4TH QUARTER 2007, VOLUME 9, NO. 4

IEEE COMMUNICATIONS SURVEYS
The Electronic Magazine of Original Peer-Reviewed Survey Articles

www.comsoc.org/pubs/surveys

# A SURVEY OF RESILIENCE DIFFERENTIATION FRAMEWORKS IN COMMUNICATION NETWORKS

PIOTR CHOLDA, AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY
ANDERS MYKKELTVEIT, BJARNE E. HELVIK, AND OTTO J. WITTNER, NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
ANDRZEJ JAJSZCZYK, AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY

## ABSTRACT

The current trend is to integrate a constantly increasing number of services in the same communication network. Some services have very high resilience requirements, while other services have lower ones. This scenario calls for frameworks capable of provisioning for multiple services in a cost-efficient manner, and already a large number of frameworks has been proposed in the literature. This article presents a comprehensive survey of research efforts related to resilience differentiation in the Internet and telecommunications networks. We present a general framework classification which is used as the basis for the subsequent review of the relevant literature. Finally, a critical evaluation of the state-of-the-art and future challenges facing operators and designers is given.

Provisioning of resilient communication services is a significant aspect of network engineering. One important research issue, which has recently gained attention, is resilience differentiation. The challenge is to enable provisioning of services with different resilience characteristics such as the availability, continuity and duration of downtimes in the same network. The different resilience requirements stem from the various kinds of services, e.g., real time services vs. background bulk data transfer, as well as from different usage of the same service, e.g., emergency handling and financial services vs. leisure activities. This research topic has gained a momentum partly due to the current all-over-IP trend in networking. Historically, each network was engineered to offer only one type of service, e.g., voice or data. Hence, only one level of resilience per network was provided. Another motivation factor for research on the topic is the increased competition between network operators and service providers with the resulting demand for cost effectiveness and the market differentiation. Building and operating a network to meet the higher requirements over the entire range, would be extremely costly. The differentiation of preparations to make before and actions to take after failures of network elements becomes a major architectural and design issue.

A proper formulation of resilience differentiation is important when establishing client-operator relations, where each communication service normally has a Service Level Specification given implicitly by a Service Level Agreement (SLA) [1]. So far, there are no methods which satisfactorily describe different issues related to fault-tolerance features of services and that can be applied in SLAs. The formulation of a set of commonly accepted resilience differentiation frameworks can help to focus work on developing methods to meet well defined resilience requirements. It would also improve harmonization of two different perspectives: the operator perspective which is focused on profit and mechanisms, i.e., particular recovery methods; and the client perspective which is focused on price and requirements stemming from a particular application, with little regard to what network technical mechanisms might help in supporting the application.

As there is a quite large variety of proposals related to resilience differentiation, in this article we strive to present the most representative of them, and show on what philoso-

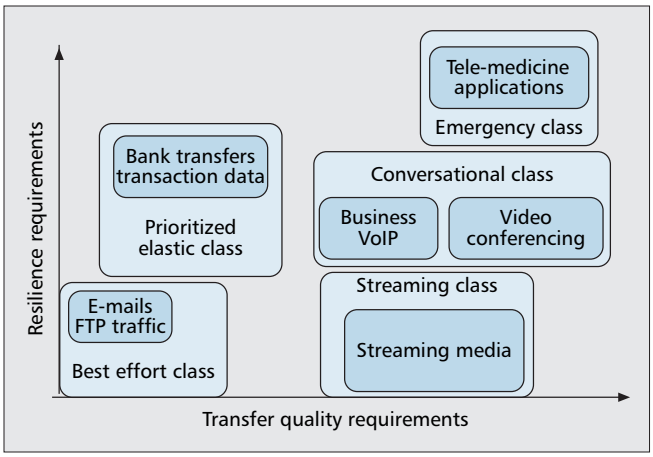


**Figure 1.** *Illustration of general resilience requirements of different service classes [3].*

phies they are founded and what their constraints or commonalities are. The article does not address other aspects related to network resilience besides the issue of differentiation. Contrary to the work of Saradhi *et al.* [2] where some of the most important research seen from an optical networks viewpoint is outlined, we have gathered practically all papers published in the area since 1991 and present a broad survey covering timeliness, topics and open issues.

A rationale for service differentiation based on resilience is presented. As some basic knowledge on recovery methods is necessary to understand the problem area of resilience differentiation, a brief tutorial is given. The most important factors which may be applied in forming resilience differentiation are given, followed up by a comprehensive classification of various concepts and examples of different framework groups. General commonalities of the different contributions presented are elaborated. The main challenges to be faced by researchers and engineers in the future are outlined, which forms a content-related conclusion of the survey.

## Need for Differentiation

Generally, the rationale for introducing resilience differentiation is related to the introduction of Quality of Service (QoS) and thus has similarities to its motivation.

First, a wide range of network dependent applications has emerged over the last decade and more will be available in the future. An application has normally a set of network service requirements, among them resilience demands. The requirements should be satisfied to enable efficient and stable operation, i.e., make the service of the application available. Requirements may differ significantly from application to application. For instance, one-way streaming applications can, by buffering, easily handle interruptions in the network service as long as recovery happens within seconds. On the other hand, an interactive tele-surgery application would require a long time of uninterrupted service. In such a case recovery time is of little interest since even a short interruption will have catastrophic consequences. Fig. 1, extracted from [3], further illustrates the diversity in services requirements. Each service has transfer/transmission quality requirements (e.g., maximum delays, etc.) and resilience requirements. A set of service classes can be distinguished.

Second, the application user's willingness to pay for network services may vary significantly. A private user may settle with low priced best effort services (no resilience guarantees) while a financial institution will require a high availability (and resilience guarantees) to avoid loss of revenue, hence it accepts higher priced services. As the number of network operators and customers is growing, increased competition forces operators to make extra efforts in attracting clients with different resilience requirements and different willingness to pay. By enabling support for the differentiated resilience in a network, both the operator as well as the clients may optimize their revenues and/or keep costs to a minimum.

Third, clients of today's networks are offered a limited set of service resilience classes. Generally speaking, only two classes can be distinguished: voice services with a very high level of reliability, and data traffic connections with best effort services and a low level of availability. This stems from the fact that historically two separate networks have existed. None of them offering resilience differentiation: it was neither necessary nor possible. We have a costly circuit switched telephone network, guaranteeing high quality for all users, and a best-effort data transmission network offering basic connectivity without strictly defined quality assurance guarantees. Today these networks are converging, and hence combining both provisioning schemes becomes a necessity.

Fourth, the majority of today's network operators offer only a single service class per network. They apply over-provisioning, e.g., having at least 50 percent free capacity, to ensure an appropriate level of service to their customers. To provision a single service class is simple, i.e., it requires little control in the network. Much due to advances within optical networks technologies (e.g., Dense Wavelength Division Multiplexing, DWDM) over-provisioning has so far been frequently used in backbone networks. However, in wireless and access networks, over-provisioning is expensive, and given the current linear growth in voice traffic and exponential growth in data traffic, operators will soon have to reconsider their resource management policies. Keeping all customers in one class combined with continued over-provisioning to handle both traffic growth and a growing need for resilience guarantees is likely to be highly uneconomical. Introducing some more complexity into network control and management planes to enable resilience differentiation and in general better utilization of network resources may very well be the only option.

Finally, we should keep in mind that resilience requirements depend not only on an application service, but also, and probably most of all, on how the user applies the service. Hence, the same type of service can have different requirements depending on the customer. For instance, a telephone service for a bank should be 99.999 percent available during the day, while for a home user such a high level of availability is typically not necessary.

## Recovery in Internet and Telecommunication Networks

As modern telecommunication networks enable transmission of large amounts of data and simultaneously many users rely on them, provisioning of fault-tolerance is essential. This goal is achieved by making a network survivable (resilient, dependable[1]), i.e., it can automatically react to failures (e.g., link cuts, software errors) by redirecting traffic from routes affected by failures to routes which are free of them. We should remember that recovery methods are related not only to the connectivity (whose disturbance is detected by Loss of Signal, Loss of Frame, etc.), but also to assumed performance (QoS) factors

[1] *Note that the notion* dependable *is sometimes perceived broader, see for instance [4].*

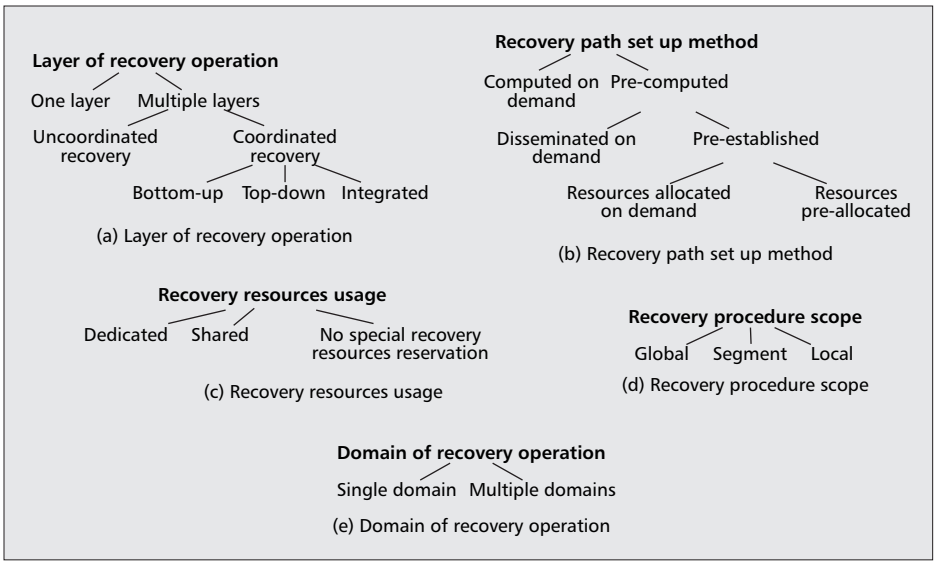


**Figure 2.** *Five classification criteria of recovery procedures: a) Layer of recovery operation; b) Recovery path set up method; c) Recovery resources usage; d) Recovery procedure scope; e) Domain of recovery operation.*

which can differ depending on the layer. In the case of the physical layer, it can be Bit Error Rate (BER), in the case of higher layers it can be throughput, latency, etc. The latter is important in the case of wireless networks as well as networks with statistical multiplexing. Bad conditions could cause delays or poor throughput which do not meet the requirements. In such a case recovery should also be triggered. The mentioned fault-reaction operations are known as resilience (recovery) procedures (schemes, methods).[2] The current trend is to make them as much as possible independent of manual configuration and dimensioning. Then, such methods are called "self-healing." As these issues are well described in literature (see, for instance, relatively new books [5–9]), only a brief introduction is provided here, aiming at achieving the following three goals:

- To introduce or remind a reader about the most typical recovery methods;
- To establish the terminology used in the article. Note that in a later section the terminology established here, and not the one used by authors of original papers, is applied;
- To show that recovery methods differ not only in the sense of architectural or protocol concepts but also generate different quality results.

Paths (routes)[3] on which traffic is carried before a fault occurs are called working (primary, normal) paths. After a failure, the data is switched onto another path (or segment, or even link). It is called an alternative (recovery, backup) path. It can be established in several ways. It may or may not be fully disjoint with the working path. Spare capacity which must be allocated in recovery paths forms so-called redundancy, that is a means to enable fault-tolerance, and not necessarily a normal operation of a network. Although redundancy is generally expensive, its usage is profitable at last: the cost of loss due to failures may be very high since clients are used to the inherited high reliability of voice systems. An operator is usually interested in the cost optimized usage. In particular, it is related to the minimization of expenses related to working/backup paths placement, i.e., the optimized traffic flow. For this reason, issues related to recovery are often perceived as a part of Traffic Engineering (TE) studies [10].

## Fault Management

A recovery procedure involves some essential phases. Commonly there are four phases after a failure [11]: Fault detection, Fault localization, Fault notification, and Recovery switching. The first three of them are known as "fault management' and are described in this subsection. The fourth one, having the largest influence on differentiating parameters, is dependent on selected recovery procedures. Therefore, this phase is described in more detail later.

***Fault Detection*** — Fault occurrence can be detected in a few ways:

- In the physical layer: by *Loss of Light*, *Loss of Signal*, *Loss of Modulation*, *Loss of Clock*
- Signal degradation: the deteriorated signal comes into the receiver during the specified period of time; it can be detected owing to Optical Signal-to-Noise Ratio (OSNR), the optical or virtual channel BER, the dispersion level, the crosstalk, or the attenuation level;
- Quality degradation: the deterioration of the transferred packets/cells: bad throughput, too long delays, etc.

Whereas in the first and second cases, the fault is detected by a node neighboring to the faulty element (link or other node), in the case of quality degradation, it can be detected by the end nodes of the connection.

***Fault Localization*** — During fault localization the point of failure is determined, i.e., a faulty item is recognized. This can be important especially in case of a local recovery method (see below) where the difference between a node and link failure is crucial.

***Fault Notification*** — Fault notification is performed to inform the intermediate nodes that there was a failure in the network as well as to inform nodes responsible for recovery triggering that appropriate procedures should be started to prevent interruption of the communication service. These functions are associated either with the end nodes of the connection (global or segment methods, see below) or with intermediate nodes, usually localized in the neighborhood of the failure (local methods). Thus, in the first case the usage of some signaling methods is necessary. As it involves some time duration, the usage of global/segment methods is essentially slower in comparison to the local ones.

## Recovery Methods Classification

There is an enormous variety of recovery methods. Although many of them live only in scientific papers or standardization documents, they potentially can be applied in a real network.

[2] *Thus,* recovery *is understood here as a means to obtain the result, i.e.,* resilience*. Sometimes 'recovery' is understood in both meanings (e.g., in Internet Engineering Task Force documents).*

[3] *Routes are related to destination-based (hop-by-hop) routing, while paths are concerned with source-based routing. As frameworks described in this article are generally related to connection-oriented techniques, we focus on the term* path*.*

Recovery procedures can be classified at least depending on the five criteria presented in Fig. 2:

1 The layer(s) in which recovery operates (related to the vertical network partitioning);
2 The recovery path set up (provisioning, establishment) method;
3 The level of recovery resources usage (sharing, overbooking);
4 The scope of a recovery procedure;
5 The number of domains over which a recovery process crosses (related to the horizontal network partitioning)

Prefix "pre-'" means that the process is done before a failure is present in a network. Criterion 1 is described for instance in [7, 12, 13]. Criteria 2–4 are covered for instance in [11]. Criterion 5 has only recently started getting attention as the recovery methods have traditionally been restricted to a single domain. All of them are characterized in the following subsections.

***Layer*** — Future networks are claimed to be essentially multilayered. The most popular scenario is the "IP-over-DWDM" model, i.e., an integrated network in which the logical (higher) layer is based on the Internet Protocol (IP), and the transport (physical, lower) layer is implemented by using optical (photonic) DWDM technology. In case of a failure in such a network, it is possible that an effect called "fault propagation" takes place. As paths established in lower layers are perceived as links in higher layers, a failure of a link (traversed by multiple paths) in a lower layer will be seen as multiple simultaneous faults in higher layers triggering many independent recovery operations. Such a competition can cause some problems and decrease spare capacity usage. The main question is: at which layer(s) should recovery procedures be performed? Lower layers have concise and quick fault detection mechanisms. Thus, they enable more rapid methods. Additionally, procedures operating in these layers are simpler, as whole wavelength channels or even wavebands are recovered. As a result, the traffic granularity becomes coarse. On the other hand, although higher layers operate slower, they must be triggered from lower layers and have slower layer specific mechanisms. Moreover, they work with finer traffic granularities, and thus, they offer better resource sharing options and allow the service or flow differentiation. The co-operation between recovery on multiple layers is a challenging issue [7].

Even in case of a single layer failure (i.e., optical layer nodes faults), it is reasonable to use procedures at multiple layers. It requires an "escalation strategy." Two main types of this strategy are [7, 12, 15]: the *bottom-up* and *top-down* strategy. Theoretically, also another attractive method is possible. The *integrated* (hybrid) strategy combines both, the bottom-up and the top-down methods and takes advantages of both of them. A centralized control system manages the recovery procedures in all layers.

When sharing in the higher layer is applied in multi-layer recovery methods, disjointness between working paths must be ensured. Otherwise such paths cannot share recovery resources. In case of multi-layer networks not only disjointness in the higher layer is necessary, but also in lower layers. If two paths established in the packet layer use two different higher layer links which are served by the same link in the optical layer, they are obviously subject to common failures and cannot be treated as disjoint. Such paths are related to the same so-called Shared Risk Link Group (SRLG).

***Set Up*** — For many years, the criterion related to path establishment was perceived as the main one. The reason was that it was strictly related to a technology and the choice based on it frequently implied general quality of recovery parameters. There is a basic partitioning to protection (pro-active) and restoration (re-active, re-routing) methods.

The protection procedures are based on backup path(s) established in advance, whereas in the restoration techniques these paths are disseminated, or established, or even computed, after a failure. Historically, protection is derived from the Automatic Protection Switching (APS) systems handling circuit-switched networks (e.g., Synchronous Digital Hierarchy/Synchronous Optical Network, SDH/SONET). Protection switching actions can be more or less predefined but the common feature is that the protection path or the set of possible protection paths is designated while the working path is established. On the other hand, restoration procedures have their origins in IP, where the path is calculated on demand to re-route traffic due to node failures or connectivity changes, when a link goes down. Although a restoration path may be known before a failure occurs, it is more frequently changed and the resources are never reserved until it is necessary. Restoration inherits all disadvantages related to the dynamic IP routing: most of all, a rather long switching time, temporary instabilities, as well as the risk of loop creation [13]. However, it ensures a good bandwidth usage. In practice, protection can be combined with restoration: first, fast but not efficient protection is started to recover the traffic, and then a better route is sought and switching is performed.

When protection is used, resources can be allocated before a failure occurs or they can be allocated on demand. The first option was used in the APS systems, while the second one can be associated with so-called "zero bandwidth paths" in Multi-Protocol Label Switching (MPLS) where backup paths are established but spare resources are not strictly allocated to them and some sort of signaling is necessary to perform it. Thus, spare resources can be effectively shared among different connections. In such a case a recovery time is considerably longer as the time is needed to convey signaling information.

***Resources Usage*** — Sharing is related to the level of exclusiveness of usage of spare resources. Selected resources can be prepared to be used to establish a backup path related to a particular working path. In such a case, the dedication takes place. On the other hand, a pool of spare resources can be shared to be used when a few (possibly disjoint) working paths are affected by a failure (sharing). If such a case is extended to all paths in a network, we do not have a reservation of any specific resources (no special recovery resources reservation). Obviously, dedication of resources makes very fast recovery possible. Nonetheless, it is very expensive as the resources cannot be taken into account as a basis for establishment of recovery paths for other working paths. On the other extreme, no special resources reservation is quite cheap, but the operation of recovery based on it takes a long time, as signaling is necessary to associate spares with a traffic coming from faulty working paths.

***Scope*** — The next classification criterion is related to the scope of a network connection which is involved with a selected recovery method operation. Here, we have two main options: the local (link, span,[4] node) and the global (path, end-to-end) recovery. The local methods bypass only a single faulty node or link. The global method safeguards the whole

*[4] When a network is presented as a graph, where nodes are cross-connects or multiplexers, we denote edges/arcs as* links*. Sometimes, the notion of link is associated with logical layers, while 'span' is associated with physical layers [6], but in this paper such a differentiation is not necessary.*

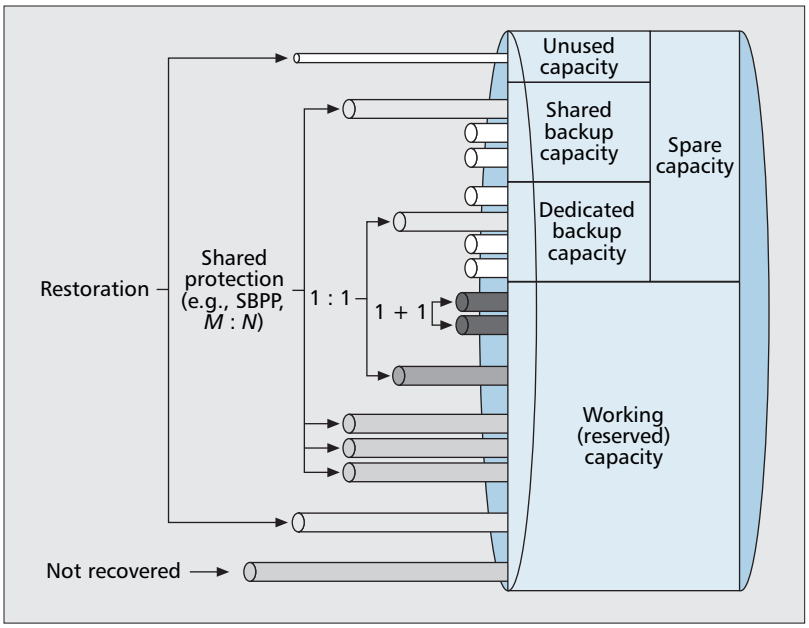


**Figure 3.** *Naming of link capacities with respect to recovery resource usage.*

working path. Obviously, faults in ingress and egress nodes cannot be recovered. By the "ingress node" we mean a node which begins a working path through which the data is injected into a network, whereas the "egress node" is a node which ends this path. The in-between method, called segment recovery, involves some of nodes in the working path but not necessarily the ingress/egress nor nodes which detect a failure (see for instance, the concept presented in [16]).

***Domain*** — We denote a domain as a fragment of a network which is owned and operated by a company or an organization. Internet connections typically span more than one domain. Thus, there is a need for multi-domain recovery methods. Historically, most research in the field takes an implicit assumption that a connection (comprising its working and backup paths) is wholly contained in one domain, i.e., it is an "intra-domain' connection. Today's sophisticated TE routing and signaling protocols are essentially intra-domain. However, works on inter-domain recovery methods are progressing. Examples include schemes for establishing disjoint paths crossing multiple MPLS domains [17] and an MPLS segment inter-domain protection schemes [18]. The main problem of applying the recovery mechanisms in existing networks is related to the fact that owners of different domains do not want to release information related to their topology, transmission and transport parameters. This fact hinders the control of a connection quality, especially from the viewpoint of resilience features.

## Different Methods and Their Influence on Quality

In the previous subsection, we presented different possibilities related to recovery methods. Many of them can be combined. Therefore, we may envisage, for instance, dedicated single layer inter-domain protection, etc. Here, we present some methods and short characteristics of them. They are only samples selected from the entire "space," but chosen to be in line with the current commonly used and standardized options.

In 1+1 *global protection* data is simultaneously transmitted on the working and node-disjoint recovery path. Thus, as both use the working capacity, a distinction between them is only conventional. The egress node selects one signal. Sometimes, path methods are enhanced to ensure their faster operation. This is the case when a "reverse backup path" or "local to egress backup path" is used. The idea is related to the instant switching of data to the modified recovery path in such a way that it begins with the node which detects the failure. Such methods are relatively fast but not cost-efficient.

*Dedicated 1+1 local protection*: operates in an analogous way to 1+1 global protection, but only a single faulty link or node (not the whole path) is bypassed. This method is known as APS in SDH/SONET systems. Although for link protection more spare resources are usually required than for path protection [19], switching is faster as the recovery decision is performed by the node located closest to the failure.

In dedicated 1:1 local protection, data is forwarded only on the working path before a failure. Again, for each of the working path links a selected recovery segment is prepared to take over data when this working link is faulty. Extra traffic may be transmitted in the link which is reserved for recovery purposes. In the case of a working path link failure this extra traffic is preempted. Switching coordination in both ingress and egress nodes is needed. In the MPLS context, this method is referred to as "fast rerouting."

According to the *shared M:N local protection idea*, working and recovery links ($N$ working, $M$ backup, $N \geq M$) are established before a failure occurrence. In the case of a working link failure, data is switched onto a recovery link. If more than $M$ working links are faulty, some fraction of the traffic is lost. Due to some reliability/cost considerations, 1:$N$ is the most common method of this type [20]. Additionally, it is quite fast and uses resources very efficiently.

*Shared M:N global protection* operates in an analogous way to shared *M:N* local protection, but whole paths are bypassed and all of them must share their ingress and egress nodes.

In Shared Backup Path Protection (SBPP) links of recovery paths are shared among different connections which are not subject to the same single failures. The difference from *M:N* path protection is essential. In *M:N* path protection, all the protected paths have the same ingress and egress node and share exclusively a common set of backup resources. In the case of SBPP, backup resources, e.g., some link capacity, are shared by primary paths, which not necessarily have common ingress and egress nodes, in a not typical way, i.e., the sharing of backup paths is partial.

When *global restoration with re-provisioning (dynamic re-routing)* is used, a recovery path is not established before a failure occurs. After the fault notification, the ingress node starts the procedure to reestablish the whole path to the egress node. However, it is not guaranteed that the process succeeds because the bandwidth is not pre-reserved and could be insufficient. The recovery path may use parts of the working path not affected by the failure. To establish a path some signaling is necessary. Thus, the time to converge after a failure occurrence can be quite long. Generally, this is a disadvantage of restoration mechanisms which, due to longer recovery times, are perceived as worse than shared protection. Nevertheless, for connections which do not convey information steadily, shared protection can be more harmful than dynamic re-routing. It is the case especially if recovery procedures are started when there is some data to transmit. Then, as shared protection uses signaling which takes some time, the data is stopped and buffered. If it is immediately sent to

nodes which independently decide of new routes (like in dynamic re-routing), it can be transferred faster.

In *global restoration with pre-signaled recovery bandwidth reservation*, the bandwidth on the links which will constitute the recovery path is reserved before a failure occurrence by using signaling protocols. However, the recovery path is not established. After a failure, the information is disseminated by using a signaling protocol again and, if necessary, the switching nodes (or crossconnects) are rearranged.

*Segment* or *local* variants of the *restoration* methods described above can be imagined. For example, [13] presents a proposal of a method called "fast topology-driven constraint-based rerouting" for MPLS networks. Other recovery methods can also be mentioned: *p-cycles* as a special case of local shared protection for mesh networks [6], *resilient routing layers* as a special case of connectionless segment protection [21]; and *flooding* as a special case of connection-oriented restoration where nodes do not have a complete view of a network [22, 23].

In an existing network, different recovery methods involve different planning schemes (e.g., see [24]) which generate a disjoint set of spare resources related to each group. Figure 3 shows how different elements of capacity in a single link are related to general groups of recovery methods.

## Quality Features related to Recovery

In this section, we enumerate different resilience-related features which determine the quality of communication services. Some features are very frequently used, while others are rare but could be useful as a basis for differentiation. We characterize them, based on how they can be quantified, what their importance is and when an operator is interested in them. Some of the parameters were characterized in [3, 7, 25, 26].

Before we characterize selected quality parameters, we must elaborate on the relation between resilience differentiation and quality parameters definition and assessment. The client or other peer to which a service is provisioned could potentially be interested in a set of features related to the service quality which can be perceived on a longer or shorter time scale. These are denoted *quality of resilience features*. The operator is more interested in the operational and implementation features related to the provisioning of a given service level in the network. Such features are denoted *operation-related features*. Although the two groups are disjoint, both the client and the operator are influenced by the former and the latter. The client must provide an operator with the objectives of a service, and a price paid by a client is somewhat related to operational conditions as they influence the cost. All features must be qualitatively or, better, quantitatively assessed to check if the offered quality meets client requirements and to check if their support is profitable for a network operator. If such a test is passed, the portfolio or a particular service will proceed. If not, the whole structure should be refined by changing the requirements or willingness to pay or maybe even through a search for new technical means or system network architectures.

### Quality of Resilience Features

The *quality of resilience* encompasses the features of a network that affect the QoS observed by the users and are related to resilience. They are partitioned into two groups. The first group is denoted *reliability attributes* and consists of three attributes adopted from classical reliability theory. The second group, denoted *recovery-related features*, includes features of communication networks that without being identical with any of the reliability attributes, still strongly influence them.

***Reliability Attributes*** — In the following, the three reliability attributes are presented, and at the end of the subsubsection, the relation between them is discussed.

**Continuity** — Continuity is related to the length of a period of time during which a service is not interrupted due to a failure occurrence. The Mean Time to Failure (*MTTF*) can serve as the measure for it. *MTTF* is defined as the average duration of time from the time instant when a service request was received, given that the service was up at this time, until the first service failure. Sometimes it can be more convenient to employ other measures that can be used as an approximation or estimate of *MTTF*. These are:

- *Mean Up Time*, *MUT*: the mean time interval from a point when a service was restored after a failure until the next service failure.
- *Mean Time Between Failures*, *MTBF*, or *Mean Time Between Interruptions*, *MTBI*; the latter concerns a service interruption caused not only by physical failures but also by a decreased quality
- *Probability of a failure or failure intensity of a working path*

Usually, a network operator can control recovery rather than a parameter such as continuity. But, obviously, continuity can be controlled through the routing of the path carrying the connection. For example, with equal link failure intensities, the lower hop count, the longer *MTTF*. A continuity-related attribute can be very important for some services. There are very sensitive applications like real-time control, which generally should not be interrupted at all. Therefore, such services should be routed on the most robust resources, recovered by hardware backups or fast physical layer schemes which do not interrupt such applications in higher layers, where relatively slow recovery is provided.

**Downtime** — Downtime is the measure describing the time period in which a service is inaccessible due to a failure in a network. The length of the outage or interruption can be assessed in many ways:

- *Mean Down Time*, *MDT*: the mean time duration from a service failure to the point when the service is restored.
- *Mean Time to Repair*, *Recover(y)* or *Restoration*, *MTTR*, are measures that are sometimes used instead of *MDT*. In the communication networks context, *MTTR* is useful because its value is relatively simple to assess. When single-failure assumptions are applicable, the network could be dimensioned so that all failures are recovered, and the Mean Time to Recovery could be used as a replacement for *MDT*. Recovery time models are presented in [27, 30].
- *Percentile/quantile of the downtime probability density function,* $T_{p\%}$.

Downtime as the basis of differentiation is especially suitable for traffic that is very sensitive to delays, i.e., related to streaming flows (movies, etc.). Such services do not require very large *MTTF*s. They tolerate failures, but they do not accept long delays related to larger recovery times, as they make the service intolerable to users. We can distinguish two types of downtime. The short one is related to the operation of recovery methods. The second one is related to an unsuccessful attempt to recover a service. Such a downtime is considerably longer as the service cannot work until the failed network element is repaired. Hence, percentiles are also useful downtime measures.

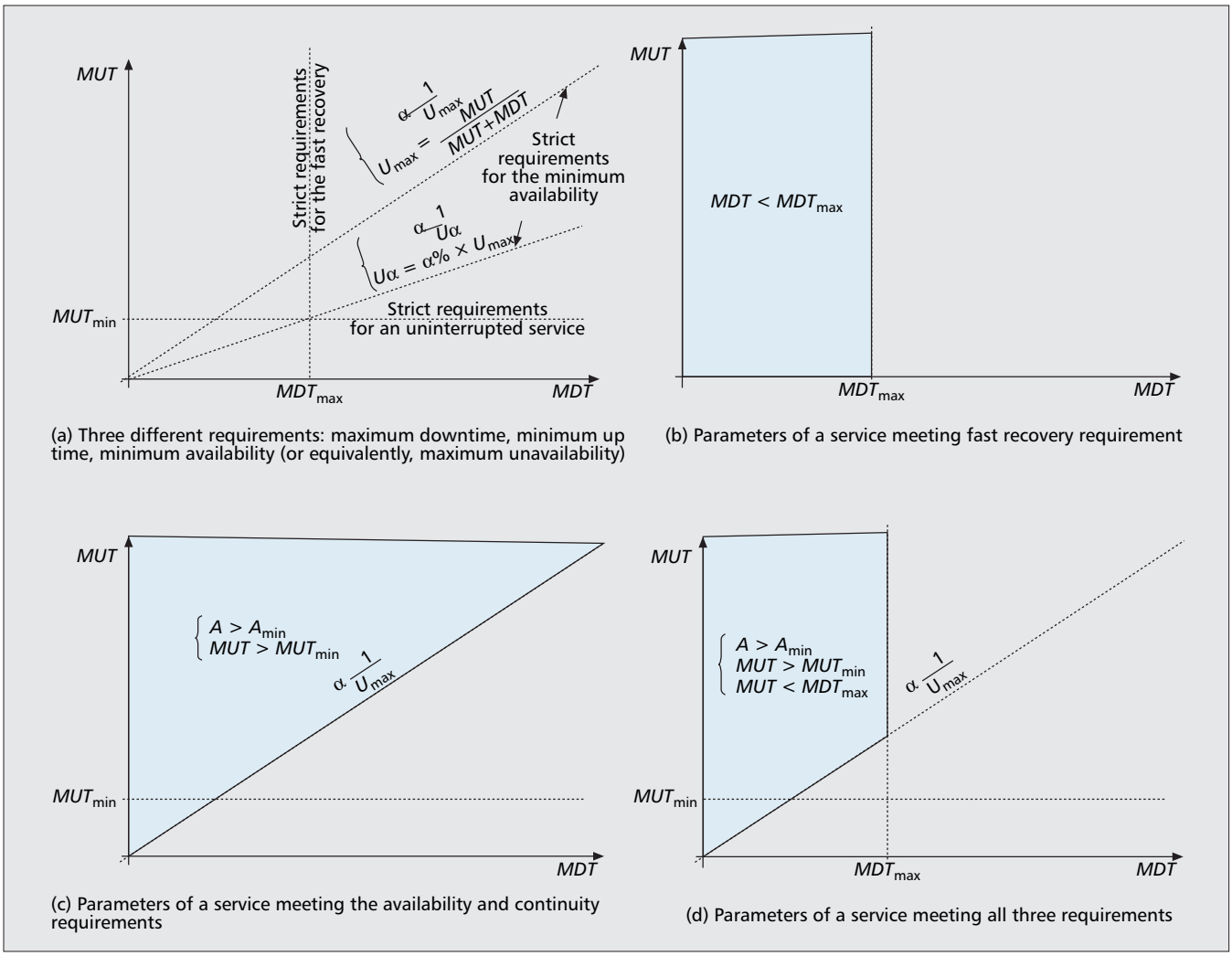


**Figure 4.** *Relationships between different requirements related to the continuity, downtime and availability of a connection [9, 35].* $MUT_{min}$*: the minimum allowable Mean Up Time,* $MDT_{max}$*: the maximum allowable downtime,* $A_{min}$*: the minimum acceptable availability,* $U_{max} = 1 - A_{min}$*: the maximum acceptable unavailability. a) three different requirements: maximum downtime, minimum up time, minimum availability (or equivalently, maximum unavailability); b) parameters of a service meeting fast recovery requirement; c) parameters of a service meeting the availability and continuity requirements; and d) parameters of a service meeting all three requirements.*

**Availability** — For practical reasons, availability is the most commonly used reliability attribute when assessing the resilience of communication networks. ITU-T [13] defines the *instantaneous availability* $A(t)$ as "the probability that an item is in an up state at a given instant of time $t$." As a service is guaranteed the availability in the long time range, we are mostly interested in the steady-state availability, $A$, which can be intuitively understood as [6]: "the probability of finding the item (system, network, connection, etc.) in an operating state at any time we want its service." Availability can be defined in a specific context, for instance for IP networks [32], and is proposed to be used to construct SLAs [33, 34]. Sometimes, the notion of the unavailability, $U$, is used. This parameter is a probabilistic complement of the availability: $U \equiv 1 - A$.

**Relation Between Reliability Attributes** — For some applications, the strict restrictions on neither continuity nor downtime are necessary. In such a case, the differentiation can be based only on the availability as a general indicator of the reliability. Availability is related to *MUT* and *MDT* according to the well known relation [31]:

$$A = \frac{MUT}{MUT + MDT} \tag{1}$$

If we take into consideration the three factors, $A$, *MUT* and *MDT*, we can imagine a resilience differentiation based on the mutual relation between them, like the examples given in Fig. 4. Note that a "working point" should be placed in the middle of the indicated areas to avoid the situation that little change in conditions make a service not meet its requirements.

***Recovery-Related Features*** — Five recovery-related features are distinguished. They are briefly described as follows.

**Quality of the Recovery Path** — Traditionally, resilience itself is perceived as a part of QoS. When resilience differentiation is considered, such a viewpoint cannot be sustained any more.

In many cases an operator can offer spare resources but they are not equivalent to those reserved for the working path. For instance, in connection-oriented networks, the back-

| | | Not-preempted connections | Preempted connections |
|---|---|---|---|
| Recovered connections | Protected | Type 1 | Type 3 |
| | Restored | Type 2 | Type 4 |
| Not recovered connections | | Type 5 | Type 6 |

Table 1. *Different possibilities of relation between preempted and recovered connections.*

up path is usually longer than the working path, i.e., has larger delays, a higher failure probability, etc. Moreover, it is possible that there is less bandwidth on the recovery path than the bandwidth reserved for the working path, resulting in a situation where only a fraction of the traffic is recovered. It is a typical case in IP networks, where the re-routing finds an alternative path, but the total capacity in the network is reduced by the failure and bottlenecks resulting in a reduced throughput. A large variety of service impairments can be taken into account depending on the network and layer of operation. A lower layer quality, related to the physical medium (e.g., lightpaths) is denoted here as "transmission quality," and encompasses BER, signal to noise ratio, severely errored seconds, etc. On the other hand, "transfer quality" is associated with QoS properties of higher layers: packet loss, jitter, delay, hop count, received capacity, throughput, fraction of the offered bandwidth, reordering and duplication of data, etc. The definitions of different transfer quality parameters and the methodology to measure them can be found in [36].

**Affected Traffic (Traffic Lost)** —Traffic which is affected by a failure can be partitioned into directly affected and indirectly affected traffic. While the former is the amount lost or disturbed due to a faulty working path (an accumulated unfinished work [37]), the latter is affected in the sense of higher congestion and higher loss probability. This amount is hard to estimate in general as it depends on specific network conditions. Directly affected traffic can be assessed as proportional to the recovery time (if buffering is neglected).

**Resilience to Multiple Failures** — For very important connections, the necessity of the uninterruptible service regardless of more than one failure in the network can be taken into account. For example, in connection oriented networks, multiple backup paths may be defined for each connection to tolerate multiple failures. Additionally, methods which are based on continuous reconfiguration after failures to improve the resilience against consecutive failures are currently studied and deployed.

**Preemption** — Preemption in this context can be understood as a process which takes away resources from one service to give them to the other, i.e., to enable the recovery of a communication connection considered as more important. Preemption can be performed on at least two levels:

- A connection can be deprived of its recovery resources because another connection has to use them; then, if there are simultaneous failures which influence both connections, the preempted one will not be recovered;
- A connection can also be deprived of its working resources, i.e., even though it is not affected by failures it can be broken because the resources are necessary to recover a more important connection; this is the case of the extra traffic in the 1:1 protection.

Such preempted connections could be offered to customers as the cheapest services. We can imagine the whole spectrum of types of the preemptive services (Table 1). Nevertheless, types 3 and 4 are not practically taken into account.

**Failure Coverage** — Failure coverage (percentage of coverage) is the fraction of traffic or connections which is recovered in a given failure scenario. It represents the efficiency of recovery, i.e., to what extent a network is resilient to a given failure scenario. It can be especially useful when packet oriented connectionless techniques are taken into account and when the resilience of the whole network is assessed. This factor, as the global one, can be also interesting for an operator as it enables the assessment of the average performance of a selected set of methods in the whole network or its domain.

## Operation-related Features

***State Overhead*** — State overhead can be understood as the amount of information which should be maintained in different network elements (e.g., nodes) to properly apply the selected recovery scheme. For instance, if disseminated on demand methods are used, the nodes should have as up-to-date as possible knowledge of the network topology. On the other hand, when dedicated recovery methods are present in the network, the state overhead is relatively small.

***Signaling Requirements*** — Some recovery methods require a sophisticated signaling, and some others do not need almost any. Additional signaling necessary, for example, in the case of sharing methods can be related to large flows which make the recovery operation slower than the dedicated schemes. They can also decrease the network throughput.

***Flexibility*** —Flexibility of a selected recovery scheme is related to the easiness of dealing with unexpected or unplanned fault events. Usually, the recovery is designed to make a network resilient to single failures. However, it appears that some methods can better react to failures which were not taken into account at the planning phase. From this standpoint, restoration often behaves better than protection.

***Scalability*** —The notion of scalability is strongly related to the planning and execution of recovery methods. It means the ability of a network to grow easily, even if complex recovery methods are applied. If procedures supporting recovery are very extensive, a great effort is necessary to add new nodes or links to the network and simultaneously sustain the same level of resilience.

***Cost of Recovery – A Quality Parameter?*** — The cost of a recovery method is very important for the operator and should be taken into consideration as a factor determining the differentiation based on resilience parameters. Recall that we claimed earlier that cost optimization is one reason for introducing resilience differentiation.

Cost is the derivative of other operational parameters. Generally, the most intuitive way is to base it on the redundancy, i.e., additional resource usage necessary to support a selected recovery method (usually the spare link capacity), one of the Capital Expenditures (CAPEX) elements. Obviously, there are some other elements contributing to the cost, e.g., additional software, the increased Operational Expenditures (OPEX) related to the new staff or higher expenses on device operation [38].

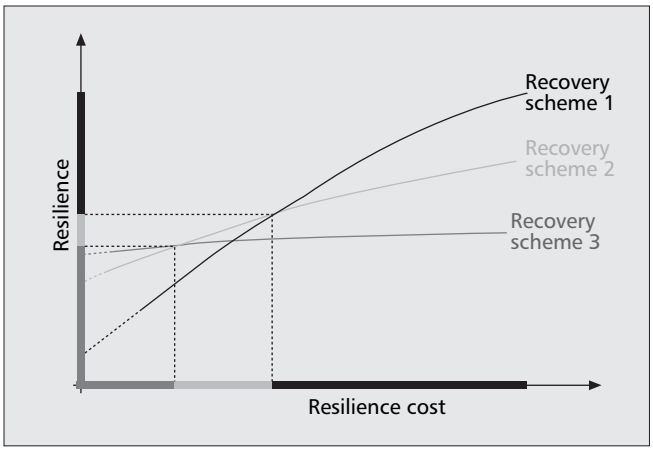


**Figure 5.** *Diagram showing how a trade-off between cost of a selected recovery method and resilience performance induced by it influences choice of a scheme, when different budgets or requirements are assumed. The optimal choice for a given range of resilience requirements/cost is symbolized by colored lines in the axes. It is assumed here that if a resilience is planned optimally, the higher cost, the better resilience.*

It can be argued that the usage of cost as a quality-like factor is a misnomer since a measure of cost is not a measure of quality itself. We can agree that cost is not a typical quality measure as the availability, delay, etc., are. However, in some sense it is only a matter of definition. If the differentiation aims at quantifying user requirements, we can say that anything what is less costly, i.e., better from the expenditure viewpoint, has also a better quality. Therefore, we argue that the cost of recovery is the equal aspect of the trade-off as other "traditional aspects of quality." It is possible to see in Fig. 5 that from this viewpoint, different recovery methods can be chosen to support different requirements because of their mutual resilience-cost relationships. In the figure we can also see the fact concluded in [39] that if a single differentiating parameter is taken into account in isolation, then the optimal usage of resources is obtained if only one type of recovery scheme is used.

Different kinds of recovery methods were briefly discussed earlier. In Table 2 we outline their relation to the quality features characterized in this section. We can see an intrinsic trade-off related to the implementation of each option. If one parameter has advantageous characteristics, we can find another, which is not fully satisfactory. Hence, complex decisions must be taken to tailor a proper recovery method for a selected service.

## Differentiation and Assessment Frameworks

In this section, the resilience differentiation frameworks that are found in the literature are presented. This topic is relatively new: the papers covered in this survey have been generally published since the mid-1990's. The prevailing fraction of the papers is related to core networks, and all frameworks assume that a service is provisioned in a single network domain. With a few exceptions the ideas are limited to a single network layer.

Some observations on the existing literature can be made. First, most works have only a preliminary character. It happens quite rarely, that a single idea is developed further, even by the original authors. Thus, a large variety of ideas can be observed. However, there is little connection between different frameworks, and many authors seem to be unaware that similar concepts to their works have already been published. With the exception of the work by Saradhi *et al.* [2] the different ideas have not been analyzed nor compared. The lack of comprehensive classification seems to be a problem as this fact does not help possible applicators to decide what kind of frameworks is interesting for them. Here, we note that the authors of [40] propose a simple classification scheme based on the recovery method in use.

In this article we propose a classification of the frameworks in which resilience has a main or large significance. Figure 6 presents the classification. The characteristics of each group are given in the relevant subsections below. A thorough literature map, with all the works and main relations and references between them, is shown in Fig. 7.

### Structural Differentiation

Frameworks in the structural differentiation category base the differentiation on structural arrangements related to the recovery of different connections. Structural frameworks map client's requirements to a specific set of recovery methods. The structural differentiation frameworks may address one or more of the *recovery-related features* from earlier, but the *reliability attributes* cannot be addressed directly. In some cases, the framework includes numerical user requirements. However, these requirements are only attributes of chosen recovery methods when the methods are applied with a certain given technology and certain given physical properties of a network. Structural differentiation frameworks are divided into two subcategories, the recovery-focused and quality-focused frameworks.

***Recovery-Focused Frameworks*** —The recovery-focused frameworks address the recovery-related aspects of the service offered to the customer. There have been two phases of publishing in this area. In the first phase, in the mid-1990's, three frameworks were proposed for Asynchronous Transfer Mode (ATM) networks. The second phase started in 2000 and addressed IP-over-(D)WDM networks.

The first published work in this category was the multiple-availability-level ATM network architecture proposed by Oki *et al.* [41]. For each connection in the network, a number of backup paths are computed. We denote these paths as permissible paths. Upon a failure of the primary path, the connection is recovered if one of the permissible paths is working and free capacity can be allocated to that path. Since the ability to choose between many backup paths will increase the probability of successful recovery, the number of permissible backup paths is used as a basis for the differentiation.

The concept of the Reliability of Service (RoS) is introduced by Veitch *et al.* [42]. The authors consider a scheme with three grades of RoS based on Virtual Path (VP) protection in ATM networks. The proposed grades of RoS are shown in Table 3. A similar framework was later presented by Gerstel and Ramaswami [43].

Yahara and Kawamura [44, 45] (and the extended journal version [46]) present a framework for different resilience levels in an ATM network. The levels are divided into two main groups: guaranteed and best-effort. The VPs in the former group receive the highest reliability levels which guarantee that 100 percent of the bandwidth can be recovered after single failures. When a failure happens, the affected VPs are attempted to be recovered in a sequence ordered by the resilience level. A higher class connection may preempt the resources assigned to a lower-resilience VP if this is necessary to recover the higher resilience level VP.

| Feature | | | Classification criterion | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Recovery path set up method | | | Recovery resources usage | | | Recovery procedure scope | | |
| | | | Computed on demand | Disseminated on demand | Pre-established | Dedicated | Shared | No special reservation | Global | Segment | Local |
| Quality of resilience | Reliability attributes | Continuity | Bad | Probably bad | Potentially good | Potentially very good | Potentially good | Probably bad | Depends on other criteria | | |
| | | Downtime | Very long | Quite long | Very short | Very short | May be short | Long | Depends on other criteria | | Short |
| | | Availability | May be low | Moderate | High | Very high | Moderate | May be low | Depends on other criteria | | Rather high |
| | Recovery-related features | Quality of the recovery path | May be poor | May be equivalent | Equivalent | May be equivalent | May be poor | Possibly poor | N/A | | |
| | | Resilience to multiple failures | Very high | High | Small | Small | Small | May be high | May be small | May be high | Very high |
| | | Preemption | Possible | Possible | Impossible | Impossible | Possible | Possible | N/A | | |
| Operation-related features | | State overhead | Moderate | Large | Small | Small | Moderate | Large | N/A | | |
| | | Signaling requirements | Moderate | Large | Small | Small | Moderate | Large | N/A | | |
| | | Flexibility | Very large | Large | Small | Small | Moderate | Large | Small | Moderate | Large |
| | | Planning | Typically distributed | Typically distributed | Typically centralized | Typically centralized | Typically distributed | Typically distributed | Depends on other criteria | | |
| | | Execution | Typically distributed | Typically distributed | Typically distributed | Typically distributed | Typically distributed | Typically distributed | Depends on other criteria | | |
| | | Spare requirements | Small | May be small | Large | Large | Moderate | Small | May be small | May be moderate | Large |

N/A: not applicable.
Affected traffic and failure coverage are not given as they are hard to estimate in general.
Scalability is partitioned into the planning and execution features.
Spare requirements are given as an estimate of a recovery cost.

■ Table 2. *Qualitative feature characteristics of different recovery options.*

Sridharan and Somani [47] propose a framework with three recovery classes: full recovery (using shared protection), no recovery and best-effort recovery (using shared protection if enough resources are free). On this basis, they optimize the revenue which is assumed to be proportional to the number of accepted connections.

Thiran *et al.* in [48] (and in the extended journal version [49]) present a framework denoted Level of Protection (LoP), where two classes are provided. The first class, Fully Protected (FP), is based on 1+1 or 1:1 protection at the WDM layer. The second class, Best Effort Protection (BP), assures only IP restoration using the spare capacity left after recovery of the FP services. Although the authors notice the necessity of service differentiation based on resilience, their main goal is optimization of the transmitted load: as not all services have to be FP, some excess capacity can be used for new working paths. The idea is extended in [50], where it is considered how to differentiate the treatment of both classes in the IP layer.

Zhang and Durresi [51] investigate differentiation in IP-over-(D)WDM networks. The authors define three resilience classes and assign appropriate recovery mechanisms to each class. The recovery mechanisms change depending on the traffic load in the network so that more resource-consuming mechanisms are assigned to a given class when the bandwidth is abundant than when the bandwidth is scarce. The scheme leads to a high utilization of network resources. Lee *et al.* [52] propose to base differentiation on the protection offered to the user. Protection bandwidth is regarded as the main cost of the recovery scheme. The authors propose three service levels based on the amount of bandwidth assigned to a user, as shown in Table 4.

Grover *et al.* [6, 53–55] propose a framework with some similarities to [52]. The framework (Table 5) considers four different recovery mechanisms in a self-healing mesh network using link restoration, and suggests four service levels based on these mechanisms. After a failure, the gold class connections will always be restored. The best-effort connections are restored if there is enough spare capacity. The bronze class is unprotected, and the capacity of the economy class may be seized by the gold and silver class. Three different schemes for realization of the differentiation are proposed. Grover *et al.* give models for optimal network design using the proposed schemes and explore the effects of the differentiation with regards to bandwidth usage. Finally they evaluate the frequency of preemption of the economy class connections.[5]

[5] *The authors denote their framework as the Quality of Protection (QoP) after the work of Gerstel and Sasaki [56]. The term QoP is in this article used to denote this earlier work which is classified as a guaranteed differentiation framework. The work by Grover* et al. *is, however, classified as a structural differentiation framework since it does not incorporate the quantitative contribution of QoP.*

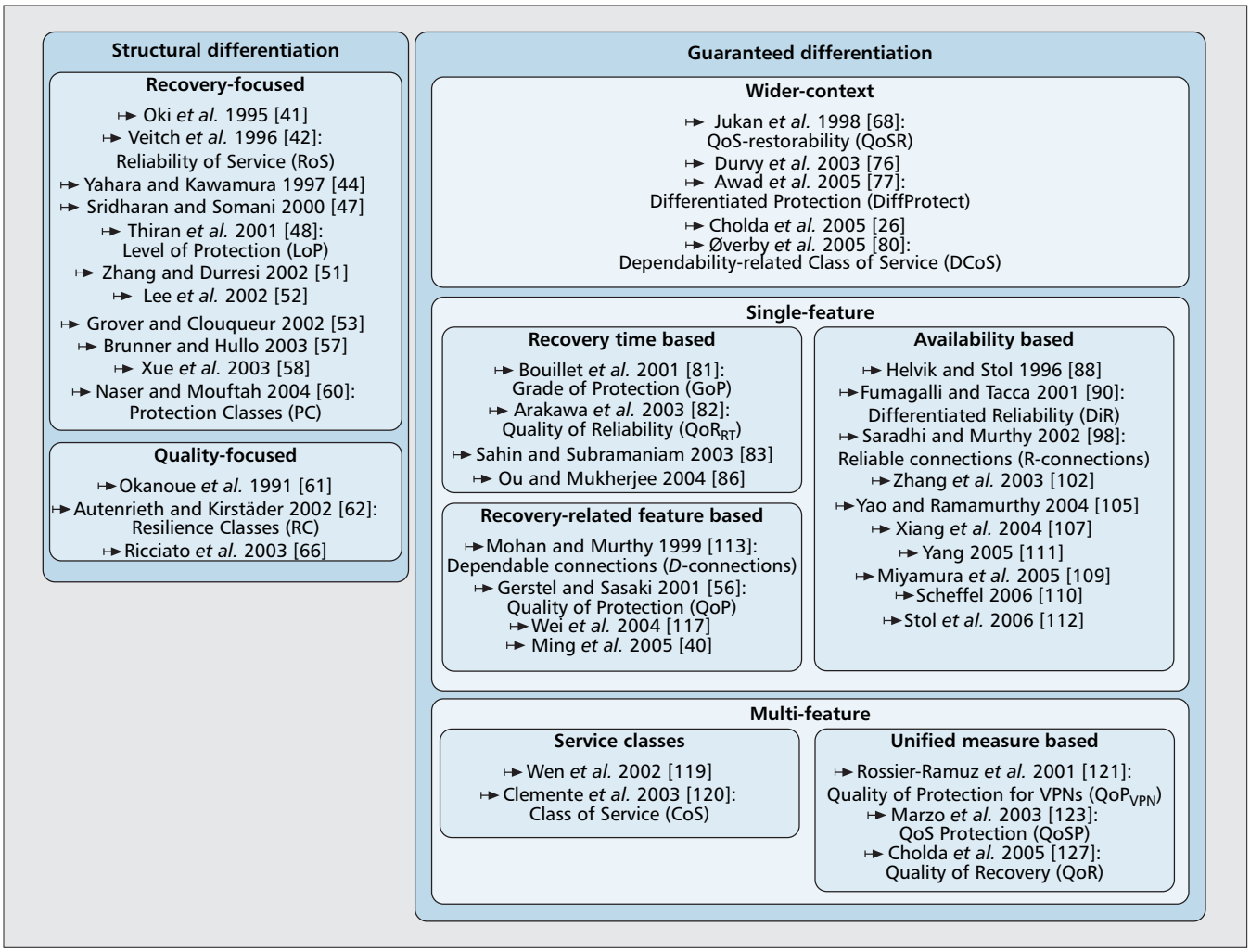


**Figure 6.** *Classification of resilience differentiation frameworks. Only the first publication related to each framework is mentioned.*

Brunner and Hullo [57] study resilience differentiation based on various schemes for shared backup paths in Generalized Multi-Protocol Label Switching (GMPLS) networks. The basis for the studies are customers' different needs for maximum time to recovery and minimum bandwidth of the restored path. These parameters could be mapped to demands for a given protection scheme, for example dedicated protection or shared path protection, although the authors do not address this mapping explicitly. The authors studied the amount of primary and backup bandwidth needed by the different protection schemes in different network topologies.

Xue *et al.* [58] present a mathematically complex scheme for routing connections based on a number of alternative paths. The authors study a concept of recovery trees [59] which enables them to design a network to be as much as possible resilient to multiple link failures that are caused by a single failure at a lower layer. However, the specific requirements for providing this resilience for selected connections are not considered in the article. Similarly, Naser and Mouftah [60] address resilience differentiation in multi-layer optical networks. Their framework is denoted Protection Classes (PC). The framework allows a connection to request protection against failures at any lower layer. The motivation for this is that a link failure at a given layer results in faults at all higher layers. Consequently, a connection protected against failures down to the lowest layer will have a better availability than connections with protection only at the client layer. The authors define a set of recovery classes defined so that PC-*n* is protected against link failures in *n* layers below the client layer. The authors introduce SRLG trees which may be used to compute disjoint paths that fulfill the requirements of the different PCs.

***Quality-Focused Frameworks*** — Frameworks in the quality-focused structural differentiation class are similar to the recovery-focused frameworks, except for the fact that they include some additional aspects with a wider context than the *quality of resilience* features presented earlier. Some aspects that are treated are *operation-related features* or features related to the QoS performance. Thus, frameworks in this group give a more thorough description of parameters related to different recovery schemes. If an SLA is constructed on their basis, the description of classes can be used to choose the most suitable recovery method.

The earliest work in this survey is the framework presented by Okanoue *et al.* [61]. The authors recognize that different survivability schemes are needed to satisfy different customer requirements. Hence, they compare three existing self-healing schemes for SONETs and propose to assign different schemes to different customers based on their demands. The framework is sketched in Table 6. Surprisingly, this article is not referenced by any other works mentioned in this survey, and it took four years before the next paper on resilience differentiation was published.

Autenrieth and Kirstädter [62, 63] present a framework, similar to the one given by Okanoue *et al.*, denoted as

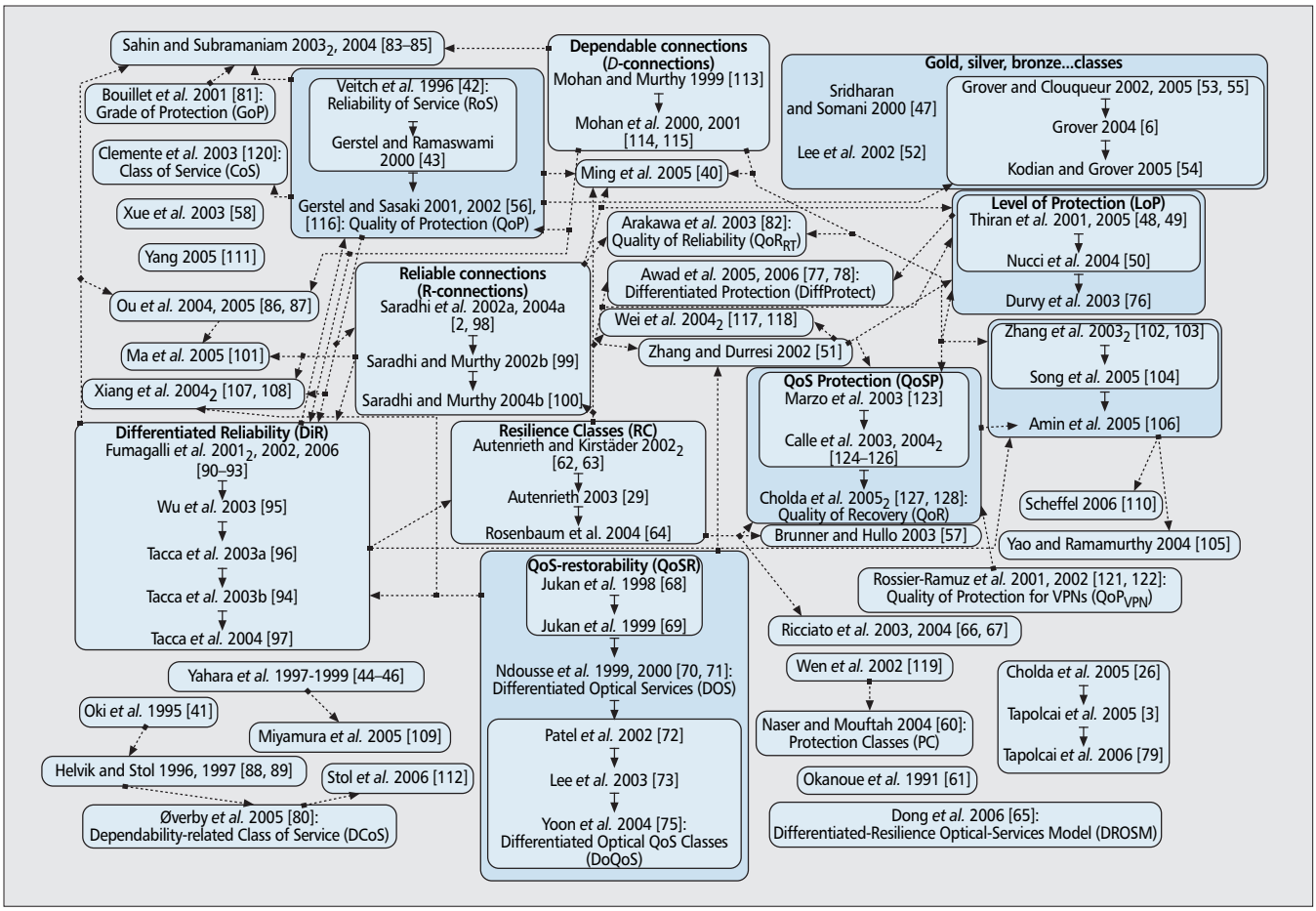


**Figure 7.** *Literature map of the resilience differentiation frameworks. Indices added to years indicate a number of papers published during a single year (only if different than one). Additionally, for more important frameworks: solid arrows denote an explicit inheritance, i.e., actual usage of a previously published idea; dotted arrows denote referenced citation with only a limited inspiration.*

Resilience Classes (RC). The framework is related to MPLS networks, and more specific data related to recovery time is given (in [29] a thorough model of recovery time, for the most representative recovery schemes, is derived). In addition to the recovery methods, the QoS of the recovery path is taken into account. The RC framework is presented in Table 7.

The RC framework is adopted for provisioning of the differentiated resilience in programmable virtual networks by Rosenbaum *et al.* [64]. Dong *et al.* [65] propose a framework denoted as the Differentiated-Resilience Optical-Services Model (DROSM) for classical circuit-switched optical networks. However, compared to the RC framework, DROSM does not offer any new ideas concerning the resilience differentiation. From the point of view of this survey, DROSM is seen as an application of the Resilience Classes.

Ricciato *et al.* [66] (and the extended journal paper [67]) propose differentiated protection against single and double failures in GMPLS networks. The QoS after the failure is taken into account since some classes use shared protection while other use dedicated protection. In the framework, the following five service classes are distinguished:

- *Unprotected*
- *Shared protection against single failures*: one backup path is established, using shared protection
- *Dedicated protection against single failures*: one dedicated backup path is established
- Shared protection against double failures: *two backup paths are established with shared protection*
- Dedicated protection against double failures: *two dedicated backup paths are established*

In case of spare capacity shortage, connections with shared protection against double failures can preempt connections with shared protection against single failures. The authors study the complexity of algorithms implementing the proposal, and analyze the effect of application of it on the recovery time, the probability of successful recovery when multiple failures occur and the cost of used capacity for the five resilience classes.

## Guaranteed Differentiation

Frameworks in the guaranteed differentiation category base the differentiation on guarantees related to the level of resilience offered to a customer. The customers choose the resilience level they need, while the provider is responsible for deciding on the necessary mechanisms that should be allocated to their connections.

| RoS grade | Reliability | Resource redundancy | Recovery time |
|---|---|---|---|
| 1+1 protection | High | High | Short |
| Shared protection | Medium | Medium | Short |
| Not recovered | Varies | Low | Long |

Table 3. *Grades of Reliability of Service (RoS) [42].*

| Service level | Protection Plan |
|---|---|
| Gold | 1+1 or 1:1 protection |
| Silver | *M*:*N* protection |
| Bronze | Restoration |

Table 4. *Differentiated service level based on recovery classes presented in [52].*

| Service level | Scheme 1 | Scheme 2 | Scheme 3 |
|---|---|---|---|
| Gold | Protected | Protected | Protected |
| Silver | N/A | Best-effort | Best-effort |
| Bronze | Not recovered | Not recovered | Not recovered |
| Economy | N/A | N/A | Preemptive |

Table 5. *Differentiated service levels proposed by Grover et al. [53].*

| Recovery scheme | Classification criteria | | | |
|---|---|---|---|---|
| | Recovery time | Flexibility | Multiple-failure handling | Spare capacity usage |
| Dedicated link protection | Excellent | Poor | Fair | Poor |
| Shared ring protection | Good | Fair | Fair | Good |
| Dynamic re-routing | Fair | Good | Good | Excellent |

Table 6. *Characteristics of different recovery schemes given in [61].*

While for the structural differentiation the definition of service classes as a mapping to some recovery mechanisms is obvious and natural, this is not the case for the guaranteed differentiation. Nevertheless, a group of the works involved here introduces a limited number of service classes. Other frameworks for guaranteed differentiation let each connection put requirements on one or multiple resilience parameters.

The frameworks for the guaranteed differentiation are divided into three subcategories. In the wider-context frameworks, differentiation is based on one or more resilience parameters, but also other parameters related to the QoS offered to the user are part of the framework. In the two other subcategories, resilience is the focus of the differentiation. The two categories are distinguished based on whether they take a single or multiple parameters related to resilience into account.

***Wider-Context Based Frameworks*** — Among frameworks for which specific requirements of services fund the basis of differentiation we have proposals that perceive resilience differentiation as a part of a more broad Quality of Service differentiation (generally provided in an integrated manner). All of these frameworks are related to either optical or IP-over-optical networks.

Jukan *et al.* [68, 69] present the idea of the QoS-restorability (QoSR). The term QoSR is related to a resilience parameter of optical connections, defined as the probability that a connection is successfully restored after a failure. In addition to the QoSR parameter, the connections may have different requirements concerning the transmission quality of the lightpaths they are assigned to, since different wavelengths have different BER. In [69] related to four service classes is given in a general, although numerical, way (Table 8). From the viewpoint of resilience, the framework can be recognized as a credible one, as the specific restorability level is given as an explicit requirement. The framework allows preemption: the lower class services can use higher class' backup resources, but when necessary they must be released to recover higher class connections.

Ndousse and Golmie [70] (and the extended journal version [71]) present an extension of the QoSR idea, denoted as the Differentiated Optical Services (DOS) framework. In addition to the fraction of recovered traffic and BER, the framework includes requirements on monitoring, security and provisioning with respect to signal generation in each node in the lightpaths (Table 9).

The idea of the QoS-restorability is adopted in three papers. Patel *et al.* [72] focus mainly on the attack-induced fault-tolerance issue. As a consequence, the importance of the resilience is decreased and the security is emphasized. Based on the Patel *et al.*'s work, Lee *et al.* [73] introduce resilience differentiation inspired by the Differentiated Services (DiffServ) model [74]. The recovery scheme and recovery time are used with the same priority as, for instance, BER or signal to noise ratio, similar to the Ndousse and Golmie's framework. This framework is adopted as Differentiated Optical QoS Classes (DoQoS) in the Yoon *et al.* paper [75] as a proposal of optical Virtual Private Networks (VPNs) provisioning foundations. Focusing mainly on implementation issues, none of the four mentioned papers present any new ideas for the resilience differentiation.

Later, the LoP framework was presented. Later, the same team published a paper [76] which we treat as an extension of LoP. Similarly as in [48] two service classes are considered, but their definitions are new: Fully Available (FA) traffic has requirements on the availability, end-to-end delay and loss while Partially Available (PA) traffic does not. PA connections are restricted to use capacity not needed by FA connections. If necessary, PA services can be preempted to free capacity for FA services.

Awad *et al.* [77] (and the extended journal version [77]) consider whether DiffServ can provide sufficient QoS to the three traffic classes, i.e., Expedited Forwarding (EF), Assured Forwarding (AF) and Best Effort (BE), when there are failures present in the network. Link failures are translated to bandwidth reductions at the IP layer, and DiffServ is used to prioritize the packets depending on their classification. As an alternative to use DiffServ as described, the authors present the Differentiated Protection (DiffProtect) model. DiffProtect is related to three protection schemes (dedicated, shared and unprotected) applied at the physical layer. In the papers, no particular resilience parameter requirements are taken into account, thus, the idea is similar to the service classes structural differentiation frameworks. However, we find this concept, as of possible practical and standard-based realization combined with existing QoS frameworks, an attractive

approach and an interesting path for further research.

Traditionally, resilience parameters were treated as a subset of QoS components. Cholda *et al.* [26] propose to partition the QoS concept into the "traditional QoS parameters,' related to the transmission quality in the fault-free state and the parameters related to resilience. The latter parameters are integrated in a concept denoted "quality of resilience" which constitutes availability, *MUT*, recovery time and traffic affected by failures. By treating the two classes of parameters as orthogonal, four service classes based on whether each of the class of parameters are guaranteed or not could be defined. However, as some sort of recovery is necessary to ensure QoS guarantee after a failure occurrence, existence of a class where transmission quality is guaranteed while resilience is not, is not possible. The resulting framework is presented in Table 10. The idea is further elaborated in [3, 79], where the locations of different QoS and resilience parameters in different layers of a network are given. Additionally, to enable resilience differentiation, the probability density function of the recovery time is defined.

∅verby *et al.* [80] also consider the idea of different resilience classes combined with different performance classes. The resilience class framework is denoted as the Dependability-related Class of Service (DCoS), and the resilience parameters involved are availability and recovery time.

| Service class | Resilience scheme | Resilience requirement | Recovery time | QoS after recovery |
|---|---|---|---|---|
| RC1 | Protection | High | 10-100 ms | Equivalent |
| RC2 | Restoration | Medium | 100 ms –1 s | May be temporarily reduced |
| RC3 | Re-routing | Low | 1 – 10 s | May have reduced QoS |
| RC4 | Preempted | N/A | Repair time | N/A |

Table 7. *Resilience Classes (RC) framework [29].*

| Service class | QoSR: fraction of recovered traffic requested | Transfer quality requested |
|---|---|---|
| $S_0$ | 100% | Low |
| $S_1$ | 80% | Medium |
| $S_2$ | >50% | Medium |
| $S_3$ | >50% | High |

Table 8. *Example of service classes distinguished in the QoS-restorability (QoSR) framework [69].*

***Single-Feature Frameworks*** — A large number of frameworks base the differentiation on a single *quality of resilience* feature. These frameworks will be interesting for a number of services for which one parameter is of a dominating importance.

As discussed earlier, there are three *reliability attributes* that could potentially be the basis for differentiation: the continuity, downtime and availability. However, there are no publications that base differentiation on continuity explicitly. Consequently, the two following paragraphs present frameworks that are related to the recovery process, which is related to the downtime, and then to the availability. In the last paragraph, frameworks that base differentiation on one of the recovery-related features given previously are characterized.

**Frameworks based on the Recovery Time Attribute** — To the authors' knowledge, four groups of works that use the time to recover as the basis for differentiation have been published. All frameworks are based on protection and can only guarantee recovery after single failures.

Bouillet *et al.* [81] introduce a framework for differentiation based on the recovery time requirements. The framework is denoted as Grade of Protection (GoP) and has three classes as indicated in Table 11. The recovery times are calculated excluding propagation delays.

Arakawa *et al.* [82] propose a framework denoted Quality of Reliability ($QoR_{RT}$). In this framework, all connections have a dedicated protection path, and the recovery time is the time it takes to switch to the backup path. There is an infinite number of classes. Therefore, we call them grades. They are determined on the basis of the maximum recovery time. In the methodology proposed so far, the grades are related to a linear function of recovery times. The best grade $QoR_{RT_1}$ is related to the recovery time equal to the minimal recovery time: $\mathrm{RT}(QoR_{RT_1}) = RT_{\min}$, whereas the subsequent grades are determined according to the following formula:

$$RT(QoR_{RT_n}) = RT_{\min} + RT_{\text{scale}} \times (n - 1) \qquad (2)$$

where $RT_{\text{scale}}$ represents the scale factor. The worst grade $QoR_{RT_\infty}$ is a special case where no recovery is provided.

| Service class | Classification criteria | | | | |
|---|---|---|---|---|---|
| | Transmission quality [BER] | Fraction of recovered traffic requested | Monitoring | Security | Provisioning |
| Class 1 | $10^{-9}$ | 90% | Intrusive | Secure | 1R |
| Class 2 | $10^{-7}$ | 70% | Intrusive | Unsecure | 2R |
| Class 3 | $10^{-5}$ | 20% | Nonintrusive | Unsecure | 3R |

1R refers to optical re-amplification only, 2R to optical re-amplification and re-shaping, and 3R to optical re-amplification, re-shaping and re-timing, respectively.

Table 9. *Optical services classification given in the Differentiated Optical Services (DOS) model [26].*

| | Resilience parameters are guaranteed | Resilience parameters are not guaranteed |
|---|---|---|
| Transfer quality is guaranteed | Class 1 | ∅ |
| Transfer quality is not guaranteed | Class 2 | Class 3 |

Table 10. *Classes based on the orthogonalization of Quality of Service and resilience parameters [26].*

| Grade | Recovery time requirement |
|---|---|
| Platinum | <50 ms |
| Gold | 50–100 ms |
| Silver | ~1–10 s |

Table 11. *Grade of Protection (GoP) levels based on [81].*

Sahin and Subramaniam [83–85] state that different recovery time requirements may need prioritized scheduling of control messages while shared protection is performed: in optical networks optical cross-connects (OXCs) in backup paths must be reconfigured. As the information related to such an operation must be processed and usually in a single OXC many such messages have to be handled, the contention forces delays due to control message scheduling. The authors propose to prioritize such a scheduling according to increasing required recovery times of recovered connections. A scheduling of control messages is proposed and simulation results showing that such an approach guarantee to meet the recovery time requirement (i.e., fraction of connections recovered in an assumed time is satisfactory) are presented. From the resilience differentiation viewpoint, the authors assume three service classes extracted from the Bouillet *et al.*'s work, i.e., platinum class: recovery time up to 50 ms, gold class: 100 ms and silver class: 1 s, respectively.

Ou *et al.* [86, 87] study WDM networks and consider provisioning of lightpaths with different recovery time requirements. The authors base their scheme on a variant of segment protection since segmented protection can provide shorter recovery times than path protection. The recovery time is mapped to a maximum hop count of the backup segments, and a method for providing backup paths that meet different recovery time requirements is proposed. The authors show that it is possible to establish more connections with the proposed approach than it would be possible if shared path protection was used.

**Frameworks based on the Availability Attribute** — The largest group of the frameworks bases the differentiation on the availability attribute. The first proposals related to this group appeared near mid-1990's and new works have been published continuously until now. From the differentiation point of view, many of the papers reiterate earlier ideas and primarily add new applications or practical implementation novelties.

The first framework in this category is proposed by Helvik and Stol [88] (and in the journal version [89]). The users of class $i$ are offered an unavailability guarantee of $U_i$ where $U_i > U_j$ if and only if $j > i$. The authors state that the probability of failure states, where all traffic of service level $i$ cannot be carried, should be smaller than $U_i$. A case study with two classes, high and low, in an ATM network, is investigated. In the study, two paths are established between any node pairs. The network is dimensioned so that all traffic can be carried in the fault-free state, and all traffic of the high class can be carried in all single failure cases. The study investigates what guarantees might be offered to the users in the case scenario.

A framework which thoroughly takes numerical availability requirements into account is proposed by Fumagalli and his team in a large number of papers. The framework is denoted as the Differentiated Reliability (DiR). The method is first applied in Wavelength Division Multiplexing (WDM) rings (with or without wavelength converters) in [90, 91] (and the extended journal version [92]). In [93, 94] is extended to mesh networks with shared path protection (thus, it is called SPP-DiR). While in the mentioned papers the framework is envisaged to be used with protection schemes, in [95] this team presents also an approach in which DiR is used in WDM networks applying restoration. Further extensions of the idea cover resilience to multiple failures [96, 97]. The last paper also uses the concept of SPP-DiR with the combination of the Routing and Wavelength Assignment problem.

In the simplest version, designed for the single failure assumption, each connection in a network has a Maximum Failure Probability (*MFP*) requirement. *MFP* is defined as "the probability that the connection is unavailable due to the occurrence of a fault in the network" [90]. A set of resilience classes is defined so that each class c is defined by a maximum acceptable level of *MFP* (MFP($c$)).[6] The authors assume that if the level of *MTBF* parameters of a network enables it, the connections do not have to use any recovery mechanisms if their working paths meet the *MFP* requirements.

The focus of many papers related to DiR is optimization: to meet the connections' *MFP* requirements and simultaneously to dimension the network using the optimal amount of resources. On the other hand, the focus of [95] is a protocol aiming at meeting the requirements when restoration is used. The authors take into account preemption: lower class connections may use spare resources reserved for higher classes, but they are preempted when necessary. The probability of preemption is taken into consideration when the acceptable level of *MFP* for preempted classes is checked. In [94, 96] DiR is extended to multiple failures by protecting the traffic with multiple shared backup paths. In this article, *MFP* is substituted by the *Maximum Downtime Ratio MDR*. The idea is related to assurance that due to multiple failures a connection will not be discontinued above the selected limit. For each connection, a recovery scheme which optimizes the cost is selected.

Saradhi and Murthy [2, 98] present an idea called the *Reliable Connections* (R-connections) which is similar to the DiR framework. All connections are associated with an availability requirement. If the working path has a satisfactory availability

[6] *Accepted level of MFP is called in further papers [92, 93] Maximum Accepted Failure Probability (MAFP) and it has an unconditional character. In [92], the authors introduce the conditional version of MFP which is called Maximum Conditional Failure Probability (MCFP). When the idea is extended to multiple failures in [94, 96], the authors use notion Maximum (Acceptable) Downtime Ratio (M(A)DT). As this is the downtime ratio, it must not be confused with the recovery time requirement. It is also called the 'reliability degree' [92, 94, 96]. However, the idea is generally the same.*

| Recovery service class | Recovery grade |
|---|---|
| Guaranteed recovery | $Q(C) = 1$ |
| Best effort recovery | $0 < Q(C) < 1$ |
| Not recovered | $Q(C) = 0$ |
| Preempted | $-1 \leq Q(C) < 0$ |

Table 12. *Service classes in the Quality of Protection (QoP) model [56].*

level, no protection path needs to be created, and thus some savings can be achieved in the network. The most original idea presented by the authors is that if a working path does not have a sufficiently high availability to meet the requirement, only a backup segment is added to the connection. If the whole structure has a satisfactory availability level, the connection is established. From the cost viewpoint, this is a large step forward in comparison to the DiR concept. While in [98] the authors present a static planning in WDM networks, in [99] they consider dynamic establishment of R-connections. In [100] the usage of R-connections in Optical VPNs is studied. Ma *et al.* [101] extend the idea of R-connections to include a hop count limitation which measures recovery time in some way, similarly as in the works by Ou *et al.*.

A large number of frameworks propose availability based differentiation in various contexts. In many cases, it seems that the concepts are formulated independently of each other. However, because of their similar nature with regards to the classification given in this article, the works published after the above presented frameworks are only reviewed shortly below. For instance, Zhang *et al.* [101–104] present a framework where, contrary to the above presented frameworks, mainly a case of shared protection is considered. In effect, the way in which the availability is calculated is changed. While [102] presents heuristics which optimize the usage of wavelength links subject to the requested connections availability, in [103] the exact, integer linear programming problem is given. In [104] the team changes focus from static to dynamic establishment of lightpaths with the selected availability levels. The idea is also adopted by Yao and Ramamurthy [105] to plan traffic grooming for differentiated availability guaranteed connections in WDM Metropolitan Area Networks and by Amin *et al.* [106] to differentiate connections by using variable routing costs. Also, two papers by Xiang *et al.* [107, 108] present a similar idea: a traffic grooming in WDM mesh networks with shared protection. The authors study situations where a backup path can carry only a fraction of the bandwidth offered by a primary path. Another idea, where the availability of a connection is a basis for the differentiation, is presented in the work by Miyamura *et al.* [109] where also the influence of SRLG on this performance parameter, as well as practical realization of multi-level resilience services in GMPLS networks (control plane considerations) is covered. Next, the work by Scheffel [110] presents the integer linear programming problem where unavailability requirements are included as constraints and where the cost (i.e., capacity usage) is minimized. An interesting fact is that the availability based differentiation has also been studied for free-space optical networks by Yang [111]. Short range links are used to meet the high availability requirements, while a lower class with less stringent requirements can use longer links. Stol *et al.* [112] consider how different levels of availability could be provided in a hybrid optical network architecture. In the network considered, both a circuit switched and a packet switched service is provided to the users. The article proposes to use an internal redundancy of the nodes in the architecture as a basis for the differentiation.

**Frameworks based on a Recovery-Related Feature** — There are three frameworks that base differentiation on a single *recovery-related quality of resilience* feature. The *Dependable connections* framework is based on the failure coverage. The *Quality of Protection* framework is related to the failure coverage or the quality of the recovery path depending on the realization of the framework. And the last framework, proposed by Wei *et al.*, is based on the quality of the recovery path.

Mohan and Murthy present a framework denoted as the *Dependable connections* (*D*-connections) in [113–115]. This proposal is published in two versions. The first version, presents a differentiation of two types of connections. A connection may either require fault-tolerance (*D*-connections) or not (*ND*-connections) [113]. This proposal is similar to the works in [48]. The second version is presented in [114] (and the extended journal version [115]) and is the main focus here.

Although the idea is only given implicitly, it inspired some further frameworks (Fig. 7). The authors deal only with *D*-connections which are formed by a pair of a working and a backup path. The framework concerns a dynamic situation where some connections (i.e., lightpaths) are established and some connections are terminated during operation. The framework uses a variant of shared protection, denoted as the backup multiplexing where only a small fraction of the capacity of a connection is reserved on the backup path. To reduce the network cost, the authors invent a new type of multiplexing where one primary path may share links with one or more backup paths. A *D*-connection having a backup path which is currently used by a primary path is denoted an orphan. If a failure happens in an orphan's primary path, the connection will be blocked, and therefore this scheme reduces the restoration probability. The authors design algorithms for achieving low blocking probabilities and at the same time keep band-

[7] *The notion QoP is used for instance in [6, 58, 86] but the works presented in mentioned papers are based on a distinct idea of the differentiation.*

| Service class | Classification criteria | | | |
|---|---|---|---|---|
| | Resil. req. | Rec. time [ms] | Fraction of rec. traffic requested | Scheme |
| High priority traffic | High | 0 ÷ 200 | 100% | Optical layer 1+1 protection |
| Low priority traffic | Low | 200 ÷ 2000 | Config.: 0 ÷ 100% | Electrical layer shared protection |
| Extra traffic | N/A | N/A | 0% | Preempted |

Table 13. *Differentiated resilience scheme given in [118].*

| | | Downtime | | | |
|---|---|---|---|---|---|
| | | 50 ms | Hundreds of ms | Seconds | Hours |
| Number of failures with guaranteed recovery | 3 | — | Platinum+ | Platinum– | — |
| | 2 | Gold++ | Gold+ | Gold– | — |
| | 1 | Silver++ | Silver+ | Silver– | — |
| | 0 | — | — | — | Bronze |

Table 14. *Class of Service (CoS) framework [120].*

width consumption low by controlling the average number of orphans in the network (the orphan threshold). The authors suggest that the orphan threshold could be used as a basis for differentiation, but do not study this issue further.

Gerstel and Sasaki propose a framework denoted Quality of Protection (QoP)[7] in [56] (and the extended version [116]). Based on the RoS, four resilience classes are defined. Each class $C$ is related to the measure $Q(C)$, as shown in Table 12. Connections from the first two classes have their own protection paths. A "not recovered' class has no reserved backup resources, and the preempted class uses the backup capacity of the first two classes in the failure-free situation. The bandwidth of a connection $C$ is denoted as B($C$). For each link in the backup path, a bandwidth equal to $B(C) \times Q(C)$ is reserved for backup capacity. The authors propose two implementation schemes based on the QoP framework. Regardless of the scheme, the guaranteed recovery class connections will always be recovered with the same bandwidth as before, while connections in the third and fourth class will never be recovered. In the *probabilistic scheme*, all best-effort recovery connections are either recovered with the same bandwidth as used before the failure (with probability $Q(C)$), and otherwise they are blocked. The preempted connections are preempted with a probability of $-Q(C)$ when a failure occurs in the path the connection borrows spare bandwidth from. In the *deterministic scheme*, each connection in the best-effort recovery class is guaranteed a reduced bandwidth which is equal to $B(C) \times Q(C)$, while the preempted class connections are guaranteed a capacity equal to $B(C) \times [-Q(C)]$. The choice of an implementation scheme must be based on the capacity of the network and the needs of the services the connections are dedicated to. In case of a failure in a connection's primary path, the guaranteed recovery class connections will always be recovered. The idea of QoP was adopted by Ming *et al.* [40] to cover schemes with backup multiplexing (without taking the preempted class into account).

Wei *et al.* [117] (and in its extended journal version [118]) present a framework for multi-layer IP-over-WDM networks. In the network considered, LSPs could be formed both at the optical and the electrical layer. The main difference between the two layers is that recovery is generally faster at the optical layer than at the electrical layer. The authors define three service classes and requirements related to each of them. The framework is presented in Table 13. It considers two parameters related to the resilience: the recovery time and the fraction of recovered traffic requested. The recovery time is determined based on the chosen resilience scheme.

***Multi-Feature Frameworks*** — In this category we find frameworks that take more than a single *quality of resilience* feature into account. In the literature, two different approaches have been pursued. In the first approach, the frameworks are based on the idea of service classes, while in the second approach, the frameworks are based on a unified measure.

**Frameworks based on the Service Class Idea** — There are two papers which propose to define service classes based on a number of strictly resilience-related parameters. Contrary to the wider-context based guaranteed differentiation frameworks, they do not combine transfer/transmission performance issues with resilience, suggesting that the resilience is a separate group of requirements related to the connections.

The first work in this group, where service classes are given implicitly, is presented by Wen *et al.* [119] and proposes a distributed algorithm for establishing protected connections in WDM networks with requirements to the availability of the assigned paths and the recovery time in case of a failure. To reduce the recovery time, the authors propose to partition a large WDM network into regions and use segmented backups within each region and between adjacent regions. The authors also present guidelines how to improve the availability of the set of paths assigned to a connection. To increase the availability of the primary path, the shortest path is used. To increase availability of the backup path, the number of backup paths sharing a link should be low, and if necessary, multiple backup paths could be used.

In the second work of this group, Clemente *et al.* [120], the idea is denoted as the Class of Service (CoS). The framework mixes quantitative and qualitative (similarly as in the structural differentiation) parameters, but begins with the enumeration of requirements. Then, after combining two of them, i.e., downtime and the number of simultaneous failures with guaranteed recovery, a set of service classes is defined, as shown in Table 14. After definition of the service classes, each class is mapped to a specific recovery mechanism which assures that both requirements will be met. It seems that the framework can be extended by incorporating more parameters than the presented two. Hence, CoS could be extended to the n-dimensional space. However, it will not always be possible to find a recovery mechanism that is capable of meeting all n requirements.

**Frameworks based on the Unified Measure Idea** — A relatively small group of frameworks proposes to numerically assess a number of resilience parameters and then apply some weighting function(s) to combine them to obtain a unified measure which is used as a basis for differentiation. The first framework in this group, which is denoted as Quality of Protection for VPNs ($QoP_{VPN}$) is presented by Rossier-Ramuz *et al.* [121, 122]. The framework takes into consideration resource sharing and dependencies between multiple domains

| Method | QoSP |
|---|---|
| Global protection | $\alpha \times PL^N + \beta \times RT^N + \gamma \times RC^N$ |
| Local protection | $\gamma \times RC^N$ |
| Protection with reverse backup path | $\beta \times RT^N + \gamma \times RC^N$ |

*PL*: packet loss, *RT*: recovery time, *RC*: amount of resources, superscript *N* denotes that a parameter is normalized, $\alpha$, $\beta$, $\gamma$: scaling parameters.

Table 15. *Quality of Service Protection (QoSP) framework [123].*

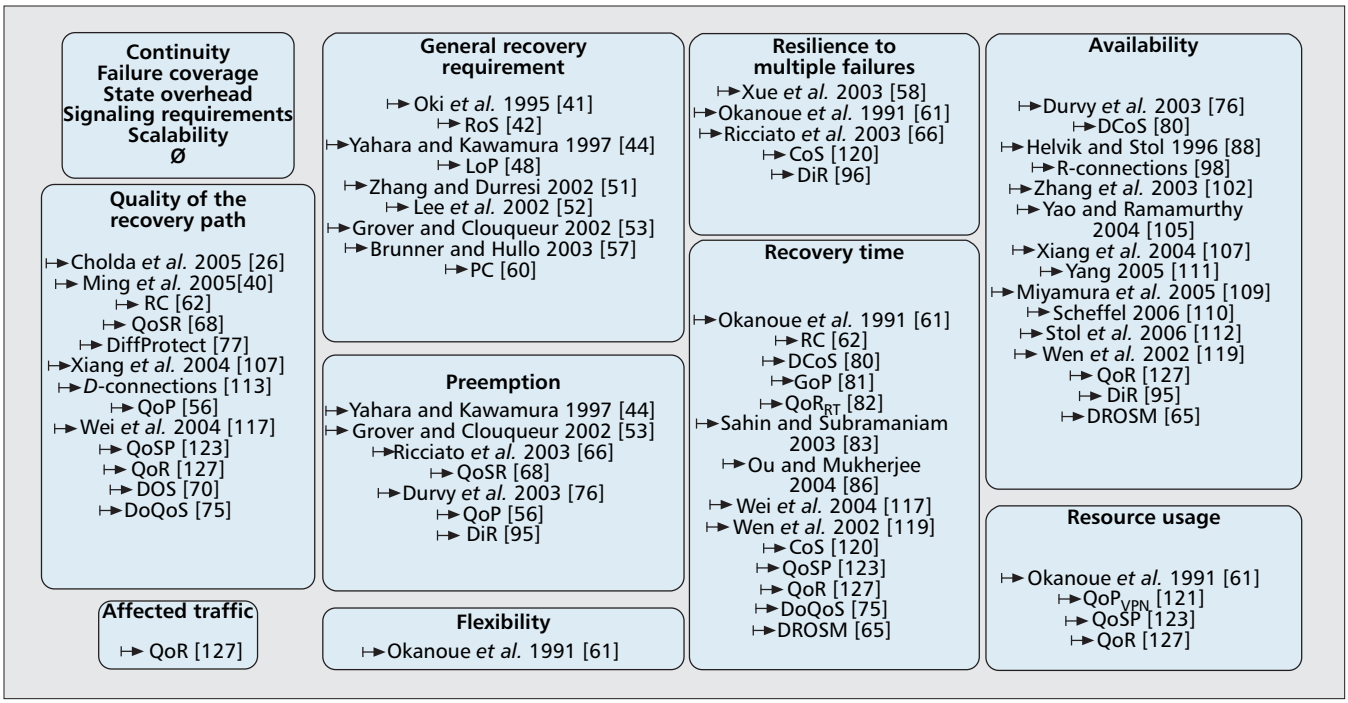


**Figure 8.** *Features used as a differentiation basis in presented frameworks. Only the first publication on each framework is mentioned. The used abbreviations are explained in Fig. 7.*

and layers. Thus, $QoP_{VPN}$ is related to a complex function. Taking into account multi-layer networks and striving to use a single combined function seems to be the aim of the efforts in this approach. However, it is not clear how to assess the multi-layer and multi-domain interdependencies, and the idea was not further elaborated by its authors.

Marzo *et al.* [123–126] propose the Quality of Service Protection (QoSP) framework which is related to the idea of gathering some factors which influence the quality of recovery and combine them to obtain a single integrated parameter. The methodology takes into account the following three factors: traffic lost, recovery time and resource consumption. The integrated parameter is a normalized weighted sum of the normalized factors. It is illustrated in Table 15. It can be noted that QoSP is calculated in a different manner for various recovery schemes. Additionally, the weights vary with different traffic classes. *QoSP* is calculated separately for each class, i.e., there are different scaling parameters. To compare parameters across different recovery methods, they are normalized.

The next idea, presented by Cholda *et al.* [127] (and the extended version in [128]), denoted as Quality of Recovery (QoR), is based on the evaluation of five recovery parameters (availability, quality of a recovery path, affected traffic, recovery time, cost of recovery proportional to spare resource usage) and on joining them to obtain a single numeric value which can describe a connection recovered by a particular scheme. The calculation consists of three phases: *abstraction*, *normalization* and *application(s)*. In the first step, values used to describe the operation of procedures in a given network (i.e., dedicated protection, path restoration, etc.) are derived. In the second stage, the values calculated in the first step are normalized, and a vector of unified parameters is obtained. Thus, the "normalization" process, based on the usage of the utility function idea, enables comparison of different recovery methods and ensures a bounded range of values. Its objective is also to introduce a desirable direction of changes of the parameter: the increase of it would mean that the quality of the procedure increases as well. Next, on the basis of a normalized vector, a unified QoR measure can be calculated. This can be done in different ways (for instance, as a weighted mean of vector elements). QoR calculated for each service can be used as an assessment when it is reasonable to use a particular scheme for a given connection, to evaluate its price, etc. in the "applications" phase.

## Summary, Comments and Assessment of Differentiation Concepts

In this section we discuss work carried out within the area. We assess the results based on the rationale given in the introduction. A point of view on future research is given in the next section.

The proposed schemes are partitioned into two main groups, structural differentiation and guaranteed differentiation. The main benefit of the structural differentiation frameworks is that the customers may easily understand how their connections are made resilient, and they may also verify, by active measurements, that the contract is met by the provider. This scenario may be attractive to some clients. The main objection against structural differentiation is that the quality of the service received by the customer does not only depend on the structural arrangements, but also on the resilience of the underlying network. Therefore, for instance, the "gold' service in one network could have the same availability as the "silver" service in another network. Additionally, if a resilience mechanism is provided and for some reasons does not work as the user implicitly expects, the operator is in principle not obliged to make an improvement to meet some "commonly expected" quality of the service delivered, as long as it is not part of the SLA. This could be a direct pessimistic consequence of the structural differentiation implementation in a network.

With guaranteed differentiation, the quality of service is the basis for an offer, so the customer should always receive the level of resilience agreed upon in the contract. In this case, the service differentiation is related to objective service requirements which are neither dependent on a network technology or topology, planning constraints, reliability characteristics of network hardware/software, operation and management, nor the size of the network.

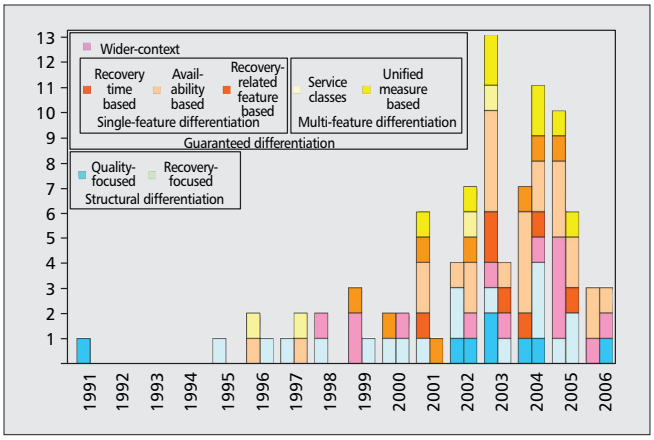


**Figure 9.** *Number of papers related to different frameworks published in years 1991-2006. For each year, the first column is related to conference/symposium/workshop papers, whereas the second column is related to magazine/journal papers, Ph.D. theses or books.*

It may, however, be difficult for the customer to verify that the received resilience level complies with the contract, as end-to-end measurements are needed. With statistical guarantees, very long measurement periods are needed to prove that the contract is violated. The providers, on their side, will need to over-provision the resources to be sure that an effective resilience is so high that the requirements agreed are met with a sufficiently high probability.

Figure 8 shows which features are and which are not covered by different frameworks. It is observed that the availability, recovery time and the quality of the recovery path are the most common. The resilience factors: failure coverage, state overhead, scalability, and signaling requirements, described earlier are not covered. This is likely due to the fact that all frameworks being characterized are related to the end-to-end (client) perspective, and not to a network operator view. Dealing with the four mentioned factors may, however, be regarded as a separate issue considered by the operator as elements in his handling of the service agreement.

A more serious omission is that the continuity feature is not taken explicitly into account, since there are services where this is the main requirement.[8] These are services where even short interrupts may be critical, and a low service failure intensity will dominate requirements and is not compensated by high availability and short downtimes. Tele-medicine applications and real-time control are mentioned as examples, and others may be foreseen. Thus, persistent connectivity will be necessary in the offering of new demanding applications.

All proposed schemes assume full insight in and control of resources and quality parameters. This is the case only when the intra-domain differentiation is studied. Nonetheless, in the inter-domain case, operators are reluctant to share information and, in general, do not accept resources in their networks to be managed by others or by an entity common to many operators. Hence, the lack of "full insight" makes the proposed schemes less useful. Furthermore, differentiation in the inter-domain case also poses a set of new questions on how the inter-autonomous-system topology should be used as well as on the interface between operators. Thus, the intra-domain differentiation issues should be addressed from this viewpoint.

From the literature map in Fig. 7 some observations can be made. First, there is a large number of publications, which may indicate that the resilience differentiation is an interesting area to researchers. However, the level of citations is relatively low. This could indicate that there are many various ways to approach the differentiation. It can also be observed that only a small number of the proposed frameworks are investigated in more than four papers. In just a few cases, a framework proposed by one research group has been adopted or extended by another group (e.g., RoS, RC, QoSR). Sometimes the authors of different publications of similar character are not aware of each other's work. We hope that this article may contribute to increased awareness among researchers about related work and encourage cooperation.

Although authors typically stress the need for resilience differentiation, their focus is often toward an assumed technique or technology, e.g., ATM, IP-over-(D)WDM or (G)MPLS networking. Thus, except for the introductory sections, the authors resign from being directed by application and user requirements and assume the perspective of what may efficiently be obtained by using recovery methods with a selected technology. As a result of this, fixed mappings between requirements and specific recovery schemes given in papers result sections are useful only for the particular network solutions studied.

All of the schemes are related to connection oriented techniques. As none of the authors discuss this issue, it is hard to adjudicate whether this stems from the fact that the works are driven by a selected technology, or if it is a general recognition that provisioning of differentiated resilience can only be performed on the basis of virtual circuits.

The majority of the papers are written from either an optimization or an operational point of view. While the first group of papers focuses on mathematical algorithms, the second group focuses on how to technically implement the presented differentiation framework or how to estimate quality parameters in a network. The techno-economical perspective, which studies how to combine technical differentiation with market opportunities, is not considered.

To conclude this section, we briefly comment on the distribution of proposals over the years, presented in Fig. 9 and in Table 16. Publications are partitioned into two groups: conference-type and journal-type, where the latter may be considered as more mature work. The idea of guaranteed differentiation appeared later than the idea of structural differentiation. The cause for this may be that the first time when the differentiation was considered, it was introduced as an additional feature when studying emerging recovery schemes. This kind of studies is still vivid, although they do not form a research mainstream. The differentiation in the wider context and especially based on availability seems to be the most popular.

## CHALLENGES

After identifying and summarizing the works carried out, this section outlines some major challenges to be solved before resilience differentiation may be successfully deployed and offered as a standard part of the portfolio.

Our position is that resilience parameters must be defined and met on the end-to-end basis. This seems to be what is implicitly or explicitly assumed by most researchers. However, the practical realization of this idea is very difficult. Communication takes place across several network segments (core, metro, access). Segments may be constituted by several (autonomous) systems, operated by different entities, applying varying technologies and having different quality characteristics. It should be kept in mind that today's access networks

[8] *It should be mentioned that Miyamura* et al. *[109] state that MTBF can also be used as a root of differentiation. However, the idea is not studied further.*

| Category | Conferences, symposia, and workshops | Journals, mags., Ph.D. theses, and books | Total |
|---|---|---|---|
| Recovery-focused | 9.68% | 9.68% | 19.35% |
| Quality-focused | 5.38% | 3.23% | 8.60% |
| Wider-context | 9.68% | 5.38% | 15.05% |
| Recovery time based | 4.30% | 3.23% | 7.53% |
| Availability based | 18.28% | 9.68% | 27.96% |
| Recovery-related feature based | 5.38% | 3.23% | 8.60% |
| Service classes | 2.15% | 2.15% | 4.30% |
| Unified measure based | 4.30% | 4.30% | 8.60% |

Table 16. *Distribution of the number of papers related to different groups of classification given in Fig. 6.*

form quality bottlenecks. This is the case for both wired and wireless access. Unless differentiation is limited to the core, resilience handling must be introduced in the access networks as well. Otherwise, differentiated services cannot be offered to end user terminals in the public domain. Moreover, differentiation in the core is constrained by the inter-domain (multi-domain) problems related to quality assurance and regulation when a connection crosses domains without common control and management planes. It is claimed that future networks will go toward the 3M characteristics: multi-layer, multi-domain, & multi-service. However, without a proper handling of the multi-domain resilience issue, the 3M postulate seems less up-to-date.

Main approaches to differentiation must be established and commonly accepted. Dealing with the following issues seems to be necessary:

- Important directions of future work could be to aim at harmonizing different schemes (e.g., availability and recovery time based) to make the resulting differentiation richer and more conforming with typical requirements. Since none of the schemes is prevailing, there seems to be a need for works which compare various approaches and try to show whether some of them are more advantageous than others. After this, standardization bodies could start to take up the issue.
- There are pros and cons related to structural and guaranteed differentiation. The former is more easily understood, and thus, operators may be more willing to use this group of frameworks. The latter forms the basis of a quantitative agreement between a provider and a user on the quality. This issue should also be studied, preferably in the real client-operator negotiation setting.
- Many frameworks propose a limited number of service classes (for instance, four). However, it is not clear if classes are necessary and whether it is a technical issue to find such classes. More studies on how to meet a strictly numerical requirement, as in the case of many availability-based frameworks, are needed. Service classes may then be determined at a later time if necessary.
- Understandable, preferably to the common user, easily predictable and measurable quality features are necessary to credibly offer the resilience differentiation. Although the means to measure some of them are established and known in the community, e.g., the availability or recovery time, there are practical challenges in performing measurements trusted by both parties of an agreement. Furthermore, the means for resilience evaluation and planning in a differentiated setting is limited and should be improved.
- Hereto, differentiation schemes focus on connection-oriented techniques. It should be evaluated if this really is a must; the issue is important as there are initiatives aiming at making the quality provisioning in Internet feasible without methods based on virtual circuits and complex signaling (e.g., flow-aware networking [129]).

Although not an issue in this article, it should be pointed out that for resilience differentiation to be put into operation, a business model is essential. The authors are not aware of any effort in this direction. Furthermore, an operational differentiation scheme must be supported by software (and possible hardware) implementations provided by different vendors. For this to become available, standardized recommendations are required. Standardization seems mandatory for the end-to-end differentiation, i.e., differentiation handling the access and multi-operational domains. With the current state-of-the-art, as outlined in the previous sections, it is seen that a substantial effort is needed.

## Conclusion

This article is the first extensive survey of ideas related to differentiation of communication services with respect to resilience. The general concept is introduced and the relevance and importance of the issue are outlined. The main approaches to provide differentiated resilience are introduced, and a description, classifications and discussions of works in the area are presented. As a result of this effort, it is seen that there are no clear candidates for consented resilience differentiation schemes. The works in the area are uncoordinated and a number of important issues like differentiation in the access network, inter-domain/-carrier co-operation and differentiation as a part of a business model, are not dealt with. In spite of its importance, the area is not thoroughly recognized and far from mature. This is unsatisfactory, and hopefully, by providing an overview of the work done, this survey will contribute in putting resilience differentiation on the agenda, and help researchers, operators and others to target their efforts efficiently.

### Acknowledgements

This work was prepared as a joint research effort within the EU FP6 NoE EuroNGI project (http://www.eurongi.org) framework. The reported work was also supported by the Polish Ministry of Science and Higher Education under grant N517 013 32/2131. The Centre for Quantifiable Quality of Service in Communication Systems (Q2S) is funded by The Research Council of Norway, NTNU and UNINETT, under the Norwegian Centre of Excellence (CoE) scheme.

## References

[1] J. Gozdecki, A. Jajszczyk, and R. Stankiewicz, "Quality of Service Terminology in IP Networks," *IEEE Commun. Mag.*, vol. 41, no. 3, Mar. 2003, pp. 153–59.
[2] C. V. Saradhi, M. Gurusamy, and L. Zhou, "Differentiated QoS

for Survivable WDM Optical Networks," Optical Communications Supplement to *IEEE Commun. Mag.*, vol. 42, no. 5, May 2004, pp. S8–S14.
[3] J. Tapolcai *et al.*, "Quality of Resilience (QoR): NOBEL Approach to the Multi-Service Resilience Characterization," *Proc. 1st IEEE/CreateNet Int'l. Wksp. Guaranteed Optical Service Provisioning GOSP 2005*, Boston, MA, Oct. 7, 2005.
[4] A. Avizienis *et al.*, "Basic Concepts and Taxonomy of Dependable and Secure Computing," *IEEE Trans. Depend. and Sec. Comp.*, vol. 1, no. 1, Jan./Feb. 2004, pp. 11–33.
[5] C. Oggerino, *High Availability Network Fundamentals*, Cisco Press, 2001.
[6] W. D. Grover, *Mesh-Based Survivable Networks. Options and Strategies for Optical, MPLS, SONET, and ATM Networks*, Prentice Hall PTR, 2004.
[7] J.-P. Vasseur, M. Pickavet, and P. Demeester, *Network Recovery, Protection and Restoration of Optical, SONET-SDH, IP, and MPLS*, Morgan Kaufmann Publishers, 2004.
[8] A. K. Somani, *Survivability and Traffic Grooming in WDM Optical Networks*, Cambridge University Press, 2005.
[9] B. E. Helvik and O. Wittner, "Network Resilience by Emergent Behavior from Simple Autonomous Agents," *Dependable Computing Systems. Paradigms, Performance Issues, and Applications*, H. B. Diab and A. Y. Zomaya, Eds., John Wiley & Sons, Inc., 2005, pp. 442–86.
[10] Y. Lee and B. Mukherjee, "Traffic Engineering in Next-Generation Optical Networks," *IEEE Commun. Surveys Tutorials*, vol. 6, no. 3, July/Sept. 2004, pp. 16–33.
[11] D. Papadimitriou and E. Mannie, Eds., "Analysis of Generalized Multi-Protocol Label Switching (GMPLS)-based Recovery Mechanism (including Protection and Restoration)," IETF RFC 4428, Mar. 2006.
[12] P. Demeester *et al.*, "Resilience in Multilayer Networks," *IEEE Commun. Mag.*, vol. 37, no. 8, Aug. 1999, pp. 70–75.
[13] D. Colle *et al.*, "Data-Centric Optical Networks and Their Survivability," *IEEE JSAC*, vol. 20, no. 1, Jan. 2002, pp. 6–19.
[14] E. Mannie, Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture," IETF RFC 3945, Oct. 2004.
[15] L. Nederlof *et al.*, "End-to-end Survivable Broadband Networks," *IEEE Commun. Mag.*, vol. 33, no. 9, 1995, pp. 63–70.
[16] P.-H. Ho, J. Tapolcai, and T. Cinkler, "Segment Shared Protection in Mesh Communications Networks with Bandwidth Guaranteed Tunnels," *IEEE/ACM Trans. Net.*, vol. 12, no. 6, Dec. 2004, pp. 1105–18.
[17] F. Ricciato, U. Monaco, and D. Ali, "Distributed Schemes for Diverse Path Computation in Multidomain MPLS Networks," *IEEE Commun. Mag.*, vol. 43, no. 6, June 2005, pp. 138–46.
[18] C. Huang and D. Messier, "A Fast and Scalable Inter-Domain MPLS Protection Mechanism," *J. Commun. Net.*, vol. 6, no. 1, Mar. 2004, pp. 60–67.
[19] S. Orlowski and R. Wessäly, "Comparing Restoration Concepts using Optimal Network Configurations with Integrated Hardware and Routing Decisions," P*roc. 4th Int'l. Wksp. the Design of Reliable Commun. Networks DRCN 2003*, Banff, Alberta, Canada, Oct. 19–22, 2003, pp. 39–46.
[20] G. Bernstein, B. Rajagopalan, and D. Saha, *Optical Network Control, Architecture, Protocols, and Standards*, Addison-Wesley, 2004.
[21] T. Cicic, A. Kvalbein, A. F. Hansen, and S. Gjessing, "Resilient Routing Layers and p-Cycles: Tradeoffs in Network Fault Tolerance," P*roc. IEEE 2005 Wksp. High Performance Switching and Routing HPSR 2005*, Hong Kong, China, May 12–14, 2005.
[22] T. Van Landegem, P. Vankwikelberge, and H. Vanderstraeten, "A Self- Healing ATM Network Based on Multilink Principles," *IEEE JSAC*, vol. 12, no. 1, Jan. 1994, pp. 149–58.
[23] H. Fujii and N. Yoshikai, "Restoration Message Transfer Mechanism and Restoration Characteristics of Double-Search Self-Healing ATM Network," *IEEE JSAC*, vol. 12, no. 1, Jan. 1994, pp. 149–58.
[24] M. Pióro and D. Medhi, Routing, *Flow and Capacity Design in Communication and Computer Networks*, Morgan Kaufmann Publishers — Elsevier, 2004.
[25] V. Sharma and F. Hellstrand, Eds., "Framework for Multi-Protocol Label Switching (MPLS)-based Recovery," IETF RFC 3469, Feb. 2003.
[26] P. Cho?da *et al.*, "Considerations about Service Differentiation Using a Combined QoS/QoR Approach," *Proc. 5th Int'l. Wksp. Design of Reliable Communication Networks DRCN 2005*, Lacco Ameno, Island of Ischia, Italy, Oct. 16–19, 2005.
[27] S. Ramamurthy and B. Mukherjee, "Survivable WDM Mesh Networks. Part II — Restoration," *Proc. IEEE Int'l. Conf. Commun. ICC'99*, Vancouver, Canada, June 6–10, 1999.
[28] S. Ramamurthy, L. Sahasrabuddhe, and B. Mukherjee, "Survivable WDM Mesh Networks," *IEEE/OSA J. Lightwave Tech.*, vol. 21, no. 4, Apr. 2003, pp. 870–83.
[29] A. Autenrieth, "Recovery Time Analysis of Differentiated Resilience in MPLS," *Proc. 4th Int'l. Wksp. the Design of Reliable Commun. Networks DRCN 2003*, Banff, Alberta, Canada, Oct. 19–22, 2003, pp. 333–40.
[30] N. S. C. Correia and M. do Carmo Raposo de Medeiros, "Protection Schemes for IP-over-WDM Networks: Throughput and Recovery Time Comparison," *Phot. Net. Commun.*, vol. 11, no. 2, Mar. 2006, pp. 127–49.
[31] "Terms and Definitions related to Quality of Service and Network Performance Including Dependability," ITU-T Rec. E.800, Aug. 1994.
[32] "Internet Protocol Data Communication Service — IP Packet Transfer and Availability Performance Parameters," ITU-T Rec. Y.1540, Dec. 2002. 24
[33] "Framework of a Service Level Agreement," ITU-T Rec. E.860, June 2002.
[34] "B-ISDN Semi-Permanent Connection Availability," ITU-T Rec. I.357, Nov. 2000.
[35] B. E. Helvik, "Perspectives on the Dependability of Networks and Services," *Telektronikk*, vol. 100, no. 3, 2004, pp. 27–44.
[36] F. Michaut and F. Lepage, "Application-Oriented Network Metrology: Metrics and Active Measurement Tools," *IEEE Commun. Surveys Tutorials*, vol. 7, no. 2, Apr./June 2005, pp. 2–24.
[37] B. Jæger and D. Tipper, "Prioritized Traffic Restoration in Connection Oriented QoS based Networks," *Comp. Commun.*, vol. 26, no. 18, Dec. 2003, pp. 2025–36.
[38] H. Lønsethagen, A. Solem, and B. Olsen, "Feasibility of Bandwidth on Demand. Case Study Approach, Models and Issues," EU FP6 IP IST-NOBEL Project internal presentation, Sept. 19–21, 2005.
[39] D. Griffith *et al.*, "Optimal Mixtures of Different Types of Recovery Schemes in Optical Networks," http://w3.antd.nist.gov/pubs/paper1.pdf
[40] C. Ming, Z. Luying, and M. Gurusamy, "Dynamic Routing of Dependable Connections with Different QoP Grades in WDM Optical Networks," *Proc. 10th IEEE Symp. Computers and Commun. ISCC'2005*, La Manga del Mar Menor, Cartagena, Spain, June 27–30, 2005.
[41] E. Oki, N. Yamanaka, and F. Pitcho, "Multiple-Availability-Level ATM Network Architecture," *IEEE Commun. Mag.*, vol. 33, no. 9, Sept. 1995, pp. 80–88.
[42] P. Veitch, I. Hawker, and G. Smith, "Administration of Restorable Virtual Path Mesh Networks," *IEEE Commun. Mag.*, vol. 34, no. 12, Dec. 1996, pp. 96–102.
[43] O. Gerstel and R. Ramaswami, "Optical Layer Survivability: a Services Perspective," *IEEE Commun. Mag.*, vol. 38, no. 3, Mar. 2000, pp. 104–13.
[44] T. Yahara and R. Kawamura, "Virtual Path Self-Healing Scheme Based on Multi-Reliability ATM Network Concept," *Proc. 1997 IEEE Global Telecommun. Conf. GLOBECOM'97*, Phoenix, AZ, Nov. 5–6, 1997.
[45] T. Yahara, R. Kawamura, and S. Ohta, "Multi-Reliability Self-Healing Scheme that Guarantees Minimum Cell Rate," *Proc. 1st Int'l. Wksp. Design of Reliable Commun. Networks DRCN 1998*, Brugge, Belgium, May 19–20, 1998.
[46] R. Kawamura and H. Ohta, "Architectures for ATM Network Survivability and their Field Deployment," *IEEE Commun. Mag.*, vol. 37, no. 8, Aug. 1999, pp. 88–94.
[47] M. Sridharan and A. K. Somani, "Revenue Maximization in Survivable WDM Networks," *Proc. 2000 SPIE Optical Networking and Commun. Conf. OptiComm 2000*, Richardson, TX, Oct. 22–26, 2000.
[48] P. Thiran *et al.*, "A Protectionbased Approach to QoS in Packet-over-Fiber Networks," *Proc. 2001 Thyrrhenian Int'l. Wksp.*

*Digital Communications IWDC'01*, Taormina, Italy, Sept. 17–20, 2001.
[49] A. Nucci *et al.*, "Increasing the Link Utilization in IP over WDM Networks Using Availability as QoS," *Phot. Net. Commun.*, vol. 9, no. 1, Jan. 2005, pp. 55–75.
[50] A. Nucci *et al.*, "Controlled Use of Excess Backbone Bandwidth for Providing New Services in IP-over-WDM Networks," *IEEE JSAC*, vol. 22, no. 9, Nov. 2004, pp. 1692–707.
[51] H. Zhang and A. Durresi, "Differentiated Multi-layer Survivability in IP/WDM Networks," *Proc. 2002 IEEE/IFIP Network Operations and Management Symp. NOMS 2002*, Florence, Italy, Apr. 15–19, 2002.
[52] S. Lee, D. Griffith, and N.-O. Song, "A New Analytical Model of Shared Backup Path Provisioning in GMPLS Networks," *Phot. Net. Commun.*, vol. 4, no. 3–4, July 2002, pp. 271–83.
[53] W. D. Grover and M. Clouqueur, "Span-Restorable Mesh Network Design to Support Multiple Quality of Protection (QoP) Service Classes," *Proc. 1st Int'l. Conf. Optical Commun' and Networks ICOCN'02*, Singapore, Nov. 11–14, 2002, pp. 305–08.
[54] A. Kodian and W. D. Grover, "Multiple-Quality of Protection Classes Including Dual-Failure Survivable Services in p-Cycle Networks," *Proc. 2nd Int'l. Conf. Broadband Networks BROADNETS 2005*, Boston, MA, Oct. 3–7, 2005.
[55] W. Grover and M. Clouqueur, "Span-Restorable Mesh Networks with Multiple Quality of Protection (QoP) Service Classes," *Phot. Net. Commun.*, vol. 9, no. 1, Jan. 2005, pp. 19–34.
[56] O. Gerstel and G. H. Sasaki, "Quality of Protection (QoP): a Quantitative Unifying Paradigm to Protection Service Grades," *Proc. Opt. Net. and Commun. Conf. OptiComm 2001*, Denver, CO, Aug. 21–23, 2001.
[57] M. Brunner and C. Hullo, "GMPLS Fault Management and Impact on Service Resilience Differentiation," *Proc. IFIP/IEEE 8th Int'l. Symp. Integrated Network Management IM 2003*, Colorado Springs, CO, Mar. 24–28, 2003, pp. 665–78.
[58] G. Xue, L. Chen, and K. Thulasiraman, "Quality-of-Service and Quality-of-Protection Issues in Preplanned Recovery Schemes Using Redundant Trees," *IEEE JSAC*, vol. 21, no. 8, Oct. 2003, pp. 1332–45.
[59] M. Medard, S. G. Finn, and R. A. Barry, "Redundant Trees for Preplanned Recovery in Arbitrary Vertex-Redundant or Edge-Redundant Graphs," *IEEE/ACM Trans. Net.*, vol. 7, no. 4, Oct. 1999, pp. 641–52.
[60] H. Naser and H. T. Mouftah, "A Multilayer Differentiated Protection Services Architecture," *IEEE JSAC*, vol. 22, no. 8, Oct. 2004, pp. 1539–47.
[61] Y. Okanoue, H. Sakauchi, and S. Hasegawa, "Design and Control Issues of Integrated Self-Healing Networks in SONET," *Proc. 1991 IEEE Global Telecommun. Conf. GLOBECOM'91*, Phoenix, AZ, Dec. 2–5, 1991.
[62] A. Autenrieth and A. Kirstädter, "Engineering End-to-End IP Resilience Using Resilience-Differentiated QoS," *IEEE Commun. Mag.*, vol. 40, no. 1, Jan. 2002, pp. 50–57.
[63] —, "RD-QoS — The Integrated Provisioning of Resilience and QoS in MPLS-based Networks," *Proc. IEEE Int'l. Conf. Commun. ICC 2002*, New York, NY, Apr. 28–May 2002.
[64] F. Rosenbaum *et al.*, "Resilience- Differentiation in Programmable Virtual Networks," *Proc. IEEE Int'l. Conf. Commun. ICC 2004*, Paris, France, June 20–24, 2004.
[65] S. Dong, C. Phillips, and R. Friskney, "Differentiated-Resilience Provisioning for the Wavelength-Routed Optical Network," *IEEE/OSA J. Lightwave Tech.*, vol. 24, no. 2, Feb. 2006, pp. 667–73.
[66] F. Ricciato *et al.*, "Performance Evaluation of a Distributed Scheme for Protection against Single and Double Faults for MPLS," *Proc. 2nd Int'l. Wksp. QoS in Multiservice IP Networks QoS-IP 2003*, Milano, Italy, Feb. 24–26, 2003.
[67] F. Ricciato, M. Listanti, and S. Salsano, "An Architecture for Differentiated Protection Against Single and Double Faults in GMPLS," *Phot. Net. Commun.*, vol. 8, no. 1, June 2004, pp. 119–32.
[68] A. Jukan, A. Monitzer, and H. R. van As, "QoS-restorability in Optical Networks," P*roc. 24th European Conf. Opt. Commun. ECOC'98*, Madrid, Spain, Sept. 20–24, 1998.
[69] —, "Service-specific Recovery of Wavelength Connections in WDM Networks," *Proc. Optical Fiber Commun. Conf. OFC'99*, San Diego, CA, Feb. 23–25, 1999.
[70] T. D. Ndousse and N. Golmie, "Differentiated Optical Services: a Quality of Optical Service Model for WDM Networks," *Proc. SPIE Conf. All-Optical Networking 1999: Architecture, Control, and Management Issues*, Boston, MA, 1999, Sept. 19–21.
[71] N. Golmie, T. D. Ndousse, and D. H. Su, "A Differentiated Optical Services Model for WDM Networks," *IEEE Commun. Mag.*, vol. 38, no. 2, Feb. 2000, pp. 68–73.
[72] J. K. Patel, S. U. Kim, and D. H. Su, "QoS Recovery Schemes Based on Differentiated MPLS Services in All-Optical Transport Next Generation Internet," *Phot. Net. Commun.*, vol. 4, no. 1, Jan. 2002, pp. 5–18.
[73] J.-D. Lee *et al.*, "Differentiated Wavelength Assignment with QoS Recovery for DWDM Next Generation Internet Backbone Networks," *Phot. Net. Commun.*, vol. 5, no. 2, Mar. 2003, pp. 163–75.
[74] S. Blake and T. Nadeau, Eds., "An Architecture for Differentiated Services," IETF RFC 2475, Dec. 1998.
[75] M.-R. Yoon *et al.*, "Optical-LSP Establishment and a QoS Maintenance Scheme Based on Differentiated Optical QoS Classes in OVPNs," *Phot. Net. Commun.*, vol. 7, no. 2, Mar. 2004, pp. 161–78.
[76] M. Durvy *et al.*, "Network Availability Based Service Differentiations," *Proc. 11th Int'l. Wksp. QoS IWQoS 2003*, Monterey, CA, June 2–4, 2003.
[77] C. Awad, B. Sansò, and A. Girard, "Network Reliability under Mixed IP and Optical Protection," *Proc. 5th Int'l. Wksp. 25 the Design of Reliable Commun. Networks DRCN 2005*, Lacco Ameno, Island of Ischia, Italy, Oct. 16–19, 2005.
[78] B. Sansò, C. Awad, and A. Girard, "Can DiffServ Guarantee IP QoS Under Failures?" *IEEE Network*, vol. 20, no. 4, pp. 32–40, July/Aug. 2006.
[79] J. Tapolcai, P. Cholda, T. Cinkler, K. Wajda, A. Jajszczyk, and D. Verchere, "Joint Quantification of Resilience and Quality of Service," *Proc. IEEE Int'l. Conf. Commun. ICC 2006*, Istanbul, Turkey, June 11–15, 2006.
[80] H. Øverby, N. Stol, and S. Bjørnstad, "Dependability Differentiation in Optical Packet Switched Networks," *Proc. 7th Int'l. Conf. Transparent Optical Networks ICTON 2005*, vol. 1, Barcelona, Spain, July 3-7, 2005, pp. 385–388.
[81] E. Bouillet, K. Kumaran, G. Liu, and I. Saniee, "Wavelength Usage Efficiency versus Recovery Time in Path-Protected DWDM Mesh Networks," *Proc. Optical Fiber Commun. Conf. and Exhibition OFC 2001*, Anaheim, CA, Mar. 19–21, 2001.
[82] S. Arakawa, J. Katou, and M. Murata, "Design Method of Logical Topologies with Quality of Reliability in WDM Networks," *Phot. Net. Commun.*, vol. 5, no. 2, Mar. 2003, pp. 107–21.
[83] G. Sahin and S. Subramaniam, "Quality of Protection Through Control-Message Scheduling in Optical Mesh Networks," *Proc. 4th Int'l. Wksp. the Design of Reliable Commun. Networks DRCN 2003*, Banff, Alberta, Canada, Oct. 19–22, 2003, pp. 39–46.
[84] —, "Online Control-Message Scheduling fir Quality of Protection (QoP) in DWDM Mesh Networks," *Proc. Optical Fiber Commun. Conf. and Exhibition OFC 2003*, Atlanta, GA, Mar. 23–28, 2003.
[85] —, "Providing Quality-of-Protection Classes Through Control-Message Scheduling in DWDM Mesh Networks with Capacity Sharing," *IEEE JSAC*, vol. 22, no. 9, Nov. 2004, pp. 1846–58.
[86] C. Ou and B. Mukherjee, "Differentiated Quality-of-Protection Provisioning in Optical/MPLS Networks," *Proc. 3rd Int'l. IFIPTC6 Net. Conf. NETWORKING 2004*, Athens, Greece, May 9–14, 2004.
[87] C. Ou, S. Rai, and B. Mukherjee, "Extension of Segment Protection for Bandwidth Efficiency and Differentiated Quality of Protection in Optical/MPLS Networks," *Opt. Switch. Net.*, vol. 1, no. 1, Jan. 2005, pp. 19–33.
[88] B. E. Helvik and N. Stol, "QoS Differentiation in ATM Networks: a Case Study," *Proc. 13th Nordic Teletraffic Seminar NTS-13*, Trondheim, Norway, Aug. 20–22, 1996, pp. 237–50.
[89] —, "QoS Differentiation in ATM Networks: a Case Study," *Telektronikk*, vol. 93, no. 1, 1997, pp. 161–68.

[90] A. Fumagalli and M. Tacca, "Optimal Design of Optical Ring Networks with Differentiated Reliability (DiR)," *Proc. Int'l. Wksp. QoS in Multiservice IP Networks QoS-IP 2001*, Rome, Italy, Jan. 24–26, 2001.
[91] —, "Differentiated Reliability (DiR) in WDM Rings without Wavelength Converters," *Proc. IEEE Int'l. Conf. Commun. ICC 2001*, Helsinki, Finland, June 11–14, 2001.
[92] —, "Differentiated Reliability (DiR) in Wavelength Division Multiplexing Rings," *IEEE/ACM Trans. Net.*, vol. 14, no. 1, Feb. 2006, pp. 159–68.
[93] A. Fumagalli *et al.*, "Shared Path Protection with Differentiated Reliability," *Proc. IEEE Int'l. Conf. Commun. ICC 2002*, New York, NY, Apr. 28–May 2002.
[94] M. Tacca *et al.*, "Differentiated Reliability in Optical Networks: Theoretical and Practical Results," *IEEE/OSA J. Lightwave Tech.*, vol. 21, no. 116, Nov. 2003, pp. 2576–258.
[95] K. Wu, L. Valcarenghi, and A. Fumagalli, "Restoration Schemes with Differentiated Reliability," *Proc. IEEE Int'l. Conf. Commun. ICC 2003*, Anchorage, AK, May 11–15, 2003.
[96] M. Tacca, A. Fumagalli, and F. Unghváry, "Double-Fault Shared Path Protection Scheme with Constrained Connection Downtime," *Proc. 4th Int'l. Wksp. Design of Reliable Commun. Networks DRCN 2003*, Banff, Alberta, Canada, Oct. 19-22, 2003, pp. 181–88.
[97] M. Tacca, P. Monti, and A. Fumagalli, "The Disjoint Path-Pair Matrix Approach for Online Routing in Reliable WDM Networks," *Proc. IEEE Int'l. Conf. Commun. ICC 2004*, Paris, France, June 20–24, 2004.
[98] C. V. Saradhi and C. S. R. Murthy, "Routing Differentiated Reliable Connections in WDM Optical Networks," *Proc. IEEE*, vol. 3, no. 3, May/June 2002, pp. 50–67.
[99] —, "A Framework for Differentiated Survivable Optical Virtual Private Networks," *Phot. Net. Commun.*, vol. 4, no. 3/4, July 2002, pp. 457–87.
[100] —, "Dynamic Establishment of Differentiated Survivable Lightpaths in WDM Mesh Networks," *Comp. Commun.*, vol. 27, no. 3, Feb. 2004, pp. 273–94.
[101] P. Ma, L. Zhou, and G. Mohan, "Reliability and Recovery Time Differentiated Routing in WDM Optical Networks," *Proc. 2005 IEEE Global Telecommun. Conf. GLOBECOM'05*, St. Louis, MO, 27 Nov. – 2 Dec. 2005.
[102] J. Zhang, K. Zhu, and B. Mukherjee, "Service Provisioning to Provide Per-Connection-based Availability Guarantee in WDM Mesh Networks," *Proc. Optical Fiber Commun. Conf. and Exhibition OFC 2003*, Atlanta, GA, Mar. 23–28, 2003.
[103] J. Zhang *et al.*, "A New Provisioning Framework to Provide Availability-Guaranteed Service in WDM Mesh Networks," *Proc. IEEE Int'l. Conf. Commun. ICC 2003*, Anchorage, AK, May 11–15, 2003.
[104] L. Song, J. Zhang, and B. Mukherjee, "Dynamic Provisioning with Reliability Guarantee and Resource Optimization for Differentiated Services in WDM Mesh Networks," *Proc. Optical Fiber Commun. Conf. and Exhibit OFC 2005*, Anaheim, CA, Mar. 6–11, 2005.
[105] W. Yao and B. Ramamurthy, "Survivable Traffic Grooming with Differentiated End-to-End Availability Guarantees in WDM Mesh Networks," *Proc. 13th IEEE Wksp. Local and Metropolitan Area Networks LANMAN 2004*, San Francisco Bay Area, CA, Apr. 25–28, 2004.
[106] M. Amin *et al.*, "Improving Survivability through Traffic Engineering in MPLS Networks," *Proc. 10th IEEE Symp. Computers and Commun. ISCC'2005*, La Manga del Mar Menor, Cartagena, Spain, June 27–30, 2005.
[107] B. Xiang *et al.*, "A Differentiated Shared Protection Algorithm Supporting Traffic Grooming in WDM Networks," *Proc. 2004 Int'l. Conf. Commun., Circuits and Systems ICCCAS 2004*, Chengdu, China, June 27–29, 2004.
[108] —, "A QoS-based Differentiated Protection Algorithm in WDM Mesh Networks," *Proc. 2004 Int'l. Conf. Commun., Circuits and Systems ICCCAS 2004*, Chengdu, China, June 27–29, 2004.
[109] T. Miyamura *et al.*, "A Disjoint Path Selection Scheme Based on Enhanced Shared Risk Link Group Management for Multireliability Service," *Proc. 2005 IEEE Global Telecommun. Conf. GLOBECOM'05*, St. Louis, MO, 27 Nov. – 2 Dec. 2005.
[110] M. Scheffel, "Adaptation of Failure Scenario based Resilience Schemes toward Availability Guarantees," vol. 5, no. 7, July 2006, pp. 521–31.
[111] X. Yang, "Availability-Differentiated Service Provisioning in Freespace Optical Access Networks," vol. 4, no. 7, July 2005, pp. 391–99.
[112] N. Stol *et al.*, "Differentiated Survivability in the OpMiGua Hybrid Optical Network," *Proc. 10th Conf. Opt. Net. Design and Modelling ONDM 2006*, Copenhagen, Denmark, May 22–24, 2006.
[113] G. Mohan and C. S. R. Murthy, "Routing and Wavelength Assignment for Establishing Dependable Connections in WDM Networks," *Proc. 29th Int'l. Symp. Fault-Tolerant Computing FTCS-29*, Madison, WI, June 15–18, 1999.
[114] G. Mohan and A. K. Somani, "Routing Dependable Connections with Specified Failure Restoration Guarantees in WDM Networks," *Proc. IEEE Conf. Computer Commun. INFOCOM 2000*, Tel Aviv, Israel, Mar. 26–30, 2000.
[115] G. Mohan, C. S. R. Murthy, and A. K. Somani, "Efficient Algorithms for Routing Dependable Connections in WDM Optical Networks," *IEEE/ACM Trans. Net.*, vol. 9, no. 5, Oct. 2001, pp. 553–66.
[116] O. Gerstel and G. Sasaki, "Quality of Protection (QoP): A Quantitative Unifying Paradigm to Protection Service Grades," *Proc. IEEE*, vol. 3, no. 3, May/June 2002, pp. 40–49.
[117] W. Wei *et al.*, "Integrated Survivable QoS Routing in Metro IP/WDM Networks," *Proc. 13th IEEE Wksp. Local and Metropolitan Area Networks LANMAN 2004*, San Francisco Bay Area, CA, Apr. 25–28, 2004.
[118] W. Wei, Q. Zeng, and Y. Wang, "Multi-Layer Differentiated Integrated Survivability for Optical Internet," *Phot. Net. Commun.*, vol. 8, no. 3, Nov. 2004, pp. 267–84.
[119] W. Wen, S. J. B. Yoo, and B. Mukherjee, "Quality-of-Service based Protection in MPLS Control WDM Mesh Networks," *Phot. Net. Commun.*, vol. 4, no. 3-4, July 2002, pp. 297–320.
[120] R. Clemente *et al.*, "A Framework for Class of Service Definition in GMPLS-based Meshed ASTN," *Proc. 4th Int'l. Wksp. the Design of Reliable Commun. Networks DRCN 2003*, Banff, Alberta, Canada, Oct. 19–22, 2003.
[121] D. Rossier-Ramuz, D. Rodellar, and R. Scheurer, "Dynamic Protection Set-up in Optical VPN Using Mobile Agent Ecosystem," *Proc. 3rd Int'l. Wksp. the Design of Reliable Commun. Net. DRCN 2001*, Budapest, Hungary, Oct. 7–10, 2001.
[122] D. Rossier-Ramuz, "Towards Active Network Management with Ecomobile, an Ecosystem-inspired Mobile Agent Middleware," Ph.D. dissertation, University of Freiburg, Freiburg, Switzerland, Oct. 2002.
[123] J. L. Marzo *et al.*, "Adding QoS Protection in Order to Enhance MPLS QoS Routing," *Proc. IEEE Int'l. Conf. Commun. ICC 2003*, Anchorage, AK, May 11–15, 2003.
[124] E. Calle *et al.*, "Enhancing MPLS QoS Routing Algorithms by Using the Network Protection Degree Paradigm," *Proc. 2003 IEEE Global Telecommun. Conf. GLOBECOM'03*, San Francisco, CA, Dec. 1–5 2003.
[125] E. Calle, "Enhanced Fault Recovery Methods For Protected Traffic Services in GMPLS Networks," Ph.D. dissertation, Universitat de Girona, Girona, Spain, Feb. 2004.
[126] E. Calle, J. L. Marzo, and A. Urra, "Protection Performance Components in MPLS Networks," *Comp. Commun.*, vol. 27, no. 12, July 2004, pp. 1220–28.
[127] P. Cholda, A. Jajszczyk, and K. Wajda, "A Unified Framework for the Assessment of Recovery Procedures," *Proc. IEEE 2005 Wksp. High Performance Switching and Routing HPSR 2005*, Hong Kong, China, May 12–14, 2005.
[128] P. Cholda, "The Reliability Analysis of Recovery Procedures in GMPLS-based Optical IP Networks," Ph.D. dissertation, AGH University of Science and Technology, Krak´ow, Poland, Dec. 2005.
[129] J. W. Roberts, "Internet Traffic, QoS, and Pricing," *Proc. IEEE*, vol. 92, no. 9, 2004, pp. 1389–99.

## BIOGRAPHIES

PIOTR CHOLDA [S'04–M'07] (cholda@kt.agh.edu.pl) received the M.S. and Ph.D. degrees in Telecommunications from AGH Univer-

sity of Science and Technology, Kraków, Poland in 2001 and 2006, respectively. He joined the Department of Telecommunications, AGH University of Science and Technology in 2006. His research interests focus on design and resilience of GMPLS based multi-layer networks as well as reliability modeling and QoR (Quality of Recovery) concepts. He is the co-author of 12 refereed technical papers and two tutorials on resilient networks. He is the recipient of the Communications QoS, Reliability and Performance Modeling Symposium Best Paper Award from ICC'06.

ANDERS MYKKELTVEIT (mykkeltv@q2s.ntnu.no) received his MSc degree in telematics from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway in 2004. He is currently a Ph.D. student at the Norwegian Centre of Excellence (CoE) for Quantifiable Quality of Service in Communication systems (Q2S). His research interests include dependability and QoS issues in communication networks.

BJARNE E. HELVIK [M'84] (bjarne@q2s.ntnu.no) was awarded the degree Dr. Techn. from the Norwegian Institute of Technology (NTH) in 1982. He is Professor at the Norwegian University of Science and Technology (NTNU), the Department of Telematics and is principal academic at the Norwegian Centre of Excellence (CoE) for Quantifiable Quality of Service in Communication systems (Q2S). His field of interests includes QoS, dependability evaluation, fault-tolerant computing systems and survivable networks as well as related communication system architectural issues.

OTTO J. WITTNER [M'98] (wittner@q2s.ntnu.no) received his Dr.Ing. degree from the Norwegian University of Science and Technology (NTNU) in 2003. He also joined the Norwegian Centre of Excellence (CoE) for Quantifiable Quality of Service in Communication systems (Q2S) in 2003 as a postdoc. His research interests focus on network control and management by emergent behavior principles, where he is looking into dependability issues, especially fault management.

ANDRZEJ JAJSZCZYK [M'91, SM'95, F'99] (jajszczyk@kt.agh.edu.pl) is a professor at AGH University of Science and Technology, Kraków, Poland. He received M.S., Ph.D., and Dr.Hab. degrees from Poznán University of Technology in 1974, 1979, and 1986, respectively. He is the author or co-author of six books and more than 200 papers, as well as 19 patents in the areas of telecommunications switching, high-speed networking, and network management. His current research interests focus on control plane architectures for transport networks and quality of service, as well as network resilience and reliability.