Chapter 4
Strategies and Concepts for Resilient Routing in Circuit-Switched Networked Systems

Providing communication possibilities to users, apart from offering storage and computation services, is one of the major aspects of any networked system. Since the locations of end users and servers in a network are characterized by a significant degree of geographical diversity, the transmission of information in these systems, both between end users themselves and between users and servers, is most often carried out via multi-hop paths, i.e., paths traversing many transit elements such as network links and nodes (routers, optical switches, etc.). Transmission paths are inevitably affected by failures of network elements traversed by these paths. Therefore, to maintain the continuity of transmission in failure scenarios, networked systems need to utilize the reconfiguration mechanisms of communication paths affected in a given failure scenario to bypass the failed network elements.

The objective of this chapter is thus to focus on mechanisms for resilient multi-hop communications able to remain operational in differentiated failure scenarios. A particular focus of this chapter is on approaches dedicated to circuit-switched wired networks providing transmission services on a per-flow rather than on a per-packet basis. Resilience mechanisms for packet-switched networks are the main topic of the next chapter of this book.

In this book, we define *resilient routing* as a routing scheme that can provide continuity of service in the presence of disruptions.

To maintain service continuity after failures, spare capacity (mostly related to link bandwidth) is commonly reserved in the network to provide the possibility to reroute the traffic along the *backup path* (also called *alternate path* or *protection path*) when the *primary* (*working*) *path* fails [19].

J. Rak, *Resilient Routing in Communication Networks*, Computer Communications and Networks, https://doi.org/10.1007/978-3-031-64657-7_4

In general, a given multi-hop path in a circuit-switched network is established as a response to a given demand d_r defined as a triple (s_r, t_r, c_r) to provide a connection between nodes s_r and t_r of a guaranteed capacity c_r. Since this capacity is meant to be guaranteed for demand d_r also after a failure affecting its primary path, in this chapter, both working and backup paths of each demand d_r are assumed to be assigned capacity c_r at all consecutive links of these paths. Therefore, the greater the capacity to be protected, the more significant the task to protect the network from failures.

Following [7, 44], and as previously mentioned in Chapter 1, after the occurrence of a failure, the recovery process is initiated with the detection of a failure. It can be recognized, e.g., by IP-MPLS mechanisms like `MPLS LSP ping` or `MPLS LSP traceroute` [25] (sent along Label Switched Paths—LSPs), which are, however, time-consuming. Another option is detecting the failure based on the `Loss of Light` or `Loss of Clock` events.

Fault detection should be followed by fault localization and isolation (i.e., determination of the faulty node/link), which is necessary to stop further transmission of information via the affected element [7]. Fault notification messages are next sent to network nodes responsible for further triggering the recovery switching to redirect the affected flows onto the related backup paths.

After the physical repair of a faulty element, the final stage is normalization, i.e., recognition of the repaired element and return to the normal operational state. Concerning routing, this would generally mean a return to transmission paths that were in use before the failure (since recovery paths are typically nonoptimal, e.g., concerning resource usage or end-to-end delay).

The ideal *recovery time* (i.e., the time from the occurrence of a failure until redirection of the affected traffic onto backup paths) should not be greater than 50 ms since the higher layers often see a disruption lasting up to 50 ms as a transmission error only. Any disruption longer than 50 ms may result at least in packet losses or unavailability of service [41]. A detailed classification of the duration of outages from [15] is given in Table 4.1.

Although utilizing protection paths to provide automatic switchover seems relatively intuitive, implementing efficient recovery schemes, being both capacity-efficient and scalable, and including multiple criteria of QoS, especially in heterogeneous mesh network environments, is difficult.

In general, characteristics of any recovery method strongly influence the values of service recovery time [7]. In the later part of this section, we will highlight the most crucial recovery techniques, focusing on restoration time characteristics and their relation with the resource efficiency objective.

In this chapter, we first outline in Sect. 4.1 the architectural properties of ring networks and describe the related resilience mechanisms in detail. The latter part of this chapter, in turn, highlights the major schemes of resilient routing in mesh networks—the most common configuration of today's communication systems. In particular, in Sect. 4.2, we explain the need to ensure the differentiated levels of resilience to match the differentiated requirements of services. The objective of

Table 4.1 Impacts of outage time from [15]

Target range	Duration	Main effects
Protection switching	≤50 ms	No outage logged; recovery of transmission control protocol (TCP) after one errored frame; no TCP fallback; no impact at all for most TCP sessions
1st type outage	>50 ms ≤200 ms	<5% voiceband disconnects; signaling system switchovers
2nd type outage	>200 ms ≤2 s	Common upper bound on distributed mesh restoration time; TCP/IP protocol back-off
3rd type outage	>2 s ≤10 s	Disconnections of all switched circuit services; disconnections of private lines; TCP sessions time-outs; Hello protocol affection; web page "not available" errors
4th type outage	>10 s ≤5 min	All calls and data sessions terminated; time-outs of TCP/IP application layer programs; users making attempts of mass redial; link-state advertisements (LSAs) sent by routers referring to failed links; updates of topology and resynchronization network-wide
Undesirable outage	>5 min ≤30 min	Massive reattempts causing heavy load of switches; noticeable Internet "brownout"; minor societal/business effects
Unacceptable outage	>30 min	Major societal impacts (societal risks: travel booking, impact on all markets); headline news; regulatory reporting often required; lawsuits; SLA clauses triggered

Sect. 4.3 is to provide a taxonomy for schemes of resilient routing in mesh networks according to the following main criteria: backup path setup method, failure model, scope of recovery procedure, usage of recovery resources, as well as the application of recovery schemes to multidomain and multilayer architectures of networked systems. The following two sections (Sects. 4.4 and 4.5) elaborate on the efficiency of recovery schemes in the two common architectures of communication networks, namely optical transport networks (OTNs) and IP networks. The summary of the chapter is provided in Sect. 4.6.

4.1 Resilient Routing in Ring Networks

A fundamental classification of resilience mechanisms based on the structure of communication networks divides the existing approaches into ring- and mesh-based. The former refers to architectures common about three decades ago, such as Synchronous Optical Networks/Synchronous Digital Hierarchy (SONET/SDH) [46] and early architectures of ring Dense Wavelength Division Multiplexing (DWDM) networks [31].

Based on flow direction, *ring networks* may be classified as unidirectional (referred to as *unidirectional path switched rings*—UPSR) or bidirectional (i.e., *bidirectional line switched rings*—BLSR), respectively. These networks consist

of *add/drop multiplexers* (ADMs) interconnected by fiber links, each fiber link providing transmission in parallel via its multiple nonoverlapping channels, each channel represented by a given wavelength λ_i [42]. The role of each ADM is to add (and drop) a certain subset of wavelengths from a given optical signal (by performing the multiplexing/demultiplexing operations) while allowing the other wavelengths to pass through the ADM.

Common variants of SONET ring networks include 2-fiber UPSR, 2-fiber BLSR (BLSR/2), and 4-fiber BLSR (BLSR/4). As shown in Fig. 4.1, both working and backup paths in ring networks are organized in rings.

In particular, as shown in Fig. 4.1a, each UPSR includes one ring for working paths and another for protection paths, configured to operate in opposite directions. In normal operational conditions, transmission in UPSRs is duplicated on both working and protection rings. The destination transmission node receives data by choosing between two signals, the one of better quality. In a failure scenario, a backup ring is used for detour purposes. Since protection rings are used in UPSRs simultaneously with working rings in a normal scenario, UPSRs are commonly considered an example of 1+1 Automatic Protection Switching (1+1 APS) [23].

A BLSR/2 structure shown in Fig. 4.1b involves two rings used simultaneously in a non-failure scenario for working paths operating bidirectionally (in opposite directions). It is important to note that in BLSR/2 only half of the capacity of each ring is used for working paths, while the other half is reserved for backup paths. The BLSR/4 ring structure illustrated in Fig. 4.1c involving four fibers between each pair of neighboring nodes consists of four rings: two rings for working paths operating in opposite directions and the other two rings configured similarly for backup paths.

Since link resources of a given ring reserved for protection paths in failure scenarios often serve low-priority traffic under normal conditions (and are preempted in failure scenarios), structures such as BLSRs can be considered as a particular form of 1:1 Automatic Protection Switching (1:1 APS) [23].

Due to differences in capacity efficiency of the considered ring systems, UPSRs gained popularity in local access networks, while BLSR configurations became important in metropolitan networks; however, they are both characterized by a relatively low level of available capacity, compared, e.g., to DWDM systems.

Ring networks are often called *self-healing rings*—SHRs. Thus, the considered variants are frequently referred to as USHR, BSHR/2, and BSHR/4, respectively [18].

In a scenario of a failure of a network element, e.g., a network link, as illustrated in Fig. 4.2, the respective detour over the failed link is formed, meaning that the traffic is switched at the node adjacent to the failed link onto the respective detour in the reverse direction. For the UPSR architecture, the non-affected parts of working and protection rings are merged to form a single ring, as shown in Fig. 4.2a. The same effect refers to the BLSR/2 architecture; however, its operational capacity is reduced by a factor of 2, as only one working ring (instead of the former two working rings) remains operational, as given in Fig. 4.2b. For the BLSR/4

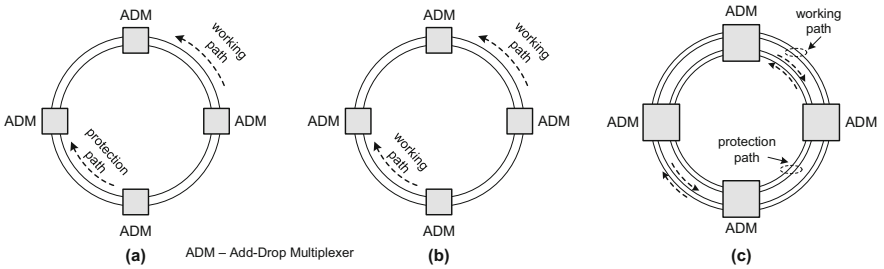

Fig. 4.1 Example of unidirectional path-switched ring (UPSR), 2-fiber bidirectional line switched ring (BLSR/2) and 4-fiber bidirectional line switched ring (BLSR/4) with add-drop multiplexers (ADMs)

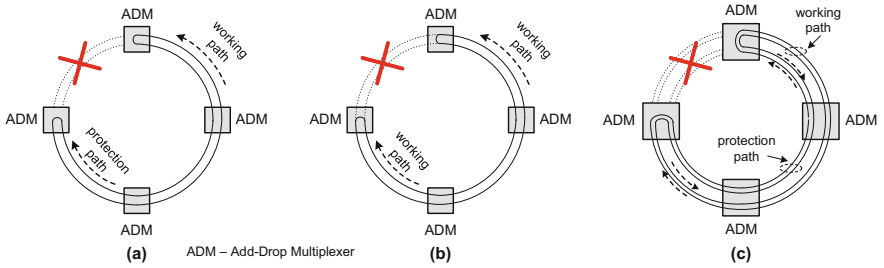

Fig. 4.2 Example operation of (**a**) UPSR, (**b**) BLSR/2, and (**c**) BLSR/4 in a scenario of a link failure

architecture, the respective working and protection rings are merged and form two operating rings after completing the recovery procedure, as shown in Fig. 4.2c.

Backup rings can thus be viewed as a preplanned protection scheme providing a short recovery switching time. However, their disadvantage is the high ratio of *network redundancy* (being the ratio of protection capacity to working capacity) of exactly 100% [19] in scenarios where every working ring is accompanied by the respective duplicate protection ring.

Architectures of ring networks commonly consist of a set of rings. The respective multi-hop transmission paths then often traverse a sequence of rings. In particular, in the case of a normal (i.e., non-failure scenario), transmission is provided using working rings. For instance, in a network consisting of two rings shown in Fig. 4.3a, the transmission path between `ADM 2` and `ADM 6` takes place via three transit ADMs: `ADM 4`, `ADM 5`, and `ADM 7`. However, in the case of a failure, e.g., a failure of a link between `ADM 6` and `ADM 7`, as given in Fig. 4.3b, the respective backup rings are activated, and transmission is redirected at `ADM 7`. Therefore, after a failure, the transmission path becomes much longer, as traffic is now forwarded five times at `ADM 4`, `ADM 5`, `ADM 7`, `ADM 5`, and `ADM 4`.

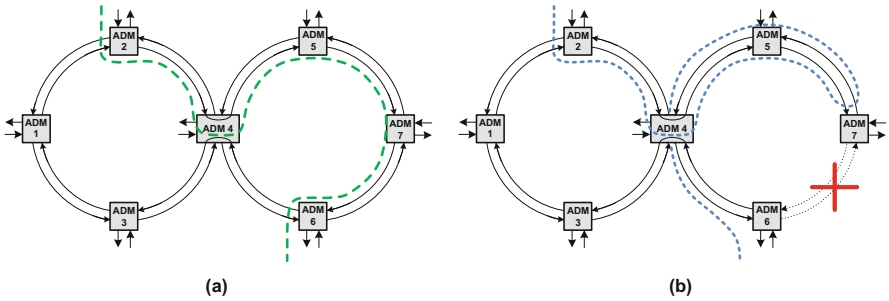

Fig. 4.3 Example multi-hop transmission in a system consisting of two bidirectional line switched rings (BLSR/2) in a normal scenario (**a**), and in the case of a single link failure (**b**)

4.2 The Need for Resilience Differentiation in Mesh Networks

The continuous traffic volume increase has triggered the evolution of ring-based topologies of wide-area optical networks toward mesh structures. Indeed, in wide-area networks where the cost of multi-hop transmission is determined by both the capacity and distance, a mesh topology of a networked system can serve a more significant number of demands compared to the capacity-equivalent ring structures [23].

In contemporary networks often characterized by a mesh topology [17], transmission paths are of end-to-end type, i.e., they do not form ring structures. As opposed to networks from the past engineered to offer a single type of service only (either voice or data), current communication networks are expected to provide a variety of services (e.g., real-time services as well as bulk data transfer) to support a wide range of applications (for example, online healthcare services based on data received from embedded sensor systems, massive content streaming, smart transportation, or emergency services) having differentiated requirements concerning resilience (sometimes referred to as the *quality of resilience* (QoR) [6]), as well as to the quality of transmission, as shown in Fig. 4.4.

This differentiation can also follow from different usage of the same application [6]. In other words, a given application can have differentiated requirements depending on how the users utilize it. For instance, even in the case of a classic telephone service, requirements on service availability for a company would be much higher than those sufficient for a home user.

Designing a communication network that consistently meets the highest requirements over the entire range of services (i.e., prepared to provide the highest level of service) by applying over-provisioning (i.e., adding an excessive amount of capacity, as in the case of optical DWDM backbone networks) would be highly costly and unreasonable. Such over-provisioning is also particularly expensive in wireless and

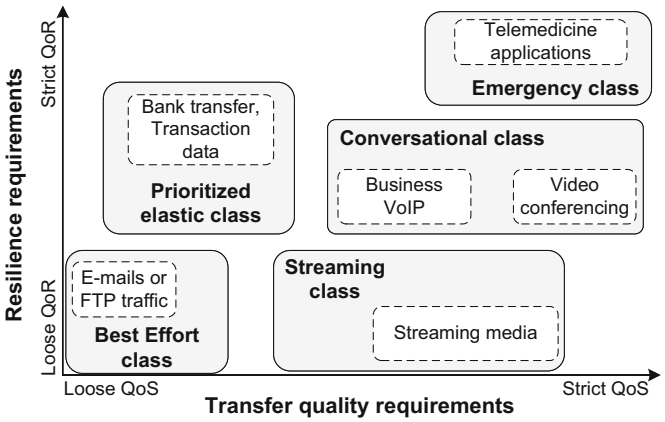

Fig. 4.4 Transfer quality vs. resilience requirements from [48]

access networks where bandwidth is limited (compared, e.g., to optical DWDM long-haul networks) [6].

Therefore, proper *resilience differentiation* (for example, as discussed in [30, 33]) is crucial in client-operator relations as an essential element of Service Level Agreements (SLAs). The operator, interested in maximizing the profit, is looking for cost-efficient resilience mechanisms tailored to specific SLA requirements. The willingness of clients to pay for the service is also differentiated. In particular, clients expect the lowest possible price for the service able to support characteristics of applications, but with only marginal regard to network mechanisms, the operator would deploy to support these applications. Utilizing multiple resilience mechanisms in the network may thus enable clients and operators to increase their profit.

Since applications are indeed characterized by a set of differentiated service requirements, including those related to service resilience, it seems reasonable to group applications into service classes and apply different models of service provisioning (as well as different resilience mechanisms) to these service classes. Indeed, the expectations of applications concerning the resilience requirements, including the level of service availability, continuity, or the maximum length of a service downtime period, vary from application to application, from almost no tolerance for service unavailability (e.g., for real-time telemedicine or financial services), via moderate tolerance of unexpected breaks in service provisioning (see, e.g., video streaming applications accepting slight changes in transmission delays due to the use of buffering) to best-effort service provisioning for the other applications with only marginal requirements on service continuity.

Several research papers also reflect this observation. For instance, in [2], four service classes are proposed based on their tolerance of the time for service downtime after a failure in a network, as summarized in Table 4.2.

Table 4.2 Requirements on service recovery time for resilience classes from [2]

Service class	RC 1	RC 2	RC 3	RC 4
Resilience requirements	High	Medium	Low	None
Recovery time	10–100 ms	100 ms–1 s	1 s–10 s	n.a.

In particular, resilience class 1 (RC1) from Table 4.2 represents high requirements on the maximum recovery time of up to 100 ms. RC2 denotes a class of medium requirements for resilience with a recovery time between 100 ms and 1 s. Low resilience requirements are characteristics of class RC3, tolerating the downtime between 1 and 10 s, while the last class (RC4) refers to the unspecified resilience-related requirements (meaning that any time for service recovery is acceptable for class RC4).

As discussed later in this chapter, different levels of service unavailability tolerance can be translated into the need to deploy different service recovery mechanisms. A general observation is that the time needed for the recovery of services and the resource cost of network resilience solutions are mutually opposing factors, i.e., the lower the acceptable time of service unavailability, the higher amount of extra resources needed (and thus, the more expensive the respective resilience scheme).

4.3 Schemes for Backup Path Resources Reservation in Mesh Networks

This section briefly overviews the most crucial resilience mechanisms proposed in the literature to provide fault-tolerant routing. Resilience differentiation can be obtained by combining several of them in a single network. Figure 4.5 outlines the most important classifications of resilience mechanisms for mesh networks, characterized in detail later in this section.

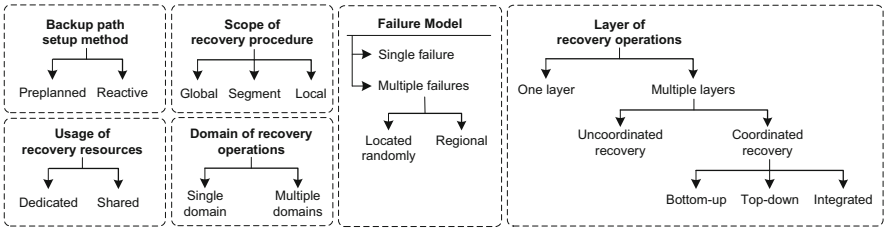

Fig. 4.5 Major classifications of resilience mechanisms in networked systems

4.3.1 Backup Path Setup Method

Concerning methodologies for setting up backup paths, these paths can be:

- Installed in a preplanned way (i.e., in advance when finding the working paths) often referred to as the *preplanned protection* in the literature [18]
- Determined dynamically (reactively) after the occurrence of a failure (known as *reactive restoration*)

The former case, historically derived from *Automatic Protection Switching* (*APS*) schemes [6], enables fast recovery of each failed transmission path (since backup paths are established in advance) [39]. Reactive restoration with its origins in IP networking [7] is, in turn, better in terms of efficiency of network resource utilization (since backup paths are installed here only when necessary, i.e., after a failure, and can reuse link capacities of failed transmission paths) [39]. However, it inherits all the disadvantages of dynamic IP routing, particularly the time-consuming recovery switching, path instabilities, and risk of loop creation. It also does not guarantee recovery due to the unpredictable amount of spare resources available after a failure [8].

In general, to provide 100% of restorability for working data flows, any backup path should not only be characterized by the same capacity as the corresponding working path, but it should also be link-/node-disjoint (i.e., have no common links/transit nodes) with the related working path—Fig. 4.6. The latter requirement is to guarantee that any failure of a link/node affecting the working path will also not disrupt the functioning of the respective backup path [19].

This disjointness is thus to assure that the two considered paths (i.e., working and backup path) of demand do not use resources of network elements belonging to the same *Shared Risk Link Group (SRLG)* defined in [18, 19] as the set of network elements, being either links, nodes, physical devices, or a mix of these, subject to a common risk of failure. Following [19], a given working path is said to be *SRLG-disjoint* with the respective backup path if both paths are not involved in any common SRLG.

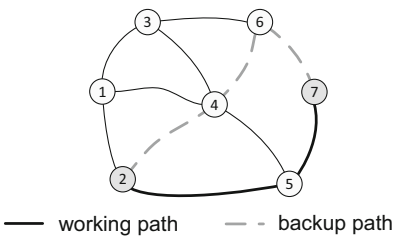

Fig. 4.6 Example of end-to-end node-disjoint pair of paths between nodes 2 and 7

4.3.2 Failure Model

As summarized in [37], failures of network elements may occur due to many reasons, including, e.g., hardware faults, non-malicious human activities, malicious attacks, or natural disasters and disruptions. These events may result in *single failures*, i.e., failures of single network elements (links/nodes) at a time or simultaneous failures of many such elements (referred to as *multiple failures*). The risk for the occurrence of certain types of failures depends on many factors, such as the type of a network (local area network vs. wide-area network), dependability characteristics of system elements, as well as the environmental properties (i.e., location, size, intensity, and frequency) of natural disasters and other weather disruptions determining their impact on networked systems.

In scenarios of failures of single elements of a system (such as failures of single communication links or single nodes), it is sufficient to configure one backup path for a given working path. For instance, as illustrated in Fig. 4.6, a single end-to-end backup path being node-disjoint with the related working path can protect that path against any single node failure. The requirement on nodal disjointness naturally refers to all the nodes of a working path except for its end nodes, as both paths are expected to operate between the same pair of end nodes. If a faulty element is one of the transit nodes of a working path, then redirection of the affected onto the related backup path occurs.

Also, it is worth noting that nodal disjointness is stronger than link disjointness of working and backup paths, as in the latter case (involving a backup path being link-disjoint with the related working path), only protection against failures of single links can be assured. A failure of a node is, in turn, equivalent to a failure of all its incident links.

In failure scenarios not affecting working paths directly (e.g., scenarios of a failure of a transit element of a backup path or failures of any other element not traversed by either of these two paths), no recovery switching operation is needed. It is worth noting here that a single backup path can protect more than one network element. However, these elements are then not any possible ones but are associated with subsets of failure scenarios affecting at most one of the two considered paths (e.g., a simultaneous failure of node 5, link (2, 5), and link (1, 3) in Fig. 4.6, where the failure of the first two elements affects the working path only, while the third one does not have any impact on either of the two paths).

Failures of single network elements are indeed among the most common failure scenarios. Single link failures happen most frequently in wide-area networks [37], where it is difficult to ensure adequate physical protection for long-haul links (e.g., undersea optical cables, which can be cut by, e.g., movements of tectonic plates or damaged by shark bites). In turn, the frequency of single node failures is higher for local area networks, where the links are shorter and can, therefore, be better protected.

Scenarios of multiple failures include failures of several network elements that occur at the same time (e.g., a simultaneous cut of several optical links placed

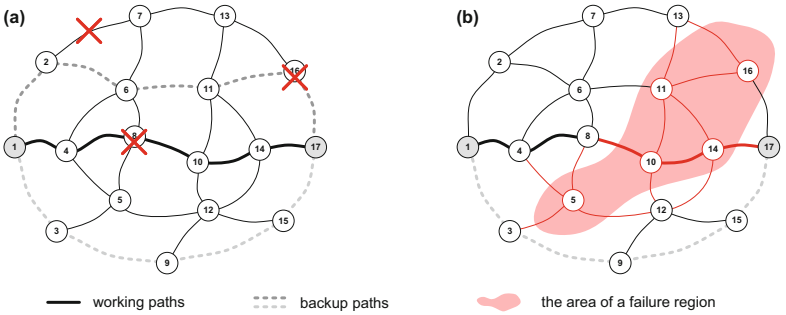

Fig. 4.7 Illustrations of scenarios of multiple random failures (**a**) and multiple failures confined to a given region (**b**)

together in a duct) or refer to failures happening sequentially before previously failed elements have been physically repaired. Among scenarios of multiple failures, we can distinguish either *multiple random failures*, i.e., failures occurring simultaneously at random locations of a system, such as failures of node 8, node 16 and link (2, 7) in Fig. 4.7a.

Another scenario of multiple failures occurring simultaneously at different locations might follow from human attacks targeted at several major nodes/links spread across the networked system. In all such cases, protection of the working path against simultaneous failures of k system elements can be achieved by installing a set of k mutually disjoint backup paths. For example, in Fig. 4.7a, a scheme involving one working path and two end-to-end backup paths being mutually node-disjoint is proper for protection against simultaneous failures of two nodes (e.g., nodes 8 and 16 as illustrated in Fig. 4.7a). It is important to note that parameter k cannot be any value, as the possibility to identify k-disjoint paths follows from the degree of system nodes. For instance, in Fig. 4.7a, since the minimum value of node degrees is 3, only three node-disjoint paths can be determined between these nodes, and as a result, protection against a simultaneous failure of at most two randomly selected (or attacked) elements of a system can be provided.

An important share of failure scenarios is linked to weather-related disruptions and natural disasters such as earthquakes, hurricanes, tornadoes, heavy wind, heavy rain causing flooding, or volcano eruptions occurring in certain geographical regions [32, 38]. As a result, they often lead to massive failures of multiple elements of a network located in a given region (referred to as *regional failures*). It is difficult to predict the occurrence of a disaster itself (e.g., earthquakes are known to be unpredictable as opposed to other disasters which are generally predictable [9]), and, in particular, to forecast the consequences of an incoming disaster such as the shape of a failure region and disaster intensity. Therefore, reactive recovery frequently turns out to be the legitimate procedure under natural disasters, where the configuration of the related backup paths is dynamically determined subject to the consequences of a disaster. In such cases, it is crucial to shape the related

backup paths in a way to make a detour over the actual failure region as presented in Fig. 4.7b, where backup path (1, 3, 9, 15, 17) provides a proper detour over the failure region.

To assure the adequate separation of working and backup paths for a given region of failures, these paths are calculated in a way to ensure their *D-geodiversity*, i.e., the geographical distance of at least *D* from any transit element of one of these paths to any other transit element of the second path [4, 13].

4.3.3 Scope of Recovery Procedure

Considering the scope of recovery, apart from *global protection* (often called *path protection*), assuming utilization of a single end-to-end backup path protecting the entire working path of a demand—Fig. 4.8a, *local protection* may be applied employing backup paths used to redirect the affected traffic over the failed link/node, as given in Fig. 4.8b [7]. The intermediate solution called *segment protection* [29] provides the existence of backup paths, each one protecting a given segment of a working path (consisting of several consecutive elements of a working path), e.g., as in Fig. 4.8c. Concerning the segment protection scheme, in scenarios of node failures, the respective neighboring segments additionally need to overlap each other by one link.

The path protection scheme is the most capacity-efficient concerning all variants of protection scope, while local protection against failures of single links is characterized by the highest amount of spare capacity needed to install the respective backup paths. As illustrated in Fig. 4.9 this is reflected by the total number of links

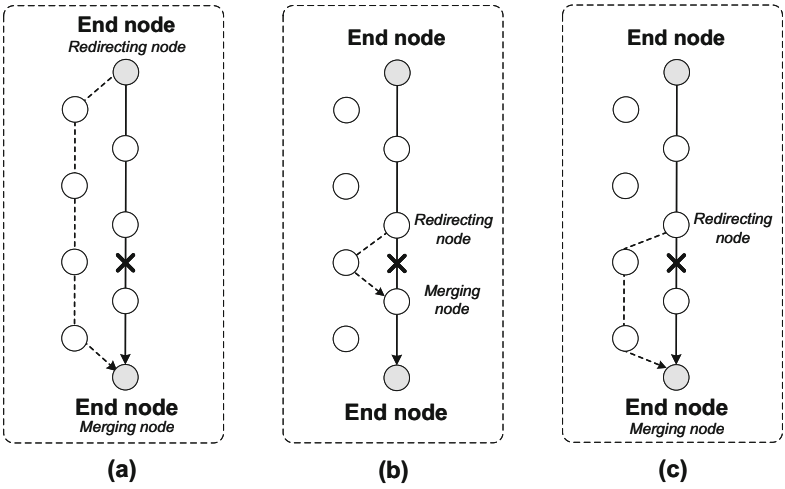

Fig. 4.8 Examples of recovery schemes: global (**a**), local (**b**), and segment (**c**)

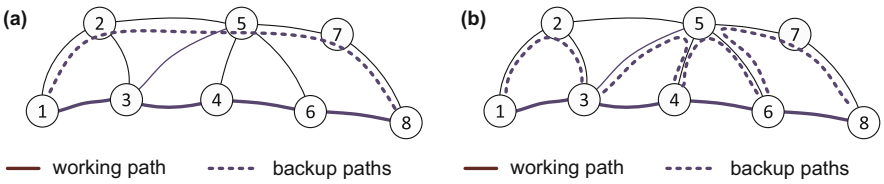

Fig. 4.9 Example illustration of a global protection scheme (**a**) and link protection scheme (**b**) for a working path between nodes 1 and 8

traversed by dedicated backup paths for the example scenario of protection of a working path (1, 3, 4, 6, 8), which is equal to four links in the case of global protection in Fig. 4.9a and nine links for the link protection scheme in Fig. 4.9b.

It is worth noting that the variants of the backup path scope analyzed in this section can coexist with the two modes of backup path setup methods. Therefore, among resilience schemes, we can identify *path protection*, *segment protection*, or *link/node protection* schemes (referring to backup paths installed in advance), as well as *path restoration*, *segment restoration*, or *link/node restoration* techniques based on installing the related backup paths reactively (after a failure).

4.3.4 Usage of Recovery Resources

Two solutions should be outlined when analyzing the schemes of assigning network resources to backup paths: dedicated and shared protection. In a *dedicated protection* scheme, resources (link capacities) of any given backup path are reserved to protect a single working path only. This technique is very costly but enables fast recovery of the affected traffic. Additionally, if preplanned protection is applied, backup paths may be either used in parallel with working paths in the normal operational state (i.e., the *1+1 protection* scheme of transmitting the signal simultaneously along both paths) or activated only for short periods to redirect the traffic affected by the failure (known as the *1:1 protection* scheme). In the latter case, capacity reserved for backup paths can be used to serve best-effort traffic under normal operation [18].

The disadvantage of a dedicated protection scheme is that, even though it provides the fastest recovery, it implies high additional cost of over 100% of the related working path cost due to the ratio of network redundancy exceeding 100% (since backup paths typically traverse more links than the corresponding working paths). Therefore, to limit the cost of a solution, the concept of *shared protection* was proposed in which several backup paths can mutually share link capacities. According to [27], the shared protection approach can limit the redundancy ratio to 35–70%.

If flows are required to be 100% restorable, sharing the link capacities by several backup paths is feasible only if the respective parts of working paths (i.e., being

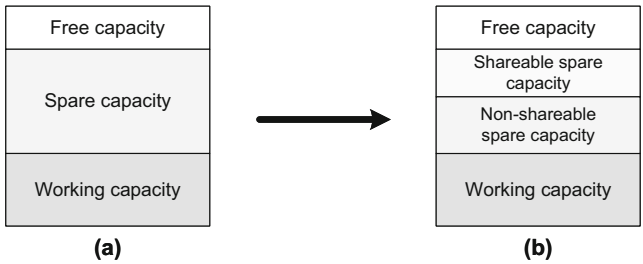

Fig. 4.10 Example link capacity classification under (**a**) dedicated, (**b**) shared protection

protected by these backup paths) are mutually disjoint, meaning that they do not share the same risk of failure (i.e., if they do not belong to any common SRLG) [19].

In resilient routing schemes, the capacity of any link can be generally classified into: (1) *working capacity* (i.e., used by existing working paths), (2) *spare capacity* (denoting capacity already reserved for backup paths), and (3) *free capacity* not used by any path (i.e., that can be allocated for either working or backup paths of new demands) [19].

As shown in Fig. 4.10, under backup capacity sharing, the spare capacity of any link is further divided into two classes: *shareable* and *non-shareable*. The former comprises backup capacity reserved for other backup paths that may be shared by the backup path to be established (i.e., when the respective part of a working path of an incoming demand is SRLG-disjoint with parts of all other working paths being protected by backup paths using this shareable capacity). The latter refers to the capacity already reserved for backup paths that cannot be shared.

Following [19, 36, 49], when finding a backup path in a backup capacity sharing scenario, the cost ζ_h of arc a_h is commonly defined as given in Eq. 4.1. According to this metric, the cost of a backup path link is thus determined only by the extra capacity that has to be reserved for a given backup path. Otherwise, if there is no need to reserve the extra capacity at a_h for this backup path (i.e., if the requested capacity is not greater than the shareable backup capacity at a_h), then ζ_h is set to a very small positive value of ε. Links with sharable capacity are thus preferred in backup path computations.

$$\zeta_h = \begin{cases} \varepsilon & \text{if } c_r \leq sh_h^{(r)} \\ (c_r - sh_h^{(r)}) \cdot \xi_h & \text{if } c_r > sh_h^{(r)} \text{ and } \overline{c_h} \geq c_r - sh_h^{(r)} \\ \infty & \text{otherwise} \end{cases} \tag{4.1}$$

where:

c_r is the capacity requested for r-th demand;
$\overline{c_h}$ is the unused capacity of arc $a_h = (i, j)$;
ξ_h is a unitary cost of arc a_h in working path computations;

$sh_h^{(r)}$ is the capacity reserved so far at a_h that may be shared with respect to the backup path of r-th demand.

Considering heuristic approaches to determine the resilient routing with shared protection, the *active path first* (*APF*) technique described in [20, 21] is typically used. In this two-step scheme, a working path of demand is found first and is followed by calculating a backup path for the topology of a residual network (i.e., with arcs traversed by the working path excluded). Numerous variants of this method have been proposed in the literature aimed at, e.g., determining the working path links in a way to get the most benefits from backup capacity sharing in the second phase [50].

However, if a backup path sharing scheme incorporates the shareability factor into the cost of a backup path link (e.g., as shown in formula (4.1), such backup paths occur to be nonoptimal concerning their length. As we showed in [36], in this case, backup paths may be even 40–50% longer compared to the results for a dedicated protection approach. For instance, for the example scenario from [36] given in Fig. 4.11, the path (2, 1, 3, 4, 7) of the total cost of $10+3\varepsilon$ is chosen to be the backup path for the working path, even though there is a much shorter candidate backup path (2, 4, 7) but of the total cost of 27.

Due to the three-way handshake procedure of backup path activation [40] including sending the `LINK/NODE FAIL` message along the working path links followed by the exchange of `SETUP` and `CONFIRM` messages along the backup path, the total time of service restoration is mainly determined by message propagation delay along the backup path. Therefore, for the classical backup path sharing scheme, improved capacity efficiency comes at the price of increased service restoration time.

Concerning the overall time needed for the recovery of the affected working paths, the respective relations among variants of recovery methods referring to the backup path setup method, the scope of the recovery procedure, and the use of backup path resources are summarized in Fig. 4.12 based on [5].

In general, there is a trade-off between capacity efficiency and recovery time, i.e., the larger the segment of the working path being protected by a given backup path, the better capacity efficiency can be obtained, but for the price of longer recovery times. A detailed analysis of service recovery time for various recovery schemes is presented in [7].

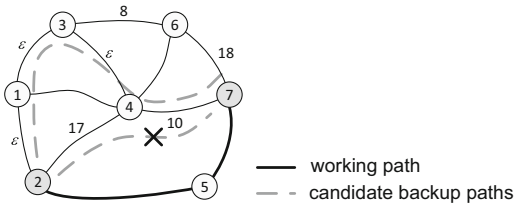

Fig. 4.11 Example candidate backup paths (backup path sharing scenario)

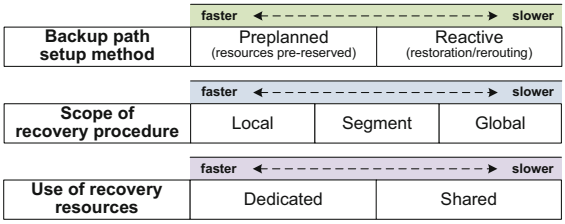

Fig. 4.12 Summary of relations among major variants of service recovery in the context of the overall recovery time

To limit the problem of increased service recovery time under shared protection, our approach introduced in [36] assumes that both working and backup paths are first determined based on the same metric of link costs (i.e., reflecting the lengths of links only). In order not to increase the length of backup paths, backup path sharing is then performed "a posteriori" by finding the solution to the problem of vertex-coloring of the respective graph of conflicts for each network link individually (i.e., to perform capacity sharing for the established backup paths to comply with SRLG constraints concerning the respective working paths). After applying our capacity sharing solution, backup paths traverse the same links as under dedicated protection. Therefore, the time needed for the recovery of the affected flows is here as short as in the case of a dedicated protection scheme.

4.3.5 Protection Cycles

Protection cycles (or shortly *p*-cycles) originally introduced in [16] are ring-like protection structures designed for mesh networks to provide backup detours for a set of working paths. They are assumed to be preconfigured, i.e., calculated and installed in the system before the occurrence of any failure (at the time of establishing the respective working paths for demands). Unlike the configuration of ring networks, where protection rings are physically associated with the respective working rings, *p*-cycles are formed using the free capacity of network links. Therefore, contrary to ring networks, *p*-cycles do not impose any limitations on establishing working paths. Also, there is no strict relation between *p*-cycles and the physical structure of a network.

Similar to common backup paths, the role of *p*-cycles is to restore services in scenarios of failures of any single network element by redirecting the affected traffic onto a backup route provided by a given *p*-cycle. For the example working path `W1` defined by the sequence of nodes (`4`, `5`, `6`, `7`, `12`, `13`) in Fig. 4.13, the related *p*-cycle can provide a detour in the case of a failure of nodes `6` or `7` as well as links (`5`, `6`), (`6`, `7`), or (`7`, `12`).

Similar to ring networks, *p*-cycles can protect segments of working paths traversing the respective *p*-cycle (referred to as the on-cycle spans), as in the case of working paths `W1`, `W2`, `W3` that share a common *p*-cycle in Fig. 4.13. However, unlike backup rings in ring networks, *p*-cycles can also be used to protect working paths

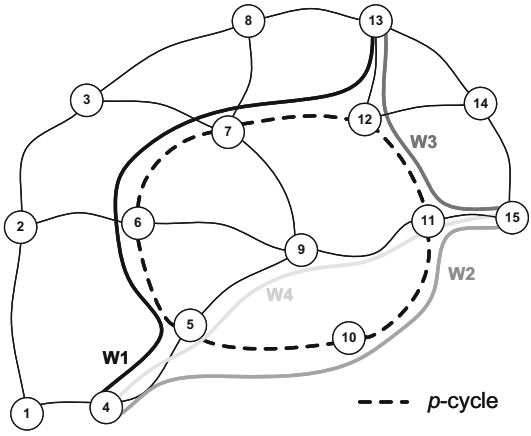

Fig. 4.13 Example configuration of a *p*-cycle for four working paths `W1`–`W4`

straddling the protection cycle (i.e., not having any common link with the *p*-cycle), as the example working path `W4` in Fig. 4.13. This additional feature improves the capacity efficiency of *p*-cycles, making it comparable to the one for shared backup path protection [1].

In general, a single *p*-cycle can protect multiple on-cycle and straddling spans if all these spans are SRLG-disjoint. For instance, the *p*-cycle from Fig. 4.13 is configured in a way to provide detours for the respective parts of four working paths `W1`–`W4`, since all these segments of the considered working paths are mutually disjoint (meaning that they will never fail simultaneously in a scenario of a single network element failure).

In the event of a failure, only two switching actions (like in ring networks) are necessary to redirect the traffic onto the protection path provided by the *p*-cycle (i.e., at the end nodes of the failed span). Therefore, *p*-cycles combine the best characteristics of mesh-based and ring-based protection methods, i.e., ring-like service restoration speed with mesh-like capacity efficiency.

Following [23], *p*-cycles are often selected either from the set of all distinct cycles for a given network graph or from a reasonably large set of candidate cycles. Regarding the combinatorial optimization issues, three major approaches have been used [1]: optimization of only spare capacity, joint optimization of working and spare capacity, and the concept of the protected working capacity envelope (PWCE) from [16] assuming routing of demands based on the information on already established *p*-cycles.

In research papers, protection cycles have been adapted to many networking scenarios. Apart from their original form focused on protecting single links of working paths (often referred to as link-protecting *p*-cycles [23]), other major variants include:

– *Path-protecting p-cycles* [22, 24] involving a single *p*-cycle to protect the entire working path, as illustrated in Fig. 4.14a. A given path-protecting *p*-cycle can protect a set of working paths, provided these paths are mutually disjoint. It is

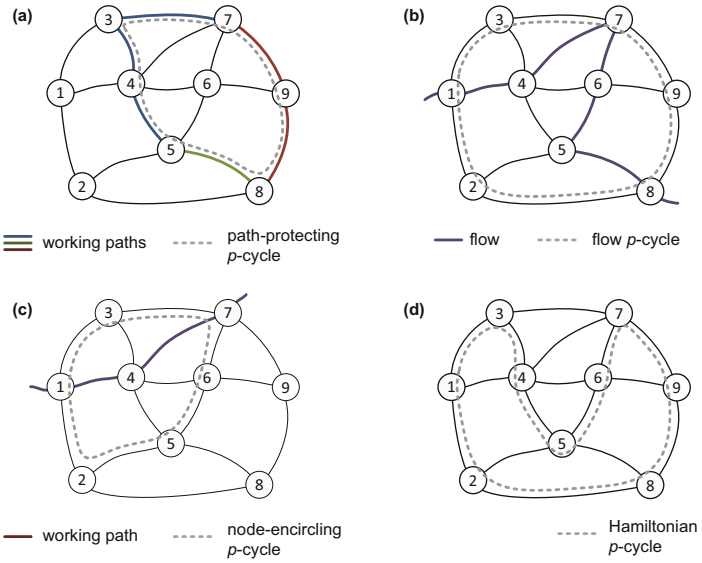

Fig. 4.14 Example configurations for major variants of *p*-cycles: path-protecting *p*-cycle (**a**); flow *p*-cycle (**b**); node-encircling *p*-cycle (**c**); Hamiltonian *p*-cycle (**d**)

worth noting that the constraint of mutual disjointness of working paths that share a given *p*-cycle enables the cycle to be fully pre-connected. That, in turn, means that there is no need for cross-connection operations after a failure (other than at the end nodes of a failed working path), which significantly reduces the time needed to activate the detours for the affected traffic [23].

- *Flow p-cycles* [14] protecting any given segment (a sequence of consecutive links) of a working path, as illustrated in Fig. 4.14b. The size of a segment protected by a given flow *p*-cycle can thus vary from a single link to the entire working path. Similar to path-protecting *p*-cycles, flow *p*-cycles can also protect against failures of transit nodes (if the related protected segments consist of at least two consecutive links).
- *Node-encircling p-cycles* [11] aimed at protecting working paths in scenarios of failures of their transit nodes. It is necessary that for any node of a given working path to be protected by a node-encircling *p*-cycle, the related adjacent nodes of that node on a working path must also belong to the *p*-cycle. Also, the protected node itself cannot be part of that *p*-cycle so that the cycle itself is not affected after a failure of a given node (see the example illustration of protection against a failure of node 4 provided for a given working path by a node-encircling *p*-cycle in Fig. 4.14c).
- *Hamiltonian p-cycles* [43]. Since there may be many *p*-cycles installed in the network to protect all the operating working paths, Hamiltonian *p*-cycles, being cycles that traverse all network nodes exactly once (see Fig. 4.14d), help reduce this number and, as a result, are characterized by even greater capacity efficiency,

compared to the scenario of using p-cycles traversing a fewer number of network nodes. Indeed, as explained in [43], for Hamiltonian p-cycles, the level of resource redundancy needed to provide protection can be as low as $1/(d_{\text{avg}}-1)$, where d_{avg} is the average node degree in the network topology.

4.3.6 Domain of Recovery Operation

End-to-end routing between distant locations frequently needs to be provided over multiple network domains, each defined based on administrative/geographical scope or network provider ownership and commonly identified with an autonomous system [7]. In the context of end-to-end routing, *multidomain routing* encounters problems related to the availability of precise routing information (i.e., following from topological characteristics of domains), which, due to confidentiality aspects, is generally not shared [44].

Another problem refers to the lack of exchanged information concerning the physical deployment of links in different domains related to SRLG disjointness. For instance, as given in Fig. 4.15, even though it may seem that the end-to-end routing using two separate paths over several domains meets the requirements of nodal disjointness, in practice links from different domains (for instance links (B1, B3) and (C2, C3) from Fig. 4.15) may be deployed in the same duct, e.g., physically routed over the same bridge, which raises the risk of a simultaneous failure of both of them. Therefore, applying *inter-domain recovery* techniques (i.e., joint actions taken in multiple domains to recover from failure) is often unrealistic.

As discussed in [26], recovery of communication paths in multidomain network configurations depends on multiple aspects. One of them refers to the location of the end nodes of a connection since both can be located either in the same domain, in separate neighboring domains, or separate non-neighboring domains. Concerning the location of a failed element, we can distinguish either intradomain failures of elements (i.e., failures of links or nodes located entirely within a given domain) or intradomain failures of border links [26, 28] such as of a link (A3, B1) in Fig. 4.16. Also, concerning the failure scenario, we can distinguish either failures of single

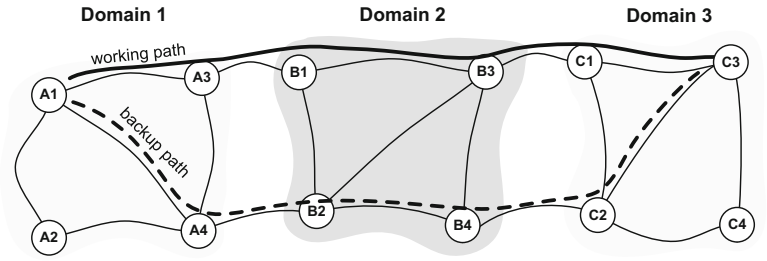

Fig. 4.15 Example scenario of multidomain routing

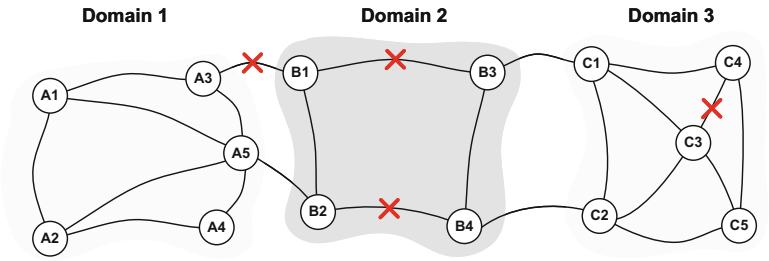

Fig. 4.16 An example illustration of a failure of an inter domain link (A3, B1); two intra domain links (B1, B3) and (B2, B4) implying, in fact, a failure of Domain 2; a failure of an intra domain link (C3, C4)

elements, failures of multiple elements located randomly, or failures of multiple elements located in a given region causing, e.g., a failure of a given domain (see the example scenario for Domain 2 in Fig. 4.16).

The simplest case from the recovery operations point of view is when both end nodes of a given affected connection are in the same domain. Then, it is common for both the working and backup paths to stay within that domain so that the recovery actions are confined to only one domain. In some cases, although both end nodes of a connection belong to the same domain, the related working path connecting them traverses another domain. In such cases, recovery actions should be kept local whenever possible to control the value of the connection restoration time to avoid propagation of recovery operations to other domains.

If both end nodes of a working path are located in different domains, and if failures occur in the domain being a transit one for a given working path, then recovery can be elastic so that the related backup path can even bypass that transit domain.

Cases described above naturally refer to unique characteristics of particular connections. In practice, it is rare to configure recovery schemes per connection. Instead, a single recovery method is deployed in the system, or certain recovery techniques are assigned to certain classes of demands. The following resilience techniques can be distinguished in the context of multidomain environments:

- Dedicated/shared protection, which implies setting up a pair of disjoint end-to-end paths (utilizing the path protection scheme) or configuring backup paths protecting smaller parts of the working path (i.e., implementing segment/link protection). Since end-to-end disjointness of working and backup paths of a connection is hard to achieve in the multidomain configuration (these paths may traverse the same element in a given domain, despite no indication of this issue in the aggregated view [26]), segment or link protection schemes seem more appropriate, especially if protection is arranged within domains.
- Restoration technique with backup paths calculated and set up after the occurrence of a failure. However, contrary to protection schemes, restoration via

multiple domains can be time-consuming and take seconds or more to determine a proper route, bypassing the failed elements.
– Adaptation of the *p*-cycles concept to serve in a multidomain configuration with the protection cycles determined as the shortest ones at the aggregated level (i.e., the level which considers only border nodes) and the protection mechanisms deployed afterward within domains, as proposed in [47].

4.3.7 Layer of Recovery Operations

Internet IP traffic is mostly carried over optical networks (e.g., in the backbone). It means that a certain kind of communication network layering is applied there. Indeed, IP links are frequently virtual, meaning they are provided, e.g., by the optical multi-hop paths. Therefore, the resulting IP virtual topology is commonly formed over the underlying optical transport network.

This simple scenario mentions only two layers, i.e., the upper IP layer (frequently enhanced with Multiprotocol Label Switching (MPLS) functionality toward QoS provisioning, often referred to as IP-MPLS) and the lower Dense Wavelength Division Multiplexing (DWDM) [7]. In this case, IP-MPLS routers are connected to lower-layer Optical Cross Connects (OXCs) ports. OXCs themselves are, in turn, interconnected in a physical mesh topology via multiwavelength optical links.

As shown in the example Fig. 4.17, a working IP layer path for a demand between nodes R1 and R4 consists of a direct virtual link (R1, R4) that is provided in the WDM layer by a lightpath (O1, O2, O4). For this demand, the backup IP layer path consists of two links (R1, R3) and (R3, R4), each one provided by a separate lightpath.

In general, this concept can be extended to the case of networks consisting of more than two layers with a client–server relationship between each neighboring pair of layers (including, e.g., Synchronous Optical Network

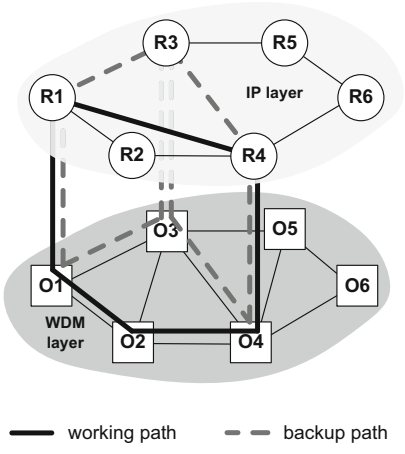

Fig. 4.17 Example scenario of a multilayer routing

(SONET)/Synchronous Digital Hierarchy (SDH) between IP-MPLS and WDM layers) [7]. The automated control of multilayer networks has been standardized in the Generalized Multiprotocol Label Switching (GMPLS) framework, including all necessary entities for use by routing and signaling protocols, in particular the User Network Interface (UNI) and the Network-Network Interface (NNI).

Considering the issue of interoperation between layers, following [7, 44], three main schemes may be distinguished, namely:

- The *overlay model* assuming that routing is performed in each layer separately (i.e., no routing information is shared between network layers)
- The *peer model* (also called *integrated model*) allowing for sharing of routing information between network layers
- The *augmented model* (or *hybrid model*) being the extension to the overlay model that makes information about nodes reachability available at the UNIs

In such a multilayer scheme, recovery actions after failures become even more complex. In general, due to the multiplexing (in the time domain) of lower-rate traffic from the upper layers into the higher-rate paths of the lower layers using time division multiplexing (TDM) [34, 35], the granularity of traffic switching becomes coarser from higher to lower layers. Therefore, more recovery actions must be performed in the higher layers (i.e., restoration of many low-rate flows) than in the lower layers (where recovery is efficient due to performing the recovery actions to the aggregate flows). Besides, recovery time in the upper layers may be additionally increased as a result of a significant number of recovery actions to be performed.

Concerning the order of layers in which recovery actions are performed, based on [7, 10], escalation strategies can be distinguished as follows:

- *Bottom-up* where recovery actions are initiated in the lowermost layer and are next propagated toward the upper layers. This technique's clear advantage is performing the recovery actions at an appropriate granularity. In particular, it means that handling the coarsest granularity actions in the lowermost layer is followed by recovery actions in the upper layers only concerning flows that could not be restored at the lower layer (e.g., a failure of the end node of the lower-layer path).
- *Top-down* where recovery is started in the uppermost layer. This approach, although allowing for better differentiation of recovery actions concerning multiple traffic classes, requires more complex signaling (since lower layers have no direct means to detect if the upper layer was unsuccessful in restoring the affected traffic).
- *Integrated* which combines characteristics of both the bottom-up and top-down strategies. In this case, the decision concerning the layer at which the recovery operations should be started depends on multiple factors such as received alarms or gathered survivability statistics.

If recovery actions are available in multiple layers, it is also essential to provide the appropriate interlayer coordination, including, e.g., determination of the sequence of layers according to which recovery actions are performed.

Such coordination between network layers is necessary to prevent multiple reactions of different layers to the same failure. This can be obtained, e.g., by the *hold-off timer* mechanism [44] used to postpone the recovery actions in the higher layer to give the lower-layer time for recovery of the affected traffic. After that, recovery actions are triggered in the higher layer for all the affected traffic that could not be restored in the lower layer.

Another proposal is to use the *recovery tokens* that help shorten the initialization of recovery actions in the higher layer. In this case, as soon as the lower layer finishes the recovery process, it sends a signal to the higher layer to start the recovery actions there.

Due to the client-server relationship, a failure of a higher-layer node (e.g., of an IP-MPLS router) cannot be restored in the lower layer. However, the reverse, i.e., recovery of a failure occurring in the lower layer (of a lower-layer link/node), is possible in the higher layer.

To perform the recovery actions, each layer must estimate the spare capacity necessary to reroute flows after failures. In particular, the IP-MPLS layer is commonly responsible for handling router failures (e.g., a failure of a router `R3` from Fig. 4.17 which cannot be dealt with by the lower layer). In comparison, the lower (optical) layer is expected to handle failures of fibers/transit OXCs. Backup resources may be shared between network layers, forming the *common pool* of resources [44] so that the respective protection paths from different layers do not share the risk of being activated simultaneously.

4.4 Analysis of Recovery Time in the Optical Layer

Optical transport networks (OTNs) utilizing wavelength division multiplexing (WDM) are considered the primary communication technology for wide-area networks due to the huge capacity of each bidirectional fiber link of several Tbps. In OTNs, each network link is formed by a pair of unidirectional fiber links with their bandwidth divided into several tens of nonoverlapping transmission channels (wavelengths), each one offering a capacity of several Gbps. This enables parallel transmission of many demands at different channels of a given optical link at different wavelengths [31].

In OTNs, every transmission demand between a given pair of source and destination optical nodes is served by an optical path referred to as a *lightpath*. The nodes of OTN are *optical cross connects* (OXCs) and are used to forward the optical signal from the respective input fiber to the related output fiber of a lightpath (with or without wavelength conversion), all in the optical domain. It is often assumed that each OXC is integrated with the *access station* (ACS) via which the traffic either enters or leaves the lightpath (at the lightpath source node and destination

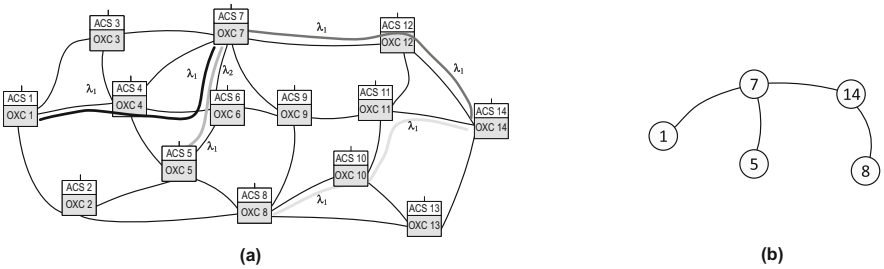

Fig. 4.18 Examples of (**a**) configuration of four lightpaths in a WDM network with wavelength conversion at node 6 for lightpath (5, 6, 7); (**b**) the related logical topology

node, respectively) [39]. At the source ACS of a given lightpath, the input signal is typically converted from the electronic to the optical form by the E/O converter and is next routed via OXCs along the consecutive links of the optical transmission path in the optical domain (i.e., without undergoing the optoelectronic conversion).

When switched from the input port to the related output port at each OXC, the signal can either remain on the same wavelength λ_i (e.g., due to lack of wavelength converters at OXCs) or be switched to another wavelength λ_j [39]. For example the two lightpaths (1, 4, 6,7) and (5, 6, 7) in Fig. 4.18a are multiplexed together at link (6, 7), however, at different wavelengths λ_1 and λ_2, respectively.

At the destination node of the lightpath, the signal is converted to the electronic form (using the O/E converter). Due to the optical signal attenuation progressing with distance, the signal needs to be periodically amplified (typically once per every 80 km of the optical link), which is done by amplifiers [31].

Data forwarding from a given lightpath to another lightpath is performed in the electronic domain, e.g., by the IP routers from the logical topology (where logical links are provided by lightpaths). For example, for the set of lightpaths from Fig. 4.18a, the related logical topology is provided in Fig. 4.18b. Therefore, transit nodes of lightpaths are not visible in Fig. 4.18b. As a result, nodes 1 and 8 are only three hops away in the logical topology, meaning that any IP datagram to be forwarded between node 1 and node 8 needs to be transmitted along three lightpaths (1, 4, 6, 7), (7, 12, 14) and (8, 10, 11, 14) with the electronic processing at each end node of each lightpath.

A single high-capacity lightpath can carry many low-rate (e.g., IP) streams by assigning timeslots concerning a given transmission channel for particular low-rate streams—the technique referred to as *traffic grooming* [51].

Due to the large distances between wide-area network nodes, optical links are at high risk of failure. Indeed, according to statistics, about 55% of cases refer to failures of single network links [37]. Since optical links undoubtedly serve large amounts of data, any failure of the optical network equipment can lead to severe data loss and thus be harmful to a huge number of end users. Therefore, upon the occurrence of any failure, it is crucial to minimize the *protection switching time*, seen as the downtime of the affected connection, and defined as the time between

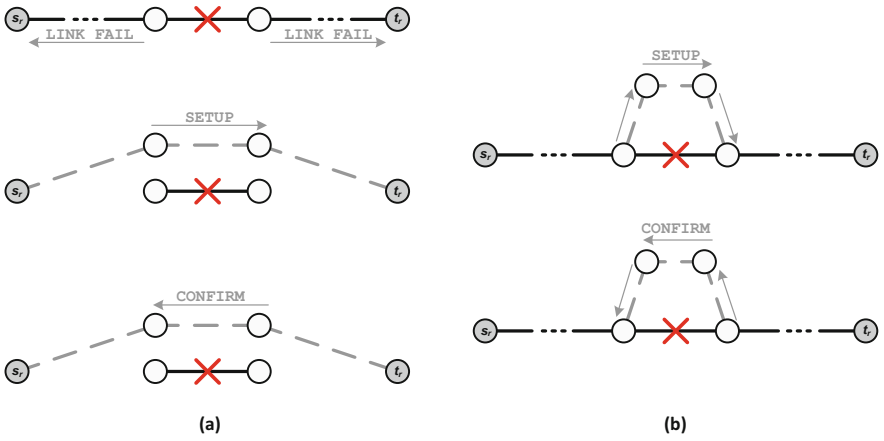

Fig. 4.19 Illustration of the recovery procedure under path protection (**a**) and link protection (**b**)

the instant an optical element (e.g., a link) fails, and the instant the backup path is activated as a detour for the affected traffic [40].

Mechanisms of resilient transmission are indeed an important part of the design of OTN architectures. Among all the resilience schemes discussed in this chapter, techniques based on path and link protection/restoration are most commonly used in practice. After detecting a failure of a network element by optical nodes being direct neighbors of the failed element (e.g., of an optical link by monitoring the levels of signal power along that link), the recovery procedure is initiated depending on the applied resilience mechanism.

In particular, in the case of dedicated protection, as illustrated in Fig. 4.19, after detecting and localizing a failure, the respective neighboring nodes of a failed element send the `LINK(NODE) FAIL` message to the respective source and destination nodes s_r and t_r of a backup path to be activated (note that this operation is not needed in the link protection scheme). After that, to activate the backup path, the respective `SETUP` message is sent along the backup path by the path source node s_r to its destination node t_r. It is followed by sending the `CONFIRM` message along the backup path from node t_r back to the source node s_r.

It is worth noting that the pair of `SETUP/CONFIRM` messages are used not only to activate the backup path but, in some cases (e.g., in the case of backup path sharing), also to apply proper configurations of OXCs along the backup path. Indeed, as illustrated in Fig. 4.20, under backup path sharing, before a failure occurs, it is impossible to configure switching at each intermediate OXC of backup paths. In particular, the configuration of optical signal switching at OXC 6 in Fig. 4.20 (i.e., at one end of the segment shared by backup paths), as well as at OXC 11 is not possible until the occurrence of a failure, since these OXCs will perform switching of the signal depending on the actual backup path activated.

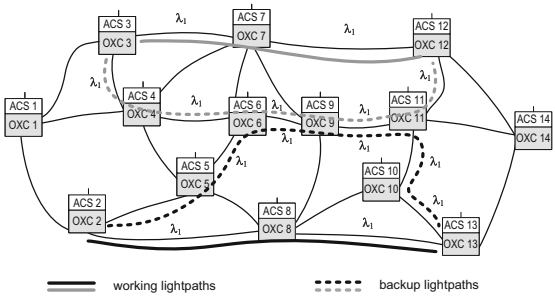

Fig. 4.20 Example configuration of two backup paths sharing the same wavelength λ_1 at links (6, 9) and (9, 11)

The total time needed to activate the backup path thus depends on the related delays occurring when sending the `LINK(NODE) FAIL`, `SETUP`, and `CONFIRM` messages. As discussed in [40], the respective delay components include failure detection time F of about 500 µs, message processing time D at a node (about 10 µs per node), link propagation delay P of 400 µs per each 80 km of the link, as well as time C to set up the OXC (up to 500 µs). The overall recovery time is thus mainly implied by the total length of links and the number of nodes along the backup path.

Indeed, as discussed in [40], assuming the number of hops along the working path for sending the `LINK(NODE) FAIL` signals and the number of hops of the related backup path for sending the `SETUP` and `CONFIRM` signals equal to n and m, respectively, the total time T_{dp} for activation of the related backup path can be calculated for dedicated path protection scheme as given in Eq. 4.2, for shared path protection as provided by Eq. 4.3, and for shared link protection as provided in Eq. 4.4. Therefore, path protection schemes are characterized by higher protection switching time values than link protection mechanisms. However, the use of backup path sharing increases the total recovery time for shared path protection (T_{sp}) and shared link protection (T_{sl}) due to two factors: (a) the need to configure the OXCs along backup paths during the backup path activation procedure, which takes additional time of $(m + 1)C$ as given in Eqs. 4.3–4.4; (b) the increased length of backup paths (see discussion in Sect. 4.3.4 of this chapter). The overall value of the protection switching time for a given failure scenario is calculated as the average time to activate the backup paths for all affected working paths.

$$T_{dp} = F + nP + (n + 1)D + 2mP + 2(m + 1)D \tag{4.2}$$

$$T_{sp} = F + nP + (n + 1)D + (m + 1)C + 2mP + 2(m + 1)D \tag{4.3}$$

$$T_{sl} = F + (m + 1)C + 2mP + 2(m + 1)D \tag{4.4}$$

As discussed in [40], under reactive (dynamic) restoration, upon the occurrence of a failure of a given element of a working path, the arrival of a `LINK(NODE) FAIL` message at the respective source node of a detour triggers the procedure of searching for a backup path for each failed working path by broadcasting the respective `SETUP` message on all its outgoing links, which also reserves resources

on links used for broadcasting. The intermediate nodes act respectively. When the `SETUP` message arrives at the destination node of a detour, that node sends back the `CONFIRM` message along the path of the original `SETUP` message and configures the OXCs along that path. Resources reserved at links not confirmed by the `CONFIRM` message are soon released by the respective canceling messages. This completes the procedure of a dynamic setup of the backup path.

Since the effects of dynamic restoration depend on link resources available after the occurrence of a failure, the *restoration efficiency* coefficient is often used to determine the success ratio of recovery defined as the number of connections that were restored divided by the total number of affected connections [40].

4.5 Recovery Time in the IP-MPLS Layer

In multilayer networks, recovery of a large subset of affected flows can be provided at the IP layer. However, there are several reasons why such a design is not efficient. Firstly, applying the IP layer recovery mechanisms at the routing level may not be fast enough and, therefore, hard to meet stringent QoS requirements. Also, recovery at the optical layer often helps reduce the number of recovery actions that would otherwise need to be performed by the IP layer. This is particularly the case for lightpaths carrying many low-rate IP flows, which, if not restored jointly at the lightpath layer, would have to be restored individually by the IP layer.

However, as already mentioned in this chapter, not all recovery actions are feasible for the execution at the optical layer. An example scenario refers to a failure of one of the end nodes of a lightpath. Since the IP layer sees every lightpath as the IP logical link, a failure of the lightpath end node can be recovered only at the IP layer in the same way as the failure of the IP router (also, it is essential to note that IP routers and OXCs are also often integrated into a single unit).

Another reason for recovery at the IP layer is the need to provide different levels of protection to different IP streams by using different protection mechanisms for several classes of high-priority and low-priority streams [12]. Under optical layer recovery, such streams merged in a given lightpath would have to be recovered jointly using the same protection mechanism, which would not be adequate for a large subset of them.

Also, when proposing a mechanism for the IP layer recovery, it is important to consider the following issues:

- Failures of some IP links may already be handled by the respective backup lightpath set up in the optical layer.
- Only a certain set of high-priority IP traffic needs to be protected, while it is often enough to serve low-priority traffic on a best-effort basis without the recovery guarantees.
- To avoid duplicate recovery operations at different layers of a multilayer network, a proper coordination mechanism (such as the one based on the hold-off timer explained earlier in this chapter) is needed.

As this chapter focuses on mechanisms of resilience for connection-oriented systems, in the context of IP transmission, we draw our attention here primarily to the IP-MPLS proactive resilience mechanisms since multiprotocol label switching (MPLS) used to forward the traffic based on labels instead of addresses can be indeed considered as a close equivalent for the wavelength-based circuit switching characteristic to optical communications.

In the IP-MPLS layer, packet forwarding decisions are made solely based on labels assigned to packets (based on criteria such as the destination node or QoS requirements). These labels assigned to packets as soon as they enter the MPLS domain can be later replaced at transit nodes of a transmission path. Labels thus enable the creation of end-to-end circuits in the form of label switched paths (LSPs), making it relatively straightforward to apply the already discussed circuit-related recovery mechanisms.

The recovery of the affected MPLS traffic is performed similarly to classical protection/restoration mechanisms. It is important to note that the recovery techniques commonly operate in MPLS unidirectionally due to the unidirectional characteristics of MPLS LSPs. The IEFT RFC 3469 document [45] provides a detailed description of MPLS recovery mechanisms according to four aspects of configuration: (1) recovery model (rerouting vs. recovery switching); (2) resource allocation (pre-reserved vs. reserved on demand); (3) scope of recovery (local repair, global repair, or, e.g., multilayer repair); (4) path setup (preestablished or established on demand). This document also defines a sequence of operations in consecutive recovery phases, including fault detection and localization, fault notification, switchover, and post-recovery operation.

The variants of MPLS recovery are also described in detail in [3]. The major ones include:

– Global protection (i.e., path protection) assuming protection of each working LSP by a single backup LSP established in advance (with backup path resources pre-reserved) and being link-/node-disjoint with the related working LSP.

  As discussed in [3], under global protection, the total time for recovery is composed of four components: time to detect the failure T_D assumed to be equal to 20 ms, the notification time T_N, the recovery switching time T_{RS}, and the restoration time T_R, as provided in Eq. 4.5.

$$T_r = T_D + T_N + T_{RS} + T_R = T_D + (fD + nD + \sum_{i=1}^{n} L_i P) + C + \sum_{i=1}^{b} L_i P \qquad (4.5)$$

  where:

  f is the flow (LSP) index;
  D is the message processing time at node assumed to be equal to 10 μs;
  n is the number of nodes between the node upstream of the failure and the source node of a working path;
  b is the number of links along the backup LSP;

C is the time to configure, test, and set up the forwarding table assumed to be between 1 ms and 10 ms;
L_i is the length of ith link in km;
P stands for a propagation delay of 5 μs per km.

In particular, since [3] assumes the sequential recovery of individual flows following their flow indices, the notification time T_N is extended by the processing time of a message at the node closest to the failure completed after the fD period.

– Local protection (often referred to as *fast reroute*), where each link of a given LSP is protected by its backup LSP. In the case of protection against a single node failure, a given backup LSP is assumed to protect two neighboring links of the working LSP. Since, under local protection, the number of backup LSPs can be large, a single backup LSP set up for a given MPLS link can be configured to protect all working LSPs traversing that link [3].

  In this case, the overall recovery time provided by Eq. 4.6 is shorter than for global protection, as it does not include the related time to send the failure notification message from the node upstream of the failed element to the source node of a working path.

$$T_r = T_D + (fD) + C + \sum_{i=1}^{b} L_i P \tag{4.6}$$

  The local protection scheme discussed above is also called the *one-to-one* backup scheme, as opposed to the *facility backup* approach allowing a single backup LSP to protect a set of working LSPs traversing the same sequence of MPLS links.

– *Rerouting/restoration* denotes a scheme of setting up the backup LSPs (and reserving the related resources for these paths) after detecting failures affecting the working LSPs. Depending on the assumed scope of the recovery scheme, we can distinguish between global or local rerouting/restoration. Due to the determination of backup LSPs after a failure, the time needed for MPLS rerouting schemes to redirect the affected traffic onto the backup LSPs is remarkably higher than for the related protection approaches. It can be measured even in seconds, compared to the millisecond values of recovery time characteristic of protection schemes.

4.6 Summary

In this chapter, we provided a detailed discussion of methods for communications resilience in circuit-switched networks. Starting with the analysis of solutions for classical ring networks, the main focus of the discussion was on the mechanisms of resilient transmission in mesh networks. Following the general classification of

transmission recovery schemes based on six criteria, including the backup path setup method, failure model, the scope of the recovery procedure, usage of recovery resources, operation in multidomain environments, and multilayer resilience, the related schemes for resilient transmission were explained. The analysis focused on the efficiency of recovery schemes assessed mainly in terms of the time needed to recover the affected paths, recovery success ratio, and resource efficiency.

A general conclusion following this analysis is that the two considered objectives, i.e., fast recovery and resource efficiency, are generally two contradicting factors. In particular, the shorter the detours (such as those in the link protection scheme), the shorter the time needed to activate the backup paths, but the higher the costs (regarding network resources). Backup path sharing schemes, although able to reduce the amount of resources needed for backup paths, generally tend to increase the time of recovery operations due to (commonly) lengthening of backup paths as well as forcing the configuration procedures of backup path transit nodes to take place no sooner than after the occurrence of a failure. Additionally, the efficiency of recovery operations can be further challenged by the configuration issues related to multidomain or multilayer routing.

Finally, it is essential to note that despite the availability of a multitude of schemes of resilient routing for circuit-switched networks in the related literature, deployment of a large subset of them has faced various problems, e.g., related to the coupling of the data and control planes common for many network system architectures. As a result, deployments of resilient routing mechanisms in commercial systems have been confined mainly to the path and link protection/restoration schemes in the last three decades. However, this situation is now changing with the growing popularity of software-defined networks—SDNs (e.g., utilizing the OpenFlow switches), where the control plane is decoupled from the data plane and localized in a logically centralized controller. Such a controller is flexible enough to implement practically any scheme of resilient routing since its operation is not confined by the constraints (as well as the life cycle) of the related data plane.

? Questions

1. Explain the principles of resilient routing in ring networks.
2. Describe the reasons for differentiation of the levels of service resilience.
3. Provide the classifications of resilience mechanisms in networked systems based on major criteria.
4. Characterize the main strategies of resilience based on the moment of calculating the backup paths.
5. Describe the major failure models considered in the design of resilient routing strategies.
6. Explain the methods of setting up backup paths based on the scope of a recovery procedure.
7. How do backup path setup methods impact the overall time for recovery of the affected flows? Provide the respective summarized view on this issue.
8. Explain the concept of *p*-cycles and describe its main features.

9. Discuss the challenges behind multidomain recovery schemes.
10. Explain the strategies and the related challenges concerning recovery in multilayer networks.
11. Discuss the main determinants of service recovery time in optical transport networks.
12. Explain the service recovery process in IP-MPLS networks and characterize the related main components of service recovery time.

References

1. Asthana, R., Singh, Y.N., Grover, W.: *p*-cycles: an overview. IEEE Commun. Surv. Tutorials **12**(1), 97–111 (2010)
2. Autenrieth, A., Kirstadter, A.: Engineering end-to-end IP resilience using resilience-differentiated QoS. IEEE Commun. Mag. **40**(1), 50–57 (2002)
3. 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)
4. Cheng, Y., Sterbenz, J.P.G.: Critical region identification and geodiverse routing protocol under massive challenges. In: Proceedings of the 2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM'15), pp. 14–20 (2015)
5. Chiesa, M., Kamisiński, A., Rak, J., Rétvári, G., Schmid, S.: A survey of fast-recovery mechanisms in packet-switched networks. IEEE Commun. Surv. Tutorials **23**(2), 1253–1301 (2021)
6. Chołda, P., Mykkeltveit, A., Helvik, B.E., Wittner, O.J., Jajszczyk, A.: A survey of resilience differentiation frameworks in communication networks. IEEE Commun. Surv. Tutorials **9**(4), 32–55 (2007)
7. Chołda, P., Jajszczyk, A.: Recovery and its quality in multilayer networks. IEEE/OSA J. Lightwave Technol. **28**(4), 372–389 (2010)
8. Colle, D., De Maesschalck, S., Develder, C., Van Heuven, P., Groebbens, A., Cheyns, J., Lievens, U., Pickavet, M., Lagasse, P., Demeester, P.: Data-centric optical networks and their survivability. IEEE Sel. Areas Commun. **20**(1), 6–20 (2002)
9. de Sousa, A., Rak, J., Barbosa, F., Santos, D., Mehta, D.: Improving the survivability of carrier networks to large-scale disasters. In: Guide to Disaster-Resilient Communication Networks, pp. 175–192. Springer, Berlin (2020)
10. Demeester, P., Gryseels, M., Autenrieth, A., Brianza, C., Castagna, L., Signorelli, G., Clemente, R., Ravera, M., Jajszczyk, A., Janukowicz, D., Van Doorselaere, K., Harada, Y.: Resilience in multilayer networks. IEEE Commun. Mag. **37**(8), 70–76 (1999)
11. Doucette, J., Giese, P., Grover, W.D.: Combined node and span protection strategies with node-encircling *p*-cycles. In: Proceedings of the 5th International Workshop on Design of Reliable Communication Networks (DRCN'05), pp. 213–221 (2005)
12. Gerstel, O., Ramaswami, R.: Optical layer survivability: a services perspective. IEEE Commun. Mag. **38**(3), 104–113 (2000).
13. Gomes, T., Santos, D., Girao-Silva, R., Martins, L., Nedic, B., Gunkel, M., Vass, M., Tapolcai, J., Rak, J.: Disaster-resilient routing schemes for regional failures. In: Guide to Disaster-Resilient Communication Networks, pp. 483–506. Springer, Berlin (2020)
14. Grover, W.D., Shen, G.: Extending the *p*-cycle concept to path-segment protection. In: Proceedings of the IEEE International Conference on Communications (IEEE ICC'03), vol. 2, pp. 1314–1319 (2003)

15. Grover, W.D.: Mesh-Based Survivable Networks. Options and Strategies for Optical, MPLS, SONET, and ATM Networks. Prentice Hall PTR (2004)
16. Grover, W.D.: The protected working capacity envelope concept: an alternate paradigm for automated service provisioning. IEEE Commun. Mag. **42**(1), 62–69 (2004)
17. Haddadi, H., Rio, M., Iannaccone, G., Moore, A., Mortier, R.: Network topologies: inference, modeling, and generation. IEEE Commun. Surv. Tutorials **10**(2), 48–69 (2008)
18. Haider, A., Harris, R.: Recovery techniques in Next Generation Networks. IEEE Commun. Surv. Tutorials **9**(3), 2–17 (2004)
19. Ho, P.-H.: State-of-the-art progress in developing survivable routing schemes in mesh WDM networks. IEEE Commun. Surv. Tutorials **6**(4), 2–16 (2004)
20. Ho, P.-H., Tapolcai, J., Cinkler, T.: Segment shared protection in mesh communication networks with bandwidth guaranteed tunnels. IEEE/ACM Trans. Netw. **12**(6), 1105–1118 (2004)
21. Ho, P.-H., Tapolcai, J., Mouftah, H.: On achieving optimal survivable routing for shared protection in survivable Next-Generation Internet. IEEE Trans. Reliab. **53**(2), 216–225 (2004)
22. Jaumard, B., Rocha, C., Baloukov, D., Grover, W.D.: A column generation approach for design of networks using path-protecting *p*-cycles. In: Proceedings of the 6th International Workshop on Design of Reliable Communication Networks (DRCN'07), pp. 1–8 (2007)
23. Kiaei, M.S., Assi, C., Jaumard, B.: A survey on the *p*-cycle protection method. IEEE Commun. Surv. Tutorials **11**(3), 53–70 (2009)
24. Kodian, A., Grover, W.D.: Failure-independent path-protecting *p*-cycles: efficient and simple fully preconnected optical-path protection. IEEE/OSA J. Lightwave Technol. **23**(10), 3241–3259 (2005)
25. Kompella, K., Swallow, G.: Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures. IETF RFC 4379 (2006)
26. Larrabeiti, D., Romeral, R., Soto, I., Uruena, M., Cinkler, T., Szigeti, J., Tapolcai, J.: Multi-domain issues of resilience. In: Proceedings of the 2005 7th International Conference Transparent Optical Networks (ICTON'05), vol. 1, pp. 375–380 (2005)
27. Liu, Y., Tipper, D., Siripongwutikorn, P.: Approximating optimal spare capacity allocation by successive survivable routing. IEEE/ACM Trans. Netw. **13**(1), 198–211 (2005)
28. Manolova, A., Ruepp, S., Dittmann, L., Calle, E., Marzo, J.: Location-based restoration mechanism for multi-domain GMPLS networks. In: Proceedings of the 2009 International Symposium on Performance Evaluation of Computer & Telecommunication Systems, pp. 304–310 (2009)
29. Molisz, W., Rak, J.: Region protection/restoration scheme in survivable networks. Lecture Notes in Computer Science, vol. 3685, pp. 442–447. Springer, Berlin (2005)
30. Molisz, W., Rak, J.: A novel class-based protection algorithm providing fast service recovery in IP/WDM networks. Lecture Notes in Computer Science, vol. 4982, pp. 338–345. Springer, Berlin (2008)
31. Mukherjee, B.: Optical WDM Networks. Springer, New York (2006)
32. Pioro, M., Fitzgerald, E., Kalesnikau, I., Nace, D., Rak, J.: Optimization of wireless networks for resilience to adverse weather conditions. In: Guide to Disaster-Resilient Communication Networks, pp. 523–556. Springer, Berlin (2020)
33. Rak, J.: Priority-enabled optimization of resource utilization in fault-tolerant optical transport networks. Lecture Notes in Computer Science, vol. 4208, pp. 863–873. Springer, Berlin (2006)
34. Rak, J.: Fast service recovery under shared protection at connection level in WDM grooming networks. In: Proceedings of the 22nd IEEE International Symposium on Computer and Information Sciences (ISCIS'07), pp. 1–6 (2007)
35. Rak, J., Molisz, W.: Fast service restoration under shared protection at lightpath level in survivable WDM mesh grooming networks. Communications in Computer and Information Science, vol. 1, pp. 362–377. Springer, Berlin (2007)
36. Rak, J.: Fast service recovery under shared protection in WDM networks. IEEE/OSA J. Lightwave Technol. **30**(1), 84–95 (2012)

37. Rak, J., Hutchison, D. (eds.): Guide to Disaster-Resilient Communication Networks, pp. 1–818. Springer, Berlin (2020)
38. Rak, J., Hutchison, D., Tapolcai, J., Bruzgiene, R., Tornatore, M., Mas-Machuca, C., Furdek, M. Smith, P.: Fundamentals of communication networks resilience to disasters and massive disruptions. In: Guide to Disaster-Resilient Communication Networks, pp. 1–43. Springer, Berlin (2020)
39. Ramamurthy, S., Mukherjee, B.: Survivable WDM mesh networks, part I—protection. In: Proceedings of the IEEE Conference on Computer Communications (INFOCOM'99), vol. 2, pp. 744–751 (1999)
40. Ramamurthy, S., Mukherjee, B.: Survivable WDM mesh networks, part II—restoration. In: Proceedings of the IEEE Integrated Circuits Conference, pp. 2023–2030 (1999)
41. Ramamurthy, B., Sahasrabuddhe, L., Mukherjee, B.: Survivable WDM mesh networks. IEEE/OSA J. Lightwave Technol. **21**(4), 870–883 (2003)
42. Ramaswami, R., Sivarajan, K.N., Sasaki, G.H.: Optical Networks: A Practical Perspective. Morgan Kaufmann, Los Altos (2010)
43. Sack, A., Grover, W.D.: Hamiltonian *p*-cycles for fiber-level protection in semi-homogeneous, homogeneous, and optical networks. IEEE Netw. **18**(2), 49–56 (2004)
44. Schupke, D.: Multilayer and multidomain resilience in optical networks. Proc. IEEE **100**(5), 1140–1148 (2012)
45. Sharma, V., Hellstrand, F. (eds.): Framework for Multi-Protocol Label Switching (MPLS)-based Recovery, pp. 1–40. IETF RFC 3469 (2003)
46. Siller, C.A., Shafi, M.: Synchronous Networking. IEEE Press, IEEE Communications Society (1996)
47. Szigeti, J., Romeral, R., Cinkler, T., Larrabeiti, D.: *p*-cycle protection in multi-domain optical networks. Photon. Netw. Commun. **17**, 35–47 (2009)
48. Tapolcai, J., Cholda, P., Cinkler, T., Wajda, K., Jajszczyk, A., Autenrieth, A., Bodamer, S., Colle, D., Ferraris, G., Lonsethagen, H., Svinnset, I.-E., Verchere, D.: Quality of resilience (QoR): NOBEL approach to the multi-service resilience characterization. In: Proceedings of the 2nd International Conference on Broadband Networks (BROADNETS'05), vol 2, pp. 1328–1337 (2005)
49. Xiong, Y., Xu, D, Qiao, Ch.: Achieving fast and bandwidth-efficient shared-path protection. IEEE/OSA J. Lightwave Technol. **21**(2), 365–371 (2003)
50. Xu, D., Qiao, C., Xiong, Y.: An ultra-fast shared path protection scheme – distributed partial information management—part II. In: Proceedings of the 10th IEEE International Conference on Network Protocols (IEEE ICNP'02), pp. 344–353 (2002)
51. Ye, Z., Cao, X., Gao, X., Qiao, C.: A predictive and incremental grooming scheme for time-varying traffic in WDM networks. In: Proceedings of the IEEE INFOCOM'13, pp. 395–399 (2013)