A unified quality of recovery (QoR) measure

P. Chołda*, †, A. Jajszczyk and K. Wajda

Department of Telecommunications, AGH University of Science and Technology, 30-059 Kraków, Al. Mickiewicza 30, Poland

SUMMARY

In this paper, carefully selected factors, playing an important role in the choice of recovery procedures applied in communication networks, are combined to form a unified quality of recovery measure. An example that shows how to use the proposed measure and how to draw conclusions from such numerical results is given. Copyright © 2007 John Wiley & Sons, Ltd.

Received 13 December 2006; Revised 13 June 2007; Accepted 9 August 2007

KEY WORDS: GMPLS; protection; quality of recovery; restoration; survivability attributes

1. INTRODUCTION AND PROBLEM STATEMENT

Many recovery procedures have been developed in recent years. A carrier aiming at guaranteeing an appropriate level of resilience faces the following problem: Which of the proposed recovery procedures should be deployed to maximize revenue? The answer requires many decisions. Some key questions can be formulated as follows: Which procedures can be deployed in a particular network?; Which of them are attractive enough to be developed by R&D divisions?; To what extent are the proposed procedures effective?; What can be the risk of providing a client with recovery procedures which are cheaper but less reliable? The precise assessment and measurement of ensured parameters and a cost calculation should be elaborated. Then, an operator is able to take a well-justified decision concerning the recovery procedures adjusted to client needs. For example, a client whose tele-medicine applications are based on a communication service will be willing to pay much more for a very high level of reliability; whereas another, which uses the network mostly for data transfer, will accept longer outages at reduced costs. To both of them an offer suitable for their requirements should be suggested in advance.

*Correspondence to: P. Chołda, Department of Telecommunications, AGH University of Science and Technology, 30-059 Kraków, Al. Mickiewicza 30, Poland.
†E-mail: cholda@kt.agh.edu.pl

Contract/grant sponsor: Polish Ministry of Science and Higher Education; contract/grant number: N517 013 32/2131

Copyright © 2007 John Wiley & Sons, Ltd.
A generic description of different levels of recovery has been proposed so far. They are reviewed in [1, 2]. In this paper we would like to broaden these ideas by introducing a new, unified measure. We would like to provide operators with an opportunity to use multiple indicators of the recovery quality. Therefore, we call our measure quality of recovery (QoR). The ideas given here were earlier described partially in [1, 3, 4].

Section 2 presents different parameters that are most frequently used to describe the QoR. Then, frameworks focused on such a goal are mentioned. The core idea of this paper, the QoR methodology, is given in Section 3. Then, in Section 4, a comprehensive numerical example is elaborated.

2. PARAMETERS USED AS THE QOR BASIS AND THE OVERVIEW OF KNOWN METHODS

Many factors are taken into account as the basis of the recovery differentiation [5]. Here, the most important factors are given.

- **Mean Time to Failure, MTTF**, serves as the measure of the period of time in which a service can be properly provided to a client.
- **Mean Down Time, MDT**, is the measure describing the time period in which a service is disconnected.
- **Mean Time to Recover, MTTR**, is the period of time between the two events: the failure occurrence and the start of the proper signal reception by the end node after a recovery; thus, when MTTR is calculated, a successful recovery is assumed.
- **Availability, A**, is the probability that an item works properly.
- **Resilience to multiple simultaneous failures** determines to which degree a network can deal with cases not conformed to the assumed single-failure situation.
- **Quality of the backup path** describes the perceived worsening of transport quality after a recovery.
- **Affected traffic/traffic loss** gives the value of the amount of data disturbed by a failure.
- **Cost of recovery** is treated in the most intuitive way as the redundancy (usually the backup capacity, i.e. link bandwidth).
- **Pre-emption** can be understood as a permission to take away resources from one entity (e.g. a connection) to provide recovery for the other.
- **Failure coverage** is the fraction of traffic or connections which is recovered in a given failure scenario.

Many approaches have been proposed as recovery differentiation or quality frameworks so far [1]. Their critical discussion with the presentation of parameters used as the QoR basis can be found in [2]. Here, we only sketch what is the novelty brought by our proposal.

Our proposal is very similar to the following three concepts but introduces some new dimensions. They are: two ideas known as Quality of Protection (QoP) and Quality of Service Protection. The first concept of QoP [6] designates each connection a number called QoP, defined as the probability of surviving a failure. We adapt this idea: a single number called QoR is calculated for all connections protected in a network. QoP given in [7] is another approach to the recovery evaluation. The framework takes into consideration sharing of resources and the dependencies between multiple domains and layers. The measure itself is related to a complex function. We
adapt this idea, as we combine different recovery-related parameters in a single index. Next, Quality of Service Protection [8] is related to the idea of gathering some factors that influence the QoR and to unify them to obtain a single integrated parameter. The methodology takes into account the following three features: traffic loss, recovery time and resource consumption. The integrated parameter is a normalized weighted sum of the normalized factors. We find this way most promising. The list of the reliability-related factors is extended, and the normalization phase is elaborated in more detail.

3. THE QOR METHODOLOGY

In this section, the QoR concept is presented in a GMPLS-like environment. This is basically an engineering-related method. It aims at providing an overall measure of quality of resilience assessment, where various aspects (called factors or parameters) of the recovery process are taken into account. According to it, a quantification of a trade-off between them is enabled: the expected values of these factors are non-linearly mapped onto the \([0, 1]\) continuous set, and then they are combined using different functions (weighted/non-weighted, linear/non-linear, etc.).

QoR treated as a single number evaluates the QoR for a selected origin–destination (OD) pairs. Each pair needs a traffic transport with its own bandwidth requirements as well as recovery requirements, i.e. it belongs to some recovery class. Those requirements are met by a set of working and recovery paths. QoR is calculated provided that different OD pairs can use in one network different recovery mechanisms whose operation may influence recovery operation and evaluation of other OD pairs. Thus, the usage of QoR is envisaged as the following situation. At some time point, an operator is required to establish a new OD pair with its bandwidth/recovery requirements. QoR suitable for a required OD pair is calculated for different alternative scenarios. Its value is dependent on the conditions related to the fact that a number of earlier established paths have existed and operated in the network. Different alternative QoR evaluations are presented to a client. The client chooses some option based on the presented evaluations, price, etc.

In Section 2 different parameters that decide on the QoR were presented. Here, we gathered those that can be easily treated in the quantitative way, namely the availability \(Q_A\), the quality of the backup path \(Q_B\), the affected traffic/traffic loss \(Q_L\), the recovery time \(Q_T\), and the bandwidth redundancy requirements \(Q_R\). The first three serve as an assessment of the robustness of a recovery procedure. The recovery time indicates the speed of the procedure, whereas the redundancy is used as a general estimator of the cost of the procedure. In our opinion, we have chosen very representative factors essential in assessing recovery procedures, but, of course, the set of indicators could be broadened with other elements.

The methodology of QoR introduction can be divided into three stages shown in Figure 1. The first step, called Abstraction, is focused on the derivation of values which describe the operation and properties of recovery procedures in a given network. Thus, the vector of real parameters is obtained: \(v_R = \{A, B, T, R, L\}\). In our proposal, these values are calculated for each individual connection, not for a whole network, since an operator and a client are usually interested in the quality of the offered connection and not in the description of a resilience of a whole network. On the one hand, this is a great advantage of our approach, as it highly simplifies the calculations and makes the interpretation of the obtained results easier to interpret. Obviously, on the other hand, it limits the usage of the solution to connection-oriented networking paradigms, i.e. QoR cannot be used for pure IP. However, as future Internet is envisaged to be based on IP/MPLS-over-DWDM
transporting scheme, which in both layers is connection orientated, it seems not to be a considerable drawback.

In the second stage, after calculating the real values, the process called Normalization is performed. As a result, the vector of unified parameters is acquired: \( \mathbf{v}_Q = \{ Q_A, Q_B, Q_T, Q_R, Q_L \} \).

The normalization process has the following goals: The normalization in the mathematical sense, i.e. modification of values after which they are placed in the \([0, 1]\) interval. It enables the comparison of different parameters 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. And the other goal is to break the dependence of the parameters calculated in the abstraction step on the primary path. The factors related to a recovered connection are strongly dependent on its primary path properties, e.g. a length. This dependence should be eliminated as much as possible in this step. It is most needed in the case of the quality of the backup path and the redundancy.

Next step, the Applications, is the main goal of the whole framework. At least three areas of applications can be proposed. The first one determines the unified QoR measure in its stricter sense. It will facilitate the comparison of different procedures and can be useful mainly for carriers. QoR in the stricter sense will also map the results of the normalization process into a subjective customer satisfaction. In [9] a measure of this kind is suggested to be used in an enhanced MPLS QoS routing. Such QoR is defined as a certain function dependent on all of the normalized parameters. The second application can be used to study the interdependencies between selected parameters. It can be helpful for operator’s purposes as well as for a better understanding of network behavior.

Figure 1. Model of the unified quality of recovery (QoR) measure.
(e.g., the relation between $Q_i$ and $Q_j$). In this case, when functions dependent on a subset of
normalized parameters are taken into account, we speak about QoR in the \emph{broader sense}. It can
also be used to study the ‘quality cuts;’ i.e., when a certain level of QoR is assumed and one of the
parameters is fixed we can check the different possibilities of other parameter values which can
be changed to obtain, e.g., optimal cost or availabilities. The third application, useful mainly in the
context of \emph{operator-client negotiations}, is related to the control of the service level agreement if
a client is able to understand and use the approach. Hence, sharing the responsibility in exchange
for more information on network behavior is possible.

The Abstraction process can be called ‘objective’ because it is related to the calculation or
estimation of some real parameters (mainly statistical values such as probabilities of particular
events, expected times, etc.). We mainly use expected (mean) values to avoid using models related
to real traffic variability. Such models either do not exist or are very complex. The abstracted values
are expressed in seconds or other commonly accepted units like Mbit/s which are independent
of the applied methodology. On the other hand, it is not true in the case of the normalization
and while defining the unified QoR measure. A carrier could arbitrarily perform a selection of
both the transformation function and factors differentiating particular parameters. The output of
normalization and application processes should represent something comparable across different
factors. Intuitively, it should be the \emph{utility} or the \emph{preference} of the abstracted parameter for a carrier
or a user. This fact should be kept in mind to prevent the normalization process from being the
blind game of numbers. Some ideas as how to justify the usage of arbitrarily chosen functions
in the QoR context can be extracted from the economics, where utility functions are theoretically
investigated. Some ideas given below are taken from [10].

The preferences’ assessment is represented with a \emph{utility function}: $u : X \to [0, 1]$. It assigns a
‘happiness score’ to each element of the consumption set $X = \{v_Q : v_Q \text{ is related to a selected}
recovery scheme for a given OD connection}\}$. Consequently, if $u(x) > u(y)$ then a consumer strictly
prefers $x$ to $y$. The arbitrary character of the utility function is most interesting from the QoR
viewpoint. In the utility theory two kinds of utilities are differentiated: the \emph{cardinal}
and the \emph{ordinal} one. The former is related to the assessment of a ‘real value’ of a utility. It is very hard or impossible
in the general case to define such a measure because the utility gained from the consumption of
one commodity may be not objective. Therefore, in the neoclassical economics an idea of ordinal
utility is more widely and successfully used. Hence, the preference of users is studied and only
the ordering of different choices is important, not the numerical value of the utility function itself.
This means that if $u(x) = 0.5$ and $u(y) = 0.25$, $x$ is preferred to $y$, but it does not mean that $x$
is twice as much preferred to $y$. The idea of QoR is aimed at providing a utility function which
Can be used as a comparison basis of different recovery schemes applied to connections carrying
traffic of selected OD pairs. Generally, on its basic level it is a simple ordinal utility function. But
a carrier using its service policy and engineering knowledge can also add to it some elements of
the cardinal utility. In such a case, by the choice of normalization functions, based on the precise
knowledge of a network and the engineering experience, we propose to define QoR in such a way
that, for example, the result ‘QoR($x$) = 0.9 and QoR($y$) = 0.1’ can be interpreted as ‘$x$ is far better
than $y$’.

3.1. Abstraction

3.1.1. Availability. The first parameter used to assess the QoR procedures measures the risk that a
connection fails after occurrence of a failure due to the lack of spare resources. There is a variety
of such indicators. In telecommunications practice, the *steady-state availability* $A$ is commonly used [11]. What should be emphasized here is the fact that we do not propose to deal with the *service availability* but with the *resources availability* only. *Service availability* can be understood as the probability that the user finds the transport service working with the desired quality of service parameters. *Resources availability* refers to the probability that a potential physical path for the service can be found working after a failure. Therefore, this factor only measures the ability of a network to recover from a failure related to a given connection, i.e. that the resources can be obtained in the moment when a failure occurs. Surprisingly, this factor will have quite good values for rerouting (restoration) schemes, which find spare resources after a failure occurrence, since they are flexible and satisfactorily deal with multiple failures.

The difficulties with the availability calculation are usually related to problems with *a priori* determination of a backup path. This means that before a failure occurs not all of a recovery path elements can be enumerated. Moreover, sometimes the backup path itself is not known. This is the case especially when reactive methods, like restoration, are used. While in the case of protection, the procedures know the backup path explicitly, restoration is based on a path sought after the failure. It is possible to calculate the availability of pro-actively protected connections by using reliability block diagrams [12, 13]. Nonetheless, in the case of restoration, only a rough estimate of the probability that a particular connection can be somehow recovered could be considered (e.g. [14]). Although this estimate is rough, it could be more complex to calculate than an exact value of the availability for pro-active methods. It can also be obtained by using simulation experiments. We are interested only in the availability assessment of a particular end-to-end connection, not in values related to the whole network. The theoretical background of availability calculation can be found in the related literature (e.g. [12, 13, 15]). Methods for availability calculation for different protection methods are introduced and analyzed in [16, 17]. Models for shared path protection are given in [18]. Real values to calculate availability on the basis of mean time between failures and mean time to repair are given in [19].

### 3.1.2. Quality of the backup path.

The carrier cannot always ensure the same expected transfer quality of the backup and primary paths. Even if the backup path has an inferior quality, such a ‘partial’ recovery can be better than none. The classical quality parameters can be partitioned into the following three groups [20]. *Additive* (linear): delay, jitter, hop count, bit error rate (BER), path length. *Multiplicative*: packet drop probability (multiplicative factors can be translated into additive ones by using the logarithmic function). And *non-linear*, i.e. min/max: link capacity (bandwidth), available bandwidth.

The relationship between quality of service and resilience parameters is generally complex [21]. Here, we take into account only $B$, i.e. the expected quality of the backup path in the state where all recovery functions are finished and the traffic is received from the alternative path at the connection end. Therefore, it concerns the period until a reversion is performed. Hence, it is related to a relatively long time in comparison with recovery switching. We are not interested here in a temporary decrease of the quality of the backup path. Such a situation is included in the recovery time.

A new unified measure of the quality of the path should be defined. If only one factor $X$ is taken into account, the quality factor can be defined as

$$B_1 = \sum_{i=1}^{k} X_{b_i} \quad \text{or} \quad B_2 = \prod_{i=1}^{k} X_{b_i} \quad \text{or} \quad B_3 = \min_{i \in \{1,...,k\}} X_{b_i}$$

(1)
where \( k \) is the number of backup path links and \( X_{bi} \) is the value of the additive/multiplicative or the non-linear quality factor for \( i \)th backup path link, respectively. The parameters defined here are real positive numbers: \( B_1, B_2, B_3 \in \mathbb{R}_+ \).

3.1.3. Recovery time. Expected recovery time \( T, MTTR \), is the mean value of the time taken from the moment of a failure occurrence to the instant when the traffic is properly received at the connection end. Therefore, it is related to the situations when recovery operates properly.\(^1\) The cases when it does not are encompassed by the resources availability factor. \( T \) includes, first of all, the failure detection time, the notification time, the delay for calculation of new paths, the switching/reconfiguration time or the routing tables convergence time [22] and the time necessary for the recovered data to reach the end node of the connection. The value is not bounded from above. The recovery time should be calculated for each failure scenario separately. If there are \( s \) scenarios related to \( s \) element failures, we will have recovery time \( T \) expressed as

\[
T = \frac{\sum_{i=1}^{s} P_i \times T_i}{\sum_{i=1}^{s} P_i}
\]

(2)

where \( P_i \) indicates the probability of scenario \( i \) for which the recovery time is equal to \( T_i \). As, in general, the calculation is very difficult, because an operator is not fully able to predict the recovery time and network behavior in the multiple failure case, we propose to perform such a computation only for the single failure case. Then, \( \sum_{i=1}^{s} P_i < 1 \). Thorough models of recovery times for MPLS networks are given in [23, 24].

3.1.4. Bandwidth redundancy. The redundancy \( R \) is the amount of spare resources reserved for recovery purposes. It can be argued that the usage of ‘bandwidth redundancy’ as a quality factor is a misnomer since a measure of cost is not a measure of the quality itself. We can agree that cost is not a typical quality measure as, for instance, a delay is. However, in some sense it is only a matter of definition. QoR aims at providing means to quantify some recovery procedures as better than others in a very broad sense. Thus, we can say that anything that is less costly, i.e. better from the expenditure viewpoint, has also a better quality. We argue that the cost of recovery is the equal aspect of the trade-off as other ‘traditional features of the quality’ and that is why we treat the parameter described here as another quality factor. Redundancy cannot be perceived as a guarantee of other recovery quality parameters. For instance, it is possible that connections that are characterized by a higher level of redundancy than some others have worse availability parameters: in the most trivial case because they use longer backup paths, which increases the unavailability and redundancy.

The aim of the \( R \) calculation is to quantify the redundancy related to one particular end-to-end connection. The main problem is to calculate the capacity reserved in possible recovery paths taking into account sharing of resources between different connections. Spare capacity that is designated for a selected LSP can be reserved, in principle, in each link of the network. This is generally the case when the restoration is used. This capacity can also be shared by recovery

\(^1\)In such a case recovery times, not outages stemming from resources shortage, dominate downtimes. Then, the linear relationship between the service availability \( A_{service} \) and the recovery time exists \( A_{service} \approx 1 - \lambda \times T \) (\( \lambda \) represents the mean failure rate). Since such a correlation exists, we decided not to use the service availability as an independent factor.
paths related to different LSPs. In our model, the redundancy related to the selected connection transporting data for a particular OD pair (Con_{calc}) is defined as

$$R = \sum_{i=1}^{l} \left( \frac{C_{i_{tot}} \times C_{i_{r}}}{\sum_{j=1}^{p} C_{ij}} \right)$$

where (cf. Figure 2) $l$ is the number of links in the network, $p$ is the number of backup LSPs which share spare resources with a backup path of Con_{calc} throughout the whole network; each backup path $j$ of all $p$ requires $C_{ij}$ units of spare capacity in link $i$. The amount of $C_{i_{r}}$ spare capacity is reserved in link $i$ for the selected Con_{calc}. Sometimes $C_{ij}$ can be less than the capacity reserved for the working path, e.g. when pure IP rerouting is considered and a carrier deliberately decides to reserve less capacity for an alternative path. The value of $C_{i_{r}} / \sum_{j=1}^{p} C_{ij}$ indicates to what extent the given connection participates in sharing of the reserved capacity. $C_{i_{tot}}$ is the total shared capacity reserved in link $i$ which could be used by Con_{calc}. For the shared protection $C_{i_{tot}} = \max_{j \in \{1,...,p\}} C_{ij}$, whereas in the case of the dedicated protection $p = 1$ and $C_{i_{tot}} = \sum_{j=1}^{p} C_{ij} = C_{i_{r}}$. It is possible that $C_{i_{tot}}$ is not the whole spare capacity in link $i$. For example, when recovery is applied in the higher layers, Shared Risk Link Groups (SRLG, [25]) should be taken into account. Then, each SRLG will be related to a separate pool of spare resources in a physical link. The connection Con_{calc} for which the calculation is performed participates in different SRLGs. $C_{i_{tot}}$ refers only to these SRLGs that are related to the Con_{calc}.

$R$, the bandwidth redundancy, is used as a measure of the cost of recovery. Of course, it could be based on more sophisticated functions than ‘bandwidth × hop’ metrics only. For example, this parameter could be enhanced by some capacity cost functions, which could be especially necessary if they are non-linear and different for distinct links.

3.1.5. Affected traffic. Expected affected traffic $L$ is determined as the amount of data that cannot be properly transported in the interval that begins with the failure occurrence and ends when recovery procedures are successfully finished. Affected traffic equals either traffic lost (e.g. in real-time applications, UDP-based transport, etc.) or data that will have to be transmitted again
Although there are methods aiming at decreasing the affected traffic to 0 (e.g. by buffering [26]), we suggest determining such traffic conservatively, i.e. in the pessimistic way, as the upper bound determined by the period of time from a failure occurrence to the point when data are rerouted to the recovery path. This fraction of the recovery time is denoted as $\tilde{T}$. Thus, the affected traffic is defined as

$$L = C_p \times \tilde{T}$$

(4)

where $C_p$ is the capacity (transmission rate or bandwidth) of the primary path of a connection for which the calculation is made. Although $\tilde{T}$ is a fraction of the recovery time, in the general case this relation cannot be given by any analytic formula since this part is dependent on the recovery scheme, the length of the recovery path and the fact whether some buffering is used.

3.2. Normalization

The process of normalization is of an arbitrary character. There are general hints that should be followed, but the choice of particular functions and parameters has to be made by a carrier. Some functions of a reasonable complexity are proposed here. An example in Section 4 will show that even such simple functions can be useful in practice. The transformation function must represent the utility of the abstracted parameter related to a studied recovery mechanism. Therefore, it should have very small values for the range of abstracted parameters that are not acceptable to discriminate them. On the other hand, quite high values are to be granted to the range of satisfactory numbers of the abstracted parameters. In the latter case, the transformation functions should not be very sensitive to the variability of the values included in the accepted range. This feature is necessary to prevent the discrimination of different values of the abstracted parameters which from the utility viewpoint are very similar. In the range of values intermediate between acceptable and not acceptable ones, there should be a slope with an angle adjusted according to the nature of the parameter. Sometimes, it can be steep if it is possible to determine a border between neighboring ‘good’ and ‘bad’ regions of a utility function. And sometimes the slope can be gentle if a relatively large range of intermediate values can be found.

Different functionals can be envisioned as transformation functions. We propose to use the set of s-shape functions. Additionally, the precise shape of the function should be defined by a set of arbitrarily chosen parameters. Note that in this context, ‘arbitrarily’ does not mean that it is performed without any justification. This should be a decision made on the basis of very thorough considerations of a carrier. It can be called ‘arbitrary’ since it is generally based on engineering and not strictly on theoretical, premises. Generally, the choice of transformation functions is also one of the degrees of freedom of the QoR framework.

3.2.1. Availability. It is clear that $A \in (0, 1)$, i.e. the availability is normalized in the mathematical sense. The main objective is to emphasize the interesting differences of its values. Hence, we define the availability component as

$$Q_A = \frac{1}{1 - A_t^q + (1 + A_t - A)^q}$$

(5)

Here $A_t$ and $q$ are the function parameters. We suggest using a power function in the denominator to change the way the availability influences the perception of the quality of recovery procedures in the proximity of 100%. The higher the value of $q$, the steeper the slope of the transformation function.
and the flatter are both plateaus (related to the accepted/not accepted regions, respectively) of the s-shaped function. The main problem is how to choose the exponent \( q \) and factor \( A_t \). Generally speaking, if the postulate of ‘four nines’ as a target is accepted, the operator will be interested in values, e.g. \( A \gg 99\% \). For typical values that are related to communication networks, see [19]. Therefore, the aim is to distinguish values of \( Q_A \) to show what is absolutely unsatisfactory and to emphasize the differences between greater availabilities, e.g. between \( A \) equal ‘only’ to 99.9% and \( A = 99.999\% \). On the other hand, the function should not emphasize the differences of values of, for instance, \( A = 99,999\% \) and \( A = 99,999,99\% \), because in this range of values all availabilities are acceptable and almost of the same utility. The parameter \( q \) will take rather ‘radical’ values to achieve our goal, as can be seen in Section 4. We propose to choose the parameters of the function according to the following method. First, the value of \( q \) should be chosen. Then, if \( q \) is high, the threshold value of \( A_t \) can be chosen as the value for which \( Q_{A_t} \approx 0.5 \).

3.2.2. Quality of the backup path. It is interesting to what extent \( B \), the value of the quality of the backup path, is worse than the value describing the primary path. We are not interested here in an absolute value of it. Therefore, first of all, we divide \( B \) by the indicator of the quality of the primary path. Thus, we obtain the relative value of the deterioration of the quality after the recovery. The determined fraction will be placed in the interval of \((0, 1)\):

\[
B_N = \min \left\{ \frac{B}{X_p}, \frac{X_p}{B} \right\}
\]

(6)

where \( B \) is the quality of the backup path calculated according to (1) and \( X_p \) is the same parameter characterizing the quality of the primary path. It should be determined in the same way as \( B \), e.g. the sum in the case of delays. We use the minimum function because we are interested in the value lower than 1 independently of the quality factor nature. Thus, we have two elements within the minimized set \( \{B/X_p, X_p/B\} \). If we assume that \( B \) is always worse than \( X_p \) we have: for cases where the quality is decreasing with the increasing value of \( B \) (e.g. the delays), the active element of the set is \( X_p/B \); for cases where the quality is decreasing with the decreasing value of \( B \) (e.g. the link bandwidth), the active element of the set is \( B/X_p \). The obtained fraction \( B_N \) will be transformed into the ‘normalized’ one, similarly as in the case of the availability. Therefore, we need to define threshold values of the \( Q_B \) component related to the required threshold value of the fraction \( B_N \). Similarly as in (5):

\[
Q_B = \frac{1}{1 - B_{N_t}^f + (1 + B_{N_t} - B_N)^f} \tag{7}
\]

Now, the values of parameters can be found in a similar way as \( A_t \) and \( q \). The difference is that the exponent \( f \) does not have to be as high as \( q \) in (5), because in the case of the quality, the range of acceptable values is not as small as in the case of the availability and inferior values do not have to be discriminated so strictly. If we want to take into consideration a large set of traditional quality indicators, we should normalize the values for each factor individually and then multiply them or find the average value. As the procedure is similar to the calculation of QoR in the stricter sense, the same formulas as shown in Section 3.3.1 can be used.

3.2.3. Recovery time. Recovery time is greater than 0 and varies within a broad range of values. We have to define the following transformation: \( T \in \mathbb{R}_+ \rightarrow Q_T \in [0, 1] \). As the formerly used
normalization functions are increasing, they are not convenient now. Here, we decided to use a simple monotonically decreasing fractional function. We define it to be almost flat for \( T \rightarrow 0 \), which means that we do not want to intensively distinguish short recovery times. Then, the function decreases to the threshold point. That represents the loss of utility with the increase of \( T \). Therefore, the function is determined as

\[
Q_T = \frac{1}{1 + r \times T^2}
\]  

(8)

While the threshold values are assumed and fixed, the scale parameter \( r \) will be calculated as the solution of the following equation:

\[
Q_{T_t} = \frac{1}{1 + r \times T_t^2}
\]  

(9)

Here \( T_t \) is the threshold value of the recovery time for which we would like the QoR component of the recovery time to be \( Q_{T_t} \). After \( r \) has been calculated from (9) and substituted into (8), we obtain the final expression for this quality contributor.

### 3.2.4. Bandwidth redundancy

In the abstraction stage, the real value of the redundancy is calculated. By normalization it should be made independent of the influence of the length and capacity of the primary path, i.e. we are interested only in the relative value. Since a client usually pays for the length and capacity of the primary path, we can define an indicator determining the assessment of the redundancy independent of these two factors. For example, if we have two primary paths protected by one backup path, we should assess that for the longer primary path this method is more advantageous than for the other, because we gain more by protecting the longer one. Therefore, first we calculate the redundancy ratio expressed as

\[
R_N = \frac{R}{k \times C_p}
\]  

(10)

where \( R \) is the redundancy calculated according to (3), \( k \) is the number of links in the primary path of a connection \( C_{\text{calc}} \) for which the calculation is performed, and \( C_p \) indicates the capacity of this path. \( R_N \) is greater than 0, not necessarily greater than 1, especially in the case of procedures with sharing. We define the redundancy component in an analogous way as in the case of the recovery time, cf. (8)–(9). The only difference is that now we use a larger exponent in the denominator (3 instead of 2) to flatten the shape of the transformation function for small and great values while we have a sharper decrease near 0.5. The function is shown in the following formula:

\[
Q_R = \frac{1}{1 + \frac{1 - Q_{R_{t}}}{Q_{R_{t}} \times R_{N_{t}}^3}}
\]  

(11)

Here \( R_{N_{t}} \) is the threshold value of the redundancy fraction \( R_N \), for which we would like the QoR component to be \( Q_{R_{t}} \).

### 3.2.5. Affected traffic

The volume of the affected traffic \( L \in \mathbb{R}_+ \cup \{0\} \) can be treated as an objective assessment factor for loss which appears to be a large burden to the network. We have to transform
its value into the range of \([0, 1]\). We will use a function similar to the function related to the component for the recovery time because they are of a similar kind. We use the following formula:

\[
Q_L = \frac{1}{1 + \frac{1 - Q_{L_t}}{Q_{L_t} \times L_t^2}}
\]  

(12)

where \(L_t\) is the threshold value of the affected traffic, for which we would like the QoR component related to the affected traffic to be \(Q_{L_t}\).

3.2.6. Additional notes on the normalization. One can argue that the normalization process is too arbitrary and that the threshold values or transformation functions are taken ‘out of the blue.’ We claim that this arbitrariness is the main advantage of our framework. Each operator has its own quality policy and should apply it in the normalization stage. What is proposed is a comprehensible methodology which should be clear to an operator when applied. Generally, thresholds should be chosen having in mind the following factors. Firstly, the type of network, an applied transport technology, protocols, architectures, layers, etc. All of them affect real values, e.g. detection times or delays. Secondly, the extensiveness of a network also has a strong impact on real values, e.g. availability values are smaller in larger networks due to larger failure intensities of longer links or a higher fault probability when there are more items involved in a connection. And thirdly, the most important factor is consistency with the requirements concerning services which are expected by clients. Thresholds should be selected after the real values of parameters have been calculated and can be compared with each other. Then, an operator can decide which values are satisfactory and how to shape the normalization functions. These remarks are also related to the choice of the QoR combination function (see Section 3.3.1).

The fact that the normalization is based on the transformation of mean (expected) values is also worth commenting. We agree that the variability of different parameters can also have a great influence on the perceived QoR. It is the truth but we have some arguments that support the usage of mean values. First, the parameters of a connection are generally constant and related to the contract agreed by its establishment. Thus, such parameters as bandwidth or delays, especially in the optical layer, where the main part of recovery will be performed, are relatively close to the mean value. Second, parameters such as the resources availability or redundancy are generally steady during a very long period of time. Recovery time, quality of the backup path and traffic loss can vary in some limits, indeed. However, there are many models that estimate these values as the means, e.g. [16–19, 23, 24]. They are confirmed and well understood. Moreover, it is far easier to estimate the expected values, because their variability cannot be studied theoretically as it strongly depends on a given network. Therefore, they cannot be found \textit{a priori}. Additionally, the variation in values is related only to reactive methods, not to the larger class of protection schemes. We claim that the complex development of the framework only to make the variation of restoration methods tractable takes a higher cost than the expected gain.

3.3. Applications

3.3.1. Formulation of the QoR measure. For protected connections, QoR in the stricter sense should be in the range of \((0, 1)\). It enables an easy orientation of how far a particular value is from
the extremes, where 0 reflects extremely bad reliability parameters and 1 would mean the ‘ideal’ recovery. At first, QoR can be determined as the product of individual components

$$\text{QoR}_1 = \prod_{i \in \{A, B, T, R, L\}} Q_i$$

(13)

The measure determined in such a way has one important feature, i.e. it does not need a definition of any additional parameters. This could simplify the usage of QoR. On the other hand, it could be a big disadvantage, because it would be very useful for the carrier to define the significance of particular components. For example, for the best-effort data transmission, longer recovery times can be more acceptable than high levels of redundancy. This disadvantage is eliminated in the second proposal, where a weighted average value is applied:

$$\text{QoR}_2 = \frac{\sum_{i \in \{A, B, T, R, L\}} z_i \times Q_i}{\sum_{i \in \{A, B, T, R, L\}} z_i}$$

(14)

Here $z_i$ indicates a weight corresponding to the particular component $i$. Some other average values (possibly weighted) can be used. In our opinion, the best results are related to the following definition. If the vector of normalized values is seen as coordinates of a point $S$ in a five-dimensional space, QoR could be calculated as the Euclidean distance from 0 to $S$. Such a measure is frequently used in poly-optimization. The distance should be again scaled down to be in the range of (0, 1). Thus, we use the so-called quadratic mean value and define QoR according to the following formula:

$$\text{QoR}_3 = \sqrt{\frac{1}{5} \sum_{i \in \{A, B, T, R, L\}} Q_i^2}$$

(15)

When the contribution of particular factors is not assessed equally, we can again use the weighted mean value according to the following equation:

$$\text{QoR}_3' = \sqrt{\frac{\sum_{i \in \{A, B, T, R, L\}} z_i \times Q_i^2}{\sum_{i \in \{A, B, T, R, L\}} z_i}}$$

(16)

The weights can be used to select a suitable recovery scheme to match customers’ requirements. We propose to perform this task by gathering customers’ needs, i.e. their application requirements, in some service/traffic classes. Here, we base our selection on classes presented and characterized in [4, 27]. In the cited sources the following classes are characterized:

- **Emergency class** (e.g. 112 phone number, government special services, real-time control/medical applications) has very high-quality requirements; interruptions are generally not tolerable and a client can afford to pay very much for the service.
- **Prioritized elastic traffic class** (e.g. bank transactions, grid networking, Storage Area Network applications) has not-so-strict QoS requirements but is characterized with quite high-recovery needs.
- **Conversational class** (e.g. videoconferencing, Virtual Private Networks) has relatively high-quality requirements but not-so-strict recovery requirements.
Table I. Proposal of weights assignment to different traffic classes.

<table>
<thead>
<tr>
<th>Traffic class</th>
<th>$\alpha_A$</th>
<th>$\alpha_B$</th>
<th>$\alpha_T$</th>
<th>$\alpha_R$</th>
<th>$\alpha_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergency</td>
<td>9</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Prioritized elastic</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Conversational</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Streaming</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Best effort</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

- Streaming class (e.g. Video on Demand) is similar to the conversational class but with not-so-stringent QoS requirements.
- Best effort class (e.g. web browsing, e-mail or File Transfer Protocol transfers) as the lowest class uses resources without any special QoS or recovery requirements.

In Table I the choice of weights that can be used with the QoR based on a mean value or the Euclidean distance is presented. The weight valued as 9 is related to the highest importance of a normalized QoR component, 5 is related to the medium importance, and 1 is related to the lowest importance, respectively.

Except for the mentioned functions, there are also other possibilities like the geometric or harmonic mean. However, the numerical results reported in [1] suggest that their usage does not prevail in comparison to the ones presented in this paper. Thus, we omitted the presentation of those possibilities here. Additionally, some other distance metrics could be used along with the Euclidean distance (e.g. Manhattan or Chebyshev distances). We decided to use that most typical, especially as it is also eagerly applied in the poly-objective optimization contexts. It is hard to decide a priori, with no given example, which of the proposed unification functions can be used or which of them are the most meaningful for comparison across different factors. Generally, the unification function should enable the interpretation consistent with the utility function. In the context of a comparison across different parameters it is the issue of a trade-off. In this sense, some sort of a weighted arithmetic or quadratic average value seems to be the most useful, as its interpretation is relatively easy and intuitive. Additionally, it ensures a choice of weights that represent the importance of different features.

Generally, when some parameters are combined into a single measure in the way presented above, it makes sense only if the mixed factors are independent (or orthogonal in the case of the quadratic measure). We claim that this condition is at least to some degree met in the case of the parameters proposed by us. This issue was partially discussed earlier when selected factors were introduced, but we summarize it here: if the availability factor is defined as the resources availability it is not correlated with the recovery time which concerns the service level. On the other hand, the affected traffic is calculated using some elements of the recovery time calculation.

3.3.2. QoR for pre-emptive and unprotected connections. In the presented framework, there is no difference in calculating QoR parameters for unprotected or pre-emptive connections in comparison with recovered connections. In the case of availability, it should be kept in mind that for an unprotected connection only the working path is taken into account, whereas in the case of a pre-emptive connection, the availabilities of connections with higher priorities must be considered. When unprotected or pre-empted connections are taken into account, some of the normalized parameters can be equal to 0. For example, $Q_B = 0$ when there is no backup path.
4. NUMERICAL EXAMPLE

We use an example of the topology of a German long-haul network [28, 29]. We assume that this is a transparent optical transport network. Lightpaths (optical LSPs), treated here as basic communication services transporting data of inter-city connections for which QoR will be calculated, are created between each pair of two nodes (cities). As there are 17 cities, 136 connections are established. The volume of traffic is estimated on the basis of the methodology given in [30]. We consider five scenarios with the following resilience schemes: dedicated path (DP) protection, shared backup path (SBPP) protection, dedicated link (DL) protection, shared link (SL) protection and path restoration with stub release (R), which means that the rerouting is based on the reserved, but shared, spare capacity. For an exhaustive description of the way these different recovery mechanisms work, see for instance [5, 25]. Additionally, the QoR values are calculated for unprotected connections (U). In each scenario, all connections are recovered by the same scheme. In the link protection schemes, node failures are not recovered. All protection schemes are planned in an optimized way, based on the cost proportional to the length of a path. The relevant methods and linear programming formulations are given in [25, 31]. The design of shared schemes is based on the single failure assumption. In the case of restoration, we assume that the amount of spare resources is the same as that designed for SBPP.

We should note that despite our defining the components, the real challenge in assessing QoR is related to the calculation of the network parameters. Some values can be calculated in a formula-based way, and we do like that if this is the case. However, for example, for schemes with sharing of resources it was not possible. Therefore, the simulation experiments were performed. All values were averaged across all simulation runs. The 95% confidence intervals have been calculated but since they are quite small, they are not presented in the figures. For clarity of presentation, the values for some fraction of connections only are given. These are (related to the length in the sense of the physical distance or the number of hops of an unprotected path): the longest distance connection Hamburg–München (indication: HM), the largest number of hops connection Essen–Ulm (EU), the shortest distance connection Essen–Düsseldorf (ED), the longest connection of one hop between Leipzig and Frankfurt (LF), the median distance connection Essen–Karlsruhe (EK), and the Dortmund–Nürnberg connection (DN), i.e. the connection characterized by the median value of distance and median value of hops. Additionally, the QoR values calculated on the basis of average values (AV) of each abstracted parameter derived from the network are shown.

The selection of reference points and thresholds for different parameters is not consistent across different parameters, because according to the fundamental idea of normalization, each factor is transformed independently of other parameters.

Figure 3 presents, on the one hand, the abstracted values calculated for the Essen–Karlsruhe connection and on the other hand, what is more important, the effects of the normalization of those values. They are related to all five quality parameters. They are subsequently used to calculate the unified values of the QoR evaluation. If some normalized values are evaluated as 0 they were not presented to improve the legibility of the figures.

The values of failure and repair rates are taken from [5]. In most of the cases it turned out that the restoration has the highest resources availability (we remind that due to a large recovery time this statement is not true in the case of the service availability). The reason is that due to the flexible nature of the restoration and free usage of quite a large spare capacity, it can recover some multiple failures. In the normalization stage we determine that the threshold value of $A_t = 99.9\%$
and \( q = 1000 \). Such a high number is necessary to distinguish availability values as similar in the arithmetic sense as, for example, 99.9 and 99.99%.

Generally, quality issues are more important when QoR is assessed for connections in higher layers. When recovery schemes are applied in the physical layer, such traditional QoS parameters like packet delay do not make much sense. But for optical networks, there is at least one quality parameter which should be taken into account, because it affects BER. This is the length of a
In a transparent optical network there is no re-amplification, re-shaping and re-timing except for the end nodes. Consequently, if a path is too long, larger values of impairments caused by diffraction are involved [32]. Therefore, a lightpath length should be minimized. The recovery lightpath is generally longer than the working lightpath of the respective connection, and it is taken into account here. The length of a path is a value of the additive nature, and \( B \) is calculated according to (1). Here, it is calculated similarly as the recovery time, i.e. as an average value of the length of the recovery path weighted by the probability of an element failure (cf. (2)). In the case of path protection, the recovery lightpath is always the same, but the mentioned average value should be calculated for link protection. For the case of path restoration, the average value is obtained by using simulation experiments. The value related to restoration is often best, since the recovery lightpath calculated for the restoration scheme does not have to be disjoint with the working lightpath, which is assumed for other schemes. It can be noted that especially for short working lightpaths, the average value of the length ratio of the recovery lightpath to the working lightpath is relatively high. As the example network is not very large and the problem of the relative lengthening of a lightpath is not a major issue, we select the thresholds not very strictly: \( f = 10 \) and \( B_{Nt} = 0.66 \). Those values mean that for the recovery path 50% longer than the working path \( Q_B = 0.5 \).

The recovery time model presented in [23, 33, 34] is used here. Since signaling is not simulated, the immediate propagation of a new network topology is assumed. To obtain example values of \( T \), the following values are taken into consideration: the failure detection time is equal to 20 ms, no hold-off time (because we do not deal with multi-layer recovery), the message processing time is equal to 10 ms, the switching time is equal to 5.5 ms (calculated as the mean value of both extremes given in [23]), and the new path calculation time is 2 ms. As the thresholds, the following values are selected: \( Q_{Tt} = 0.9 \), \( T_t = 37.9 \) ms (this is the worst recovery time for the DP protection). Thus, the scale indicator \( r = 7.73 \times 10^{-5} \). In case of the unprotected lightpath, \( Q_T = 0 \).

It is assumed that if we reserve the same amount of spare capacity on the backup path as the working capacity on the primary path in the dedicated scenario, the threshold value of the component will be 0.2 (i.e. \( Q_{R_t} = 0.2 \), \( R_{N_t} = 1 \)). Such a choice of thresholds stems from the fact that 100% redundancy is assessed here as very bad. It results in the scale indicator equal to 4. One issue should be noted here. Although from the global viewpoint the cost of restoration is the same as the cost of SBPP, it is not the case when the redundancy is calculated for individual connections. This bases on the fact that it is assessed according to (3) and (10). In most cases the redundancy factor for SBPP is inferior to the factor calculated for restoration. This is especially the case when a connection that is characterized by a short working lightpath uses a long recovery lightpath in the SBPP case where redundant resources are calculated along a rigid recovery path. This is not a case when redundancy is calculated for restoration. According to (3) the sharing of spare resources for restoration assumes that all spare resources are shared among all 136 connections. The fact that the redundancy of a selected connection is calculated along all links could suggest that the redundancy is greater than in the SBPP case. Nevertheless, it is usually not true, as according to (3) also the contribution related to the connection for which the calculation is performed is taken into account. As a large number of connections in the example network use not much capacity (and only some of them use very large capacity according to the used traffic matrix model) the redundancy factor is only in a few cases worse for restoration than for SBPP. This is also the reason why the average AV has a character similar to the majority of connections.

Affected traffic was calculated on the basis of the recovery time calculation (switching time is not taken into account). Here, the network is quite small; therefore, recovery times are not so long.
and in the consequence the loss is not harmful. The choice of the thresholds was inspired by the SDH/SONET ‘50 ms recovery requirement’ [25]. In the example, all lightpaths carry OC-48, which means that an acceptable loss, related to 50 ms outage, should be less than 120 Mbits. For such a value, the threshold of QoR factor is set to 0.9 (i.e. $Q_{lt} = 0.9$, $L_t = 120$Mbits). Consequently, the scale factor equals $7.7 \times 10^{-6}$. Again, for the unprotected lightpaths, $Q_L = 0$.

The results for the unified QoR measure calculations are illustrated in Figure 4, where $QoR_i$ values were calculated according to (13)–(15), respectively. The values are presented for the selected connections enumerated above. For each connection, QoR is calculated assuming that all connections in the network use only one recovery method simultaneously. We can note that $QoR_1$ calculated as a product assesses all the schemes very rigorously. Only for the SBPP, SL protection and path restoration, it is possible to differentiate the quality among connections. Values determined in such a way tend to decrease intensively: all components are below one, so each additional multiplier drastically reduces the whole value. For instance, connections that use a very short working lightpath, like the ED connection, have for this reason a very poor evaluation. For

---

Figure 4. Comparison of QoR (in the stricter sense) values: (a) $QoR_1$, product; (b) $QoR_2$, arithmetic mean value; and (c) $QoR_3$, quadratic mean.
them, backup paths are much longer in comparison with the paths used before a failure occurrence. Owing to that, the relative quality of the backup path is assessed as very low and consequently QoR\(_1\) is very poor. If only one value of the normalized components is equal to zero, the whole value is 0. Thus, we can see that for unprotected or pre-empted connections, which are characterized by some normalized parameters, \(Q_i\) equal to 0, QoR\(_1\) should not be used as it does not carry any useful information. In the QoR\(_1\) case, when values related to means are compared, SBPP seems to be the best recovery scheme and dedicated link protection the poorest one (not mentioning the unprotected lightpaths). The average assessment of different connections does not necessarily amount to the assessment of selected LSPs. For instance, the Hamburg and München connection recovered by the SL protection appears to be best from the lightpaths presented for QoR\(_1\). We can see that QoR\(_2\) calculated as the average value \(\frac{\sum Q_i}{\sum 1} = \frac{\sum Qi}{\sum n}\) does not scatter the values computed for particular schemes. It seems to be more intuitive and compatible with our purposes, although its drawback lays in the fact that the values are very similar across different schemes. QoR\(_2\) does not change the general assessment of the procedures: again SBPP is the best one, the next is the SL protection, and then restoration, dedicated path and link protection. Meanwhile, the assessment for particular lightpaths in the same schemes is not changed in comparison with QoR\(_1\), but the differences between assessments of lightpaths are not so strong now. More interesting effects are obtained if some differentiated weights are used (see below). The quadratic mean QoR\(_3\) generally repeats the order obtained by the product metric.

Now, we can pay attention to the issue of the QoR adjustment to different recovery classes. In Figure 5 we present the results obtained for a selected (EK) connection when the quadratic mean is applied with weights proposed in Table I. We can see that although the results of normalization are identical (i.e. the values of \(Q_i\) are obtained independently of a traffic class), the choice of weights for different classes implies that the order of the suggested recovery schemes varies significantly across classes. For the emergency class the usage of the dedicated link protection is suggested, which conforms to a common feeling that this recovery method is related to the highest quality parameter values (especially the recovery time and usually the availability) but it...
is only recommended for clients who can afford very expensive connections. On the other hand, the SBPP, relatively cheap but offering a decent quality, is recommended for prioritized elastic traffic and for the streaming class. A scheme that is a little bit better from the quality viewpoint, i.e. the SL protection, is proposed for the conversational class. This is intuitively understandable as the quality expected for this class is higher. And at last, a cost-effective restoration with the poor recovery time parameter is perfectly suitable for the best effort traffic. Another interesting issue related to the figure lays in the fact that within schemes proposed for selected classes we can distinguish some groups of methods that have similar QoR values. These are groups of schemes that for a particular traffic character seem to be equivalent from the standpoint of used weights.

The proposed set of measures can also be used to study interdependencies between particular quality components, which can also help to decide which scheme should be selected. Such an analysis can be even more precise than those presented above. It is illustrated in an example of the two dependencies (Figure 6): between $Q_R$ and $Q_A$, and between the quadratic mean of $Q_T$ and $Q_A$ (QoR, this average is treated as QoR in the broader sense) and the cost of the procedure expressed as a linear function of the redundancy factor $R_N$.

In the analysis of Figure 6(a) (the Hamburg–München connection) we try to find, first of all, points that have the greatest values of both components and simultaneously the points that are related to evidently disadvantageous schemes. Here, the best scheme, most distant from the origin point $(0,0)$, an indisputably superior solution, is path restoration that, on the one hand, has the best availability and simultaneously is most cost-effective from the redundancy standpoint. Other shared schemes are very close to it. On the other hand, it is possible to eliminate very poor dedicated schemes that have normalized availability factor similar to SBPP, but unlikely to have a very poor redundancy. We can see that for SBPP satisfactory redundancy values go along with quite good values of the availability. This means that here a reasonable trade-off between the availability and the redundancy is easy to obtain. Such analyses of interdependencies enable to eliminate disadvantageous schemes in advance.

One can argue that the dependencies observed in Figure 6(a) are not related to the proposed methodology and they can be noted without it as well. It is only partly true, because the introduction...
of the normalization step not only simplifies the interpretation, but also shapes the character of differences due to introduction of the *cardinal utility* concept. For example, it can decrease the difference between the four and three nines availability in the $Q_A$ factor, whereas it can strongly increase the difference between the redundancies ($Q_R$ factor) related to them. By some of the operator’s policies (e.g. a greater interest in the cost than in the reliability issues) it can be assessed that such an increase of the resources availability is not justified. Figure 6(b) illustrates some other interesting facts and for sure could not be prepared without the introduction of the QoR methodology. The $x$-axis shows QoR in the *broader sense* (QoR), whereas the $y$-axis shows the cost of the scheme. One can suppose that the function describing the relation between the quality of the recovery procedure and the cost would be monotonically growing. It is generally not true, although the Hamburg–München connection is an example of such a relationship. Also, the characteristics for another connection, between Leipzig and Frankfurt, are shown. Note that they are very similar, but they differ in the shift related generally to the length of the working lightpaths, i.e. shorter ones have usually better quality indicators. What can be surprising, the most typical cost vs QoR characteristics are represented by the values presented here for the set of averaged parameters (AV), but the difference in cost between the restoration and SBPP varies considerably. Here, the most typical effect is a decrease of the quality with a higher cost between the SBPP and the SL protection. The same situation takes place when the DP and dedicated link protections are compared. This fact would suggest that local methods could be eliminated, naturally, only in the case where the speed of recovery is not the main requirement. In the case of the Leipzig and Frankfurt connection, the SL protection which has a similar cost as the dedicated link protection can be suggested as an evidently inferior scheme because the almost negligible difference in the cost is related to an increase of QoR for the dedicated scheme. Then, almost without any trade-off, the latter scheme can be chosen. If a cheaper scheme is sought for the same connection, a similar reasoning can be performed in relation to the choice between SBPP and the DP protection. They both have a very similar quality, but the cost of the latter is higher. This suggests the use of SBPP. Restoration appears to be the worst recovery scheme for all of the three presented groups. This fact conforms with the common intuition that the poorest quality goes with the cheapest method. With the example of the Leipzig–Frankfurt connection we can once again emphasize the difference between the QoR understood in the * stricter* and in the *broader sense*. In Figure 6(b) we observe that the restoration has the worst quality when QoR is treated in the *broader sense*. This is related to the fact that in the case of QoR understood in the stricter sense all five normalized parameters are taken into account, the crucial $Q_R$ factor among others. For QoR understood in the *broader sense*, only a subset of those parameters is used (here, $Q_A$ and $Q_T$). The presented Leipzig–Frankfurt connection has a very advantageous value of the $Q_R$ parameter and not very attractive value of the $Q_T$ factor, and this fact is the root of this difference.

5. CONCLUSIONS

The intention of this paper was to combine the basic and commonly accepted performance measures describing recovery methods into a single index which could be used as a means to compare different schemes. Although the problem has been dealt with earlier, this is the first so comprehensive approach to the quality of recovery and as such it proposes the ideas primarily. In our opinion, the proposed solution can be useful in practice as well as in a subsequent research. Our new three-step methodology enables an operator to assess recovery methods used in its network. Additionally,
as shown, it helps to choose a proper method for a selected customer and a service class as well as to eliminate evidently disadvantageous schemes. Moreover, the relationship of the recovery assessment to the utility concept is presented. On this basis, the normalization process is proposed and recommendations on how to perform it according to particular requirements and expertise are thoroughly presented.

REFERENCES


AUTHORS’ BIOGRAPHIES

**Piotr Cholda** received his MS and PhD degrees in Telecommunications from AGH University of Science and Technology, Krakow, 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 quality of recovery (QoR) concepts. Dr Cholda is the co-author of 13 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.

**Andrzej Jajszczyk** is a professor at AGH University of Science and Technology in Krakow, Poland. He received his MS, PhD, and Dr Hab degrees from Poznan University of Technology in 1974, 1979 and 1986, respectively. He spent several years at the universities in Australia and Canada. He is the author or co-author of six books and over 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, quality of service and network reliability. He has been a consultant to industry, telecommunications operators and government agencies. He was the founding editor of the *IEEE Global Communications Newsletter* and Editor-in-Chief of *IEEE Communications Magazine*. He has been involved in organization of numerous conferences. He is a member of the Association of Polish Electrical Engineers and a Fellow of IEEE.
Krzysztof Wajda received his MSc and PhD in telecommunications from AGH University of Science and Technology, Krakow, Poland, in 1982 and 1990, respectively, and currently is an assistant professor at Department of Telecommunications, AGH-UST. Since 1991 he was involved in few international projects: COST 242, Copernicus ISMAN, ACTS BBL, IST LION, NOBEL and NOBEL2, e-Photon/ONe and e-Photon/ONe+ and many national projects. He has been a consultant to private telecommunication companies and Polish Telecom. His research interests are traffic engineering, architecture and implementation of broadband packet networks, multimedia services, network reliability issues.