flow, it marks the last packet sent along the affected primary LSP and stores the subsequent packets in its queue to avoid packet reordering. These queued packets are next released by that LSR from that queue and forwarded along the backup LSP (right after receiving the marked packet again from the downstream LSR and forwarding it along the backup LSP).

Apart from mechanisms involving the use of one working LSP for transmission for a certain FEC between a given pair of end nodes in the system, there are also schemes available that involve a set of disjoint paths utilized in parallel in a normal state. For example, the scheme of *self-protecting multipaths* (*SPMs*) from [20] uses a set of k preestablished mutually node-disjoint multipaths for data transmission between a given pair of end nodes, as illustrated in Fig 5.11.

In the event of a failure affecting, e.g., one of these paths, as the remaining k-1 paths continue their operation, the flow from the affected path is redistributed onto all other (operational) paths. Such a switchover can be indeed fast since there is no need for setting up any new path after a failure. An additional advantage of the SPM scheme is its ability to ensure adequate load distribution across the network.

5.3.2 *Reactive Approaches to Resilient Routing in MPLS Networks*

In the case of using the reactive schemes for resilience in MPLS networks, the determination of backup LSPs for all affected working LSPs is triggered after the occurrence of a failure. Therefore, compared to proactive schemes, the overall time needed to switch the affected traffic onto the alternate paths is extended by the time to determine the backup LSP [2], which commonly denotes the time needed for the delivery of the `PATH` message to the egress LSR and the time to send the related `RESV` message back to the ingress LSR. Compared to protection schemes, the improved capacity efficiency characteristic of reactive schemes comes at a price of increased recovery time.

Therefore, as requirements on service availability are often differentiated for different demands, as discussed in [3], reactive recovery schemes involving rerouting of the affected traffic seem to be proper for services for which the acceptable recovery time is between 100 ms and 10 s. In [3], such services are identified as belonging to classes RC2 and RC3 with medium and low resilience requirements, followed only by the "best-effort" service class RC4, for which the recovery time upper limit is unspecified, and, therefore, no specific resilience mechanism is assumed. Any service requiring the recovery time to be lower than 100 ms (class RC1), in turn, calls for the use of preconfigured backup LSPs discussed in Sect. 5.3.1.

The validity of using reactive recovery methods in class-based resilience approaches is also confirmed in several other works, including, e.g., [10] introducing a proposal of a differentiated resilience scheme for serving anycast flows in MPLS networks in a way to survive failures of single links and failures of single replica servers. The three considered classes include Class 1 with working LSPs protected by the preestablished dedicated backup LSPs, each backup LSP leading to another replica server than the corresponding working LSP, Class 2 with working LSPs protected by the preestablished shared backup LSPs, and Class 3 with backup paths determined reactively after the occurrence of a failure using the free capacity of links available after a failure. The results of performance evaluation presented in [10] confirm that apart from the resource efficiency of the reactive recovery, in such a class-based approach, the existence of Class 3 (with no backup paths installed in advance) allows for reducing the blocking probability for demands from higher service classes.

5.4 Summary

In this chapter, we discussed the properties of mechanisms for the resilient operation of packet-switched systems. Our analysis focused on IP networks, particularly the resilience of IP Layer-2 Ethernet mechanisms, IP Layer-3 routing, and IP-MPLS switching. As Layer-2 frames do not include fields similar to the Layer-3 Time-to-Live (TTL) to prevent forwarding loops in failure scenarios, in this chapter, we highlighted the properties of selected spanning tree algorithms designed for fast recovery of spanning trees affected by failures. In the middle part of this chapter, we discussed major schemes of IP fast reroute, namely, LFA, rLFA, Not-Via, and FIR, to enable fast and loop-free recovery of the affected transmission routes using local detours. Despite operating in a connectionless manner, these mechanisms can indeed be efficient in restoring the affected traffic, as they focus on adopting preplanned local detours determined proactively by link-state routing algorithms. The IP-MPLS recovery mechanisms described in the final part of this chapter can also operate efficiently in failure scenarios, mainly if their proactive variants are deployed.

? Questions

1. Explain the properties and the operation of the STP protocol.
2. Characterize the differences between the operation of the RSTP and the STP protocols.
3. Discuss the scenarios for the use of the MSTP protocol.
4. Explain the challenges in ensuring fast recovery in IP networks.
5. Describe the main features of the LFA technique.
6. Explain the recovery-related advantages of the rLFA scheme over the LFA approach.
7. Characterize the main features of the Not-Via scheme concerning the failure recovery aspects.
8. Discuss the difference between the FIR and the Not-Via scheme.
9. Discuss the main features of proactive mechanisms supporting the resilient operation of IP-MPLS networks.
10. Explain the operation of selected reactive mechanisms of resilience for IP-MPLS networks.

References

1. Atlas, A., Zinin, A.: Basic specification for IP fast reroute: Loop-free alternates (2008). https://tools.ietf.org/html/rfc5286
2. Autenrieth, A.: Recovery time analysis of differentiated resilience in MPLS. In: Proceedings of the 4th International Workshop on Design of Reliable Communication Networks (DRCN'03), pp. 333–340 (2003)
3. Autenrieth, A., Kirstadter, A.: Engineering end-to-end IP resilience using resilience-differentiated QoS. IEEE Commun. Mag. **40**(1), 50–57 (2002)
4. Alicherry, M., Bhatia, R.: Simple pre-provisioning scheme to enable fast restoration. IEEE/ACM Trans. Netw. **15**(2), 400–412 (2007)
5. Bryant, S., Previdi, S., Shand, M.: A framework for IP and MPLS fast reroute using Not-Via addresses. RFC6981 (2013). https://tools.ietf.org/html/rfc6981.
6. Callon, R.: Use of OSI IS-IS for Routing in TCP/IP and Dual Environments. RFC 1195 (1990). https://www.ietf.org/rfc/rfc1195.txt
7. Chiesa, M., Kamisinski, A., Rak, J., Retvari, G., Schmid, S.: A survey of fast-recovery mechanisms in packet-switched networks. IEEE Commun. Surv. Tutorials **23**(2), 1253–1301 (2021)

8. Cicic, T, Hansen, A.F., Apeland, O.K.: Redundant trees for fast IP recovery. In: Proceedings of the 2007 Fourth International Conference on Broadband Communications, Networks and Systems (BROADNETS'07), pp. 152–159 (2007)
9. Csikor, L., Retvari, G,: IP fast reroute with remote Loop-Free Alternates: the unit link cost case. In: Proceedings of the 2012 International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT'12), pp. 663–669 (2012)
10. El-Gorashi, T.E.H., Elmirghani, J.M.H.: Differentiated resilience for anycast flows in MPLS networks. In: Proceedings of the 2009 11th International Conference on Transparent Optical Networks (ICTON'09), pp. 1–5 (2009)
11. Elmeleegy, K., Cox, A.L., Ng, T.S.E. : On count-to-infinity induced forwarding loops in Ethernet networks. In: Proceedings of the IEEE INFOCOM 25th IEEE International Conference on Computer Communications (IEEE ICCC'06), pp. 1–13 (2006)
12. Francois, P., Filsfils, C., Evans, J., Bonaventure, O.: Achieving sub-second IGP convergence in large IP networks. SIGCOMM Comput. Commun. Rev. **35**(3) 35–44 (2005)
13. Haskin, D.L., Krishnan, R.: A method for setting an alternative label switched paths to handle fast reroute (2000). https://datatracker.ietf.org/doc/html/draft-haskin-mpls-fast-reroute-05
14. Hundessa, L., Domingo-Pascual, J.: Reliable and fast rerouting mechanism for a protected label switched path. In: Proceedings of the 2022 IEEE Global Telecommunications Conference (GLOBECOM'02), pp. 1608–1612 (2002)
15. IEEE: IEEE Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges, pp. 1–281 (2004)
16. IEEE: IEEE 802.1Q-2014 IEEE Standard for Local and metropolitan area networks–Bridges and Bridged Networks (2014)
17. IEEE: 802.1s-2002, Amendment to 802.1Q Virtual Bridged Local Area Networks: Multiple Spanning Trees (2002)
18. IEEE: 802.1w-2001, Part 3: Media Access Control (MAC) Bridges: Amendment 2–Rapid Reconfiguration, pp. 1–116 (2001)
19. Jin, D., Chen, W., Xiao, Z., Zeng, L.: Single link switching mechanism for fast recovery in tree-based recovery schemes. In: 2008 International Conference on Telecommunications (ICT'08), pp. 1–5 (2008)
20. Menth, M., Reifert, A., Milbrandt, J.: Self-Protecting Multipaths—A simple and resource-efficient protection switching mechanism for MPLS networks. Lecture Notes in Computer Science book series, vol. 3042, pp. 526–537 (2004)
21. Moy, J.: OSPF Version 2. RFC 2328 (1998). https://www.ietf.org/rfc/rfc2328.txt
22. Pallos, R., Farkas, J., Moldovan, I., Lukovszki, C.: Performance of rapid spanning tree protocol in access and metro networks. In: Proceedings of the 2007 Second International Conference on Access Networks & Workshops, pp. 1–8 (2007)
23. Pan, P., Swallow, G., Atlas, A.: RFC4090 - Fast reroute extensions to RSVP-TE for LSP tunnels (2005). https://tools.ietf.org/html/rfc4090.
24. Qiu, J., Gurusamy, M., Chua, K.C., Liu, Y.: Local restoration with multiple spanning trees in metro Ethernet networks. IEEE/ACM Trans. Netw. **19**2, 602–614 (2011)
25. Qiu, J., Liu, Y., Mohan, G., Chua, K.C.: Fast spanning tree reconnection for resilient Metro Ethernet networks. In: Proceedings of the 2009 IEEE International Conference on Communications, pp. 1–5 (2009)
26. Qiu, J., Mohan, G., Chua, K.C., Liu, Y.: Handling double-link failures in metro Ethernet networks using fast spanning tree reconnection. In: Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM'09), pp. 1–6 (2009)
27. Rosen, E., Viswanathan, A., Callon, R.: Multiprotocol label switching architecture (2001). https://tools.ietf.org/html/rfc3031
28. Shan, D.M., Chiang, C.K., Mohan, G., Qiu, J.: Partial spatial protection for differentiated reliability in FSTR-based metro Ethernet networks. In: Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM'11), pp. 1–5 (2011)

29. Sharma, S., Gopalan, K., Nanda, S., Chiueh, T.: Viking: A multi-spanning-tree Ethernet architecture for metropolitan area and cluster networks. In: Proceedings of IEEE INFOCOM'04, vol. 4, pp. 2283–2294 (2004)
30. Wang, D., Li, G.: Efficient distributed bandwidth management for MPLS fast reroute. IEEE/ACM Trans. Netw. **16**(2), 486–495 (2008)
31. Xin, Ch., Ye, Y., Dixit, S., Qiao, Ch.: An agent-based traffic grooming and management mechanism for IP over optical networks. In: Proceedings of the 11th International Conference on Computer Communications and Networks (IEEE ICCC'02), pp. 425–430 (2002)