Universitat de Girona

Department of Electronics, Computer Science and Automatic Control

PhD Thesis

Enhanced fault recovery methods for protected traffic services in GMPLS networks

Author : Eusebi Calle

Thesis advisor: Josep Lluís Marzo

A THESIS SUBMITTED IN FULFILLMENT
OF THE REQUIREMENTS FOR THE
PhD in Computer Science

Girona, February 2004

Dr.	
President	

Dr.	
Secretari	

Dr.	
Vocal	

Dr.	
Vocal	

Dr.	
Vocal	

Data de la defensa pública	
Qualificació	

To Laura whose patience, love, support and
encouragement enabled me to
complete this thesis.

To my parents for their interest, support and encouragement
of my academic and profesional success.

Abstract

New network technology enables increasingly higher volumes of information to be carried. Various types of mission-critical, higher-priority traffic are now transported over these networks. In this scenario, when offering better quality of service, the consequences of a fault in a link or node become more pronounced. However, IP based networks still do not have the same degree of reliability offered by traditional telecommunication networks (for instance telephone networks). In our opinion, achieving such reliability will be crucial in the success or failure of a wide range of services that should eventually substitute current telecommunication networks.

Moreover, an initial design of a network may not be satisfactory due to changes in offered load, traffic characteristics, etc. Network resources also vary due to resource reservations and topology changes (such as node or link failures). An important part of designing the next generation QoS network concerns its reliability, which can be provided through fault management mechanisms, applied at different network levels and time scales. Multiprotocol Label Switching (MPLS) and the extended Generalized MPLS (GMPLS) provide fast mechanisms for recovery from failures by establishing redundant Label Switch Paths as backup paths. With these backups, traffic can always be redirected in case of failure.

The main objective of this thesis is to improve some of the current MPLS/GMPLS fault recovery methods, in order to support the protection requirements of the new Internet services. Therefore, the definition and evaluation of the quality of protection provided by each fault recovery model is a main objective of this work. Some parameters, such as fault recovery time, packet loss or resource consumption, all within the scope of this quality of protection, are considered. Throughout this work, different analyses and experimental results supporting these decisions are presented.

In this thesis a review and detailed comparison of the MPLS fault recovery

methods are presented. Path protection methods (global backups, reverse backups and 1+1 methods), as well as segment protection and local methods are included in this analysis. The fault recovery process is also analyzed, differentiating how path protection methods are created and managed in case of a failure. The extension of these mechanisms to optical networks using GMPLS control plane is also taken into account.

Although this thesis is mainly focused in pre-reserved and dedicated bandwidth models, other models are also taken into account (such as no bandwidth allocation or shared bandwidth allocation). In the same way, this thesis is focused on single network component (link) failure protection models. However, an approach to multiple failure protection is also introduced.

In the first phase (Chapter 1) MPLS fault recovery methods are analyzed without taking into account resource or network topology constraints. This analysis reported a first classification of the best protection methods in terms of packet loss and recovery time. This first analysis cannot be applied to real networks. In real networks, bandwidth or network topology constraints can force a change in the a priori optimal protection choice. In this new scenario, current routing algorithms must be analyzed.

In Chapter 2, the main aspects of the QoS routing methods are introduced, and some of these mechanisms are described and compared. Chapter 2 points out that major QoS routing algorithms do not include protection as a main objective and, moreover, the same QoS objectives for selecting the working path are used for selecting the backup path.

In order to evaluate the quality of protection, two novel concepts are introduced and analyzed in Chapter 4: the *network failure probability* and the *failure impact*. The physical network provides an initial value of the network protection level in terms of network reliability and availability. In Chapter 3 a proposal to evaluate network reliability is introduced, while in Chapter 4 a formulation to calculate the failure impact (the QoS degradation in terms of packet loss and delay) is presented.

In Chapter 4 a proposal to reduce the failure probability and failure impact as well as the enhancement of some current routing algorithms in order to achieve better protection are explained. A review of the traffic services protection requirements and a new classification, based on the failure probability and failure impact values, is also provided in this work.

Results, in Chapter 5, show that path protection schemes improve network reliability. Segment/local protection schemes reduce the network failure impact. Minimum impact with maximum reliability can be achieved using local protection throughout the entire network. However, it is not scalable in terms of resource consumption. In this case our failure probability evaluation model can be used to minimize the required resources. Results demonstrate the reduction of the failure impact combining segment protection and our network reliability evaluation model in different network scenarios.

The time required for transmitting the failure indication and backup activation signals is crucial for the failure impact minimization. The main components for reducing the time to transmit these signals are also analyzed. Nodes delay (queuing and processing time) and link delay (transmission and propagation time) for signaling and flooding-based schemes are formalized and experimentally compared. Results show that network design components minimize the fault recovery time.

In summary, an in-depth analysis is carried out and a formulation to evaluate the network protection level is presented. This evaluation is based on network reliability maximization and failure impact reduction in terms of QoS degradation. A scalable proposal in terms of resource consumption, detailed and experimentally analyzed, offers the required level of protection in different network scenarios for different traffic services.

Resum

Les noves tecnologies a la xarxa ens permeten transportar, cada cop més, grans volums d' informació i trànsit de xarxa amb diferents nivells de prioritat. En aquest escenari, on s'ofereix una millor qualitat de servei, les conseqüències d'una fallada en un enllaç o en un node esdevenen més importants. D'altra banda, els serveis de xarxa IP encara no ofereixen el mateix grau de fiabilitat que ofereixen les tradicionals xarxes de telecomunicacions (per exemple les xarxes de telefonia). En la nostra opinió, l'èxit o el fracàs dels nous serveis, amb majors necessitats de QoS, que vindran a substituir les actuals xarxes de telecomunicació, dependrà del nivell de fiabilitat que pugui proporcionar la nostra xarxa.

Un òptim disseny inicial de la xarxa es pot degenerar degut a canvis en la càrrega de la xarxa, noves característiques del trànsit de la xarxa, etc. Els recursos de la xarxa també varien degut a noves reserves o canvis en la topologia (com els produïts per fallades de nodes o enllaços). Una part important del disseny de la nova generació de xarxes amb qualitat de servei serà la fialibitat de la xarxa. Aquesta es pot obtenir mitjançant mecanismes de control de fallada, aplicats a diversos nivells de xarxa i diverses escales temporals. Multiprotocol Lavel Switching (MPLS), juntament amb l'extensió a MPLS generalitzat (GMPLS), proporcionen mecanismes ràpids de recuperació de fallada establint camins, Label Switch Path (LSPs), redundants per ser utilitzats com a camins alternatius. En cas de fallada podrem utilitzar aquests camins per redireccionar el trànsit.

El principal objectiu d'aquesta tesi ha estat millorar alguns dels actuals mecanismes de recuperació de fallades MPLS/GMPLS, amb l'objectiu de suportar els requeriments de protecció dels serveis proporcionats per la nova Internet. L'altre objectiu ha sigut la definició i avaluació del nivell de protecció proporcionat per cadascun dels mecanismes de protecció. Per tal de fer aquesta avaluació s'han tingut en compte alguns paràmetres de qualitat de protecció com els temps de recuperació de fallada, les pèrdues de paquets o el consum de recursos.

En aquesta tesi presentem una completa revisió i comparació dels principals mètodes de recuperació de fallada basats en MPLS. Aquest anàlisi inclou els mètodes de protecció del camí (backups globals, backups inversos i protecció 1+1), els mètodes de protecció locals i els mètodes de protecció de segments. El procés de recuperació de fallada (com, quan i on es creen i s'utilitzen els mecanismes de recuperació de fallades) és analitzat àmpliament per cadascun dels mecanismes de protecció. També s'ha tingut en compte l'extensió d'aquests mecanismes a les xarxes òptiques mitjançant el pla de control proporcionat per GMPLS.

Principalment aquesta tesi s'ha centrat en els mecanismes de reserva de banda amb pre-reserva i banda dedicada. No obstant, també hem considerat altres tècniques com les de banda compartida, banda dedicada o les de no reserva de banda. D' altra banda ens hem centrat en els mecanismes de recuperació de fallades d'un sol component de xarxa (enllaços). No obstant, presentem una aproximació pel tractament de fallades múltiples.

En una primera fase (Capítol 1) d'aquest treball, cada mètode de recuperació de fallades és analitzat sense tenir en compte restriccions de recursos o de topologia. Aquest anàlisi ens dóna una primera classificació dels millors mecanismes de protecció en termes de pèrdues de paquets i temps de recuperació. Aquest primer anàlisi no és aplicable a xarxes reals. Per tal de tenir en compte aquest nou escenari, en una segona fase, s'analitzen els algorismes d'encaminament on sí tindrem en compte aquestes limitacions i restriccions de la xarxa.

En el Capítol 2 revisem alguns dels principals algorismes d'encaminament amb qualitat de servei i alguna de les principals propostes d'encaminament per xarxes MPLS. La majoria dels actual algorismes d'encaminament no tenen en compte l'establiment de rutes alternatives o utilitzen els mateixos objectius per seleccionar els camins de treball i els de protecció.

Per millorar el nivell de protecció introduïm i formalitzem, en el Capítol 4, dos nous conceptes: la *Probabilitat de fallada de la xarxa* i *l'Impacte de fallada*. Un

anàlisi de la xarxa a nivell físic proporciona un primer element per avaluar el nivell de protecció en termes de fiabilitat i disponibilitat de la xarxa. En el Capítol 3 introduïm la nostra proposta per avaluar la probabilitat de fallada. En el Capítol 4, formalitzem l'impacte d'una fallada, quant a la degradació de la qualitat de servei (en termes de retard i pèrdues de paquets).

En el Capítol 4 també expliquem la nostra proposta per reduir la probabilitat de fallada i l'impacte de fallada. Detallem l'aplicació d'aquesta proposta per la millora d'alguns dels actuals mecanismes d'encaminament amb qualitat de servei. Per últim fem una nova definició i classificació dels serveis de xarxa segons els valors requerits de probabilitat de fallada i impacte.

Un dels aspectes que destaquem dels resultats d'aquesta tesi (Capítol 5) és que els mecanismes de protecció global del camí maximitzen la fiabilitat de la xarxa, mentre que les tècniques de protecció local o de segments de xarxa minimitzen l'impacte de fallada. Per tant podem assolir mínim impacte i màxima fiabilitat aplicant protecció local a tota la xarxa, però no és una proposta escalable en termes de consum de recursos. Nosaltres proposem un mecanisme intermig, aplicant protecció de segments combinat amb el nostre model d'avaluació de la probabilitat de fallada.

Al llarg d'aquesta tesi demostrem que la minimització del temps de transmissió de les senyals d'indicació de fallada i activació dels backups són crucials en la minimització de l'impacte. Els principals aspectes per reduir els temps de transmissió d'aquestes senyals s'analitzen detalladament. En els Capítols 4 i 5, presentem una detallada formalització i experimentació del retard en els nodes (temps de procés i encuament) i el retard en els enllaços (temps de transmissió i de propagació) per a les dues principals estratègies de notificació de fallades ('signaling' i 'flooding'). Els resultats d'aquest anàlisi, ens han permès definir nous objectius en la millora dels algorismes d'encaminament.

Resumint, aquesta tesi presenta diversos mecanismes per l'anàlisi del nivell de protecció de la xarxa. Els resultats dels models i mecanismes proposats milloren la fiabilitat i minimitzen l'impacte d'una fallada en la xarxa.

Acknowledgments

I would like to express my sincere thanks to my thesis advisor Dr. Josep Lluís Marzo, for giving me the opportunity to work in his research group, for trusting in my commitment, for his guidance and encouragement, and for his stimulating suggestions and discussion in many topics which helped me throughout this thesis. I would also like to acknowledge Dr. Teo Jové for guiding me to the appropriate line of investigation in the first phase of this thesis.

I wish to express my gratitude to all members of Broadband Communications and Distributed Systems group. Special thanks go to Pere Vilà, Ramon Fabregat and Anna Urra with whom I have discussed several issues in this thesis.

I wish also to express my gratitude to Dr. Caterina Scoglio and Tricha Anjali from the Broadband and Wireless Networking Laboratory of Georgia Institute of Technology, for their comments, paper reviews and help in many parts of this thesis.

Last but certainly not least, I wish to give again my deepest thanks to my parents and to Laura. Without their love, support and encouragement I doubt that this thesis would ever have been written.

Finally, I wish to thank the many people who have, in one way or another, made this thesis possible. I apologize for not listing everyone here.

This work has been partially supported by the Ministry of Science and Technology of Spain under contracts MCYT TIC2003-05567, MCYT TIC2002-10150-E, CICYT TEL99-0976, and by UdG research support fund (UdG-DinGruRec2003-GRCT40).

Contents

List of Figures

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

List of Tables

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

Enhanced fault recovery methods for protected traffic services in GMPLS networks

Eusebi Calle

CHAPTER 1

1

Fault Recovery in GMPLS/MPLS networks

1.1. Introduction

An initial design of a network may not be satisfactory due to changes in offered load, traffic characteristics, and so on. Network resources also vary due to resource reservations and topology changes (such as node or link failures). An important part of designing a QoS network concerns the reliability of the network. This reliability can be provided with fault management mechanisms, applied at different network levels and time scales.

Multiprotocol Label Switching (MPLS) and, recently, Generalized Multiprotocol Label Switching (GMPLS) provide fast restoration methods for recovery from

failures by establishing redundant Label Switch Paths as backup paths. With these backups, traffic can always be redirected in case of failure.

In this section the main characteristics of the design and application of the GMPLS/MPLS recovery methods are discussed. First, MPLS and GMPLS are briefly reviewed. Then, the main components involved in a network-protected scenario are presented and their principal functions are described. Different options for classifying the recovery models, such as the M:N model, are also presented in this section. Finally, the recovery cycle is analyzed and a comparison of the main recovery models is presented.

1.2. Introduction to MPLS and GMPLS

Multiprotocol Label Switching (MPLS) has two main objectives: to speed up packet forwarding, and to provide traffic engineering in IP networks. To achieve a fast forwarding MPLS uses a label-swapping scheme rather than address matching to determine the next hop for a received packet. In order to provide traffic engineering IP networks operate like connection-oriented networks. MPLS separates the control (signaling and routing) and data label (forwarding). Generalized MPLS (GMPLS) extends the devices (packet-based, time-based, wavelength-based and fiber-based) where the switching operation can be deployed onto a single common control plane.

1.2.1. MPLS background and operation

Multiprotocol Label Switching emerged from the evolution of routing/forwarding protocols. MPLS delivers a solution that integrates the control of Level 3 routing with the simplicity of Label 2 switching. Basically MPLS provides the separation of control and forwarding components and the Label-swapping forwarding algorithm [ROS98] and [DAV00].

The control plane has two main functions: path discovery (routing), which involves creating the routing tables, and the signaling function (to signal a routed

path). The routing protocol exchanges information with other routers to build and maintain a routing table, using standard level 3 routing protocols, such as Open Shortest Path First (OSPF) [MOY98], Intermediate System to Intermediate System (IS-IS) [ORA90], or Border Gateway Protocol (BGP). The forwarding table is maintained from the control engine and is distributed along network nodes from a signaling protocol, such as the Resource Reservation Protocol (RSVP) [BRA97] and [AWD01] or Label Distribution Protocol (LDP) [AND01].

The forwarding component is based on a label-swapping forwarding algorithm (the same algorithm used to forward packets in ATM and Frame Relay switches). Signaling protocol and label distribution allows the creation of the Label Switch Paths (LSP) similar to Asyncronous Transfer Mode (ATM) Virtual Paths and Virtual Circuits (VPI/VCI).

One major feature of MPLS is its ability to place IP traffic on a defined path through the network. For each specific service, a table of Forwarding Equivalence Class (FEC) is created to represent a group of flows with the same traffic-engineering requirements.

This is one of the key elements that make MPLS so useful. FECs allow MPLS 'flow aggregation', assigning a single label to different flows with the same FEC. Flow aggregation reduces the number of labels needed to handle a particular set of packets, and also reduces the amount of label distribution control traffic needed. This improves scalability and reduces the processing time.

At the ingress node of an MPLS network, incoming IP packets are examined and this node, called Label Edge Router (LER) assigns a label. The Label packets are forwarded along the path, called Label Switch Path (LSP), where each Label Switch Router (LSR) makes a switching decision based on the packet label field. The Label Information Base (LIB) provides an outgoing label and the outgoing interface. Figure 1.1 shows the MPLS architecture.

Once the LSP is selected (routed) and the signaling protocols have assigned the labels, this LSP can be used. When an IP packet arrives, the ingress node (LER1)

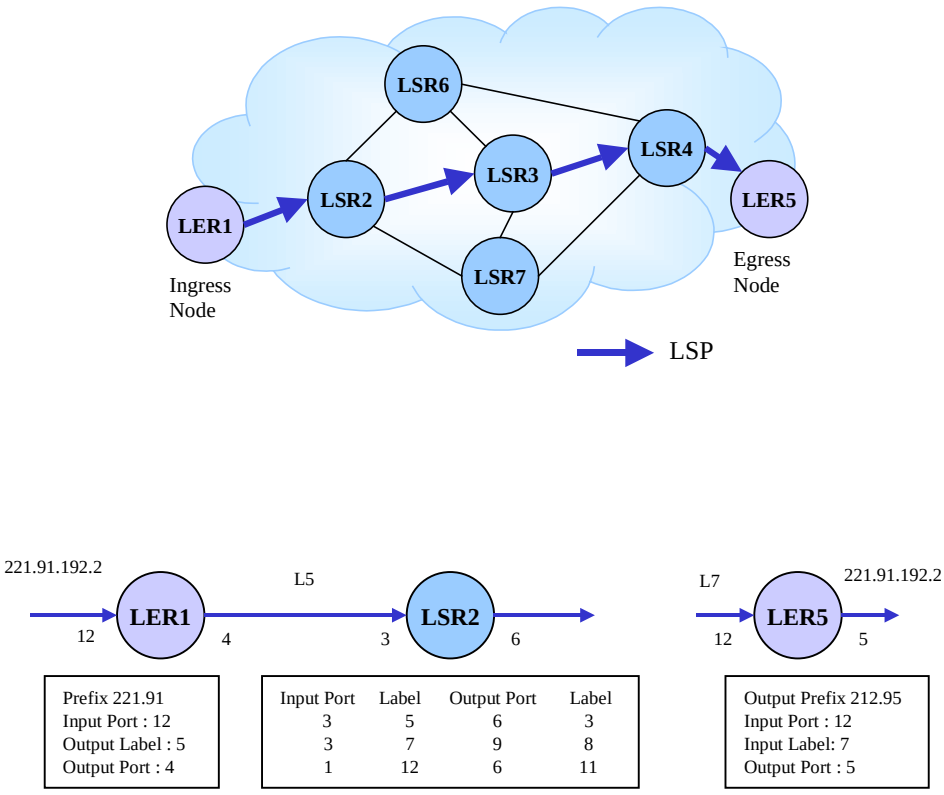


Figure 1.1: MPLS architecture

analyzes the IP address and input port and assigns the outgoing label 5. The MPLS core label switch routers, such as Label Switched Router 2 in the Figure 1.1, only use the LIB (Label Information Base) to execute the forwarding process. A packet that arrives with a Label 5 using input port 3 is forwarded to the output port 6 with the label 3. At the egress node the LER recovers the IP address.

Figure 1.2 shows the MPLS header (MPLS 'shim' header) where the label field (20 bits) carries the actual value of the MPLS header. The EXP (experimental field) (3 bits) is for QoS provisioning. The Time To Live (TTL) field (8 bits) provides conventional IP TTL functionality. The S (Stack) field (1 bit) supports a hierarchical label stack, which is a sequence of labels on the packet organized as

a last-in, first-out stack. A label stack enables a packet to carry information about more than one FEC, allowing it to traverse different MPLS domains or LSP segments within a domain using the corresponding LSPs along the end-to-end path. Note that label processing is always based on the top label. This property of MPLS is essential in the context of optical networks because the number of wavelengths (which act as labels) is not very large.

Figure 1.2: MPLS header

Extending OSPF and IS-IS allows nodes to exchange information about network topology, resource availability and even policy information [APO99]. This information is used by the routing protocol, in combination with the RSVP-TE [BRA97], [AWD01] and CR-LDP [ASH02] signaling protocol extensions, to compute and create paths subjected to specified resource and/or policy constraints. This allows MPLS to fulfill, in a suitable manner, two main purposes: traffic engineering and fast rerouting.

1.2.2. Generalized MPLS

The Internet Engineering Task Force (IETF) has extended MPLS protocols to include other switching mechanisms via Generalized MPLS (Table 1.1 summarizes these switching technologies). The development of GMPLS requires modifications to current signaling and routing protocols. New protocols such as the Link Management Protocol (LMP), have also been developed to manage and maintain the functionality of the control and data planes between two neighboring nodes. The current IETF GMPLS standardization efforts are summarized below:

- A new Link Management Protocol (LMP) designed to address issues related to link management in optical networks.

- Enhancements to OSPF and IS-IS routing protocols to advertise availability of optical resources in the network (e.g. bandwidth on wavelengths, link protection type, or fiber identifiers).

- Enhancements to RSVP-TE and CR-LDP signaling protocols for traffic-engineering purposes that allow LSPs to be explicitly specified across the optical core.

- Scalability enhancements, such as hierarchical LSP formation, link bundling and unnumbered links.

Switching Domain	**Traffic Type**	**Forwarding Scheme**	**Example of device**	**Nomenclature**
Packet or cells	IP or ATM	Label (shim header) or virtual channel connection (VCC)	IP router ATM switch	Packet switch capable (PSC)
Time	TDM/SONET	Time slot in repeating cycle	Digital cross-connect system (DCS), ADM	TDM switch capable (TSC)
Wavelength	Transparent	Lambda	Dense wavelength-division multiplexing (DWDM)	Lambda switch capable (LSC)
Physical Space	Transparent	Fiber	Optical Cross-connect (OXC)	Fiber switch capable (FSC)

Table 1.1: GMPLS switching technologies

There are some aspects to take into consideration while using MPLS to control optical networks. For instance, the MPLS label space is comparatively large (one million per port), whereas there is a relatively limited number of lambdas and Time-division Multiplexing (TDM) channels (tens to hundreds per port today, scaling to thousands in the next few years). Optical networks will deploy hundreds of parallel fibers, each carrying hundreds of lambdas between a pair of network elements. This means that the overall number of links in optical/TDM networks can be several orders of magnitude larger than that of an MPLS network.

1.3. Fault recovery models architecture

In this section the main aspects to develop are reviewed and the use of GMPLS/MPLS fault recovery methods is presented.

A GMPLS/MPLS Protection Domain is defined as the set of Label Switch Routers (LSRs) over which a working path and its corresponding protection path are routed. The protection domain is denoted as working path and backup path.

Figure 1.3 shows a simple MPLS protected domain, formed by a Working Label Switch Path (or a segment of the WP), which is the protected segment and the Backup Label Switch Path (or the Recovery Path) where the traffic is switched once a failure is detected. The Protection Switch LSR (PSL) and Protection Merge LSR (PML) components are two Label Switch Routers with the protection function. All the components involved in the MPLS fault control will be further detailed in this section.

Some definitions used throughout the following sections and chapters are listed below:

- *Working or Active Label Switch Path*: An LSP established to carry traffic from a source LSR to a destination LSR under normal conditions, i.e.

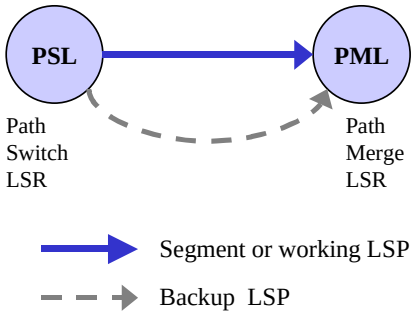


Figure 1.3: MPLS protection domain components.

in the absence of failures. In other words, a working LSP is an LSP that contains traffic streams requiring protection. The portion of a working LSP that requires protection is denoted as a protected segment.

- *Protection or Backup Label Switch Path*: An LSP established to carry the traffic of a working path (or paths) following a failure on the working path (or on one of the working paths, if more than one exists) and a subsequent protection switch by the PSL. A protection LSP may protect either a segment of a working LSP or an entire working LSP.

- *Label Switch Router (LSR)*: The MPLS term, Label Switch Router (LSR) is used in this document to describe a circuit-switch node such as an optical cross-connects (OXCs) in optical networks.

- *Protection Switch LSR (PSL)*: A PSL is the LSR that is the origin of both the working path and its corresponding protection path. Upon learning of a failure, either via the Fault Indication Signal (FIS) or via its own detection mechanism, the protection switch LSR switches protected traffic from the working path to the corresponding backup path.

- *Protection Merge LSR (PML)*: The PML is the LSR that terminates both a working path and its corresponding protection path, and either merges

their traffic into a single outgoing LSP, or, if it is itself the destination, passes the traffic on to the higher layer protocols.

- *Selector and Bridge nodes*: In GMPLS networks there is a specific terminology referring to PSL and PML nodes. In this case they are called Bridge and Selector nodes, respectively ([PAP03] and [LAN03]).

1.4. Fault recovery models classification

In the recent literature many different ways of classifying the fault recovery models have been presented. In this section three classification models are depicted. The first model is based on the number of working and backup paths, the second model is based on when the backups are computed and the resource allocation, and the last model is based on fault recovery cycle (from the moment the failure is detected to when the traffic is restored).

1.4.1. The M:N model

Probably the most well-known fault recovery classification method is the M:N model. It is based on the number of backup and protected working paths. In this model M is the number of backup LSPs used to protect N, the working LSPs. Using this model the feasible protection models could be:

- 1:1: 1 working LSP is protected/restored by one backup LSP.
- M:1: 1 working LSP is protected/restored by M backup LSPs.
- 1:N: 1 backup LSP is used to protect/restore N working LSPs (shared backups).
- N:M : N working LSPs are restored by M backup LSPs

- 1:0 : No protection (for instance, Best effort traffic)

- 1+1: Traffic is sent concurrently on both the working LSP and the backup LSP.

1.4.2. The path provisioning and resource allocation model

In this section another way to classify the fault recovery methods based on when the alternative (backup) paths are computed and their type of resource allocation is presented.

The path provisioning can be categorized, as shown in Figure 1.4, according to when the backup path is computed (pre-computed or computed on demand), or when the path is established (pre-established or established on demand) and when the resources are allocated to this path (pre-allocated or allocated on demand).

There are many mechanisms mentioned in the literature to reserve resources. In all these cases the level of resource reservation, as shown in Figure 1.5, can be classified as dedicated (such as 1:1, 1+1), shared (1:N, M: N) or no reservation (1:0, best effort), in which case the failure is recovered only if the resources are available.

Under shared restoration preemptable traffic (preempting low priority connections in case of resource contention) and non-preemptable traffic are supported.

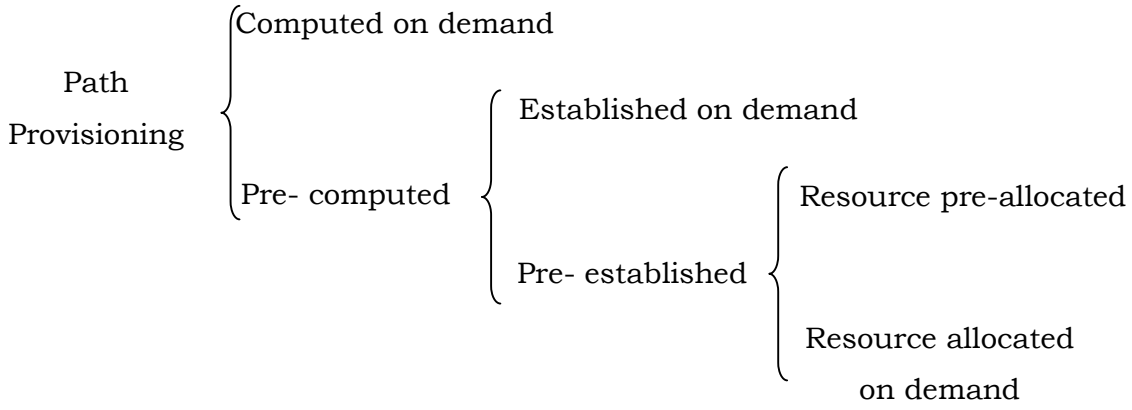


Figure 1.4: Path provisioning classification

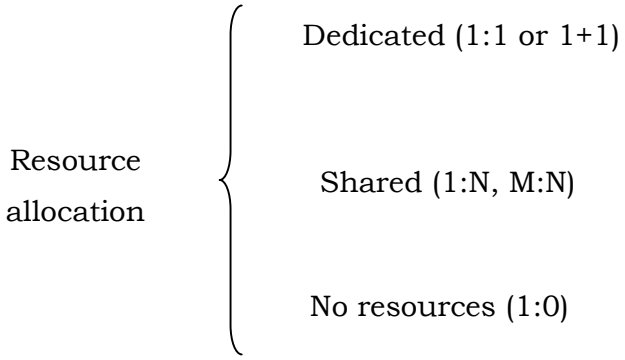


Figure 1.5: Resource allocation classification.

In the case of pre-allocated backup paths, there is the question of what use these resources may be put to when the backup path is not in use. According to [SHA03], there are two options:

- *Dedicated-resources*: If the backup path resources are dedicated, they may not be used for anything except carrying the working traffic. For example, in the case of 1+1 protection, the working traffic is always carried on the backup path. Even if the backup path is not always carrying the working traffic, it may not be possible or desirable to allow

other traffic to use these resources.

- *Extra-traffic-allowed*: If the backup path only carries the working traffic when the working path fails, then it is possible to allow extra traffic to use the reserved resources at other times. Extra traffic is, by definition, traffic that can be displaced (without violating service agreements) whenever the recovery path resources are needed to carry the working path traffic.

1.4.3. The fault recovery cycle model

Another way to classify the fault recovery schemes is based on the fault recovery cycle. Fault recovery methods begin with fault identification and end with link recovery. This cycle of events contains various components:

a. A method for selecting the working and protection paths (routing algorithm).
b. A method for signaling these paths setup, (for instance, LDP/RSVP or CR-LDP/RSVP-TE).
c. A mechanism for fault detection.
d. A hold off time.
e. A fault notification method.
f. A switchover mechanism to move traffic from the working path to the backup path (possibly requiring a complete recovery method to activate the backup path)
g. A repair detection mechanism (optional), to detect that a fault along a path has been repaired.
h. A switchback or restoration mechanism (optional), for switching traffic back to the primary working path, once it is discovered that the fault has been corrected or has been repaired.

This is the general cycle of events that describes the establishment and

utilization of a protection method; however some recovery methods do not need all of these components, or the sequence order can be modified. For instance, fault notification (e)) in local backups or switchover in 1+1 methods are not required. In case of dynamic (i.e. computed on demand) fault management methods, steps a) and b) for selecting and signaling the backup path come after step e).

Table 1.2 shows some fault recovery model examples and their recovery cycle.

Fault recovery model	**Fault Rec. Cycle**
No local protection, Pre-computed, Pre-establish, Res. Pre-allocated, (1:1,1:N,M:N).	a,b,c,d,e*,f, [g,h] **
Local protection, Pre-computed-, Pre-establish, Resource Pre-allocated, (1:1,1:N,M:N).	a,b,c,d,f, [g,h] **
No local protection, Computed on demand, (1:1,1:N,M:N)..	c,d,e*,a,b,f [g,h] **
Local protection, Computed on demand, (1:1,1:N,M:N).	c,d,a,b,f [g,h] **
Pre-computed, Pre-establish, Resource Pre-allocated, (1+1)	a,b,c,d,f, [g,h] **

Table 1.2: Fault recovery models and fault recovery cycle.

* e (fault notification) occurs when the LSR which detects the failure is not responsible for the switchover. No local protection schemes. (see section 1.5.)

** g and h steps (primary WP restoration) are always optional (normalization [SHA03]).

Each step involves a completion time. A more detailed explanation of each phase and the main aspects related to that recovery time are presented in the following.

Fault Detection

This phase involves the time between the occurrence of a network fault and the moment the fault is detected by MPLS or GMPLS-based recovery mechanisms.

According to [SHA03] there are four classes of impairments: Path Failure, Path Degraded, Link Failure, and Link Degraded.

- *Path Failure (PF)* is a fault that indicates to an MPLS-based recovery scheme that the connectivity of the path is lost. This may be detected by a path continuity test between the PSL and PML. Some, and perhaps the most common, path failures may be detected using a link probing mechanism between neighboring LSRs. An example of a probing mechanism is a liveness message that is exchanged periodically along the working path between peer LSRs.

- *Path Degraded (PD)* is a fault that indicates to MPLS-based recovery mechanisms that the path has connectivity, but that the quality of the connection is unacceptable. This may be detected by a path performance monitoring mechanism, or some other mechanism for determining the error rate on the path or some portion of the path. This is local to the LSR and consists of excessive discarding of packets at an interface, either due to label mismatch or due to TTL errors.

- *Link Failure (LF)* is an indication from a lower layer that the link over which the path is carried has failed. If the lower layer supports detection and reporting of this fault, i.e. any fault that indicates link failure for example SONET Loss of Signal (LoS), this may be used by the MPLS recovery mechanism. In some cases, using LF indications may provide faster fault detection than using only MPLS-based fault detection mechanisms.

- *Link Degraded (LD)* is an indication from a lower layer that the link over which the path is carried is performing below an acceptable level. If the

lower layer supports detection and reporting of this fault, it may be used by the MPLS recovery mechanism. In some cases, using LD indications may provide faster fault detection than using only MPLS-based fault detection mechanisms.

In some mechanisms it is required that the node upstream of the faults be able to detect the failure. If LSPs are unidirectional, some failures (those no reported by the lower layer) can not be detected. In this case a 'Hello protocol', should be used [HUA02]. In [HUA02] a 'hello protocol', similar to the Open Shortest Path First (OSPF) [MAY98] is proposed. However, timers in routing protocols are typically set to relatively large values compared to what is needed for a recovery mechanism. Also, the fault detection mechanism must provide the trigger for generating the Fault Indication Signal (FIS). Their 'hello protocol' proposal provides a mechanism which is complementary to all existing mechanisms such as physical layer fault detection through liveness messages exchanged between neighboring LSRs. Each LSR sends a liveness message periodically to its neighbors. A liveness message will carry the identification (ID) of the LSR and the IDs of its neighbors discovered through the liveness messages sent by its neighbors. An LSR can learn if a bi-directional link is working properly if it sees its own ID in the liveness message sent by the LSR at the other end of the link.

In GMPLS it is important to identify the data plane errors and the control plane errors, due to the fact that both topologies (data and control plane) do not have to coincide.

Hold-Off

Hold-off corresponds to the configured waiting time between the detection of a fault and the taking of MPLS-based recovery action, to allow time for lower layer protection to take effect. The hold-off time in the case of GMPLS over the SONET can be set to 50 ms such that SONET protection scheme can be activated before the MPLS layer recovery mechanism is triggered. However, if the network is not able to recover faults at lower layers the hold-off time is not activated. For instance, in the case of WDM-based recovery, this time should be zero since

there is no underlying layer recovery.

The Fault Hold-Off Time may occur after the Fault Notification if the node responsible for the switchover, the Path Switch LSR (PSL), rather than the detecting LSR, is configured to wait.

Fault Notification

The time between initiation of a fault notification message (e.g Fault Indication Signal (FIS)) by the LSR detecting the fault and the time at which the Path Switch LSR (PSL) begins the recovery operation. Zero if the PSL detects the fault itself.

In order to achieve a faster recovery time, it is important that the intermediate nodes do not process, unnecessarily, the fault indication. RSVP-TE [BRA97], [AWD01] and [LAN03] provide the 'notify' message to indicate if a failure has occurred. In GMPLS the 'notify' message provides the advantage of indicating data plane and control plane errors. This is very important in GMPLS, due to the fact that the data and control plane can be physically separated, and could fail independently. When a control plane error occurs and, the data plane is still working, the 'notify' message indicates which LSP has failed and what resources have been affected.

Some nodes (ingress or egress nodes) have the capability of detecting some types of failures (for example, a Loss of Light (LoL)). However the fault notification message should be sent to all concerned nodes on the recovery path in order to solve the worst-case scenarios.

Switchover

A switchover is the process of switching the traffic from the path through which the traffic is flowing (working path) onto the alternate path (backup path). This phase starts after the responsible entities (PSL and PML nodes) are notified of the failure and the backup path is activated.

1.5. GMPLS/MPLS main fault recovery models

In this section the main fault recovery models are presented. These models were first defined for MPLS networks in [HUA02] and [SHA03]. Afterwards, some of these models were adapted to optical networks (via GMPLS control plane). This section explores the failure recovery functionality of these models and a comparison, based on some QoS parameters, is made. Basically, the parameters used in this comparison are: recovery time, packet loss, packet reordering and resource consumption. While this section only introduces an intuitive comparison, in the next chapters a more formalized analysis is detailed.

1.5.1. The global/centralized backup model

In this model [HUA02], the Ingress Node takes the responsibility for fault recovery as the Fault Indication Signal (FIS) arrives. This method requires the establishment of an alternate disjoint backup path for each active path (working path).

In this model protection is always activated at the Ingress Node, irrespective of where a failure occurs along the working path. This involves propagating the failure information all the way back to the source node before a protection switch

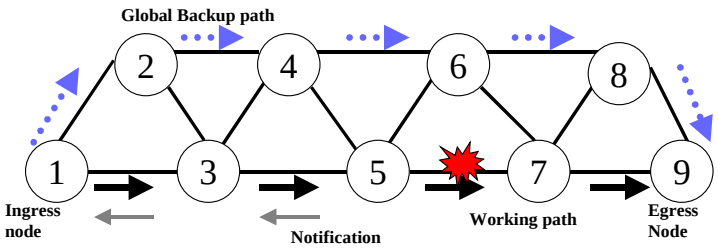


Figure 1.6: The global/centralized backup model

is activated. If no reverse LSP is created, the fault indication can only be activated as a Path Continuity Test (the ingress node monitors the path).

This method has the advantage of setting up only one backup path per working path. This is a centralized protection method that means that only one LSR has to be provided with PSL features and another LSR with the PML functions. On the other hand, this method has an elevated cost (in terms of recovery time), in particular if a Path Continuity Test (a monitoring technique to detect link or node failures) is used as a fault indication method. During this time there is a packet loss proportional to the required recovery time. Moreover, those packets that were circulating on the failed link at the time of the failure will also be lost. This is a common drawback in all recovery models, but, there are currently some proposals such as [HUN02] that avoids such packet loss in the failed link by applying tagging and buffering techniques.

Figure 1.6 shows a simple scenario formed by nine LSRs where a working path (i.e: LSR1-LSR3-LSR5-LSR7-LSR9) and an LSP backup recovery path (i.e: LSR1-LSR2-LSR4-LSR6-LSR8-LSR9) are pre-established. In normal operation, traffic from ingress router LSR1 to egress router LSR9 is carried through the LSP working path. When a link fault is detected (for instance between LSR5 and LSR7) a failure notification signal (FIS) is sent to the ingress node (LSR1). When the notification arrives at LSR1, traffic is switched to the LSP global backup path (see Figure 1.6).

1.5.2. The reverse backup model

Pre-established alternative paths are essential where packet loss due to an LSP failure is undesirable. Since it may take a significant time for a device on a label switched path to detect a distant link failure, packets could continue to be sent along the primary path. As soon as such packets reach a switch that is aware of the failure, the switch to an alternative path away from the failure must immediately reroutes them if the loss of data is to be avoided.

The main function of this method is to reverse traffic at the point of a failure of

the protected LSP back to the source switch of the protected path (Ingress Node) via a Reverse Backup LSP.

As soon as a failure along the protected path is detected, the LSR at the ingress of the failed link reroutes incoming traffic. It redirects this traffic into the alternative LSP traversing the path in the reverse direction of the primary LSP. The traffic and notification signal are both sent to the ingress node. As soon as the FIS arrives the ingress node stops sending traffic to the working path and switches the traffic to the alternative (global backup) path.

This method is especially suitable in network scenarios where the traffic streams are very sensitive to packet losses. For example, in voice transmission, delay is common, but if a file is being transmitted, packet losses could be critical. If the link segment or the node where the failure occurs is allocated far from the ingress node and the transmission rate is very fast, the number of packets lost could be very high if a centralized backup is used. Reverse backup utilization allows the recovery of packets as the failure occurs, rescuing lost packets if a centralized method is applied.

Another advantage is that the fault indication mechanism is simplified, since the reverse backup offers, at the same time, a way to transmit the fault indication signal to the ingress node and the traffic is recovered via reverse backup (see

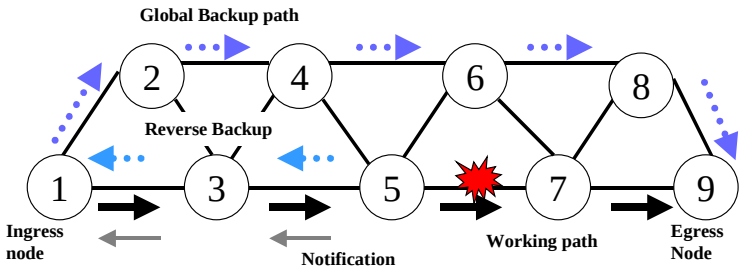


Figure 1.7: The reverse backup model

Figure 1.7). One disadvantage is poor resource utilization. Two backups per protected domain (segment or protected path) are needed. Another drawback is the time required to reverse fault indication to the Ingress Node as in the Centralized model. Regardless, a reverse backup can be established in association with the working path, simply by making each LSR along a working path remember its neighbor. Another problem of this scheme is that packets arriving from the reverse direction are mixed with incoming packets, resulting in packet disordering through the alternative LSP during the recovery period.

Figure 1.7 shows an example of reverse backup utilization. LSP working and recovery paths are established as in the centralized model. In addition there is also a reverse path from LSR5 (LSR5-LSR3-LSR1) which reaches the ingress node. When a link failure is detected in LSP (LSR5-LSR7), the traffic is switched back to LSR1 (ingress node) through the reverse backup LSP, and then carried through the LSP recovery path as in the centralized/global model.

1.5.3. The local/segment backup model

With this model fault recovery starts from the point of the failure. It is a local method and is transparent to the Ingress Node. The main advantage of this model is that it offers lower recovery time than the global/centralized model and it avoids packet loss.

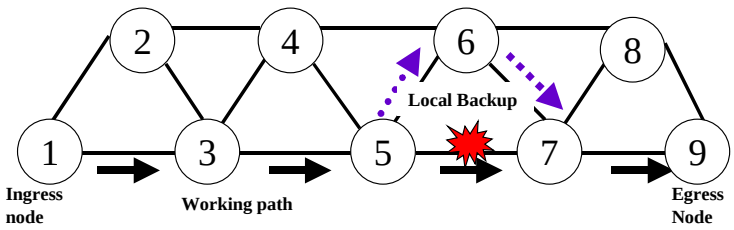


Figure 1.8: The Local backup model

An added difficulty, of the local model, is that every LSR, where protection is required, has to be provided with switchover function (PSL). A PML should be provided too. Another drawback is the maintenance and creation of multiple LSP backups (one per protected domain), meaning low resource utilization and a high management complexity. On the other hand, this method offers transparency to the Ingress Node and faster restoration time than centralized mechanisms.

Figure 1.8 illustrates this case. The same working path as in the global model is used (i.e: LSR1-LSR3-LSR5-LSR7-LSR9). The LSP local backup path is formed by LSR5-LSR6-LSR7 that is shorter than the LSP recovery path in the centralized method. The LSR5 should be a PSL node and the LSR7 a PML node. When a link failure occurs, traffic is switched from (LSR5-LSR7), which is a segment of the working path to the LSP backup path (LSR5-LSR6-LSR7).

An intermediate solution could be the establishment of local backups, but only for segments where a high degree of reliability is required, supplying protection for those path segments only.

Figure 1.9 depicts this case, the protected segment is formed by LSR5-LSR7-LSR9. If a failure occurs (i.e. LSR7-LSR9) a FIS is sent to the PSL node (note that in this case the PSL is not the ingress node). After the FIS arrives the traffic is recovered using the alternative segment backup (LSR5-LSR6-LSR8-LSR9).

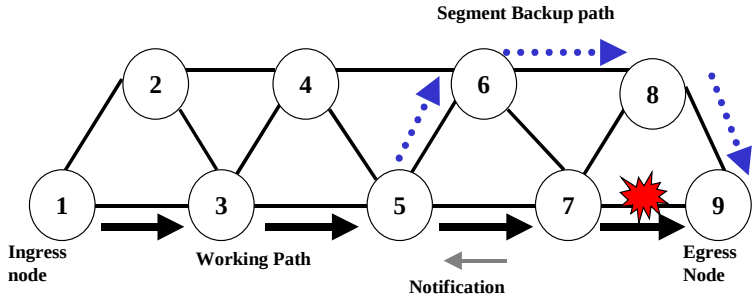


Figure 1.9: The segment backup model

There are several versions of the local/segment model depending on where the PSL and PML nodes are allocated. For instance the PML node could match with the ingress node. However segment protection offers better recovery time than global/reverse models where there is a failure notification time and in the case of the example, packet loss can occur.

1.5.4. The 1+1 model

This fault recovery model uses two working paths (LSR1-LSR3-LSR5-LSR7-LSR9 and LSR1-LSR2-LSR4-LSR6-LSR8-LSR9). In this case the PML/Selector LSR is monitoring the best working path (for instance selecting the best signal). After a failure, the PML/Selector detects that there is only one path and selects this path as the working path. This method is fast and does not lose packets, but it consumes a lot of resources since both paths need to be reserved a priori. Furthermore, a PSL/Bridge LSR also has to be set up so as to be able to send the traffic over both paths. Figure 1.10 depicts this model.

This model avoids failure notification times and packet loss. However there is high resource consumption because both paths need pre-allocated resources. This model is commonly used in optical networks.

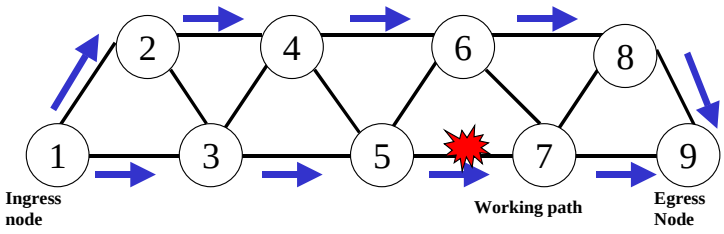


Figure 1.10: The 1+1 model

1.5.5. Protection cycles

Another fault recovery model is the protection cycles (p-cycles). P-cycles can be regarded as pre-configured protection cycles in a mesh network. A p-cycle allows the protection of those links that have their end points in nodes, which belong to the same p-cycle. Consequently, the links belonging to the p-cycle (1-2, 2-3, 3-4 and 4-1 in Figure 1.11) and some that do not ('straddling links', link 1-3 in Figure 1.11) are protected.

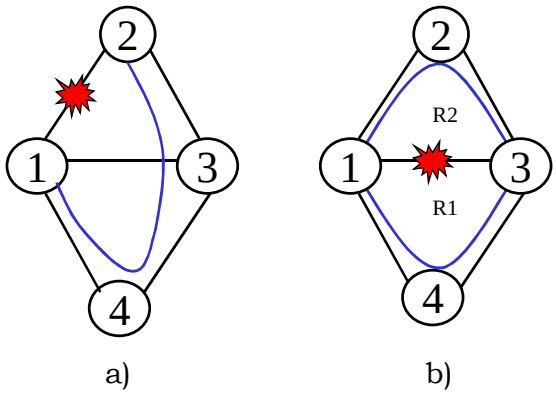


Figure 1.11: Protection Cycles

In the first case, Figure 1.11 (a), the potential faulty link is protected by the remaining links of the p-cycle. In the second case, Figure 1.11 (b), a straddling link can be protected by the two alternative paths provided by the p-cycle (R1, R2), hence more than one working path sharing the same straddling link can be protected.

In this framework, protection-switching decisions can be made quickly because they are carried out in the faulty link. Note that one p-cycle cannot protect more that one link fault at the same time. However, the use of multiple p-cycles in a network decreases the impact of multiple failures.

1.6. Multiple faults

Despite focusing in this work on single failure recovery schemes, in this section we briefly introduce the multiple fault problem. Two schemes to recover multiple failures are also described.

1.6.1. Priority-Based Recovery

Fault recovery schemes typically assume single failure events. However, multiple failures may occur in some short time interval. Protection against occurrences of failure scenarios requires large amounts of spare capacity. Ideally, the network should at least recover some of the working paths in this situation [RS03].

For example, consider Figure 1.12, where two failures occur at the same time. In this case there are two working paths: WP-LSP1: [1-2-3-4] and WP-LSP2: [7-8-9-10] and their respective backup paths BP-LSP1: [1-5-6-4] and BP-LSP2: [7-5-6-10].

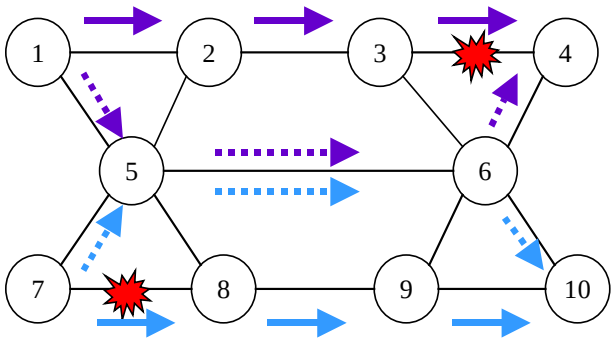


Figure 1.12: Multiple failures: Priority-based recovery

One failure is a link failure between LSRs 3-4, and the other failure is a link failure between LSRs 7-8. In this example, LSR3 detects a failure and sends a fault notification message to the ingress node LSR1. At almost the same time, LSR7 detects a failure. If no prioritization is used, at LSR6, a recovery path switches traffic from LSR6 to LSR4 because the fault notification message is for WP-LSP1. On the other hand, at LSR5, a recovery path switches traffic from LSR5 to LSR6 because the fault notification message is for WP-LSP2. As a result, an invalid recovery path is set to follow (7-5-6-4).

Priority-based control is an effective method and recovers specific working paths under the condition of multiple failures. In the above example, if the priority of WP-LSP1 is higher than WP-LSP2, then the fault notification messages for WP-LSP1 are preferred. In other words, the system checks the priority of the protection path and changes the priority setting. In such a case, the switching traffic from LSR6 to LSR4 takes precedence over switching traffic from LSR6 to LSR10.

By adopting priority-based control, such behavior can be avoided. As a result, the high priority recovery path is activated. Priority in general should be set according to a network operator's policy and/or network service.

1.6.2. Multilevel protection

In [MAR03] more than one protection system is maintained to achieve different protection levels depending on the traffic class-type. Therefore, a multilevel protection scenario is dynamically set up using the main features of the QoS on-line approaches. (For more details see the following chapter).

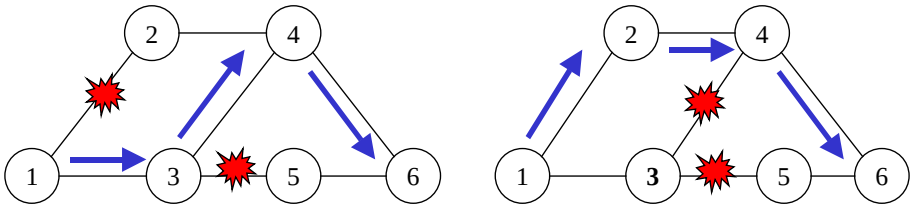
In network scenarios with a high degree of protection requirements, the application of multilevel fault management could improve performance, in comparison with single level management. Nonetheless, complete scenario construction is very costly (in terms of time and resources), so intermediate

scenarios could be built instead.

For example, the protected domain could start with just a global path method and as the protection requirements grow (e.g., a node fails repeatedly), a new local backup path could be established, thus providing a new protection level.

One advantage of using the multilevel protection approach is seen in scenarios with multiple faults. Figure 1.13-(a) shows an example, where the working path is (1-3-5-6). If link (3-5) fails, the global backup path (1-2-4-6) is used at first. Then, if link (1-2), which part of the global backup path, also fails, the local backup path (3-4-6) is used. Therefore, in this multiple fault case, traffic could be routed through path (1-3-4-6) avoiding broken segments. Other link (or node) faults can be overcome in a similar way.

Another application of multilevel protection is shown in Figure 1.13-(b). Again, the working path is (1-3-5-6) and link (3-5) fails. In this case, the local backup path (3-4-6) is used at first. Then if link (3-4) fails too, another backup mechanism (global backu path model) is applied and both faults are overcome.



a) b)

Figure 1.13: Multiple failures: Multilevel protection

1.7. Fault notification models

The IETF-CCAMP has dedicated a great number of efforts to formalize and minimize recovery time. One of the most challenging aspects of minimizing recovery time within the recovery cycle is fault notification. In this section a more detailed review of the main fault notification models is presented. A comparison between different fault notification techniques is also presented. After the fault detection step, there are several options that can be used to transmit information about the fault.

The failure notification could be "per-failure" or "per-LSP" [RSA03]. The main difference between "per-failure" and "per-LSP" notification is in the number of notification mechanisms that have to occur at the same time. Per-failure fault notification allows one mechanism to notify all relevant nodes of the fault. On the other hand, per-LSP notification requires activating as many mechanisms as the number of failed LSPs (for example, all LSPs that failed due to a link failure). In an optical network carrying possibly hundreds of wavelengths per fiber, per-LSP notification can be taxing on the hardware and resource-intensive.

An example of “per-LSP” failure notification is making use of control plane signaling (sending RSVP-TE notify messages as per [BER03]). This is the approach used in [LAN03] where each LSP end-node sends a notify message to its corresponding end-node and receives an ACK back. An example of “per-failure” notification is flooding, where the detecting node floods the network with information about the fault.

1.7.1. Signaling-based notification

In the case of signaling, link failure recovery occurs as part of a process. In the case of a node detecting a failure and notifying the LSP sources, the steps of the process are as follows:

- Detect all LSPs that are affected by a link failure.
- Send a failure indication message to the source of each identified LSP.
- Intermediate nodes that receive the message forward it to on to the LSP source node.

When each LSP source node receives the failure indication message, the following occurs:

- The LSP source node sends a failure acknowledgement message to the detecting node.
- Upon receiving that message, intermediate nodes send it on to the originating node.
- The LSP source node sends an end-to-end switchover request message to the LSP destination node along the protection path, with information about the LSP that is to be recovered.
- The LSP destination node sends an end-to-end switchover response message back to the LSP source node along the protection path.
- Upon receipt of the response message, the LSP source node starts sending data along the protection path.

This process is shown in Figure 1.14. In this case a failure occurs in link 3-4. After the failure is detected, node 3 sends a failure indication message to the ingress node (node 1). Intermediate nodes (node 2) receive this message and send it to the upstream nodes. Once the failure indication message arrives at the ingress node, node 1 sends a failure acknowledgement message to the detecting node (node 3). The Ingress node sends an end-to-end switchover request message to the LSP destination node (node 6) along the protection path (1-7-8-9-10-11-12-6), with information about the LSP that is to be recovered. The LSP destination node (node 6) sends an end-to-end switchover response message back to the LSP source node along the protection path. Once the ingress node receives the response message, the LSP source node starts the switchover.

Figure 1.15 shows the problem of queuing delay in case of using signaling-based techniques. In this case there are 3 LSPs in the broken link. After the failure is detected, node 3 sends as many failure indication messages as the number of LSPs to their ingress nodes. Each message arrives at the intermediate node buffers and is sent to the upstream nodes. In the case of a high number of LSPs

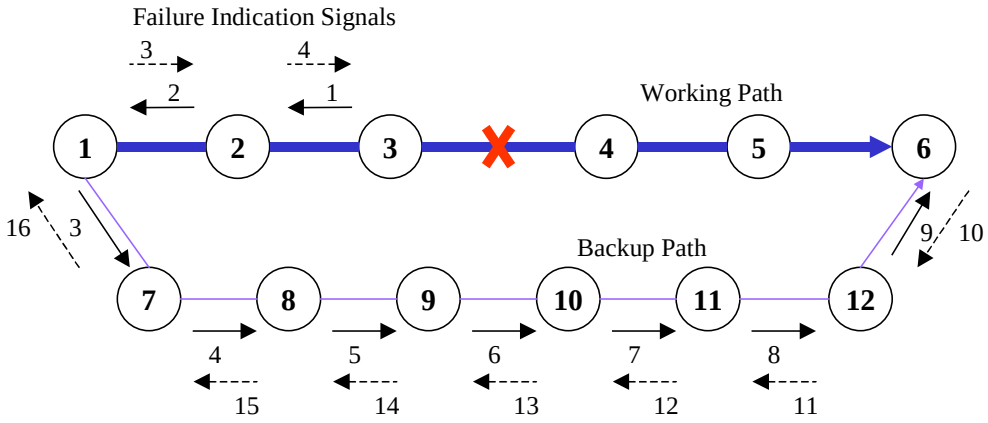


Figure 1.14: Signaling-based failure notification

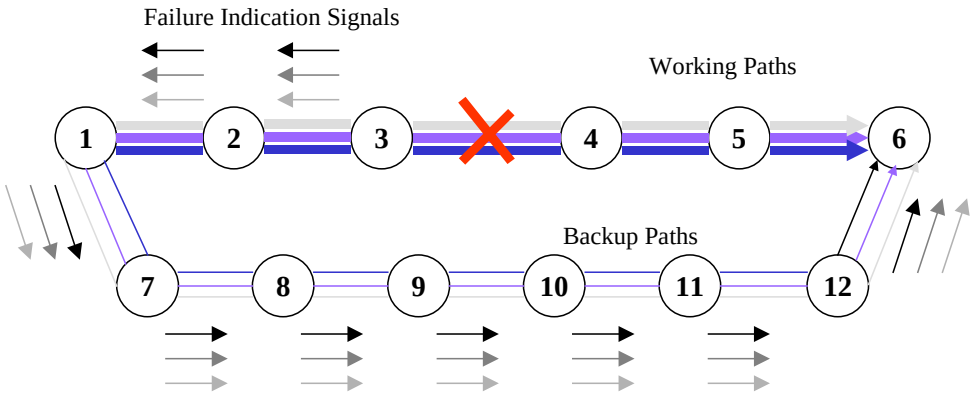


Figure 1.15: Queuing delay in signaling-based failure notification

in the broken link, the message indication messages could experience a high delay due to queuing delay. A similar case occurs when the alternative paths are signaled before the switchover. The worst case is shown in Figure 1.15, when all LSPs have the same ingress and egress nodes. In this case all nodes in the alternative path (and also in the upstream segment, between the detecting and the ingress nodes) have to queue the messages.

In such a case the number of messages in the network is proportional to the length of the notification path and double the backup path length.

And the maximum queuing delay is proportional to the number of protected LSPs in the failed link and the number of messages.

1.7.2 Flooding-based notification

An alternative approach to address the issue of messaging is to use flooding. Instead of sending per-LSP notification and initiating per-LSP recovery at each LSP source node, the node that detects a failure (e.g. fiber cut) notifies all nodes of the network. Nodes that are concerned with the recovery take the actions required of them while others forward the messages on with no extra action but knowledge about the resource failure in order to maintain an accurate picture of resource availability.

One such implementation of flooding is OSPF-based flooding. The usual link-state protocol floods advertisements periodically. In fact, OSPF requires that Link State Advertisements (LSAs) be refreshed every 1800 seconds [MOY98] and that they expire in 3600 seconds. Flooding frequency is crucial to the stability of the network, since increasing it may lead to excessive messaging and a larger number of retransmissions and ACKs. In the case of recovery from link failure in data networks, this may not be a problem and using OSPF-based flooding could be a good solution that decreases the amount of messaging related to signaling.

This process is shown in Figure 1.16. The detecting node sends a failure indication message to its neighbors (in this case only node 2). Intermediate nodes send this message to their neighbors (avoiding replicated failure indication messages). Obviously, the failure indication message arrives at all ingress nodes in the network following the shortest path (in terms of delay). In the figure,

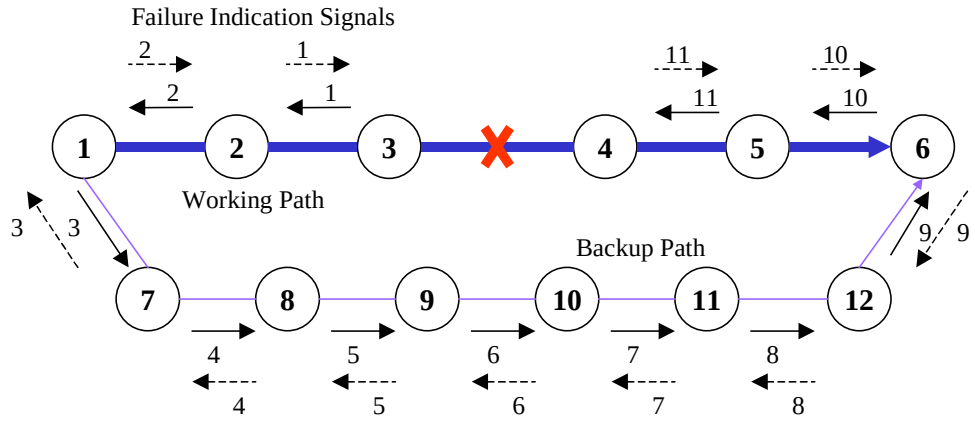


Figure 1.16: Flooding-based failure notification

however, flooding can only follow the same path as signaling. Intermediate nodes send, asynchronously, the failure indication acknowledgment messages. After the egress node (node 6) has received the message, no further action is required (though the notification message is forwarded to the remaining nodes). The ingress node can start sending the traffic to the alternative paths (these nodes could then know at what time to start the switchover).

Figure 1.17 shows a queuing time zero in flooding-based techniques. If multiple LSPs are affected in the failed link, only one message is sent, so no queuing delay occurs.

In this case the number of messages in the network is proportional to the length of the notification path and of the backup path. And the maximum queuing delay is zero.

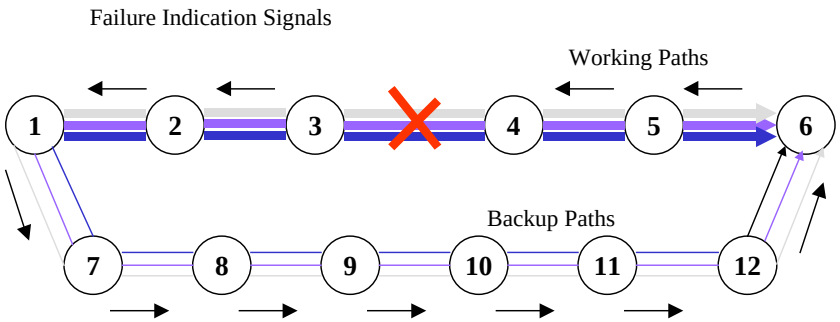


Figure 1.17: Queuing delay in flooding-based techniques

Advantages of flooding techniques over signaling-based techniques:

The advantages of flooding with respect to signaling-based techniques can be resumed in minimum notification delay, due to the following:

- Buffer delay: The queuing time is zero.
- Node processing delay: Minimum time processing the notification messages. In signaling an RSVP message, the node processing time is about a few hundred of microseconds [YL02].
- Minimum path delay: In flooding techniques notification messages follow the minimum reverse path (in terms of delay). Messages are disseminated in all directions to all nodes in the networks, guaranteeing minimum delay.

This complete dissemination (in the case of flooding) also improves the routing. All nodes are notified of the failure and this can improve the evaluation of future routing requests, thereby avoiding signaling failures.

1.8. Extending the MPLS fault recovery models to optical networks

A great deal of work have been done in MPLS in developing fault management mechanisms. Although extending MPLS fault management to optical networks appears feasible, there are some points that deserve particular attention. In this section some of this points are reviewed in more detail.

In MPLS networks, the control and data planes (in which the packets are processed) share the same transmission media. This means that a single fault affects both equally. However, in optical networks the control and data planes can have different topologies, hence control messages can be sent through an "out-of-band" path (for instance, a dedicated wavelength). In other words, two OXCs that are neighbors on the data plane are not necessarily neighbors on the control plane. In this scenario, faults should be considered independent on each plane [DHA03]. This is shown in Figure 1.18.

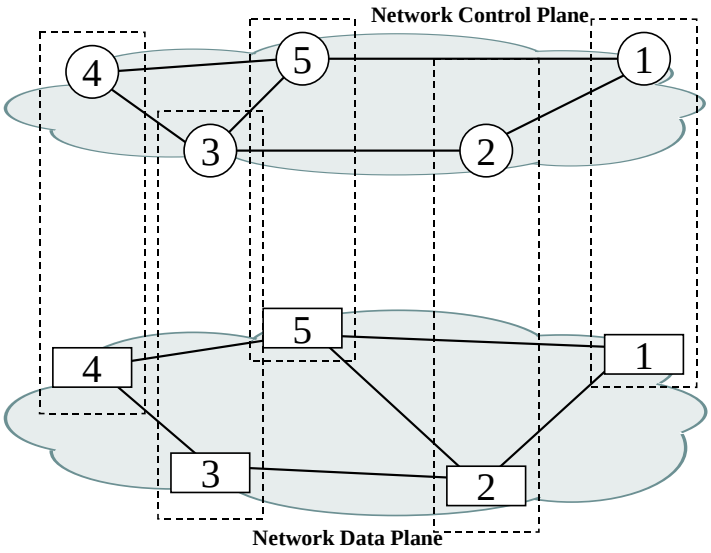


Figure 1.18: The GMPLS architecture.

Another difference between a lightpath, in transport optical networks, and an LSP, in MPLS packet networks, is that an LSP may have a reserved allocation of zero resources (such as bandwidth), but whenever a lightpath is routed, the corresponding wavelengths have to be reserved at the same time. This also affects the recovery methods because some of the "fast restoration" techniques developed for MPLS networks are based on pre-computed backup path with zero-resource (bandwidth) allocation.

In the case of shared backup [RAB03] restoration in MPLS networks, multiple labels are assigned, one for each of the backup LSPs transiting a node (corresponding to link and/or node disjoint working LSPs that they protect) and using the shared backup path. But in this case, only one set of resources (buffers, bandwidth) needs to be reserved.

When we consider optical networks, the situation is different. In such networks a backup LSP can be pre-signaled but not pre-reserved (unless simple 1+1 protection is desired). This is because, once an LSP in a transport optical network is established (that is, it is cross-connected), the full bandwidth of the LSP is automatically consumed, irrespective of whether traffic actually flows on that LSP. For this reason, when implementing shared backup schemes in optical networks (or allowing extra-traffic between endpoints other than the source-destination of a backup LSP), a backup LSP cannot be cross-connected until after the specific failure for which this LSP was pre-signaled has occurred.
Thus, for transport optical networks an additional step of reconfiguration is required at all the nodes that lie along the path of a backup LSP corresponding to a working LSP.

Figure 1.19 shows this difference. Figure 1.19 a) shows a restoration scenario in a packet-based MPLS network. There are two working LSPs, WP1 and WP2, with a single shared backup LSP BP1/BP2. The label assignments have been made as shown (L1 and L2 for WP1 at node 2, L1' and L2' for BP1 at node 5, and L3' and L4' for BP2 at node 5, and so on). When a fault affecting WP1 occurs (Figure 1.18.b), node 1 immediately performs a protection switch upon learning of the

failure and begins transmitting working traffic from working path WP1 down the backup LSP BP1 with the label L1'. Node 5 now simply label switches the traffic arriving on link 1-5 with label L1' by placing it on the outgoing link 5-3 with label L2'. Node 5 may drop the low-priority traffic (or any extra traffic in LSPs E1 and E2 respectively) being carried when the backup LSP was not active by simply purging it from its outgoing queues.

By contrast, for an optical transport network, where the LSPs in question are lambda LSPs, we assume a single lambda per link for ease of exposition. Here the intermediate node 5, upon learning of a fault along working path WP1, has to first drop any extra-traffic (or low priority) LSPs using the bandwidth (lambda) reserved for the backup LSP BP1/2. It then reconfigures its cross-connect matrix to connect the incoming lambda on link 1-5 to the outgoing lambda on link 5-3 (that is, it changes its configuration from 1-5 -> 5-4 and 4-5 -> 5-3 to 1-5 -> 5-3). So, in optical networks more steps are needed to activate and use a backup path in the case of shared backup fault recovery.

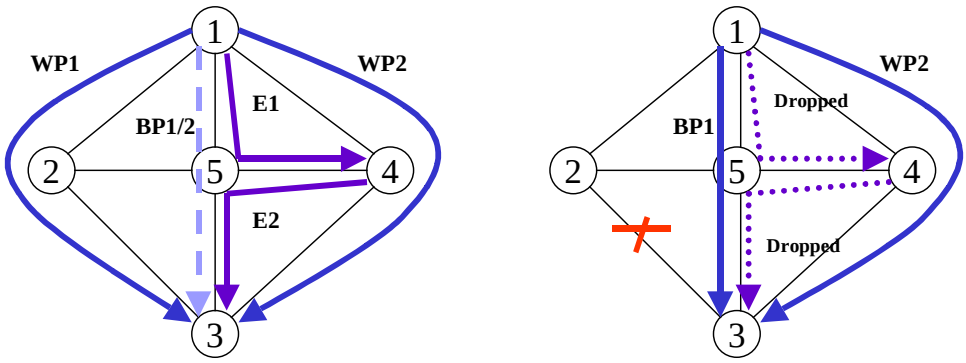


Figure 1.19: MPLS restoration over optical restoration

1.9. Summary and motivation

MPLS and GMPLS technology have introduced a new framework into traffic-engineering of current and future networks. Fault management is one of the main issues to be developed within this framework. Many efforts and proposals have been made to create new fault recovery schemes in MPLS-based networks. More fast and suitable methods have been deployed under this new technology. The current extension of these protocols to support other switching technologies, such as those in optical networks, has been made through a new Generalized MPLS control plane. The IETF-CCAMP working group is developing several Internet Drafts related to recovery in networks featuring a GMPLS control-plane. They cover the topics of terminology [MAN03], requirements [RAB03b], functional specification and mechanism analysis [PAP03].

Both MPLS and GMPLS fault recovery methods can be classified and compared from different points of view. The number of backup paths to protect a working or segment LSP or the way that a backup path can be created, and the resource assigned to this backup path, are some of these ways. In order to compare different fault recovery models, QoS parameters, such as the recovery time or packet loss can be used. The failure recovery cycle, or the steps/phases deployed by each fault recovery model, allows not only the classification of each model, but a way to intuitively compare some features, such as recovery time and packet loss.

The number of simultaneous failures that can be recovered by a fault management scheme and the type of failure is another way to categorize a model. In this paper we have mainly focused on single link failures. However, node failure and multiple failure have also been taken into account throughout this study.

Actually, this first analysis and classification of the different fault recovery schemes has highlighted the first conclusions and issues of this paper. The need to evaluate the level/degree of protection provided by each recovery scheme in

different network scenarios has been one of the main objectives of this work. The evaluation and formulation of this level requires, on the one hand, defining what are the QoS protection parameters, and on the other hand, a way to evaluate this grade in different network scenarios.

We have based our QoS protection level formulation mainly on the delay experienced in the network from the moment a failure occurs to the moment the connection is restored, and the packet loss during this process. However, the most suitable schemes (analyzing these QoS parameters) cannot always be applied to all network scenarios. Resource consumption and topological constraints avoid creating the most suitable protection in many network scenarios. On the other hand, traffic constraints involve different QoS requirements that should be satisfied by the network design but some of which can be incompatible with the QoS protection degree.

The next chapter introduces one of the main steps in designing a network with certain QoS requirements. The concept of QoS routing and different MPLS routing proposals will be presented in this chapter.

CHAPTER 2

2
QoS restorable routing methods

2.1. Introduction

In the previous chapter the main components of implementing and managing fault recovery methods were reviewed. One of these components, the first step in the recovery cycle, is the selection of the working and, optionally, the backup path. In this chapter a review of QoS routing algorithms is presented. Different mechanisms to select the most suitable path depending on the QoS requirements

are compared, and the options for protecting the active paths with segment or path protection are also presented.

Routing algorithms attempt to find a feasible path. These algorithms could be divided according to what type of routing information is used to compute path routes and when this computation is applied. First it is important to take into account that the classification of routing algorithms can be static or dynamic. Static algorithms only use static network information, while dynamic algorithms use link load information that is periodically updated. Secondly, routing algorithms can be on-line (or on-demand) routing or off-line (or pre-computed) routing. Depending on when paths are computed, with on-line routing algorithms path requests are attended sequentially (i.e., one by one), while off-line routing algorithms do not allow new path route computation (because they are already pre-computed).

2.2. QoS routing. Principles and Previous work

The main goal of a routing algorithm is to find a feasible path (a path with enough QoS, bandwidth) that achieves efficient resource utilization. To optimize network performance, QoS routing algorithms use two techniques. The first is to pick the minimum hop count path in order to reduce resource consumption, and the second is to balance the load of the network (i.e., the least loaded path is selected). This optimization of network utilization, reducing resource consumption and balancing the network load, is not easily achieved using only one routing algorithm, since these two objectives are usually incompatible. A path with the least number of hops does not necessarily have to be the path with the best resource consumption. Consequently, developing a suitable QoS algorithm involves taking into account more than one aspect. A suitable way to develop a QoS routing algorithm, keeping the objectives of load balance and resource consumption in mind, is to apply new routing criterias or to mix several QoS criteria. These QoS criteria could be: minimum hop count, maximum residual bandwidth, minimum path cost based on the link utilization, etc.

Recently in the literature, several proposals for QoS routing, taking into consideration these criteria or mixing many of them, have been developed and experimented with ([GUE97], [MA97]).

A common routing method is to use a min-hop algorithm (MHA). This algorithm only chooses the feasible path with the lowest number of hops (links) as a single routing criteria. In [GUE97] a Widest-Shortest Path (WSP) algorithm based on the Bellman-Ford algorithm is proposed. Two criteria are mixed: the first one is to pick the path with the minimum hop count amongst all feasible paths, and the second, if more than one path is chosen, is to select the one with the maximum reservable bandwidth (MRB). The MRB of a path is the minimum amount of the reservable bandwidth of all links on the path. Another routing proposal is exactly the opposite of WSP. In this case, the first priority is selecting the path with the minimum bandwidth and, if more than one is feasible, the path with the minimum hop count is then selected. This algorithm is called the Shortest-Widest Path (SWP). WSP gives the highest priority to resource utilization while SWP gives it to balancing the network load. Other proposals define a cost function and apply a shortest-path computation based on such cost.

Nevertheless, the above algorithms present several drawbacks when selecting a path with a longer number of hops (in the case of WSP) or a path with a critical bandwidth allocation, both of which could become congested points. To avoid this, other proposals impose some constraints which act to ease these drawbacks. In Dynamic-Alternative Path (DAP) [MA97], a hop count restriction is used to avoid selecting paths greater than a threshold (n) number of hops computed by MHA. This is basically a WSP algorithm with a hop limitation.

Several proposals making use of MPLS network capabilities to develop new path selection algorithms with QoS guarantees have been proposed in the recent literature ([KOD00], [KAR00], or [SUR01]). Unlike the above QoS routing algorithms, in these proposals the use of ingress-egress nodes knowledge is the common denominator.

2.3. MPLS QoS on-line routing algorithms

MPLS has some capabilities which facilitate the implementation of QoS parameters to route new paths (LSPs). In this section a review of several MPLS QoS on-line routing proposals is presented. Their advantages and disadvantages are also highlighted.

2.3.1. Dynamic Routing of bandwidth guarantees tunnels with restoration

This is one of the first proposals [KOD00] to consider the MPLS aspects in designing a routing algorithm. They deployed an on-line routing algorithm of bandwidth-guaranteed LSPs to route backup and working paths as requests arrive. In this algorithm, if sufficient bandwidth to set up both the active and the backup paths is not available, then the request is rejected. Only protection against single link/node failures is considered. Multiple backup establishment and thus the possibility of sharing backups, one of the main points of the present paper, are not considered.

Different routing algorithms, based on the information available to path computing, are proposed. These methods compute, basically, an integer linear programming problem. An algorithm with only aggregated link bandwidth usage information (called Dynamic-Routing with Partial-Information DR-PI) is proposed as a good solution in terms of computing cost and network performance.

The main goal of this proposal is to develop an on-line routing algorithm to minimize bandwidth usage. Unlike other proposals, this method does not consider minimizing the request rejection ratio as a primary goal. Nevertheless, a study of the blocking rate between the proposed algorithms (with partial-complete-no network information) prove, similar to [MA97] experiments, that if the routing algorithm has better knowledge of the actual network parameters, less rejected requests are computed. The main conclusion of this proposal is

that, in terms of bandwidth allocation, an algorithm with only aggregated link bandwidth usage information performs as well as algorithms with more complete information. The main drawback of this proposal is that the request rejection count or the request for multiple backups (or simply an LSP request) are not taken into account. This drawback is overcome in the next proposal.

The same authors recently proposed Dynamic Restorable Routing [KOD02], which enhances some aspects of [KOD00] by adding local/segment protection.

2.3.2. Dynamic Restorable Routing Algorithm

In this proposal [KOD02] setting up bypass (backup) paths for every link or node traversed by the primary active path is presented. The use of local restorability (local backups) with shared resource consumption is proposed. A comparison between different network information scenarios demonstrates that a partial scenario, which uses aggregated and not per-path information, provides sufficient information for efficient selection of local bandwidth guaranteed backups paths. This algorithm only knows what fraction of each link's bandwidth is currently used by working/active paths and what portion is currently used by backup paths.

2.3.3. Minimum Interface Routing Algorithm

The "Minimum Interface Routing Algorithm" (MIRA) [KAR00] is another proposal that takes into consideration particular aspects of the MPLS architecture to design an on-line routing scheme. In this case, ingress and egress nodes are taken into account. Kodialam and Lakshman introduce the concept of interference and develop a multiple max-flow computation to determine the path of least interference.

Interference: The main idea is to establish paths that do not interfere "too much"

with future LSP setup requests, considering pre-established values of ingress-egress pairs. Figure 2.1 shows an example of this "interference" effect. Consider the maximum flow (maxflow) value 1 between a given ingress-egress pair (S1, D1). Note that the maxflow value 1 decreases whenever a bandwidth demand is routed between S1 and D1. The value of 1 can also decrease when an LSP is routed between some other ingress-egress pair. The amount of interference on a particular ingress-egress pair, for example (S1, D1), is defined and an LSP is routed between some other ingress-egress pair as the value of 1 decreases.

Existing LSP1 (S1,D1) and LSP2 (S2,D2) and LSP3 are required between S3 and D3. If the MHA (Minimum Hop Algorithm) is used, the route between (S3,D3) will be 1-7-8-5. This route produces a blocking path between S2 and D2 as well as between S1 and D1. In this example it is better to choose route 1-2-3-4-5 even though the path is longer.

Minimum Interference Paths: The minimum interference path for an LSP between, for example (S1, D1), is the explicit route which maximizes the minimum

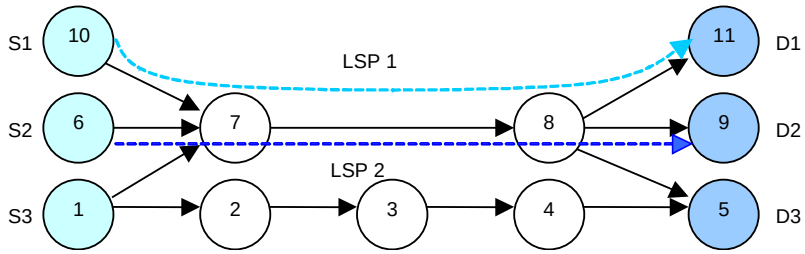


Figure 2.1: Minimum Interference Paths

maxflow between all other ingress-egress pairs. In other words, this can be thought of as choosing a path between (S1, D1) which maximizes the minimum residual capacity between every other ingress-egress pair.

The objective might be to choose a path that maximizes a weighted sum of the maxflows between every other ingress-egress pair. This algorithm not only makes capacity available for the possible arrival of future demands, but also makes capacity available for rerouting LSPs in case of link failures.

Critical Links: Critical links are links characterized by a decrease in the maxflow value of one or more ingress-egress pairs whenever an LSP is routed over them. This is the criteria to create a weighted graph.

The Path Selection by Shortest Path Computation is developed using the well known Dijkstra or Bellman-Ford algorithms for computing the present explicit route. They do this by generating a weighted graph where the critical links have weights that are an increasing function of their criticality.

The increasing weight function is picked to defer loading of critical links whenever possible. The actual explicit route is calculated using a shortest path computation as in other routing schemes.

The algorithm has an input graph G (N,L) and a set B of all residual link capacities. A flow of D units has to be routed between an ingress node a and an egress node b, generating an output route between a and b having a capacity of D units of bandwidth.

An experimental analysis of MIRA [SUR01] points out that in a set of network scenarios MIRA does not work as expected. Two main drawbacks are highlighted in the following.

MIRA focuses exclusively on the interference effect on single ingress-egress pairs. For example, Figure 2.2 illustrates this effect. In [SUR01] this network is called “The concentrator topology”.

One node C acts as a concentrator for n ingress nodes S1..Sn. Node C is connected to a high capacity link of capacity n+1, whose endpoint is an egress node D. A high bandwidth ingress node S0 is also connected to the concentrator, through an n capacity link. S0 is also connected to D via an alternative 3-hop path, of capacity n. In this example the MIRA checks the LSP requests one by

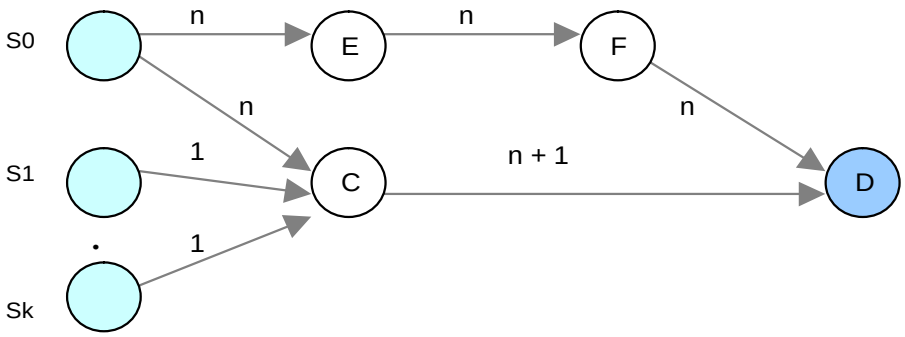


Figure 2.2 : The concentrator topology

one. The first request (S0,D) has two possible paths, a 2-hop path (S0,C,D) and a 3-hop path (S0,E,F,D). The first one is not so critical because it is not considered to be a minimum cut for any individual ingress-egress pair. This permits a residual bandwidth 1, enough for any individual request. Therefore, MIRA chooses the path (S0, C, D) which is an incorrect path in this scenario. An optimal algorithm would route the (S0,D) request along the top on the alternative path (S0,E,F,D), and it would use the (C,D) link to route the n 1-unit request from Si to D. More examples of this drawback are shown in [SWW01]. Other examples of this effect are shown in [SUR01].

Another drawback is that MIRA is computationally very expensive. MIRA performs hundreds of maximum flow computations, each of which is several orders of magnitude more expensive than shortest path computations

2.3.4. Profile-Based Routing

Suri, Waldvogel and Warkhede introduce, in [SUR01], the idea of using a "traffic profile" of the network, obtained by measurements or service level agreements (SLAs), as a predictor of the future traffic distribution. The objective is that the algorithm could anticipate a flow's blocking effect on groups of ingress-egress pairs. (MIRA only considers one ingress-egress pair at a time.)

The ability of MPLS networks to specify explicit paths for any flow provides an important tool in engineering how traffic is routed, and thereby improves network utilization by minimizing the number of requests that are rejected when the network becomes overloaded. A traffic profile can be as simple as an average bandwidth requirement over a certain time period.

The Profile-Based Routing (PBR) uses quasi-static information in a preprocessing step (one multi-commodity flow computation) to determine certain bandwidth allocations on the links of the network. The on-line phase of the routing algorithm then routes LSP requests using a "shortest path" (SPF)-like algorithm but with additional information provided during the preprocessing phase. The multi-commodity-preprocessing phase allows the on-line algorithm to exercise admission control by rejecting some requests because of their blocking effects in the network.
The multi-commodity flow formulation permits a cost function, which is minimized to achieve optimal routing. In order to minimize the number of rejected requests, the simple "linear cost function" is used. A variety of non-linear cost functions can be used to handle features such as minimum guaranteed bandwidth or fairness across multiple flows.

One drawback of this proposal is the no explicit recovery treatment. As in the case of MIRA, only ingress-egress nodes are considered. In MIRA only the case of a centralized backup establishment (one backup along a path formed by a source ingress-node and a destiny egress-node) is considered. No local or reverse backups are considered. In PBR all types of backup establishment are considered.

Comparison between PBR and MIRA

PBR is computationally less expensive than MIRA and performs better than MIRA in certain network scenarios. Consider a topology where a large capacity link is shared by many ingress-egress pairs when small bandwidth requests arrive. However, the request between one node pair is very large and this node pair is also connected through a longer set of links with large capacity. Under MIRA tThe shared high capacity link is not in the minimum cut for any individual ingress-egress pair in the network. Thus, if the large bandwidth request arrives first, the high capacity link is utilized, leading to rejection of numerous subsequent requests. PBR, on the other hand, limits the maximum amount of the request of a single commodity that can be mapped to a link. Thus, PBR would choose the longer path to route the request, leaving the high capacity link to the smaller requests between the other node pairs. However, it has been shown that MIRA performs better than PBR when the traffic profile information does not give a correct depiction of the bandwidth requests.

The multi-commodity flow problem in the preprocessing phase of PBR is an optimization problem and the traffic of a source destination pair can be split among multiple paths. Sometimes the splitting of flows may not be allowed, leading to request rejection. Consider a scenario where the class profiles detail that an aggregate of x bandwidth units of traffic is allowed between a node pair. If the path that has x units available is considerably larger than paths with an available bandwidth of less than x units, the preprocessing phase divides the x units among more than one shorter paths. However, if the request between the node pair demands x bandwidth units and the flow cannot be split, the PBR algorithm rejects the request. On the other hand, MIRA prunes the network to remove the links with less than x units of available bandwidth and thus the bandwidth request is routed on the longer path with enough resources to hold the request without splitting. Thus, PBR and MIRA each have their advantages depending on network scenarios.

2.4 Routing Information

The basic information needed by any routing protocol to make appropriate path selection decisions is the state of the network. Every routing protocol uses this information to forward packets. The information about the state of the network includes the network topology along with resource availability for QoS purposes. Each change in the state of the network should be detected and disseminated to all the routers in the same Autonomous System (AS) and also propagated across AS boundaries until all ASs have been informed of this change. Since the topology variations are less frequent, the main cause for state change is the resource availability variation in the network.

The large amount of information exchange required for the state update can compromise the scalability of the routing schemes. To reduce this amount, two approaches are possible: reducing either the frequency of updates or the details in the updates. The former is achieved by using various mechanisms such as class-based, threshold-based and periodic updates. The latter is achieved by aggregating the network state information. In the case of MPLS networks, a centralized network manager can also be used for the network operation, making the problem of information dissemination redundant.

2.5 Differences in establishing the working and backup paths

The backup path should be selected so that it is ready to transport traffic between the LSP source and destination whenever any one link/node along the primary route fails. An important consideration is to select the primary and backup routes in such a manner that the maximum number of LSP requests can be accommodated in the future. Thus, the route selection process should, as far as possible, avoid those physical links that are of critical importance to a large number of source-destination pairs.

The primary and backup paths should not undertake the same risks of failure, otherwise the same fault may cause both paths to fail. If a resource is already taken by a protection path, that resource should be shared as much as possible by other protection paths, up to the maximum number allowed on that resource. Multiple protection paths sharing common resources should not be activated simultaneously. To achieve this, the routing algorithm must disallow protection paths from sharing resources if their primary paths have common failure factors. Shared protection offers higher network utilization than dedicated protection. However, only the paths with the most strict protection requirements need to be dedicatedly protected. The other paths can be protected under shared protection that would free up network resources.

2.5 QoS on-line routing algorithms comparison

Table 2.1 and Table 2.2 show a comparison of the reviewed on-line routing QoS and MPLS on-line QoS routing approaches. Their main features and drawbacks are included in these tables.

Algorithm	Main objective	Routing Information	Route computation	Drawbacks
WSP Widest-Shortest Path [GUE97]	Efficient resource utilization.	Maximal reservable bandwidth (MRB).	MHA over feasible paths first and the path with the maximum-reservable bandwidth.	May select a path with a larger number of hops (only in the case of the WSP).No limit is established. May select a path that could become a congestion point (no request rejection aspect is considered). No recovery treatments are considered.
SWP Shortest-Widest Path [GUE07]	Balance the network load.		The path with the MRB first and the MHA path over the MRB results.	
DAP (Dynamic Alternative Path) [MA97]	Improve WSP limiting the path hop/link number.		A WSP with a hop count restriction.	

Table 2.1: QoS on-line routing algorithms qualitative comparison.

Algorithm	Main objective	Routing Information	Route computation	Drawbacks
DR-PI Dynamic Routing with Partial-Information [KOD00]	Optimize the bandwidth usage [KOD00] / Local Protection [KOD02].	Ingress-Egress Nodes and the aggregated link bandwidth usage.	An integer linear programming problem.	The number of rejected requests is not taken into consideration in [KOD00]. Considerable computational complexity for on-line implementation. No local/segment backups are considered in [KOD00].
Dynamic Restorable Routing [KOD02]				
MIRA Minimum Interference Routing Algorithm [KAR00]	Optimize the bandwidth usage and minimize the number of rejected requests.	Ingress-Egress nodes and link bandwidth usage.	The concept of the interference to generate a weighted graph with the critical links (as a cost) and a SPF algorithm to pick the path.	Cannot detect critical links in topologies with clusters of nodes Computationally expensive. No pre-established backups are considered.
PBR Profile-Based Routing [SUR01]	Optimize the bandwidth usage and minimize the number of rejected requests.	Ingress-Egress nodes. Current residual capacity. Traffic class (service type).	A pre-processing step (multi-commodity flow computation) to determine certain BW allocation and an on-line phase using a SPF algorithm.	No explicit recovery treatments are considered.

Table 2.2: MPLS QoS on-line routing algorithms qualitative comparison.

2.6 Conclusions and motivation

In this chapter some of the main QoS routing algorithms have been reviewed. The objective of a routing algorithm is to select a route between two nodes of a network. However, this route should satisfy some QoS parameters, such as bandwidth or delay. The first QoS routing proposals, as we have explained, selected those routings with minimum hops or maximum reservable bandwidth. Usually, these two objectives are incompatible. This means that it is not easy to develop a suitable routing algorithm.

The introduction of MPLS facilitates the development of new routing algorithms with some of the advantages of MPLS, such as TE facilities: explicit routing and aggregation.

On the other hand, major proposals reviewed do not take into account the creation of protection routes. Neither backup methods nor selecting working paths with maximum availability have been extensively considered in major routing proposals. Some of them choose restoration methods to protect their paths, while most select only one backup method to protect the network. Path protection is the primary method selected by these proposals. However, the protection or restoration methods are usually a secondary objective of the routing methods. This means that no level of protection is evaluated to choose the most appropriate fault recovery strategy.

Our objective during this research, as explained in the first chapter, is not only to evaluate what the suitable protection scheme is, but to evaluate what the level or degree of protection provided by these schemes is. Furthermore, selecting the suitable working path, the path with certain availability characteristics, is a principal objective of this work.

In the next chapter the main reliability and availability formulation is presented. The most well-known models are introduced in this chapter.

In order to improve the network protection level a novel model to evaluate the network reliabilty (component and path failure probability) is also introduced in the following chapter.

CHAPTER 3

3
Network reliability and availability

3.1. Introduction

In the previous chapters a review of the main fault recovery methods and the state of the art of some QoS routing algorithms were discussed. As has been pointed out, major methods (MPLS QoS routing algorithms) do not include, in their objectives or QoS parameters, offering QoS protection. In many cases this protection is offered as a single backup model (for instance, global backup

paths), without considering the real protection requirements of the network or the protection traffic service requirements. Other methods only include restoration models, offering a certain degree of protection.

However, many proposals avoid including protection or some kind of protection (1+1 or local protection), due to the high level of resources needed to deploy these protection models over the whole network. In this paper we have taken into account new objectives in routing methods to offer the desired degree of protection with the suitable resource consumption.

A first study of the network can contribute to discovering what is the sensitivity of the network or the probability that it will fail. This sensitivity is considered to offer an easier or more complex protection of the network. If the network is more likely to fail in some segments, specific protection for these segments can be added. If some parts of the network are transporting high priority protected traffic (traffic very sensitive to packet loss or delays caused by a failure), the routing method can transport this traffic using other zones of the network with lower failure probability.

In this chapter some terminology related to reliability, availability, and failure probability is presented. The recovery cycle is reviewed highlighting the main components affecting the time needed to recover from a failure. An easy proposal to evaluate the failure probability in the GMPLS or MPLS networks is also presented.

3.2. Measures of network reliability

When performing a reliability prediction analysis, there are several metrics that provide measures of reliability. These metrics include failure rate, mean time between failures (MTBF), reliability and availability [OGG01], [WIL98] and [WZZ01].

3.2.1. Failure Rate (FR)

Failure Rate is the number of failures experienced or expected for a device divided by the total equipment operating time. The Failure rate can be characterized by a bathtub curve (Figure 3.1). The initial region that begins at time zero when a customer first begins to use the network link is characterized

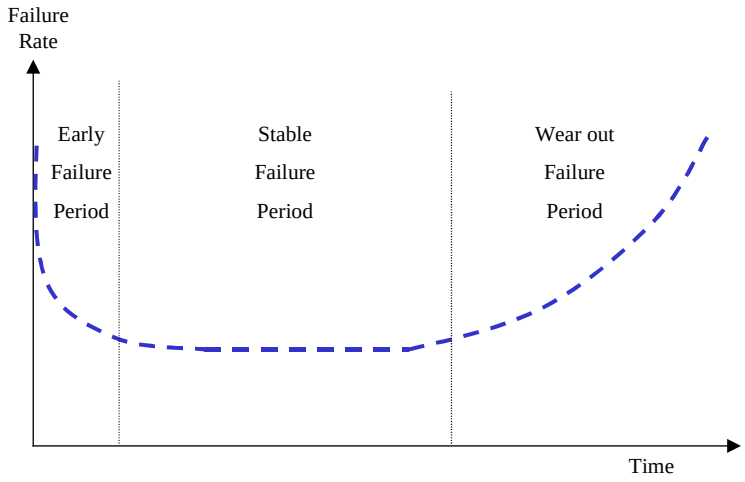


Figure 3.1: Failure rate. The bathtub curve

by a high but rapidly decreasing failure rate. This region is known as the Early Failure Period. This decreasing failure rate typically lasts from several weeks to a few months.

Next, the failure rate levels off and remains roughly constant for (hopefully) the majority of the useful life of the link. This long period of a level failure rate is known as the Stable Failure Period. Note that most systems spend most of their lifetimes operating in this flat portion of the bathtub curve.

Finally, if units from the population remain in use long enough, the failure rate

begins to increase as materials wear out and degradation failures occur at an ever-increasing rate. This is the Wear out Failure Period.

3.2.2. Mean Time to Repair (MTTR)

The MTTR is the total amount of time spent performing all corrective maintenance repairs divided by the total number of those repairs (ITU-T E800/4260) [ITU800].

3.2.3. Mean Time Between Failures (MTBF)

The MTBF is the mean time expected between failures, typically measured in hours. MTBF (ITU-T E800/4238) [ITU800] is a statistical value and is meant to be the mean over a long period of time and a large number of units. For constant failure rate systems, MTBF is the inverse of the failure rate (FR):

$$MTBF = 1/FR \tag{3.1}$$

If the failure rate is measured in failures/million hours, MTBF = 1,000,000 / Failure Rate for components with exponential distributions. Technically, MTBF should be used only in reference to repairable items, while MTTF (Mean Time to Failure) should be used for non-repairable items, but MTBF is commonly used for both repairable and non-repairable items.

3.2.4. Mean Time to Failure (MTTF)

The MTTF is the mean time expected before the first failure of a piece of equipment. It is a statistical value and is meant to be the mean over a long period of time and a large number of units.

3.2.5. Reliability (R)

Reliability is the probability that a device will perform without failure over a specific period of time. This is determined by finding e to the power of the negative value of the period of time divided by the MTBF.

$$MTBF = \int_0^\infty R(T)\,dt \tag{3.2}$$

In other words,

$$R(T) = e^{(-T/MTBF)}. \text{ (T: number of hours.)} \tag{3.3}$$

For example, given the MTBF for a network link is 1,000,000 hours, what is the probability the link will operate without failure for five years? To answer this question, divide five years, 43,800 hours, by the MTBF (43,800/1,000,000 = 0.0438). Then find the value of e raised to the power of the negative value of that number ($e^{-0.0438} = 0.9571$).

There is a 95.71% probability that the link will not fail in a five-year period.

Note that $R(T) = e^{-(FR*T)}$.

3.2.6. MTBF and R for multiple components

Once the Failure Rate (inverse of MTBF) is determined, MTBF for multiple items (or components) is easily calculated as the inverse of the sum of each component's failure rate.

$$MTBF = 1/(FR1 + FR2 + FR3 + \ldots FRn) \tag{3.4}$$

n : Number of components in the system.

FR is the failure rate of each component of the system up to n, all components.

Therefore:

$$R(T) = \prod_{i=1}^{N} Ri(T) \tag{3.5}$$

3.2.7. Availability (A)

Availability is the probability that a system will be operational when called upon to perform its function. It is quantified as a percentage.

The equation for A is:

$$A = MTBF/(MTBF + MTTR) \tag{3.6}$$

Note that (see Fig. 12) : A=MTTF/MTBF = MTTF/(MTTF + MTTR)

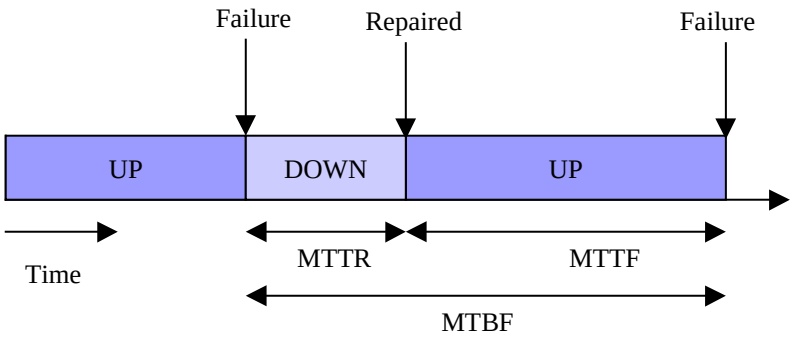


Figure 3.2: The renewal process

3.2.8. Unavailability (U)

Unavailability is the probabilistic complement of availability (i.e., U = 1 – A) and

is defined as the probability (fraction of time) the system/network will be in the failed state. When reporting system/network performance, unavailability is usually converted to minutes per year or, if the mean time to repair (MTTR) from a nonsurvivable failure of the system/network is known, to the mean time between failures (MTBF), usually in years, where

$$U = MTTR/MTBF \tag{3.7}$$

3.3. An approach for computing failure probabilities

In this section our proposal to compute an LSP failure probability is presented. First, the model to compute link failure probabilities based on component failure probability models, geographical conditions and failure statistics is explained. Then, a formulation to evaluate an LSP failure probability based on the link failure probability knowledge is detailed.

3.3.1. Link Failure Probability Evaluation

It is normally difficult to calculate the exact failure probability of a given segment of the network. However, an approximate value can be obtained based on certain information available before faults. The calculation can be approximated based on known probabilities regarding certain aspects of transmission technology, for instance, the type of physical link, the node characteristics, the geographical distribution of the network segments, etc. This initial value can be updated to a more realistic value using actual failure statistics.

However, in some cases the Internet Service Providers (ISP) or the network system manager can change or modify these values based on their own experience.

Figure 3.3 shows our proposal to compute the link failure probability. We propose characterizing the initial link failure probability by using one of the

failure probabilities. There are two models: MIL-HDBK-217 and Bellcore/Telcordia Issue 1. Both are accepted standards developed over several years.

The MIL-HDBK-217 [MIL217], also known as the Military Handbook for "Reliability Prediction of Electronic Equipment", is published by the Department of Defense based on work done by the Reliability Analysis Center and Rome Laboratory at Griffiss AFB, NY.

The non-military alternative to MIL-HDBK-217 is Bellcore/Telcordia Issue 1 [TEL99]. This is a reliability prediction model which was originally developed by AT&T Bell Labs by modifying the equations from MIL-HDBK-217 to better represent what their equipment was experiencing in the field.

The value obtained from these failure probability models should be weighted by geographical conditions. It is well known that some links have higher failure probabilities, and that this can be attributed to their physical-geographical situation. For instance, transoceanic links need better-protected installations due to their importance for the network and their geographical risk of failure. We

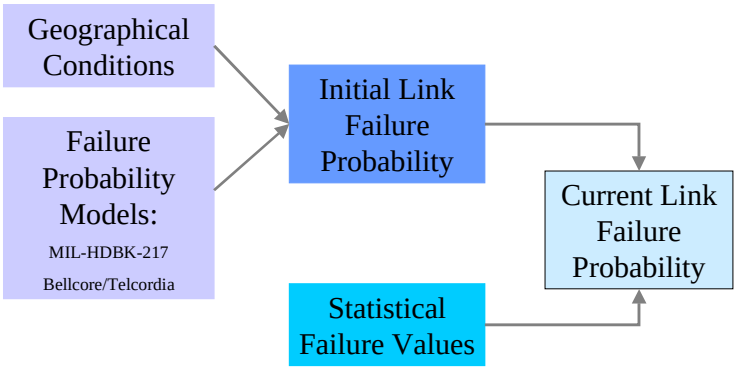


Figure 3.3: Link Failure Probability evaluation model

propose assigning a weighted failure probability value depending on physical-geographical situation and on failure probability models (see Figure 3.3).

Finally, statistical failure values (the latest failure rates) can be used to compute the current link failure probability. Each ISP can record the latest link failure to evaluate future failure behaviors. Different statistical tools can be used to compute these values. However, these tools, like the other phases of this probability computation, are beyond the scope of this work.

3.3.2. Label Switch Path Failure probability formulation

In this section we present the LSP failure probabilty formulation presented in [CAL04]. LSPs can cross through different links (see Figure 3.4) each with their own Link Failure Probability (LFP). In this work it is assumed that all LFP's are known (as explained above) and they are also independent of each other. These values are normally very small (LFP<< 1).

Label Switch Path Failure Probability (LSP_FP) represents the overall fault probability of an LSP and it means that an LSP fails if any segment (i.e., a single link or a combination of links) along the path fails. However, it is easier to

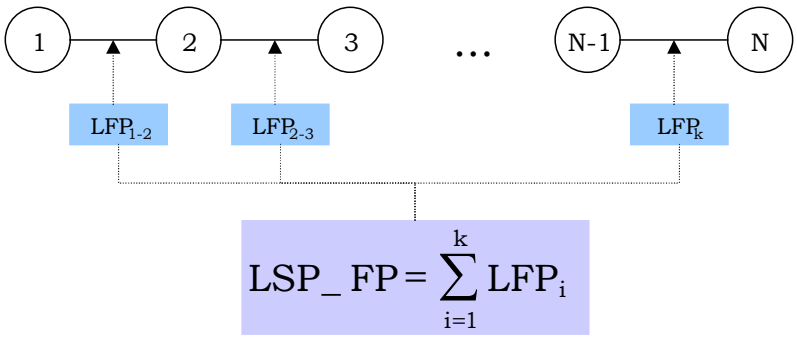


Figure 3.4: Label Switch Path failure probability

evaluate the inverse probability 1-LSP_FP (i.e., the probability that all the links of the path will work fine) which can be calculated as:

$$1 - LSP\_FP = \prod_{i=1}^{k} LFP_i^{-1} = \prod_{i=1}^{k} (1 - LFP_i)$$

k : Number of links of the LSP (3.8)

By considering the hypothesis of LFP<< 1, the product of this term is transformed into the following:

$$1 - LSP\_FP = \prod_{i=1}^{k} (1 - LFP_i) \approx 1 - \sum_{i=1}^{k} LFP_i$$

k : Number of links of the LSP (3.9)

The LSP Failure Probability can be calculated as the inverse of LSP_FP^{-1}, therefore:

$$LSP\_FP = 1 - LSP\_FP^{-1} \approx 1 - (1 - \sum_{i=1}^{k} LFP_i)$$

k : Number of links of the LSP (3.10)

Finally:

$$LSP\_FP \cong \sum_{i=1}^{k} LFP_i$$

k : Number of links of the LSP (3.11)

As expected, the probability of failure of an LSP is approximated by the addition of the failure probabilities of its links.

3.4. Conclusions and motivation

In this chapter the main network availability and reliability formulation has been presented. This formulation allows an approximated evaluation the level of protection provided by the network. Our proposal to evaluate the reliability of a segment and a path of a GMPLS/MPLS network has also been presented. Despite not being completely exact, this model allows us to make more intelligent decisions and develop new routing proposals with a higher level of protection than the current major proposals.

In the next chapter both the evaluation of network reliability and the evaluation of the backup models are included, creating a new framework to deploy new fault management strategies.

CHAPTER 4

4

Reducing the failure probability and failure impact

4.1. Introduction

To calculate the level of protection required for a given segment of a network, we consider an a priori factor: the probability of failure somewhere in the network; and an a posteriori factor: the impact on traffic (in terms of QoS degradation, i.e. recovery delay and packet losses) in the event of a failure.

In this chapter the relationships between the network failure probability, the failure impact and the different protection methods are introduced. A case study to evaluate what the tradeoffs are between the main protection proposals

(Chapter 1) for the reduction of failure probabilities and the failure impact is also presented.

One of the main issues in network design is the optimization of network resources. This is the main reason to evaluate the amount of network resources when applying each fault recovery method proposed in chapter 1. In the first section, the evaluation of backup resource consumption (in terms of bandwidth) is introduced. The following section presents the concept of failure impact.

The reduction of failure impact by optimizing recovery time (and consequently packet loss) is analyzed in section 4.4. Different case studies are presented in order to propose which are the optimal protection methods in terms of failure impact and failure probability reduction with the suitable resource consumption.

In this chapter we propose the application of some techniques to improve current QoS routing algorithms and the characterization of the traffic services based on the above reduction techniques.

4.2. Resource consumption in backup paths

Resource Consumption (RC) in protection methods is evaluated depending on the repair method used. For simplicity, we propose the utilization of allocated bandwidth per link. The resource consumption is computed on a per link basis by computing the number of links on a path and the allocated bandwidth on each link. Resource allocation is assumed to be bandwidth in the rest of this thesis.

$$RC = NL \cdot RB \tag{4.1}$$

Where:

RB	Reserved Bandwidth
NL	Number of Links

The above general formulation has to be adapted to the different backup path methods described in chapter I. The resource consumption for the global method (RC_G) depends on the number of links in the backup path (NL_G.) The resource consumption for the reverse repair method (RC_R) is the sum of the RC_G plus the resources required for the reverse path ($NL_R \cdot RB$). The resource consumption for the local repair method (RC_L) depends on the reserved bandwidth and the number of links (NL_L). In the case of local backup, it should be noted that more than one local backup can be created to protect several links in the working path. In short, the RC for the different methods is evaluated by:

$$RC_G = NL_G \cdot RB \quad (4.2)$$

$$RC_R = RC_G + NL_R \cdot RB \quad (4.3)$$

$$RC_L = NL_L \cdot RB \quad (4.4)$$

Where:

RC_G , RC_R , RC_L	Resource consumption (Global, Reverse and Local respectively)
NL_G , NL_R , NL_L	Number of links (Global, Reverse and Local respectively)

A particular case using the reverse backup method is proposed by Haskin [HUA02]. In this case the resource consumption is $RC_G + NLw \cdot RB$ (where NL_W is the number of links in the working path). Selecting protection methods with bandwidth allocation implies a combination of different methods (local, global or reverse) in order to achieve the requested protection level with a balanced resource consumption cost.

Figure 4.1 shows an experiment published in [CAL04]. More details of the implementation of this experiment can be found in [CAL04] or in section 5.4. The percentage of resources used by the three pre-established, pre-allocated protection methods are shown in Figure 4.1. As expected, the results show that

reverse backups consume more resources and local backups obtain the smallest percentage. In this case, only 20% of the network links are protected. However, if the network protection percentage increases, local backups may consume more resources than global or reverse backups (about 0.15 for reverse backups and 0.1 for global backups and local backups). In [CAL04] the relationship between the local backup resource consumption and the number of links to be protected is pointed out. The results also show that reverse backups always use more resources than global backups. However, in each trial there is a different proportion between global and local backups. This is due to the fact that the establishment of the reverse backups begins on the last node to be protected, minimizing the resource consumption when this node is near the ingress node.

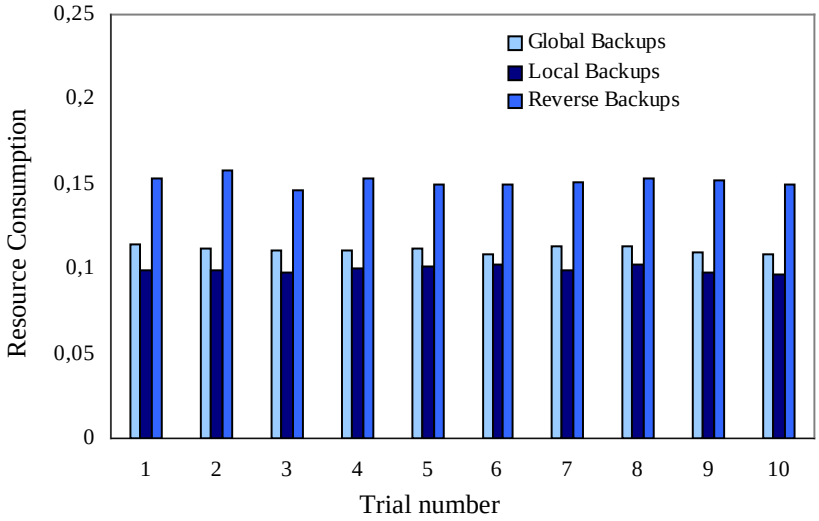


Figure 4.1: Backup resource consumption

4.3. The Failure Impact

The guaranteed quality of service (QoS) of the traffic is a crucial aspect of evaluating the failure impact. We suggest dividing it into two components: recovery time and packet loss. Other QoS components, such as increasing delay

or packet reordering, are not considered in this work. However, in other works, such as [Lemma], these components are compared and analyzed.

Each fault protection method offers a different recovery time. In [CAL04] and [CAL04b] we propose the following classification (Table 4.1):

Level of protection	**Recovery Time (T_{REC})**
Very low	> 1 min
Low	200 ms – 1 min
Medium	50 ms – 200 ms
High	20 ms – 50 ms
Very High	< 20 ms

Table 4.1: Recovery Time and Level of Protection

Some MPLS (and GMPLS) fault recovery mechanisms have large and very large recovery times.

The reduction of the fault recovery time is one of the main aspects to take into account in order to reach the level of protection required by many current traffic services.

4.3.1. Failure Recovery Time in GMPLS/MPLS networks

In MPLS-based networks the usual method of recovering a failure is the utilization of an alternative and disjoint path to the working path. The establishment and the use of this path can be deployed in different ways. Figure 4.2 shows the recovery phases in the MPLS schemes. In Table 4.2 a brief explanation of each phase is introduced.

These methods can use pre-established (pre-routed and pre-signaled) backup paths or establishing these backup paths dynamically (i.e. after the failure is detected). Resources can also be allocated a priori or, in the case of dynamic schemes, after the backup routing phase [MAR03]. The complete recovery cycle (i.e. non pre-established backups) starts when a failure is detected and finishes when the traffic is recovered (after the failure is fully repaired) back to the initial working path (normalization process).

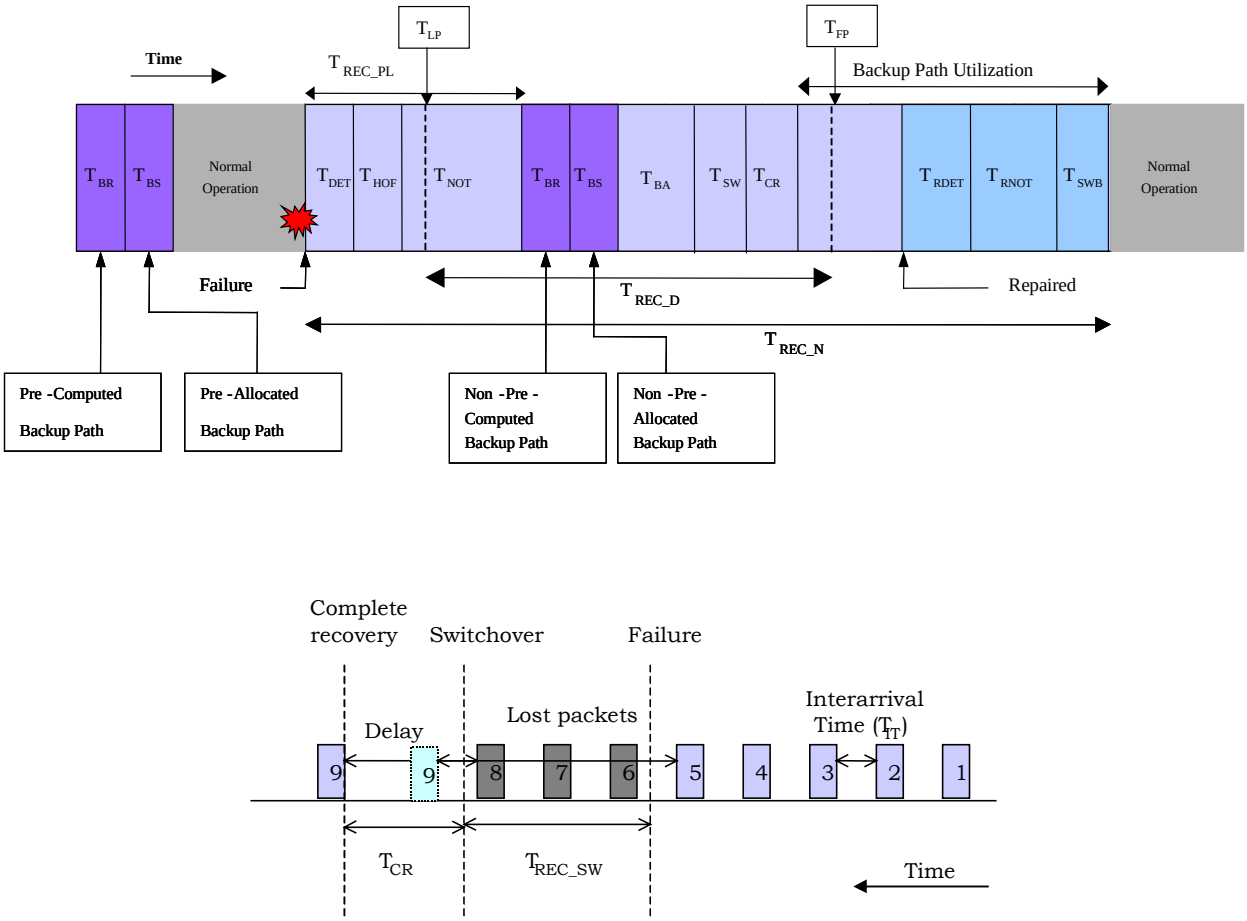


Figure 4.2: The failure recovery time process.

Acronym	Component	Description
T_{DET}	Failure detection time	The time required to detect the fault (for instance, an alarm from lower levels or a 'hello' protocol)
T_{HOF}	Hold-off time	The time required to allow failure recovery at lower layer mechanisms (if necessary)
T_{NOT}	Failure notification time	The time required to inform (i.e. signaling-based or flooding-based notification) the node responsible for switchover
T_{BR}	Backup routing time	The time required for new backup creation, routing (TBR) and signaling (TBS)
T_{BS}	Backup signaling time	
T_{BA}	Backup Activation	The time required to activate (signaling/cross connection) the backup path before the switchover
T_{SW}	Switchover time	The time required for traffic switchover from the active path to the backup path
T_{CR}	Complete recovery time	The time required to complete the fault recovery (the time it takes the first packet to arrive from the backup path to the egress node)
T_{RDET}	Initial path recovery detection time	The time required to detect the working path restoration (time for the WP recovery detection)
T_{RNOT}	Initial path recovery notification time	The time required to notify of the working path recovery (time for the WP recovery notification)
T_{SWB}	Switchback time	The time required to switch the traffic back from the backup path to the working path

Table 4.2: Failure recovery component description

Therefore, the recovery time (T_{REC_N} in the figure 4.2) from the moment the failure occurs to when the traffic is restored to the initial working path can be evaluated by simple addition, as the following expression shows:

$$T_{REC\_N} = T_{DET} + T_{HOF} + T_{NOT} + T_{BR} + T_{BS} + T_{BA} + T_{SW} + T_{CR} + T_{RDET} + T_{RNOT} + T_{SWB} \quad (4.5)$$

During part of the this recovery process there is a proportional packet loss. However this packet loss is not proportional to formula 4.5.

Once a failure occurs, packets are lost until the traffic is switched to the backup path. This time is denoted T_{REC_PL} (see Figure 4.2) and can be evaluated as follows:

$$T_{REC\_PL} = T_{DET} + T_{HOF} + T_{NOT} \quad (4.6)$$

The time required to repair the failure is normally long or very long (hours or days) with respect to the time to recover from a failure using a backup path (ms to minutes).

The time required for activating the switchover is proportional to:

$$T_{REC\_SW} = T_{DET} + T_{HOF} + T_{NOT} + T_{BA} \quad (4.7)$$

However, the delay experimented by the Path Merge Label Switch Router (PML), for instance an egress node in the case of path protection, is different from the T_{REC_PL}. The reason is that there are packets that still arrive to the PML node from the initial working path after the failure occurs.

The delay between the last packet (T_{LP}) from the working path (packet 5, in figure 4.2) and the first packet (T_{FP}) arriving from the backup path (packet 9 in figure 4.2) is called as T_{REC_D}.

The Interarrival time (T_{IT}) (i.e. the time between packet arrivals) is also taken into account in this formula, therefore the accumulated delay to the restored traffic is:

$$T_{REC\_D} = (T_{FP} - T_{LP}) - T_{IT} \quad (4.8)$$

The time required for the first packet (T_{FP}) to arrive from the backup path to the PML node involves all the phases of the recovery process depicted in table 4.2., except the normalization process (the recovery of the initial working path). However, T_{LP} and T_{FP} are not easy to evaluate because of their dependence on the current traffic conditions.

Packet Loss (P_{LS}) is proportional to the T_{REC_PL} and to the Transmission Rate (R_{TR}). Therefore, taking into account the formula 4.6, packet loss can be evaluated. Packet loss in the fault link P_{FL} (i.e. those packets being transported in the physical link at the moment of failure) should be also taken into account.

The resulting expression for packet loss is:

$$P_{LS} = R_{TR} \cdot T_{REC\_PL} + P_{FL} \quad (4.9)$$

Losses cannot be totally avoided by the protection mechanisms presented in Chapter 1. However, there are some proposed mechanisms, (such as the one presented in [HUN02]) which overcome this drawback.

4.3.2. Components to reduce the recovery time

The main aspects of reducing the failure recovery time are presented in this section. Table 4.3 sums up the options for reducing failure impact by reducing the time needed for each phase of the fault recovery. Note that the normalization process (the traffic restoration to the initial working path) is not included in this table because its effects are not considered in this work.

Recovery phase	Features	Time Reduction
Fault detection (T_{DET})	Depends on the technology	Cannot be reduced (except in the case of monitoring techniques; see section 1.4.3)
Hold off time (T_{HOF})	Depends on the lower layers	Setup (0-50 ms)
Notification time (T_{NOT})	Depends on the Failure Notification Delay and notification method	Minimizing the Failure Notification Distance and optimizing the process
New Backup creation (T_{BR} + T_{BS})	Depends on the routing and signaling method applied	Pre-establishing the backup
Backup Activation (T_{BA})	Depends on the backup distance and signaling cross-connection process	Minimizing the backup distance and optimizing the process
Switchover (T_{SW})	Depends on the node technology	Cannot be reduced
Complete recovery (T_{CR})	Depends on the backup distance	Minimizing the backup distance

Table 4.3: The fault recovery cycle and the failure impact reduction.

All protection mechanisms do not strictly follow the recovery cycle as defined in section 1.4.3. For instance, 1+1 protection mechanisms overcome the failure impact by drastically reducing the fault recovery time. However, these 1+1 protection mechanisms cannot always be applied. This is due to high resource consumption or, in those cases, to not finding two disjoint paths.

Reducing fault detection and switchover time depends on the technology, hence it cannot be easily modified. For instance, in some nodes lower layers report the failure detection via alarm indication. In other cases, the failure detection

process is carried out using monitoring techniques as explained in chapter 1. In those cases, the monitoring time can be increased in order to achieve faster detection. However, this could result in scalability problems [HUA02].

On the other hand, the time to establish on-demand backup paths (once the fault is detected) depends on the routing and signaling methods used. In MPLS, a backup path can be pre-established with no allocated resources (bandwidth). This technique is also known as “fast restoration”.

In network scenarios with high traffic loads and no packet prioritization techniques, a no bandwidth reservation could result in a large notification time [CAL04] and [CAL04b]. In optical domains, the bandwidth (at least a wavelength) must always be allocated. In the case of pre-established backup paths, this delay can be avoided.

When the protection level, in terms of recovery time, has to be fast or very fast, pre-established and pre-allocated backup paths should be used. In these cases the reduction of the notification time, backup activation, and complete recovery are probably the most challenging aspects of designing the protection methods for a network.

In order to minimize these times, the process failure notification and backup activation process are crucial. In this work, the signaling-based and flooding-based processes are taking into consideration to optimize these phases.

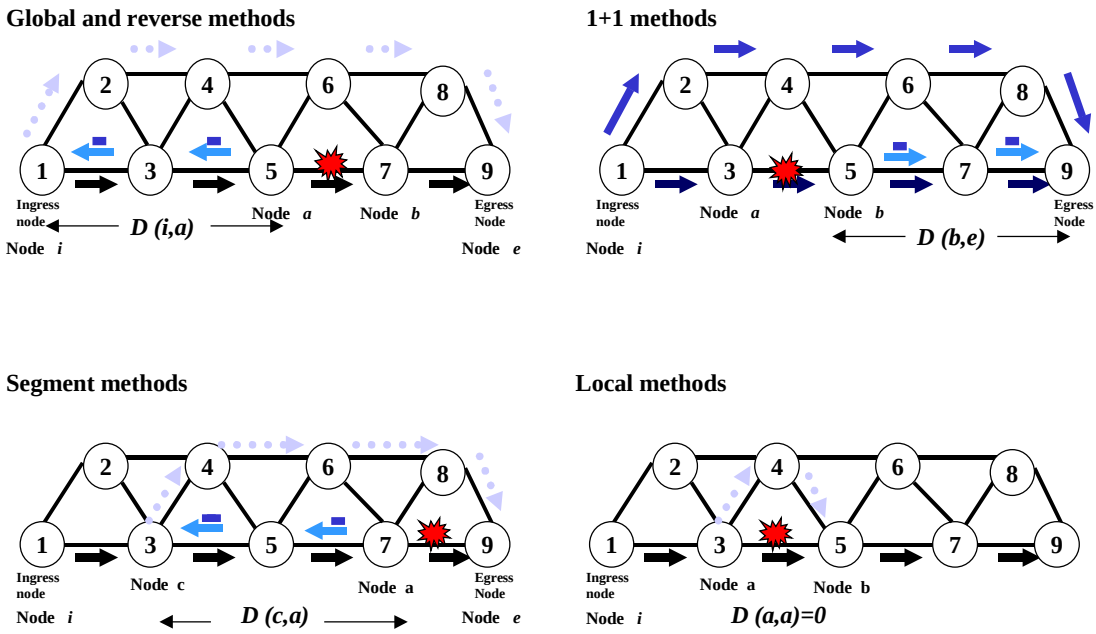
In the next section, the failure notification distance is defined for the different recovery models. In these cases the failure impact is considered in order to reduce the packet loss, following formula 4.6.

4.3.3. Recovery Time and Failure Notification

In this section the relationship between the failure notification process and the reduction of the failure recovery time is presented.

Currently, there are different failure notification strategies, depending on whether: a) they are designed to notify a link/node failure or an LSP failure, b) these techniques are able to notify about data plane and control plane failures and c) these techniques are time constraint techniques or not.

The notification time is the sum of the time to propagate the fault indication signal and the distance D_{NOT} (in the rest of this document D_{NOT} is also referred to as D). D_{NOT} is defined as the number of links/nodes between the node detecting



i : ingress node
e : egress node
a: initial node of the failed link
b : next node of the failed link
c : initial node of a segment or local backup path

Figure 4.3: Failure notification depending on the protection method

the failure and the node responsible for the switchover.

In this section a detailed analysis of the failure notification time is introduced, considering different notification strategies. However, this distance D depends on the fault recovery method applied (see Figure 4.3).

In the global and reverse methods this distance is equal to D(i,a), where the node detecting the failure is the node previous to the broken link (node a), and the node responsible for the switchover is always the ingress node (node i).

In segment recovery methods this distance is equivalent to D(c,a). In this case the node responsible for the switchover is not always the ingress node. It can be any other node (node c) along the working path, setup with PSL functions.

In the case of 1+1 methods, this distance is D(b,e). Whenever the egress node is not able to detect the failure by itself (it is not a selector node as defined in optical networks), the egress node must be notified of the failure, in order to execute the switchover (selecting an alternative path as the active path). Consequently, the distance is calculated between the node detecting the failure, in this case the next node to the failed link (node b), and the egress node (node e).

Finally, if local backups are used, the distance is zero, because in this case the node detecting the failure and the node responsible for the switchover is the same node (D(a,a)).

Therefore:

$$T_{NOT} = T_{PR} \cdot D_{NOT} \tag{4.10}$$

Where:

D_{NOT} : Notification distance (number of hops)
T_{PR} : Time to propagate the signal on each hop.

This formula is just an approximation. The distance and the time required to propagate the failure indication signal are not the same in each hop. In the following section a more detailed formulation of the main factors affecting the time to transmit a failure indication signal is presented.

In [RAB03] other factors which affect the failure notification time are pointed out. First, the time needed to traverse each link and the delays incurred at the nodes are considered. The time to traverse each link is the addition of the transmission time and the link propagation time. The Link Propagation Time (T_{PROP}), or the latency in the propagation of the packets along links, is proportional to:

$$T_{PROP} = L_{LINK} / S_{PROP} \quad (4.11)$$

where L_{LINK} is the physical Length of the link and S_{PROP} is the Link Propagation Speed. The link propagation speed is usually approximated by a light speed in a fiber at 2/3 of its speed in free space (about 200,000 km/s.)

The Transmission Time (T_{TRANS}) is calculated based on the link capacity as follows:

$$T_{TRANS} = P_{SIZE} / S_{LINK} \quad (4.12)$$

where the P_{SIZE} is the Packet Size (i.e. number of bits) and S_{LINK} is the Link Speed (expressed in bits/sec).

On the other hand, two delays are important at nodes: the processing time and the queuing/buffer delay. The Node Processing Time (T_{PROC}) is considered in the literature [LI01] as a few tenths of a millisecond in the case of a Reservation Protocol (RSVP) object. This value is smaller in the case of a Link Management Protocol (LMP) message requesting the activation of an LSP path.

The Buffer/Queuing Processing Time (T_Q) depends on the failure notification scheme used. For a flooding-based method this time is negligible and for a

signaling-based method T_Q is proportional to the number of protection paths and to the number of failure notification messages.

An upper bound of the queuing time can be evaluated as follows:

$$\text{Max\_}T_Q \text{ (signaling)} = N_{LSP} \cdot (\text{Packet size} / \text{Link BW}) \quad (4.13)$$

where

N_{LSP} : Number of protected LSPs established in the failed link.

In the absence of a priority queuing, the maximum queue delay can be calculated at node A assuming fair queuing at the FIFO buffers of all control channels and input buffers only:

$$\text{Max\_}T_Q \text{ (signaling)} = \text{Number of queues} \cdot (\text{Queue size} / \text{Link BW}) \quad (4.14)$$

This value depends on the hardware of the buffer implementations. In this scenario some failure indication messages can be lost. This case is not considered in this work.

The maximum number of messages is:

$$\text{Max\_Number of msg (signaling)} = (N_{LSP}\text{-}1) \cdot [D_{NOT} + 2 \cdot D_{BP}] \quad (4.15)$$

where

D_{NOT} : Distance (number of hops) of the notification path
D_{BP} : Distance (number of hops) of the backup path.

In formula 4.6 the time while packets are lost is presented (T_{REC_PL}). Note that in this case the T_{NOT} is proportional to the time for transmitting the failure indication message from the node which detects the failure to the node responsible for the switchover (PSL node).

In summary, the delay to transmit a packet from a node X to a node Y can be expressed as follows:

$$\text{Packet\_Delay} = \sum_{i=X}^{Y} (T_{PROPi} + T_{TRANSi} + T_{Qji} + T_{NPi}) \quad (4.16)$$

where

T_{PROPi} : Propagation time link i
T_{TRANSi} : Transmission time link i
T_{Npi} : Node Processing time (node i)
T_{Qji} : Queuing time of the failure indication message j in node i.

In the case of signaling-based protocols, T_{Qji} is calculated as follows:

$$T_{Qji} = N_{MSGji} (\text{Packet size} / \text{Link BW}_i) \quad (4.17)$$

Where

N_{MSGji} : Number of failure notification messages before failure notification message of LSP j in node i.
Link BW_i : Link Bandwidth of link i.

N_{MSGji} depends on the queuing algorithms and the technology. A more detailed formulation and analysis of the queuing time, and even the node processing time, are beyond the scope of this work. The presented formulation only gives an approximated boundary for evaluating the recovery time, enough to achieve the objectives proposed in this thesis.

4.4. Reducing the failure probability and failure impact

In this Section, we present an analysis of the use of the proposed paradigms, such as fault probability, notification time and resource consumption. The objective is designing (and managing) networks in order to minimize fault probability and fault impact. This is not an easy goal, sometimes there is a trade-off between reducing the impact and reducing the failure probability. For instance, reducing fault probability may imply increasing the distance D(i,a), and therefore increasing the potential impact of a fault. On the other hand, reducing both simultaneously could imply excessive resource consumption. In addition, the class of traffic to be protected can also be crucial in making the right decision. In the next section these aspects are discussed in detail.

4.4.1. Residual Failure Probability (RFP) and Failure Impact

In chapter 3 and in section 4.2, respectively, the failure probability and the failure impact have been defined.. However, if the network links/segments are protected (using a backup path) the residual probability and impact values can be reduced or eliminated. In a simple scenario (shown in figure 4.4.a) the working path (formed by the LSRs 1-3-5-7) contains two links with different failure probabilities ($1 \cdot 10^{-4}$ and $4 \cdot 10^{-4}$).

If the working path is not protected, the Residual Failure Probability (RFP) of this path is the sum of the path link failure probabilities (as seen in section 3.3). However, if the path is protected, using segment or local backups (fig. 4.4.c and 4.4.d respectively), the residual failure probability is negligible (for simplicity it is assumed to be zero in formulation). On the other hand, if the backup policy is to protect just one link, the residual failure probability is the sum of all the non-protected link failure probabilities (fig. 4.4. b).

On the other hand, the failure impact is evaluated as the degradation of the QoS after a failure occurs in a protected segment. This degradation is proportional to the failure recovery process (as explained in the above section). We propose to

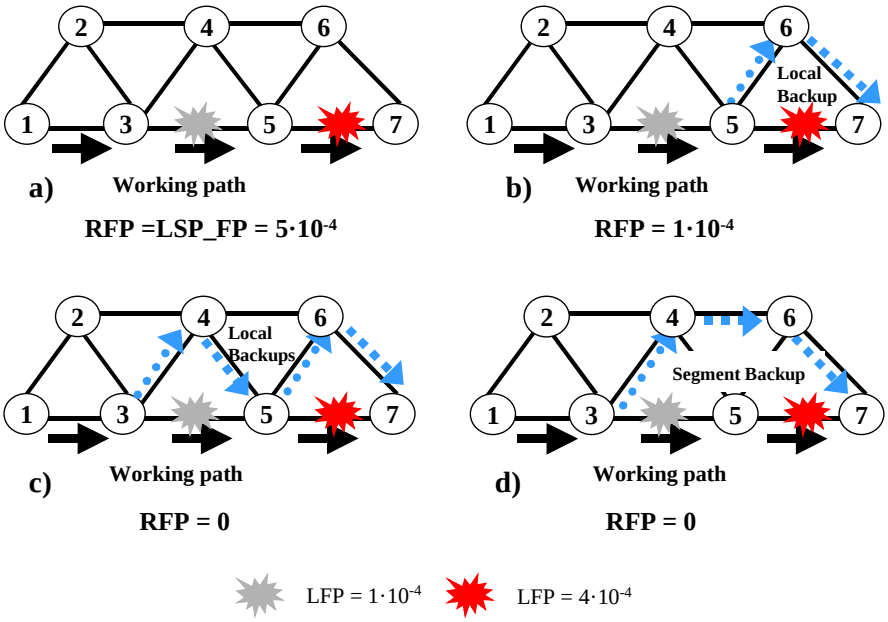


Figure 4.4: Residual Failure Probability in the working path

evaluate this impact based on the failure recovery time. The time since the failure occurs to the PSL node stops sending packets along the working path (the time while some packets are lost) is the failure notification time (see section 4.2).

Therefore, if the links are protected using local backups (D(i,a)=0) the failure impact is virtually null (fig. 4.4.c). If segment or global backups are used, the failure impact would depend on the distance of each link (to be protected) respect to the first node (PSL node) of the LSP backup path.

4.4.2. Reducing the network failure probability and impact: a case study

In this section, the reduction of the network impact and the failure probability are analyzed in some cases using different backup techniques. We assume that a two-step routing method is applied, which means that the working path is selected first and then the suitable backup path is chosen based on the working path protection requirements.

Using a two step rouitng give us a better resource consumption respect to one-step routing algorithms [CAL04]. Refer to Section 4.5 for more details.

For a given working path, the most suitable backup method is selected depending on its protection requirements and traffic service. In the next section we present a classification of the traffic services offering the most suitable protection for each case presented.

Case 1: The fault probabilities of all links of the route are zero or negligible: there is no need for protection (see Fig. 4.5). No backup paths are used and the resource utilization is optimal.

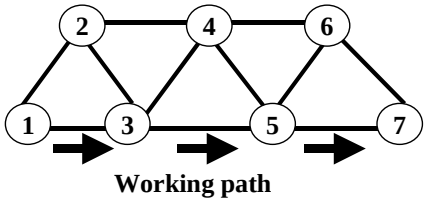


Figure 4.5: *Case 1:* LFP=0 for all the working path

Case 2: There is just one link to be protected (no zero fault probability) in the route: just one local backup associated with this link is needed (see Fig. 4.6). If the class of traffic carried does not need protection

(best effort, low priority, and so on), no backup path would be established (Case 1).

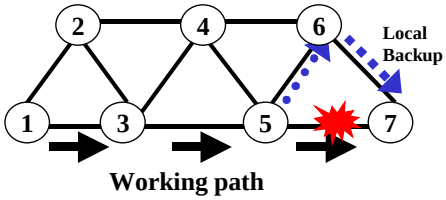


Figure 4.6: *Case 2:* Only one link to be protected

Case 3: The working path has some consecutive links, forming a segment of links, to be protected. Segment backup paths (Fig. 4.7.a) or global backup path protection techniques (Fig.4.7.d) can be used to protect this segment. Path protection techniques (such as the global

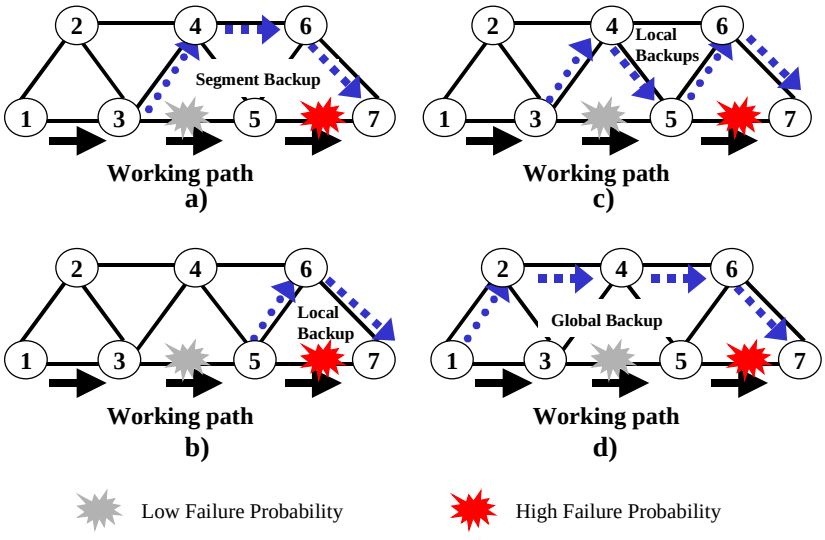


Figure 4.7: *Case 3:* Consequtive links to be protected.

.

backup paths) eliminate the Residual Failure Probability (RFP = 0). However, the Failure Impact for the link with high failure probability is proportional to the distance (D(3,5)=1), in the case of segment protection, and D(1,5)=2, in the case of global backup protection). If the carried traffic is sensitive to the recovery time or the packet loss and there are links with large fault probability, the segment cannot be protected using segment or global backups. Hence, local backup protection should be used in order to avoid a large failure notification delay (Fig. 4.7.c). The Failure Impact can be also eliminated using two local backups, but this can result in larger resource consumption. An intermediate solution can be achieved when just the link with high failure probability (link 5–7, see Fig. 4.7 b) is protected with a local backup. In this case (case 4.7.b) the failure probability and impact for the link with high protection requirements are eliminated. Regarding case 4.7.c (two local backup paths), the amount of resources is also reduced. However, the link 3-5 (with a certain failure probability) is not protected.

Case 4*:* The working path has some links to be protected, but they are separate so a segment backup cannot be used. In this case, the protection method to be applied depends on the level of the desired protection and on the traffic class. If the number of links to be protected is large, a global backup, which includes all links (large and small fault probabilities), can be used (Fig. 4.8.a). This involves eliminating the residual failure probability (RFP=0), but could increase the distance (as shown in Figure 4.8.a) thereby introducing greater packet loss and longer recovery times in the case of failure (a high failure impact). In this example, the distance for the high failure probability link is 2 (D(1,5)). For higher levels of protection, local backups should be established for each link (Fig. 4.8.b). At least those with high fault probabilities should be protected in order to offer a balance between the final protection degree and resource consumption (similar to case 3). This last case is shown in Figure 4.8.c.

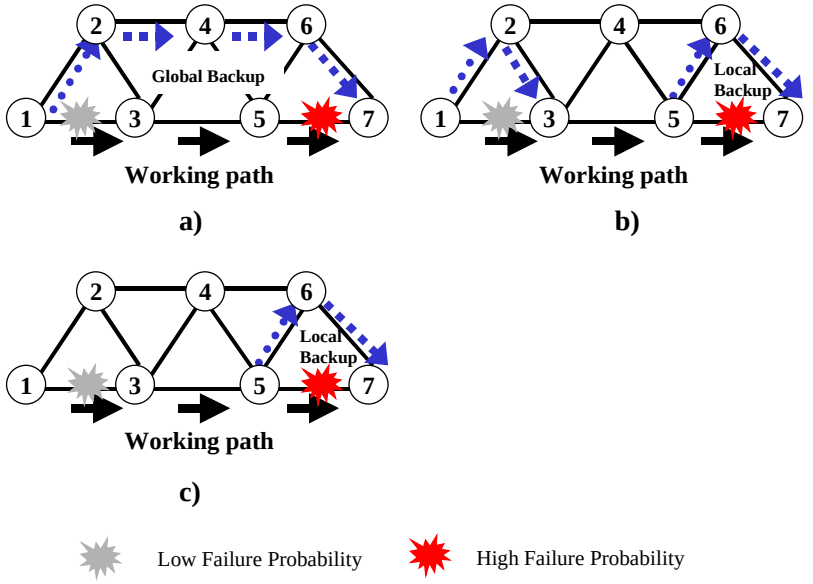


Figure 4.8: *Case 4:* Separated Links to be protected

Case 5*:* Hybrid cases: depending on the fault link probabilities and distances D(i,a*)* (for notification), a specific choice between local, segment or global protection should be made. Even a "no protection" policy may be chosen, depending on the traffic class.

4.5. Protected traffic services in GMPLS networks

In previous sections the main components of implementing and managing a fault recovery method were reviewed and compared. However, each fault recovery method offers different QoS aspects. For instance, local backups offer high recovery times and low-none packet loss, but on the other hand, in paths with a high number of links to be protected, applying local backups could result in large resource consumption. For a given network scenario, it is not simple to

determine what is the most suitable fault recovery method.

Often the major decision parameter used to select the protection scheme is the type of traffic transported by the network. By characterizing each traffic type, according to protection constraints, the selection of the most suitable fault recovery scheme could be simplified.

In the following section different traffic class characterizations in terms of their protection requirements are introduced.

4.5.1. Traffic class protection requirements - the DiffServ example.

In [MAR03] a proposal for mapping different recovery methods based on a differentiated services classification 'diffserv' is introduced. Let us consider a diffServ scenario where four Class-Types are defined according IETF's RFC [FAC02] and [BLA00]. A Expedited Forwarding (EF) class is defined to transport real-time traffic, two Assured Forwarding (AF1 and AF2) classes are used by traffic with two different flavors for losses and, as is usual, a Best Effort class for traffic with no QoS requirements.

Following the QoS requirements, we propose different protection strategies as shown in Table 4.4. Local recovery protection is assigned to EF due to the restoration time constraint, which should be short for real time traffic. As very low losses are required, for AF1 the reverse backup is chosen. The protection domain for AF2 and BE can be global or local depending on link reliability. The next three columns in Table 4.4 (LSP setup, resource allocation and bandwidth) are the protection parameters defined in [BLA98]. LSP setup concerns the initiation of the recovery setup; in the pre-established case, a recovery path is established prior to the link failure, while for the on-demand LSP setup the recovery path is established after the failure. The pre-established scheme for setup is obviously faster, and therefore it is proposed for EF and AF1 traffic

classes. Resource allocation, in the next column, indicates if network resources (normally bandwidth) are allocated to LSP before the failure (pre-reserved) or

Traffic Class	QoS requirements	Protection domain	LSP setup	Resource Allocation	Bandwidth
EF	Real-time[1]	Local recovery	Pre-established	Pre-reserved	Equivalent
AF1	Very low losses	Reverse recovery	Pre-established	Reserved on demand	Equivalent
AF2	Low losses	Global/local	On-demand	Reserved on demand	Limited
BE	No requirements	Global/local	On-demand	Reserved on demand	Limited

Table 4.4: Protection assignment for DiffServ Classes Types

after the failure, noting that LSP can be established with no specific bandwidth allocated. As the last column shows, there are two strategies to allocate bandwidth to LSPs: to allocate equivalent bandwidth (the same amount as the working path) or limited bandwidth (less than the working path). For EF and AF1 equivalent bandwidth is allocated, hence no significant QoS degradation is expected.

The QoS routing performance enhancements could use the traffic-profile concept to characterize the probability and/or the sensibility of a traffic-profile in the case of failure, in terms of packet losses, restoration delay, and so on. Therefore, the routing algorithm could provide different solutions depending on the traffic type. There are some proposals [AUT02], [AUT02b], [CHE99] and [ZHA02] aimed at mapping different traffic classes with the protection methods described in Section 2.

4.4.2. Differentiated resilience services proposal

We propose in [CAL04] characterizing different resilience services by setting bounds to the network fault probability and the failure impact requirements.

In this way, we characterise the following traffic classes:

- *High-resilience requirement traffic services*: Traffic that is very sensible to network faults (like EF diffserv traffic). Residual Failure probability and Failure Impact values should be set up at zero. 1+1 or local backup paths can be used in order to accomplish these values.

- *Medium-resilience requirement traffic services*: Traffic that is sensible to network faults (like AF1 or AF2 diffserv traffic). However, resource consumption should be taken into account to route the working and backup paths. Residual failure probabilities and failure impact values should be bounded in order to achieve the desirable QoS with appropriate resource consumption. Segment and global backups can be used to protect these services.

- *None-resilience requirement traffic services*. No protection requirements are needed (BE traffic).

It is not a simple task to select the most suitable protection mechanism for every working path. Although all the information about traffic class, available network resources, and so on is available, a decision mechanism, more or less sophisticated, is desirable to select the most suitable protection method.

In the following section a discussion of current restorable QoS routing is presented and contrasted with some of the enhancing mechanisms described in Chapter 3.

4.6. Enhancing the QoS routing algorithms

In previous sections the reduction of the network failure probability and the network impact if a failure occurs were introduced and analyzed. In this section the application of these techniques, enhancing the level of protection of some routing algorithms is introduced.

The enhancement of the QoS routing algorithms could be done adding new objectives to their algorithms to compute the suitable working and backup paths at the same time or adding the most suitable backup scheme after the working path has been selected.

In our first proposal [MAR03a], a backup decision module was introduced to select the most appropriate backup technique depending on the traffic class. However, although this technique offers the possibility of combining different backup techniques, this module does not take any decision about selecting the working path (for instance the path with minimum failure probability). On the other hand, this involves developing a very costly (in terms of computing time) module. We have added new objectives to the QoS routing algorithms to compute both the working and backup paths, reducing the impact and the failure probability.

In this section the backup decision module is presented. Secondly, the new objectives for the enhanced routing methods are introduced which depend on the techniques used for the backup path. Then, the main advantages of using two-step routing algorithms over one-step routing algorithms are discussed. Finally, the new routing information to deploy these algorithms in a GMPLS control plane is presented.

4.6.1. The backup decision module

In [MAR03a] a backup decision module was introduced. This module performs after the working path computation. The backup decision module selects the

most suitable backup model (in this case only global, reverse or local backup paths are considered), depending on the working path features and the traffic class supported by this path.

Backup model	**QoS_Protection (QoSP) formulation**
$QoSP_{Global}$	$\alpha * P_{LS}^{N} + \beta * T_{REC}^{N} + \lambda * RC_{G}^{N}$
$QoSP_{Local}$	$\lambda * RC_{L}^{N}$
$QoSP_{Reverse}$	$\beta * T_{REC}^{N} + \lambda * RC_{R}^{N}$

Table 4.5: QoSP formulation

Traffic Class	**QoS requirements**	**α**	**β**	**λ**
EF	Very low P_{LS} and T_{REC}	0,5	0,45	0,05
AF1	Very low P_{LS}	0,5	0,3	0,2
AF2	Low P_{LS}	0,33	0,33	0,33
BE	No requirements	0,05	0,05	0,9

Table 4.6: QoSP and traffic class assignment

The QoS protection (QoSP) concept was introduced . The QoSP is based on the packet loss, failure recovery time and resource consumption parameters. The utilization of the failure notification distances is also considered in the QoSP evaluation. A formulation based on these tree parameters (packet loss, failure recovery time and resource consumption) computes the QoSP value. However, the packet loss and the recovery time have a direct relationship; as explained, resource consumption is totally independent in them. A normalization process is introduced to evaluate an approximate value of QoSP according to Table 4.5.

Each traffic type (diffserv model) weighs each QoSP parameter (packet loss,

recovery time and resource consumption) based on their own protection requirements to compute the final QoSP value.

The backup decision module uses this value choosing the most suitable protection scheme (see Fig. 4.9). This module computes each QoSP value (Table 4.5) for the traffic classes and decides on the backup path.

The backup decision module is a first approach to evaluate the level of protection and compute the suitable protection scheme based on the QoSP value. However, the proposal presents several drawbacks. The first problem is that the number of backup or fault recovery schemes is very limited (only global, reverse and local pre-established, pre-allocated schemes). This involves a very complex process (in terms of time computation) to scale this proposal to all (or more) protection models. On the other hand, the weighted values for each traffic class (Table 4.6)

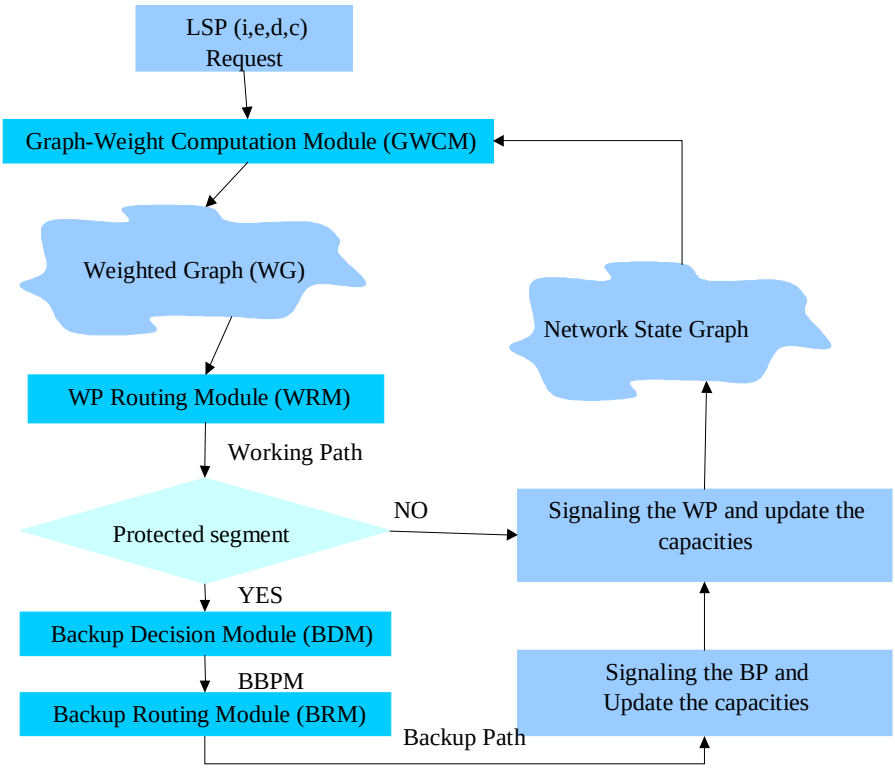


Figure 4.9: The backup decision module proposal.

have to be tuned and this is not a simple process. Finally, the working path features (such as the probability of link failure) are not considered in this scheme.

In the next section we introduce a new proposal, less complex (in terms of scalability), to enhance the level of protection provided by the current QoS routing algorithms without a backup decision module. This is based on adding new objectives to the routing algorithms.

4.6.2. Adding new routing protection objectives

In Section 4.5 different methods to reduce the failure probability and the failure impact were introduced. In the current literature, as explained in Chapter 2, several QoS routing algorithms use different backup methods to offer a certain level of protection. However, some of these methods do not reduce the residual probability or the failure notification distance.

In this section we present new objectives to improve the protection level of some current QoS routing algorithms. In Table 4.7 a method to enhance these algorithms introducing new objectives in the current QoS routing algorithms is shown.

The first column in Table 4.7 describes the protection methods used by the routing method. For instance, a routing method could protect the LSPs using segment protection methods for those links with a certain level of failure probability. However, this method would not protect those links with lower levels of link failure probability. In the second column, the objective of reducing the residual failure probability for each case is described. For example, if no backup method is used, the residual failure probability (RFP) reduction is equivalent to the LSP failure probability (LSP_FP) reduction. If segment backups are used, those unprotected segments need to be reduced, hence RFP reduction is a new goal in the routing methods.

Backup method	Failure Probability Reduction	Failure Notification Distance Reduction
No backups	RFP = LSP_FP	-
1+1 backups	-	D(b,e)
Global/Reverse Backups	-	D(i,a)
Segment Backups	RFP	D(c,a)
Local Backups	RFP	-

Table 4.7: QoS routing algorithms. New objectives.

Some mechanisms (such as path protection methods) achieve a RFP = 0, as explained in Section 4.5, hence no RFP or LSP_FP have to be added to their objectives. Finally, the corresponding failure notification distance, with the objective of being reduced in order to minimize the failure impact, is introduced in the last column. Again, some methods, such as those routing methods with local backups, do not need to introduce this objective, because their notification distances are zero.

The application of these objectives needs to modify some of current routing algorithms in two different ways. On one hand, some of the backup techniques, such as the 1+1 methods, need to compute two disjoint routes. If a one-step routing method is used to compute these paths, the evaluation and reduction of the RFP and notification distances cannot be done. On the other hand, some new routing information has to be added to the routing information entities.

The next sections present these problems and their solutions in more detail.

4.6.3. Two-step routing algorithms versus one-step routing algorithms

The process of establishing the working and backup paths can be done in two steps by first calculating the working path (the shortest path meeting the QoS constraints) and then calculating the backup path (the shortest disjoint path). In some cases, the working path of the two-step algorithm blocks all the possible global backup paths (see Fig. 4.10.a). There are some proposals that establish the shortest cycle algorithm in order to avoid this disadvantage. However, as explained in Section 4.5, it is more useful to take into account the working path properties (with respect to the failure probabilities) in order to select the working path with the optimum protection requirements. This allows us to select a path with less failure probabilities, and, in some cases, allows for better resource consumption (using local/segment backups or even no backups at all). Figure

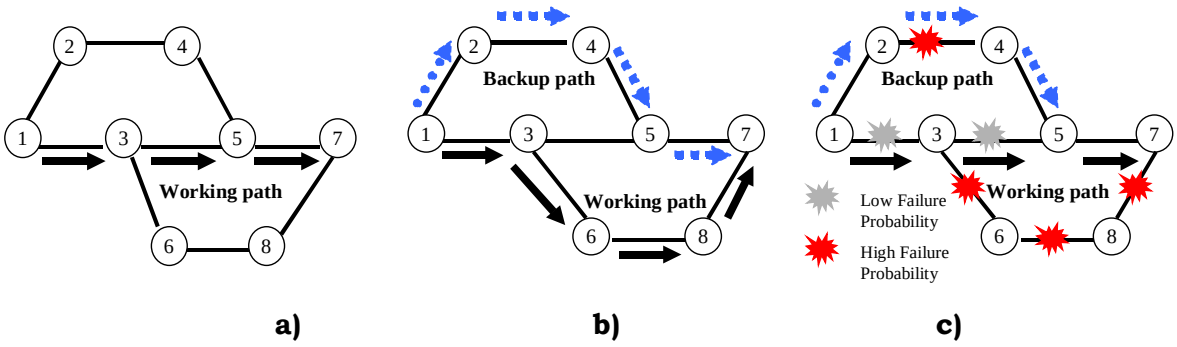


Figure 4.10: Disjoint routes computation

4.9.c) depicts a case where a two-step algorithm allows a working path with less failure probabilities to be selected. Furthermore, it can be protected with a segment backup path that results in less total (working and backup path) resource consumption than in case b).

4.6.4. Routing information

Routing protocols are used to communicate the resource properties and compute the paths. All routing proposals should maintain information concerning link failure probabilities, maximal reservable bandwidth, etc. Tree databases are proposed in [ZHA02] in order to support the information and routing computation at the GMPLS control level.

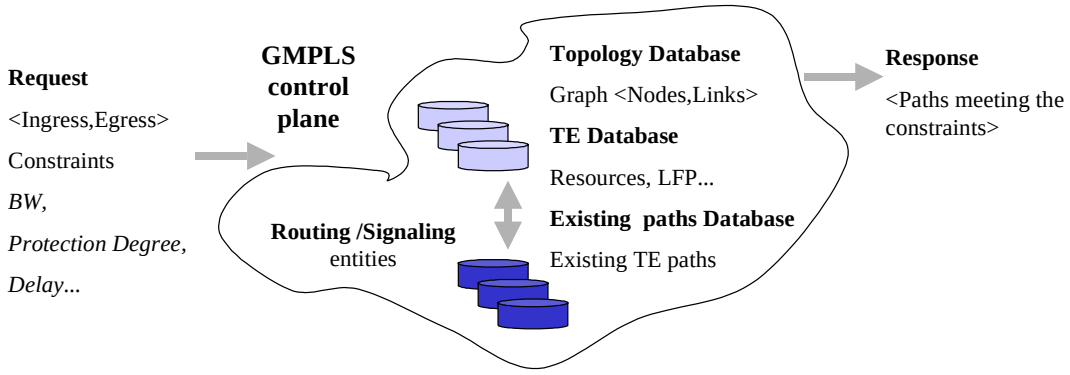


Figure 4.11: Interfaces with the Routing Algorithm Module

- Topology Database: Contains the information about the network graph.
- TE Database: Contains information about the network constraints used by the routing protocols.
- Existing Path Database: Contains information about the current working and backup paths.

Figure 4.11 depicts the routing process of an LSP request using the databases. The frequency to update this information and the development of QoS routing protocols under different routing information scenarios is out the scope of this thesis. More information can be found in [MAS03], [KOD00] or [KOD02].

4.7 Summary and Conclusions

In this chapter the main points for reducing the failure probability and failure impact in a multiservice network have been presented. A deep analysis and formulation of the failure recovery time has been provided. Segment, local and path protection techniques have been considered for this analysis. Several network scenarios have been explored and different traffic services have been characterized taking into account the failure probability and failure impact values. Two proposals to enhance the level of protection of some current routing algorithms have been also presented.

In the next chapter some experimental and analytical results are shown. These results demonstrate the correctness of this chapter's formulation and support the main choices made.

CHAPTER 5

5
Analytical and experimental results

5. Introduction

This chapter introduces the results of several analytical and experimental tests carried out to verify the performance of the algorithms proposed in this work. The chapter is organized into sections for each of the four sets of test. The first set of experiments was carried out in a simple mesh topology evaluating the behavior of some network parameters which affect failure recovery time and, consequently, packet loss (refer to the formula in Section 4.3). This set of experiments was deployed using an NS-2 simulator [FAL] AND [FALa]. Global, local and reverse

backup techniques were taken into account in this first evaluation.

In the second set of experiments a backup decision module (depicted in Section 4.6.1) was implemented and executed in different network scenarios (with different traffic services and bandwidth LSP requests). Several analytical results are explained, pointing out the most suitable backup techniques to achieve the required QoS protection level. However, the backup decision module involves a costly process (in terms of computing time). On the other hand, some parameters, for each of the traffic classes, have to be tuned to achieve more accurate results.

The next set of experiments modified a QoS routing algorithm to reduce the network failure probability. A typical ISP network was used to carry out this set of experiments. Two cases were deployed: the first case minimized the LSP failure probabilities, without using any backup technique; the second case implemented different backup techniques, reducing the residual failure probability (as explained in section 4.4.2).

This set of experiments was carried out in both static and dynamic environments. The static case considers long-lived LSPs (the LSPs are not deleted during the experiment). The dynamic case sets up a rate and a holding time for each experiment.

The last set of experiments was carried out using another ISP network topology (NSF-Net). These experiments included all the decisions and formulations of this thesis. The failure probability and failure impact (in terms of recovery time) reductions were considered. Different failure notification and backup activation techniques (signaling and flooding) are included in this section. Path and segment protections were also experimented with.

Table 5.1 summarizes the different network topologies, simulators and objectives of each set of experiments. References to where these experiments have been published are also included.

Experiments and refs.	Objectives	Network Simulator	Network Topology
Section 5.2 [CAL03b] [CAL04]	Evaluating formulation of section 4.3. Failure recovery time and packet loss. (Different propagation and transmission time.) Different traffic rates and loads. Explicit routing and global, local and reverse protection.	NS-2, (MPLS module) for ns2.8, modified by us: Global, reverse and local backups. CBR and VBR traffic.	NS-Mesh topology
Section 5.3 [MAR03a]	Evaluating the backup decision module (section 4.6.1). Global, local and reverse protection are considered.	Analytical results	-
Section 5.4 [CAL03] [CAL03b] [CAL03c] [CAL03d] [CAL04c]	Enhancing some current QoS routing algorithms. Network protection degree. (Chapter 3 formula, and section 4.6). 5.4.1. Modified K-WSP. No protection (no backups). LSP failure probability evaluation. Notification distances. 5.4.2. Modified K-WSP. Path, segment and local protection. Residual failure probability evaluation. Notification distances.	Call simulator	KL-Net Topology. Different link capacities and link failure probabilities. Fixed physical link lengths.
Section 5.5. [CAL04b]	Adding failure notification and backup activation techniques (signaling and flooding). Failure recovery time analysis. Formula of section 4.6).	Call simulator	NSF-Net topology. Different physical link lengths and link failure probabilities. Fixed link capacities

Table 5.1: Experiment and analytical results.

5.1. Network Topologies

Different topologies have been used to develop the experiments. In this section a brief description of each topology is presented.

5.1.1. The NS-Mesh topology

The NS-Mesh network was implemented in the NS-2 network simulator [FALa] MNS2.0 (MPLS module) for ns2.8 [GAE99], [GAE00] and [GAE02]. This module

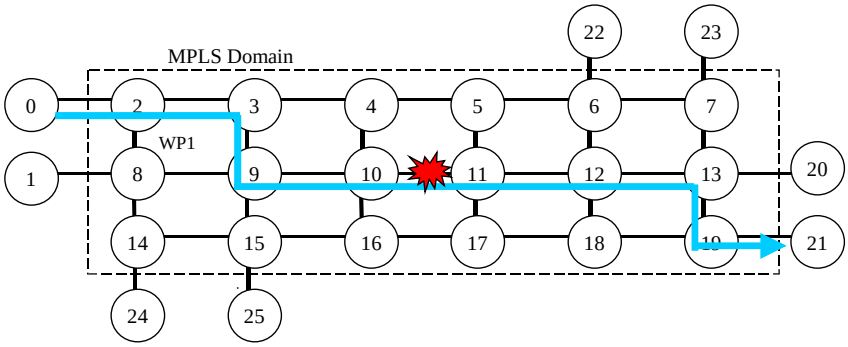


Figure 5.1: NS-Mesh Test Network Topology

Network parameters	Traffic parameters (CBR)	Background Traffic parameters (VBR)
Link BW : 2Mb **Link Delay (T_{PROP}):** Variable {1-10ms} **Queues :** DropTail **Network Load :** Variable {0-40 %}	**Traffic Rate :** Variable {0.25Mb,...0.5Mb} **Interval time:** 10 ms **Packet size** = 500 bytes	**Rate :** Variable {0.25Mb...,0.5Mb} **Burstiness :** 0.5 **Time_dev and Rate_dev :** 1.0 **Packet size :** 500 bytes

Table 5.2: Test NS2 Network parameters

was modified to enhance certain features, such as providing background traffic (VBR, Variable Bit Rate) in scenarios with different network loads [CAL04]. We also tried out different protection methods described in Chapter 1.

The first topology is formed by 25 nodes distributed in a 2-D matrix form (shown in Figure 5.1). There are 4 pre-established working paths:

WP1	0-2-3-4-10-11-12-13-19-21
WP2	1-8-2-3-4-5-6-7-23
WP3	24-14-8-2-3-4-5-6-22
WP4	25-15-9-10-11-12-13-20

Traffic load is made up of 4 cbr (Constant Bit Rate) flows between nodes 0, 1, 24, 25 to nodes 21, 23, 22, 20 respectively, and vbr (Variable Bit Rate) background traffic. Background traffic has been introduced to simulate a more realistic scenario (by varying the vbr, different network loads are simulated). A more detailed list of the NS parameters used by the cbr and vbr flows is described in Table 5.2.

Failures are introduced in different segments of the network to simulate the influence of the distance between the node that detects the failure, and the node responsible for taking the switchover actions.

5.1.2. The KL-Net and NSF-Net topology

The KL-Net topology (see Fig. 5.2) is a typical Internet Service Provider (ISP) network topology. This network has been used in many recent papers, such as [KOD00] and [KAR00]. More recently, in [SUR01], this scenario is referenced to as the KL-graph. There are 15 nodes and 28 links. The capacity of the links are 12 and 48 (bolded lines) units, but they are scaled by 100 in order to experiment with thousands of LSPs. Each link is bi-directional (i.e. it acts like two unidirectional links of half of the capacity). There are four Ingress-Egress node pairs (1-13. 2-9, 4-2 and 5-15).

The NSF-Net topology is another typical ISP network topology (Figure 5.3) used by Rabbat [RAB03] to evaluate failure recovery time. In this topology the physical link lengths are known. This network is used in Section 5.5 to evaluate the propagation time (using the physical link lengths) and recovery times.

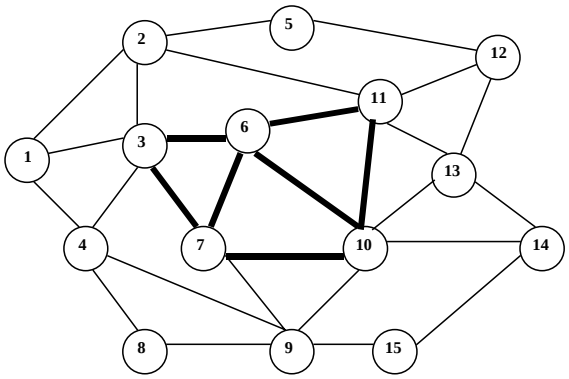


Figure 5.2: The KL-Net network topology

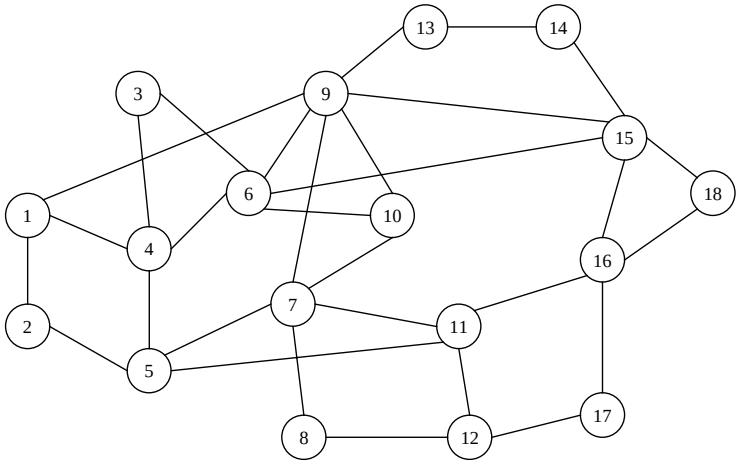


Figure 5.3: The NSF-Net topology

5.2. LSP recovery time and packet loss evaluation

These experiments were carried out using the NS-Mesh topology described in Section 5.1. In Table 5, different link propagation delays are analyzed evaluate their influence on P_{LS} and T_{REC_PL} (Formula 4.6) when the propagation time (including different link propagation delays T_{PROP} with fixed T_{PROC}, and T_Q) varies. In this case, global backup paths are used to protect the path. Results also reveal that the propagation time (link delay) is the most relevant parameter for both P_{LS} and T_{REC}, when the notification distances are large.

In Table 5.3, the influence of the notification distance for different traffic rates is shown. The objective of this analysis is to point out that the notification distance

T_{PROP} (ms)	**Failure Notification distance**							
	D(i,a) = 2		**D(i,a) = 3**		**D(i,a) = 4**		**D(i,a) = 0**	
	T_{REC}	*P_{LS}*	*T_{REC}*	*P_{LS}*	*T_{REC}*	*P_{LS}*	*T_{REC}*	*P_{LS}*
20	40.2	10	60.4	14	80.7	24	0.2	2
10	20.2	5	30.4	8	40.5	12	0.2	1
8	16.2	4	24.4	6	32.5	9	0.2	1
2	4.2	1	6.4	2	8.5	3	0.2	0-1

Table 5.3: Influence of failure notification distances and the link propagation time.

is also a crucial aspect when selecting the protection method in scenarios with different traffic rates. A distance equal to zero means that a local method is chosen; otherwise the global or the inverse method can be selected. Results reveal that the T_{REC} is directly proportional to the distance. Table 5.3 also shows how the different traffic rates influence Packet Loss (P_{LS}) (according to Formula 4.9).

The same traffic rate has been considered for all experiments. In this scenario,

for large notification distances (for instance D(i,a)=4) the propagation time between all links is revealed as the crucial aspect. For instance, when the link propagation time increases from 2ms to 20ms, the recovery time is almost 100% worse (see Table 5.3). The same occurs for packet loss. There are 24 packets lost for link delays of 20 ms compared to 3 packets for delays of 2 ms.

Note that when the traffic rate is high (for instance, 0.002 packets/sec, see Table 5.4), the notification distance may dramatically affect the recovery time (for instance, T_{REC} for D(i,a)=4 is twice as much as for D(i,a)=2). The same occurs with packet loss: there are 62 packets lost for D(i,a)=4 in comparison with 26 packets lost in the case of D(i,a)=2.

As has been shown by the above experiments, the reduction of the recovery time (and also of the packet losses) can only be achieved by reducing the distance

Traffic Rate (Packets/sec)	**Failure Notification distance**							
	D(i,a) = 2		**D(i,a) = 3**		**D(i,a) = 4**		**D(i,a) = 0**	
	T_{REC}	P_{LS}	T_{REC}	P_{LS}	T_{REC}	P_{LS}	T_{REC}	P_{LS}
0.02	20.2	2	30.4	3	40.5	6	0.2	1-0
0.01	20.2	5	30.4	8	40.5	12	0.2	1
0.008	20.2	6	30.4	9	40.5	15	0.2	1-2
0.004	20.2	13	30.4	18	40.5	30	0.2	2-3
0.002	20.2	26	30.4	36	40.5	62	0.2	5

Table 5.4: Influence of failure notification distances and traffic rate in packet loss and recovery time.

(D(i,a)). Other components, such as the propagation time, depend on the physical link technology and cannot be reduced.

The optimal case is the use of local backups (D(i,a) = 0). However, their main drawback is that the distance D(i,a) is not known in advance because the link which is going to fail is not yet known. Nevertheless, the use of link fault probabilities (as described in Chapter 3) can be used to estimate these distances.

The next section discusses the tradeoffs of reducing the fault probabilities, the failure impact and the network resources.

5.3. The backup decision module results

The following experiments calculate the expression for QoS of protection QoSP (Table 4.5) to search for the best method to apply according to the QoS requirements of the request. Different scenarios are considered, varying traffic classes (EF, AF1, AF2, BE), required bandwidth, number of segments to be protected in the working path and the distance to the first node of the protected segment. These experiments have been published in [MAR03a]. For simplicity, in a multiple-protected segment scenario, concatenated segments and a single distance measure are assumed. α, β, λ parameters are assigned according to Table 4.6.

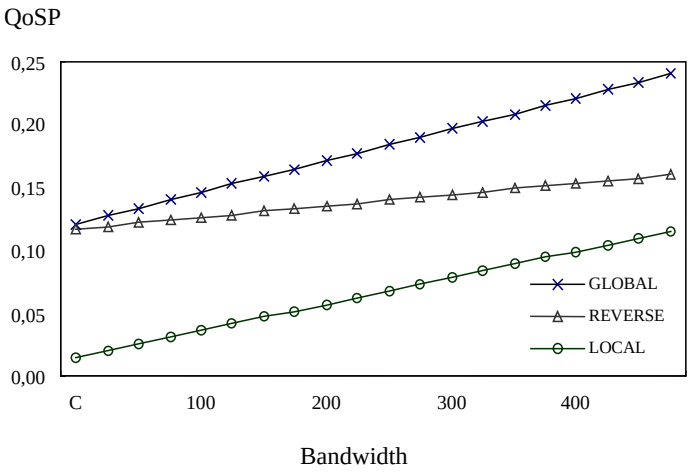


Figure 5.4: Backup decision module analysis. QoSP values and bandwidth requirements for the EF traffic.

In Figure 5.4 the analysis of QoSP and the bandwidth influence for the EF traffic class, with failure notification distance (D(i,a)) 2, are depicted. For this

experiment, the number of segments/links to Protect (NP) was fixed at 6.

Results, in Figure 5.4, show the QoSP values for different bandwidth requirements. For EF requests, the Backup Decision Module (BDM) gives priority to the local method, which ensures that the requirements for packet loss and recovery time will be reached. The second option is the reverse method, although the difference between the two methods increases with changes in the required bandwidth, since it affects packet loss. The greater the bandwidth request, the worse the packet loss in the case of a failure.

In Figure 5.5 the QoSP values for different failure notification distances for an EF traffic class service are analyzed. For this experiment, NP is 2 and the BW requests are constant. Results show the QoSP values for different distances

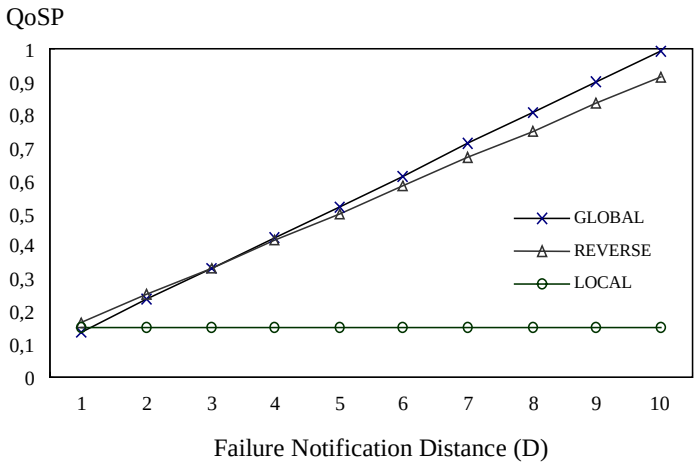


Figure 5.5: Backup decision module analysis. QoSP values and failure notification distances for EF traffic.

D(i,a). As expected, the BDM first selects the local backup method as the option that best suits the characteristics of EF traffic. More interesting is that the BDM second option varies according to the distance. For short distances, a global backup is suggested, with lesser resource consumption than the reverse method. For longer distances (2 or longer in Figure 5.5), a reverse backup is better. This is

because in the case of EF traffic, RT and PL are crucial in comparison with resource consumption.

In this last experiment the QoSP and failure notification distance influence is analyzed for AF2 traffic. Figure 5.6 shows the influence of the distance with a high number of segments/links to protect (NP=5). For shorter distances, the global method is chosen, providing a complete working path protection with values of P_{LS} and relatively adequate T_{REC}. However, for longer distances (D≥4) the local method (low P_{LS} and T_{REC}) is the method of choice. If the distance is greater than 5, we see that the second option of the BDM is the reverse backup and the global method becomes the worst choice.

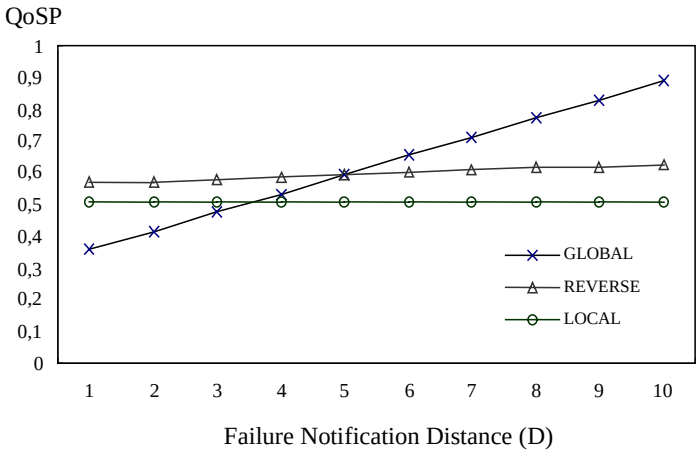


Figure 5.6: Backup decision module. QoSP and failure notification distances for AF2 traffic.

5.3 Network Failure Probability evaluation

This set of experiments, carried out using the KL-Net topology, focused on enhancing (reducing) the failure probability degree of some current QoS routing algorithms. Some of the experiments explained in this section have been published in [CAL03], [CAL03b], [CAL03c], [CAL03d] and [CAL04c].

Link	Failure Probability (10^{-4})
1-2	5
2-3	8
2-11	2
3-4	1
3-6	4
4-9	2
6-11	2
7-10	3
8-9	6
9-15	1
13-14	4

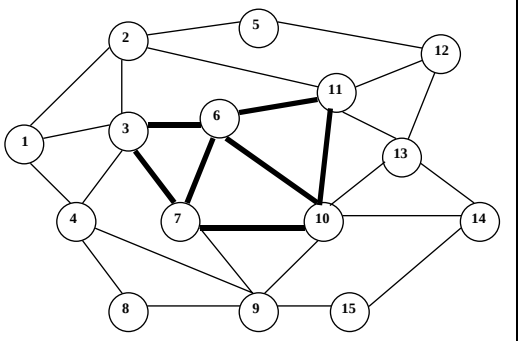


Table 5.5: Link Failure Probabilities (KL-Net topology)

This section describes two types of experiments. In the first set LSP requests arrived randomly, at the same average rate for all node pairs. All LSPs were long life (i.e. "static case"). For each experiment, 10 trials (with 3000 LSP demands)

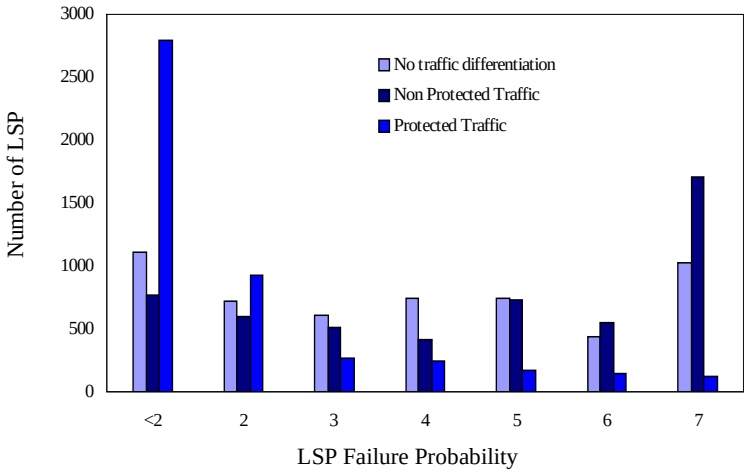


Figure 5.7: LSP Failure Probability. Long-lived LSPs (No protection)

were conducted. The bandwidth allocation for the LSPs was uniformly distributed between 1, 2 and 3 units. We have analyzed the well known Widest Shortest Path (WSP) routing algorithm behavior which has been compared with a modified WSP. There are two types of LSP requests, Protected Traffic requests and non Protected Traffic requests. Half of the requests were protected and the remaining 50 % were non protected. LSPs with less LS_FP probabilities were selected for the protected traffic.

In the 'dynamic case' simulation experiments, label switch path requests arrived randomly, at the same average rate for all ingress-egress node pairs. Label switched paths arrived between each ingress-egress pair according to a Poisson process with an average rate λ, and the holding times were exponentially distributed with a mean value of 1/μ. In this set of experiments, λ/μ = 150. 10 independent trials were calculated over a window of 10,000 *LSP* set-up requests.

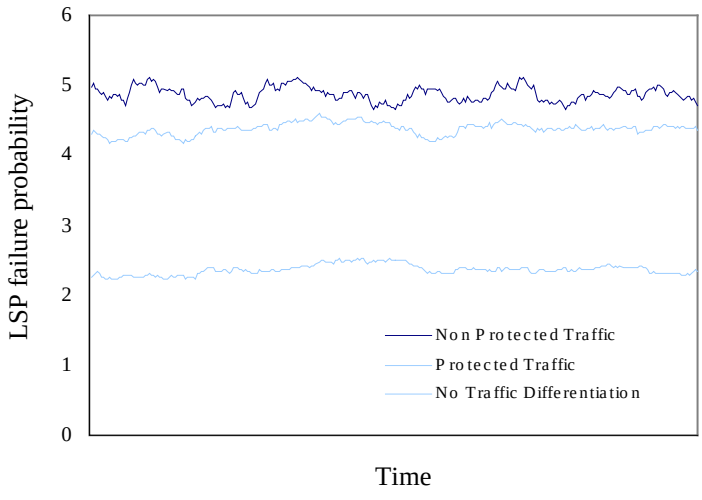


Figure 5.8: Label Switch Path Failure Probability evaluation (no protection.) Traffic differentiation versus no traffic differentiation.

5.3.1. LSP Failure Probability

In this set of results the minimization of the Label Switch Path Failure Probabilities (LSP_FP) is considered. No protection schemes are considered, so the residual failure probability is the same as the LSP failure probability.

In this experiment the number of LSP with a specific failure probability value was evaluated in a 'static' case simulation environment. Results in Figure 5.7 (published in [CAL03a]) show that the WSP (without traffic differentiation) distributes their LSPs along all probabilities without any pattern. On the other hand, the WSP modified to select those paths with minimum failure probabilities (in case of protected traffic) distributes their number of label switch paths following an approximated logarithmic pattern. As expected, for protected traffic, LSPs are accumulated to lower failure probability values.

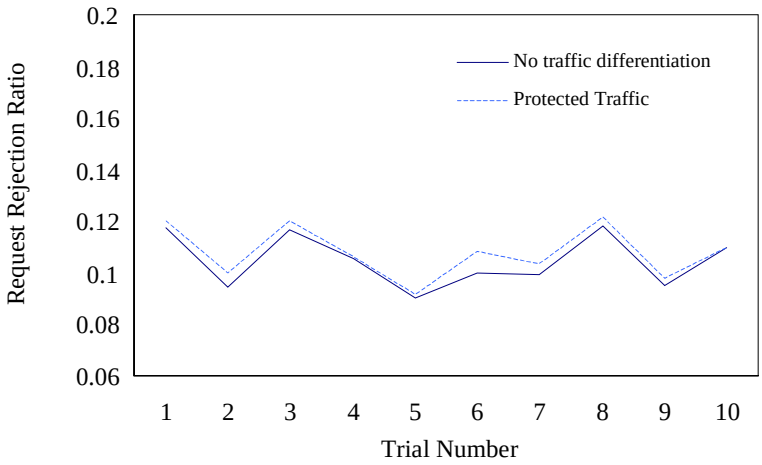


Figure 5.9: Request rejection ratio analysis. Traffic differentiation versus no traffic differentiation

Figure 5.8 shows a similar experiment, but in the 'dynamic' case scenario. In this case, the Label switch path failure probability (LSP_FP) average was calculated. Paths for protected traffic were also chosen with the minimum LSP_FP. Results

(published in [CAL04]) show that traffic with no protection accumulates large LSP_FP (i.e. $5 \cdot 10^{-4}$). On the other hand, protected traffic gets low LSP_FP values. If the minimization of the LSP_FP is not considered, there is an accumulation of large values for all traffic (about $4 \cdot 10^{-4}$).

Results in Figures 5.7 and 5.8 demonstrate that taking into account the LSP_FP and the traffic classes, the network protection degree can be improved. However, it is important to remark that these algorithms do not deteriorate the number of requests accepted. Figure 5.9 shows that all these algorithms keep a Request Rejection Ratio over 8-12%. This behavior is similar to that of executing a simple WSP. This experiment has been evaluated in the 'dynamic' case but similar results can be obtained in the 'static' case (see [DRCN2003]).

5.3.2. LSP residual failure probability

In this second set of experiments the residual failure probability was evaluated. In this case different fault recovery schemes were used to analyze the influence of the residual failure probability on the resource consumption. The KL-Net topology and the link failure probabilities (see Table 5.5) were used to deploy these experiments. In Figures 5.10 and 5.11 each point in the charts represents the network residual failure probability. This value is computed every 100 new LSP requests as the accumulation of all current LSP residual failure probability values. Results in Figures 5.10 and 5.11 show similar behavior. The Network Residual Failure probabilities for the protected traffic are accumulated close to zero, while the non protected traffic values are more dispersed across higher network failure probabilities.

The residual failure probabilities can be reduced by using Local backups or Segment backups. The failure impact and resource consumption are affected in different ways. In Figure 5.12 the resource consumption for the local and segment backups experiment is depicted. Using segment backups with different notification distances (D= 0, 1 and 2) distributes the protection of the network. In this experiment, major backups were local backups and there were a few

backups with large notification distances. However, these LSPs (with D=2) can result in a high number of packet losses if a failure occurs in them.

Results show that there is a strong relationship between the residual failure probabilities and resource consumption. Local or segment backups can be used in order to reduce the residual failure probability. Local backups avoid notification distances while providing better protection in terms of packet loss and recovery time, but there is higher resource consumption than in segment backup protection. There is also a trade-off between the number of the used resources, the selected protection mechanism (local, segment or global backups), and the residual failure probability and failure impact.

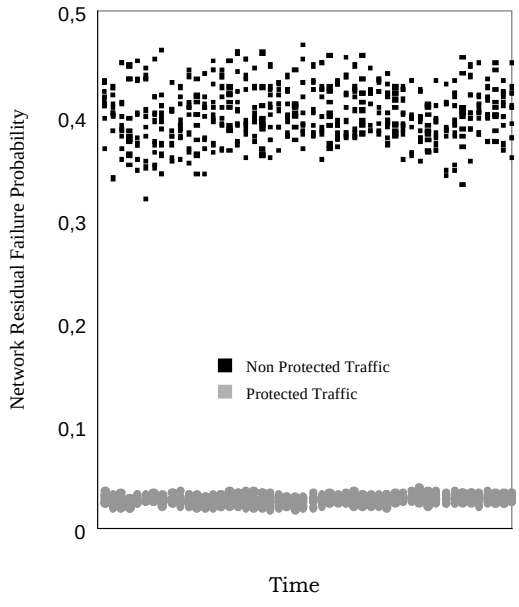


Figure 5.10: Residual Failure Probability evaluation. Segment Backups and traffic differentiation.

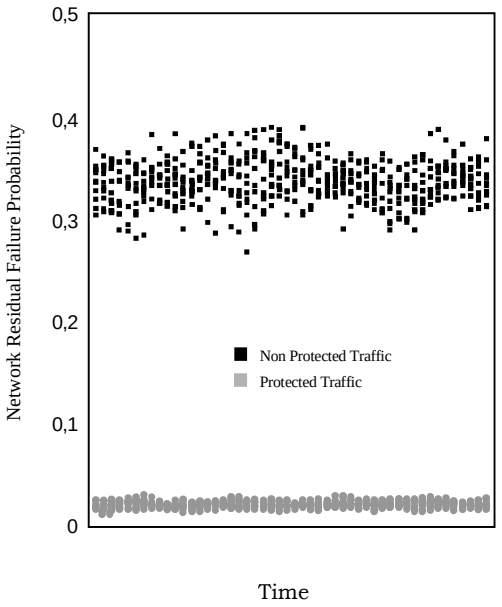


Figure 5.11: Residual Failure Probability evaluation. Local Backups and traffic differentiation.

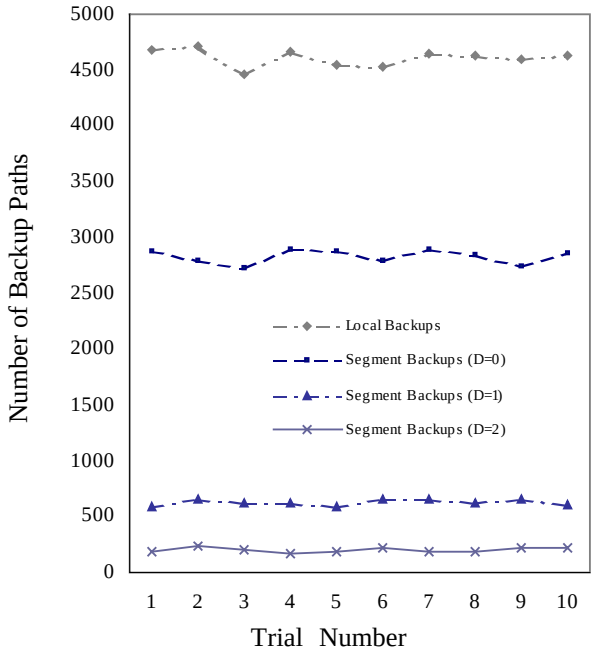


Figure 5.12: Backup Resource Consumption

5.4 Reducing the failure probability and the failure impact

The influence of the fault notification method and the failure impact in terms of recovery time is evaluated in this set of experiments. The scalability of the fault notification scheme has also been also analyzed in terms of the number of packets (failure notification messages). Both signaling-based and flooding-based schemes have been used to make the comparison. The NSF-Net is used in this case, with the physical link length assigned in Table 5.5. The links are bi-directional links of 12 units (1,2 Gb/s) of bandwidth. There are 6 links with LFP > 0, making up 20 percent of the network. The failures occur at link 11-12.

In the simulation experiments LSP requests arrive randomly, at the same average rate for all ingress-egress node pairs. LSPs arrive between each ingress-egress pair according to a Poisson process with an average rate λ, and the holding times

are exponentially distributed with mean $1/\mu$. In this set of experiments, λ/μ = 150. Ten independent trials are calculated over a window of 10,000 LSP set-up requests.

In these experiments the failure recovery time formulated in Chapter 4 is reviewed. Two failure notification schemes are compared. Signaling-based schemes and flooding-based schemes are considered in this experimental analysis.

Two backup models are also compared: one backup path model, the global backup path model; and one segment backup path model. The routing algorithm is the modified k-Widest Shortest Path ([CAL03a] and [CAL03]). The WSP, in the case of segment backup paths, creates and protects those segments (consecutive or adjacent links) with a certain Link Failure Probability.

The compared parameters are the failure notification time, the backup path activation time, the number of protected LSPs and the number of failure recovery messages.

The failure notification time (T_{NOT}) is the time required to notify the failure from the node which detects the failure to the node which is responsible for the switchover and is computed using Formula 4.15. D_{NOT} is the distance between D(i,a) for the global backup path and D(c,a) in the case of segment backup paths (according to the formulation provided in Section 4.3).

The backup activation time (T_{BA}) is the time required by the PSL node to start the switchover after the failure has been notified. This time (T_{BA}) is computed using Formula 4.15), along the path between the PSL node and the PML node.

Formula 4.15 is only used in the case of signaling backup path techniques. In this case, the backup activation distance is proportional to the D(i,e) in the case of global backup paths and D(c,b) in the case of segment backup paths.

Link	Length (Km)
1-2	400
1-4	500
1-7	1000
2-5	600
3-4	300
3-6	500
4-5	700
4-6	400
5-8	800
5-11	1200
6-7	200
6-10	300
6-15	1300
7-8	800
7-10	300
7-13	200
7-15	900

Link	Length (Km)
8-9	300
8-10	600
8-11	700
9-12	500
11-12	200
11-16	600
12-17	700
13-14	400
14-15	200
15-16	500
15-18	300
16-17	700
16-18	400

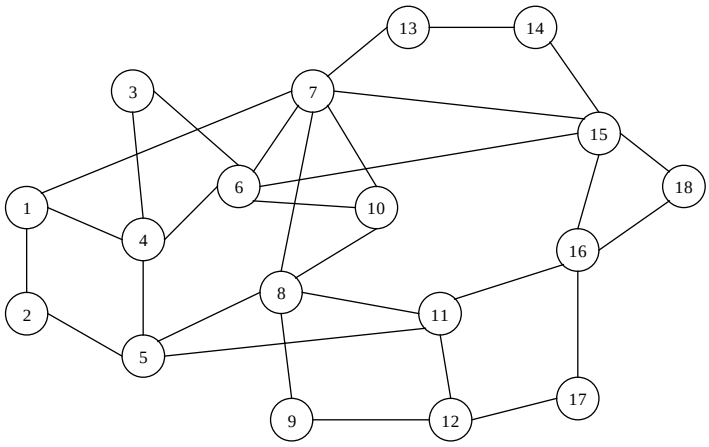


Table 5.6: Physical Link Length (NSF-NET topology)

In the case of flooding techniques, T_{BA} is computed considering the maximum distance between all the nodes of the backup path. The time required to notify the last node is an upper bound value, which is used to compute the backup activation time with flooding-based algorithms.

The number of messages required to complete the failure notification and backup activation is also addressed in this analysis. Formula 4.14 is used to evaluate the number of messages generated by the signaling-based method. In the case of flooding, only the number of messages (proportional to the number of nodes in the network) is considered.

Finally, the number of protected LSPs in the broken link, at the moment when the failure occurs, is evaluated. The number of LSPs is crucial because, in the case of signaling-based methods, this number concerns the scalability of the method, not only for the computation time, but also for the number of messages.

Case 1: Failure notification and backup activation time when the node processing time is proportional to the number of messages.

In Figure 5.13 the failure notification time and the backup activation time are

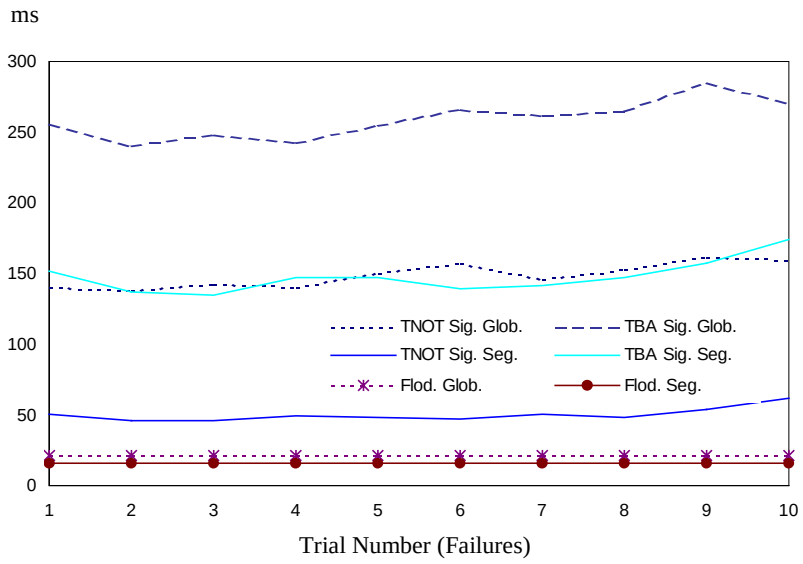


Figure 5.13: Failure notification time (T_{NOT}) and backup activation (T_{BA}) time analysis, when the node processing time (T_{PROC}) is proportional to the number of messages.

presented. In this case the propagation time is very large due to the link distances (km, see Table 5.6). But the influence of the node processing time (0,3 ms for each RSVP packet) is the major factor in the results. In this case nodes process each packet individually and sequentially.

Results in Figure 5.13 show that the signaling-based schemes are always worse, in terms of delay, than the flooding-based schemes in terms of T_{NOT} and T_{BA}. This is because the number of LSPs to be protected in the broken link involves a high number of messages, which are sequentially processed in each node, accumulating large delays on each hop. Although the propagation time is significant (about ms) in this experiment, it is negligible in relation to the node processing time in the case of signaling-based techniques. The transmission time

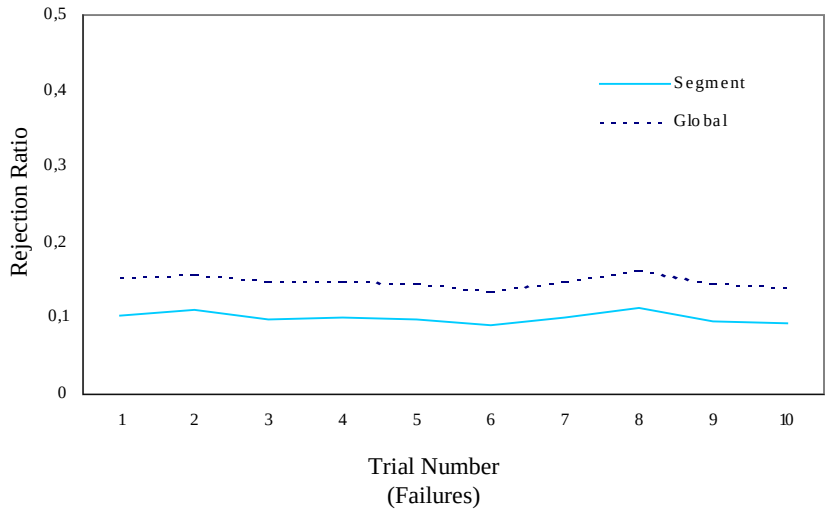


Figure 5.14: Rejection Ratio (Sig. Segment Vs Sig. Glob. Backups)

is also very small (the magnitude of the link capacity is Gb/s) and the queuing time is not considered, because high priority queues are assumed for the failure notification packets. Therefore, in this case the major difference between signaling-based and flooding-based schemes is the number of paths (LSPs) to be protected and the number of fault notification and backup activation messages.

Each point in the figure represents the same link failure in different times (the same single link failure each 1000 LSPs requests is generated). The Figure 5.15 is arranged according to the messages generated on each trial. This allows us to see what the effect is when the number of messages increases. Note that, regardless of whether the notification time (and the backup activation time) increases sharply, it is not very uniform. For instance, in trial 5 (see Figure 5.13) there is a higher increment (in terms of T_{BA}) than in 6. The main reason is that the number of LSPs to be protected and the number of messages do not increase proportionally (see Figure 5.15 and Figure 5.16).

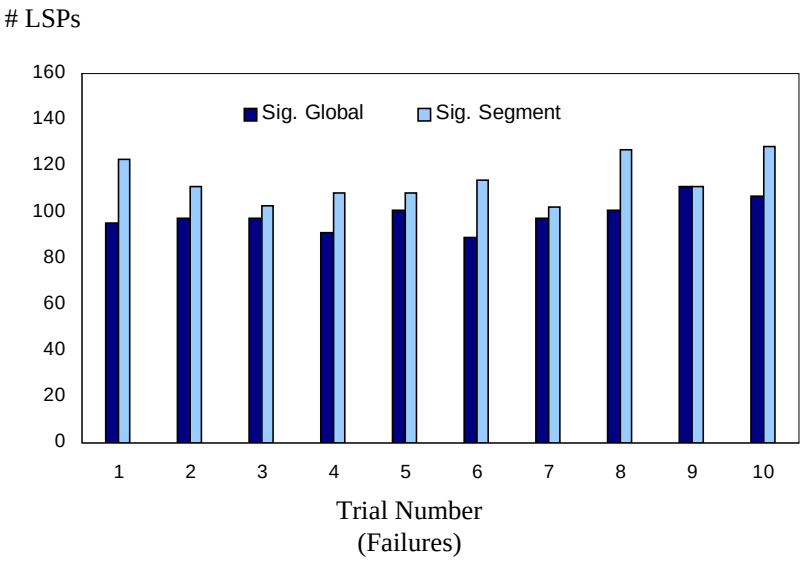


Figure 5.15: Number of protected LSPs

The algorithm used (the modified k_WSP) selects the path depending on the residual capacity and the number of hops. Consequently, the protected LSPs have neither the same number of hops (D_{NOT}) nor the same inverse path (paths to the PSL nodes), and therefore in some cases these factors can increase or decrease the final failure notification time. This results from the difficulty of calculating the recovery time precisely in signaling-based techniques.

On the other hand, the segment backup model in the case of signaling-based techniques significantly improves the global backup path performance. However, when the network is protected using segment backups, there are more free resources (see Figure 5.14), hence more LSPs can be created and protected, as explained in [CAL03]. When using segment backup paths there are more

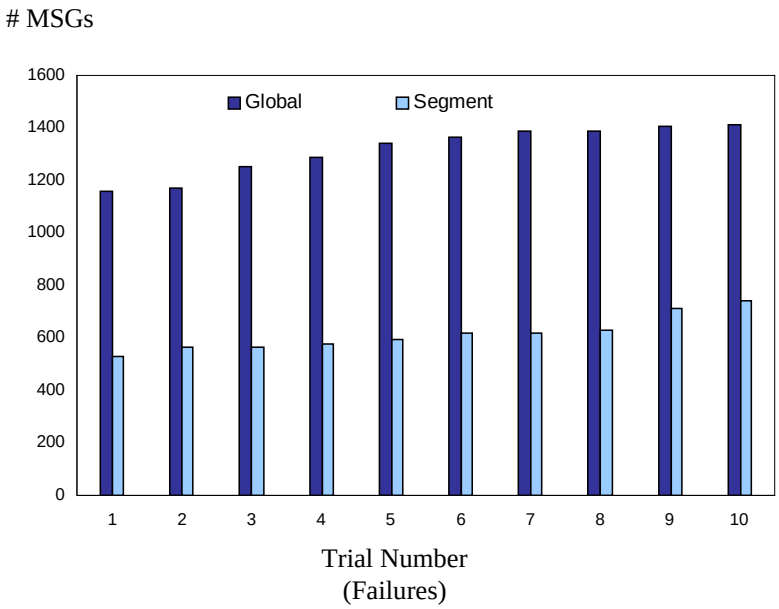


Figure 5.16: Number of messages.

established LSPs (Figure 5.8), and consequently, in the case of signaling-based techniques the failure notification time can increase if the node processing time for each packet is significant.

The conclusion is that by reducing the number of hops in the notification path, the failure notification time can also be reduced in both signaling and flooding techniques.

In Figure 5.16 the number of messages generated on each trial to notify and activate the backup paths is shown. Results show that the number of messages for notifying and activating the segment backup paths is larger than the number for the global backup paths. However, as explained above, it is difficult to

compute the final number of evaluated messages (Rabbat also states this in [RAB03]). The reason is the number of protected LSPs and the number of hops are not the only parameters used to evaluate this value. The location of these paths has to be considered in this computation. Obviously, this location cannot be known a priori, because it depends on the routing method used. Each node can generate a different number of messages (as opposed to the maximum number of messages proposed in [RAB03]), avoiding a pre-evaluation of this number. For the flooding methods the number of messages is not represented, because it is a fixed number proportional to the number of nodes in the network (18 in the case of NSF-NET).

Case 2: Time to start the switchover when the nodes have a fix node processing time.

In the next set of experiments a fixed node processing time is assumed. In this case nodes are able to compute hundreds of messages with a fixed time of 3 ms. Again the flooding messages are transparent to the nodes (no processing time is considered for these packets). In Figure 5.17, both the notification time T_{NOT} and

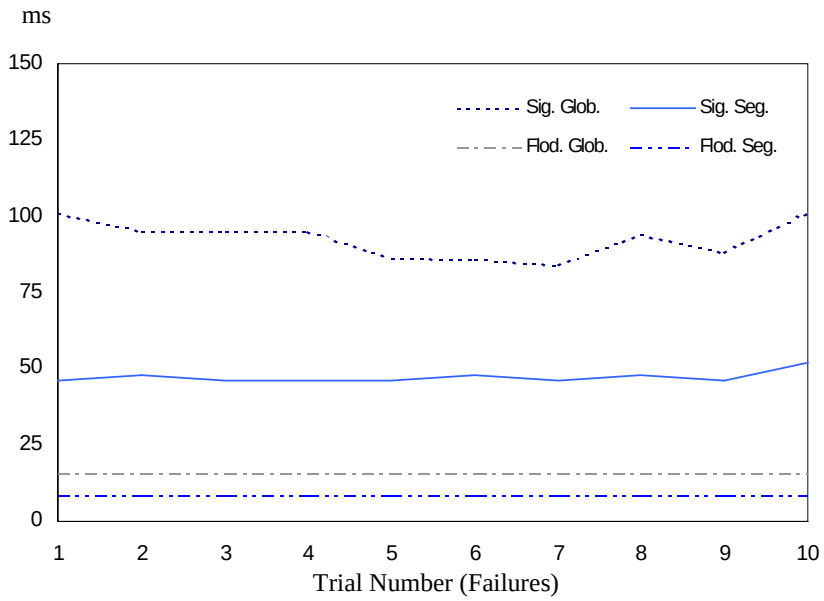


Figure 5.17: Time to start the switchover with a fixed node processing time (T_{PROC} = 3 ms)

the time to activate the backup T_{BA} are included, according to Formula 4.7. This time represents the time required by the PSL node to start the switchover operation.

Signaling methods spend some time activating the backup paths, after the PSL node is notified of the failure, due to the implementation of the RSVP protocol. As has been explained in Chapter 1, signaling methods involve sending a message from the PSL to the PML, activating the backup path and an acknowledgement message from the PML to the PSL, verifying this activation.

In this case, there is again a large difference between signaling-based and flooding-based techniques. However, the number of LSPs to be protected does not affect the final result. As expected, segment backup techniques (in signaling-based methods) are better than in the previous experiment. Here the failure notification time and the backup activation time perform up to a 50% better in segment than in global backups. The main factor in reducing the failure recovery time is the distance between the number of hops. The node processing time (T_{PROC}) is the most costly factor (in terms of time). However, in this case a more predictable value of the final recovery time can be evaluated. If this time (T_{PROC}) is reduced (as node technologies improve day by day), the only factor to be considered would be the propagation time (T_{PROP}).

Case 3: Reducing the node processing time.

In this experiment a very small fixed processing time is considered to project the behavior in future network scenarios. In this case the crucial aspect is the propagation time (T_{PROP}) which cannot be reduced because it is directly proportional to the physical length of the links and the light speed in a fiber.

Figure 5.18 shows the amount of time necessary to start the switchover using the signaling-based and flooding-based techniques. In this experiment the segment backup methods achieve very low recovery time (less than 30 ms).

Results show that the distance (physical link length) is dominant in this

experiment. Consequently, the propagation time is the crucial factor to be taken into account. This factor cannot be reduced by limiting the number of hops (nodes); the only way is to reduce the physical distance between the node detecting the failure and the PSL node. The physical distance between the PSL node and the PML node has also to be considered if the time to start the switchover, and consequently, the total delay related with the failure occurrence, is a critical factor.

In summary, the main factors affecting the recovery time, independently of whether this time is the time to notify of the failure or the time to activate the backup path, are the delays associated with the nodes and links. When the failure notification packets are prioritized and the link capacity is very large (GBs), the node processing time and the propagation time are crucial.

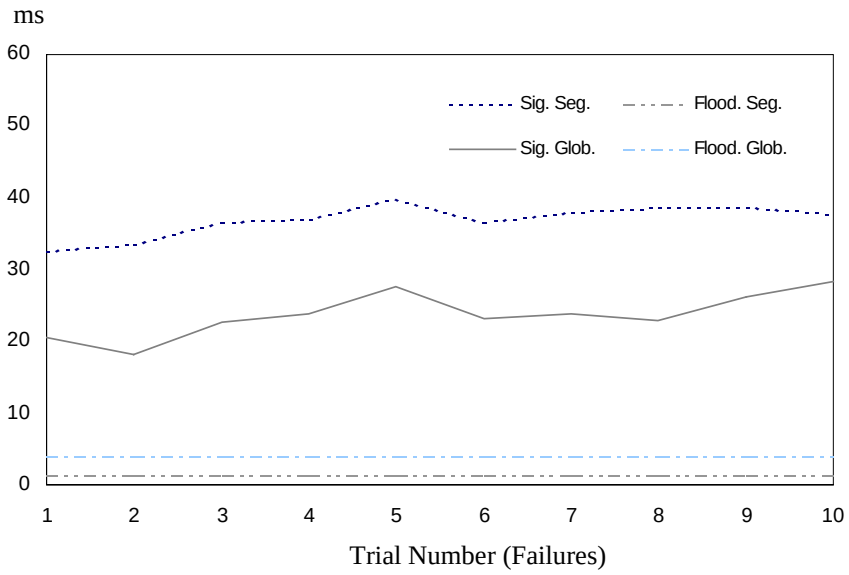


Figure 5.18: Time to start the switchover reducing te node processing time (T_{PROC} = 0.3 ms)

It is expected that the node processing time will tend to be reduced in future optical generation networks. However, the propagation time cannot be reduced because it is proportional to the physical link length. In networks with large links (hundreds of km), this will be the most crucial aspect. In order to reduce the

impact of a failure, the distance (physical distance) between the nodes responsible of the failure recovery has to be reduced. The paths should be selected according to the minimum physical length (shortest paths) in order to reduce the failure impact.

Segment backup paths can be considered to offer more effective protection. However, the application of segment/local protection can only be deployed if the segments to be protected are known a priori. This involves evaluating the probability of failure. Results have shown that combining segment protection and failure probability models not only improves the recovery time, but also improves the resource consumption, allowing the establishment of more protected working paths instead of using global backup paths.

CHAPTER 6

6
Conclusions and future work

6.1. Introduction

The objectives set out for this thesis have been achieved and a novel and in-depth study of the network protection level has been realized. Several different and new methods to evaluate and enhance the protection of current and future MPLS and GMPLS-based networks have been presented and tested.

This chapter summarizes the objectives achieved and the major conclusions of this work. It also exposes many other research issues. Some of them will be considered for future work.

6.2. Summary and Conclusions

This thesis was aimed at developing fault recovery methods capable of providing reliable and fast recovery from network component failure in a GMPLS/MPLS-based network for traffic services with high protection requirements. The objectives have been achieved. The main contributions are summarized in the following paragraphs.

MPLS fault recovery method characterization and comparison. In the first phase of this work the main fault recovery methods in MPLS-based networks have been analyzed and compared. A first set of conclusions has been drawn, comparing these methods individually in very simple scenarios. Fault recovery time, packet loss, packet reordering and resource consumption have been identified as the principal parameters with which to analyze the performance of these methods.

Extension of MPLS fault recovery methods to optical networks using. Despite being initially focused on MPLS-based mechanisms, the extension of these methods to optical networks using Generalized MPLS (GMPLS) has been considered. Some obstacles to extending MPLS fault recovery methods to optical networks have been pointed out in and considered throughout this work. Current CCAMP (IETF) efforts to standardize GMPLS and the main issues related to those efforts have been considered in this thesis proposal.

QoS On-Line Routing and MPLS Protection review. A review of the main QoS routing algorithms and some of the current MPLS routing proposals have been presented in Chapter 2. The main drawbacks of these proposals, related to improving network protection, have been also discussed. This study has been published in the prestigious journal 'IEEE Communication Magazine', pointing out the main aspects of enhancing the protection level provided by the current major routing algorithms.

A novel network reliability model. One aspect not usually considered in the network design is a previous analysis of the network reliability. This study is

crucial to optimize the network design, offering more suitable protection depending of the failure probabilities. A proposal to evaluate the network reliability level has been presented here. A novel model for link failure probability evaluation has been proposed in Chapter 3. Path failure probability and residual failure probability have been formalized and considered to design new routing objectives.

Definition of the impact of a failure. All the steps to restore the traffic from the moment a failure occurs to when the traffic is restored through a backup path have been analyzed in depth. A detailed explanation of the fault recovery process has been presented in Chapter 4. Each component to reduce the time requiered to complete each step in the recovery process has been also presented.

Formulation of the time to recover a failure. Three different times have been identified as crucial in the fault recovery process. The time affecting the packet loss, the time required to start the switchover and the complete traffic delay. Each time has been analyzed and formulated considering different protection methods. This study has identified the most suitable protection strategies to reduce the impact of a failure in different network scenarios.

A review and comparison of the main failure notification techniques. The IETF-CCAMP has considered the reduction of fault recovery time crucial. Signaling-based and flooding-based techniques have been chosen as the two main failure notification techniques. Our proposal has been extended to take into account these techniques. Results (presented in Chapter 5) have concluded that flooding-based techniques are better than signaling techniques in many aspects such as the time to notify the PSL nodes and activate the backup path. Another important advantage of the flooding-based techniques is the capability to pre-evaluate the delays with a high level of accuracy. On the other hand, signaling-based techniques are used currently. In networks with low numbers of LSPs and low notification distances, they can be used to notify each LSP.

Segment and local protection over path protection techniques. Independently of the failure notification techniques, results have shown that segment and local

backup techniques improve the failure recovery times and minimize the impact of a failure. In this work a combination of the failure probability evaluation and the utilization of segment and local backup paths have enhanced path protection techniques. The failure impact (delay and packet loss) has been reduced and the network resource consumption optimized.

Reducing the recovery time in different network scenarios. Failure indication signals and backup activation messages must be sent as fast as possible in order to reduce the fault recovery time. Node and link delay has been taken into account in this analysis. Nodes introduce a delay proportional to the node processing time and the buffering time. Links add a delay proportional to the propagation time and the transmission time.

In order to consider the behavior of different network scenarios, large, medium and small node processing times have been analyzed. In this study priority failure notification packets have been considered. If nodes process each packet in a fixed time, or if there is a large fixed node processing time, the distance, in number of hops (nodes), has been considered as the aspect to be reduced to achieve the minimum failure recovery time objective. If the node processing time is very small, the propagation time, and consequently, the link length, have been identified as the major components to be reduced. Selecting paths with minimum length, the failure impact is reduced.

Evaluation of the protection design components in future network. In future networks the node processing time is expected to be reduced and the link bandwidth extended, reducing the transmission time. Also expected is the use of priority queues for the failure notification packets, reducing the buffering time. Nevertheless, in large ISP networks, with large links (hundreds of km), the propagation time is the only aspect that cannot be reduced. In this case segment protection is chosen as the most suitable protection technique.

A novel definition of protected traffic services. Our proposal improves the network reliability and reduces the failure impact by applying different fault recovery schemes. However, not all the current and future traffic services will require the

same level of protection. Moreover, in some cases improving the protection level involves costly fault recovery methods (in terms of resource consumption) that cannot be deployed in the whole network. In this work we have characterized the traffic protection requirements creating different traffic service categories. Results have shown a high level of protection with rational resource consumption using protection-differentiated services.

6.3. Future work

The main issues for future work are presented in this section.

In Chapter 3 a model to calculate the link (or component) failure probability has been introduced. However, this model is based on different standards. An in-depth analysis of these values for the current network technology should be considered for future work. This model is also based on tuned and statistical values and offers only approximate failure probabilities. Other models should also be considered for this evaluation.

This work starts with the MPLS-based technologies. GMPLS control planes have been considered to extend MPLS to optical networks. However, an in-depth analysis of current and future optical network technology should be considered for future work. Some aspects, such as the number of lambdas, have not been taken into account in this thesis. Optical node architectures are another crucial aspect for calculating the fault recovery time, as explained in Chapters 4 and 5. A more detailed study of these technologies should also be included.

This thesis has been mainly focused on dedicated bandwidth allocation protection methods. Shared backup schemes should also be considered to extend this work. Better resource consumption can be achieved by extending some of the proposed methods with shared backup paths.

Other emerging areas in network protection, such as p-cycles and shared risk

groups (nodes and links), have not been considered in this work. However, many of the proposed schemes and the proposed network protection level evaluation can be easily applied.

Enhancing some current QoS routing algorithms to offer better network protection has been one of the main objectives of this thesis. Adding new protection objectives to the current routing normally involves the utilization of network optimization models. Further work in the analysis of current network optimization models and the application of these models to the proposed schemes is considered for future work.

References

[AND01] L. Andersson, P. Doolan, N. Feldman, A. Fredette, and B. Thomas. ”LDP specification”. IETF RFC 3036, January 2001.

[APO99] G. Apostolopoulos, D. Williams, S. Kamat, R. Guerin, A. Orda, and T. Przygienda “QoS Routing Mechanisms and OSPF Extensions”. IETF RFC 2676, August 1999.

[ASH02] J. Ash, Y. Lee, P. Ashwood-Smith, B. Jamoussi, D. Fedyk, D. Skalecki,and L. Li.”LSP Modification Using CR-LDP”. IETF RFC 3214, January 2002.

[AUT02] A. Autenrieth, A. Kirstädter, “Engineering End-to-End IP Resilience Using Resilience-Differentiated QoS”, IEEE Communications Magazine, January 2002.

[AUT02a] A. Autenrieth, A. Kirstädter: “RD-QoS - The Integrated Provisioning of Resilience and QoS in MPLS-Based Networks.” IEEE International Conference on Communications (ICC 2002), New York, USA, April 28 - May 02, 2002

[AWD01] Daniel O. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow. ”RSVP-TE: Extensions to RSVP for LSP Tunnels”. IETF RFC 3209, December 2001.

[BER03] Berger, L. (Editor) et al., "Generalized MPLS Signaling - RSVP-TE Extensions," IETF RFC 3473, January 2003.

[BLA00] D. Black. “Differentiated Services and Tunnels”. IETF RFC 2983, October 2000.

[BLA98] S. Blake, D. Black, , M. Carlson, E. Davies, Z. Wang, and W. Weiss. “An Architecture for Differenciated Services”. IETF RFC 2475, December 1998.

[BRA97] R. Braden, L. Zhang, S. Berson, S. Herzog, and S.Jamin.”Resource ReSerVation Protocol (RSVP)”. IETF RFC 2205, September 1997.

[CAL01] Eusebi Calle, Teo Jové, Pere Vilà, Josep L Marzo "A Dynamic Multilevel MPLS Protection Domain" In Proceedings of DRCN 2001, Budapest (Hungary), 7-10 October, 2001. Edited by Tibor Cinkler

[CAL02] Eusebi Calle, Pere Vilà, Jose L. Marzo, Santiago Cots "Arquitectura del sistema de gestión de ancho de banda y protección para entornos de redes MPLS (SGBP)" Symposium on Informatics and Telecommunications, SIT 2002. September 25-27, 2002. Sevilla, Spain

[CAL03] Eusebi Calle, Jose L Marzo, Anna Urra, Pere Vila "Enhancing MPLS QoS routing algorithms by using the Network Protection Degree paradigm." In proceedings of GLOBECOM 2003, San Francisco (USA)

[CAL03a] Eusebi Calle, Jose L Marzo, Anna Urra "Evaluating the probability and the impact of a failure in GMPLS based networks" In proceedings of DRCN 2003, Alberta (Canada) IEEE 2003

[CAL03b] Eusebi Calle, J L Marzo, Anna Urra "Protection performance components in MPLS networks." In proceedings of SPECTS 2003, Montreal (Canada)

[CAL03c] Anna Urra, Eusebi Calle, Jose L Marzo "Sistema multi-agent per al control de la protecció en xarxes GMPLS" Accepted in CCIA 2003, Palma (Spain)

[CAL04] Eusebi Calle, Jose L Marzo, Anna Urra "Protection Performance Components in MPLS Networks" Accepted in Computer Communications Journal, Elsevier 2004

[CAL03d] Eusebi Calle, Anna Urra, Jose L Marzo "Multiagent system for controlling GMPLS network protection" Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications. IOS Press 2003.

[CAL04b] Eusebi Calle, Anna Urra, Jose L. Marzo " Network Reliability and Failure Recovery Time Evaluation in GMPLS-Based Networks using Segment and Path Protection". (Submitted to) SPECTS 2004. San Jose. USA.

[CAL04c] E. Calle, J.L. Marzo, A. Urra, LL. Fàbrega "Enhancing fault management performance of two-step QoS routing algorithms in GMPLS" (to be presented) ICC 2004, 20-24 June, 2004, Paris (France).

[CHE99] Thomas M. Chen and Tae H. Oh. "Reliable Services in MPLS", IEEE Communication Megazine, Volume: 37 Issue: 12, pages 58 –62, December 1999.

[DAV00] B. Davie and Y. Rekhter. "MPLS Technology and Applcations". Morgan kaufmann publisher Inc. ISBN 1-55860-656-4, May 2000.

[DHA03] Sudheer Dharanikota, Raj Jain, Protection and restoration in DWDM networks: Recent developments and Issues, Invited Paper to SPIE Conference.

[ELW01] A. Elwalid, C. Jin, S. Low, and I. Widjaja. "MATE: MPLS Adaptive Traffic Engineering". IEEE INFOCOM 2001, pages 1300 –1309 vol.3,April 2001.

[FAC02] F. Le Facheur et al. "Requirements for support of Diff-Serv-aware MPLS traffic engineering". IETF RFC 3270 May 2002.

[FAL] K. Fall and K. Varadhan. "The network simulator ns-2" . The VINT project. UC Berkeley, LBL, USC/ISI, and Xerox PARC, http://www.isi.edu/nsnam/ns/.

[FALa] K. Fall and K. Varadhan. “The ns Manual”. The VINT project. UC Berkeley, LBL, USC/ISI, and Xerox PARC, http://www.isi.edu/nsnam/ns/nsdocumentation.html.

[GAE00] A. Gaeil and C. Woojik.”Design and Implementation of MPLS Network Simulator (MNS) supporting LDP and CR-LDP”. Proceedings of the IEEE International Conference on Networks (ICON’00), September 2000.

[GAE02] A. Gaeil, J. Jang, and C. Woojik. “An Effcient Rerouting Scheme for MPLS-based Recovery and its Performance Evaluation”. Telecommunication Systems, vol.9, pages 441 –446, March-April 2002.

[GAE99] A. Gaeil and C. Woojik.”Overview of MPLS Network Simulator:Design and implementation” Technical report, Department of Computer Engineerig, Chungnam National University, Korea, December 1999.

[GUE97] R. Guerin, D. Williams, A. Orda “QoS Routing Mechanisms and OSPF Extensions”, Proceedings of Globecom. November 1997.

[HUA02] Changcheng Huang, Vishal Sharma, Ken Owens, Srinivas Makam, “Building reliable MPLS Networks using a path protection mechanism”, IEEE Communications Magazine, March 2002

[HUN02] L. Hundessa, J. Domingo-Pascual, “Reliable and Fast Rerouting Mechanisms for a Protected Label Switched Path”, Proceedings of Globecom, 2002

[ITU800] ITU-T E.800 “Terms and definitions related to quality of service and network performance including dependability”

[KAR00] K. Kar, M. Kodialam and T. V. Lakshman, "Minimum interference routing with applications to MPLS traffic engineering" Proceedings of IEEE Infocom 2000, March 2000.

[KOD00] M. Kodialam, T.V. Laksman, “Dynamic Routing of Bandwidth Guaranteed Tunnels with Restoration” Proceedings of IEEE Infocom 2000, March 2000.

[KOD02] M. Kodialam, T.V. Lakshman, “Restorable Dynamic QoS routing”, IEEE Communications Magazine, June 2002

[LAN03] J. P. Lang, B. Rajagopalan, “Generalized MPLS Recovery Functional Specification”. Internet Draft. Work in progress. May 2003

[LAN03a] Lang, J., et al (eds.), "RSVP-TE Extensions in support of End-to-End GMPLS-based Recovery", work in progress, September 2003.

[LI01] Li, G., J. Yates, et al, “Experiments in Fast Restoration using GMPLS in Optica/Electrònic Mesh Networks”, OFC 2001, Anaheim, CA, March 2001.

[MA97] Q. Ma and P. Steenkiste “On Path Selection for Traffic with Bandwidth Guarantees”. Proceedings of IEEE International Conference of Network Protocols. October 1997

[MAN03] Mannie, E. And D. Papadimitriou, "Recovery (Protection and Restoration) Terminology for GMPLS", Internet Draft, (work in progress), February 2003.

[MAR03] J. L. Marzo, E. Calle, C. Scoglio, T. Anjali, “QoS On-Line Routing and MPLS Multilevel Protection: a Survey”. IEEE Communications Magazine, October 2003

[MAR03a] J. L. Marzo, E. Calle, C. Scoglio, T. Anjali "Adding QoS Protection in Order to Enhance MPLS QoS Routing" In proceedings of ICC 2003. Anchorage, Alaska (USA). IEEE 2003, ISBN 0-7803-7804-4

[MAS03] X. Masip, S. Sànchez, J. Solé, and J. Domingo. “QoS routing Algorithm under Inaccurate Routing Information for Bandwidth Constrained Applications” In proceedings of ICC ’03.

[MOY98] Moy, J., "OSPF Version 2", IETF RFC 2328, April 1998.

[MIL217] MIL-HDBK-217, the Military Handbook for "Reliability Prediction of Electronic Equipment". Published by the Department of Defense, based on work done by the Reliability Analysis Center and Rome Laboratory at Griffiss AFB, NY.

[NS200] The Network Simulator – ns-2, http://www.isi.edu/nsnam/ns

[OGG01] Chris Oggerino “High Availability Network Fundamentals” Published by Ciscopress 2001.

[ORA90] D. Oran. OSI “IS-IS Intra-domain Routing Protocol.” RFC 1142, February 1990.

[PAP03] D. Papadimitriou, E. Mannie, “Analysis of Generalized MPLS-based Recovery Mechanisms (including Protection and Restoration)”. Internet Draft. Work in progress. Setembre 2003

[RAB03] R. Rabbat, V. Sharma, N. Shinomiya, C. Su, P. Czezozowski “Fault Notification Protocol for GMPLS-Based Recovery. Internet Draft (work in progress). February 2003.

[RAB03a] R. Rabbat and T. Soumiya "Extensions to LMP for Flooding-based Fault Notification," Internet Draft (work in progress), June 2003.

[RAB03b] Rabbat, R. and T. Soumiya, "Optical network failure recovery requirements", Internet Draft, (work in progress), June 2003.

[RAB03c] R. Rabbat, V. Sharma, A. Ali “Expedited Flooding for Restoration in Shared-Mesh Transport Networks”, Internet Draft, (work in progress), October 2003.

[SHA03] V. Sharma, B. M. Crane, S. Makam, K. Owens,C. Huang, F. Hellstrand, J. Weil, L. Andersson, B. Jamoussi, B. Cain, S. Civanlar, A. Chiu. "Framework for MPLS-Based Recovery". RFC3469. February 2003.

[SUR01] S. Suri, M. Waldvogel, P. Ramesh Warkhede. “Profile-Based Routing: A new Framework for MPLS Traffic Engineering”. Proceeding QofIS. September 2001.

[TEL99] Telcordia document "Reliability Prediction Procedure for Electronic Equipment" (document number SR-332, Issue 1) AT&T Bell Labs 1999.

[WIL98] Mark R. Wilson, “The Quantitative Impact of Survivable Network Architectures on Service Availability” IEEE Communications Magazine. May 98.

[WZZ01] “Mean Time Between Failure (MTBF) And Availability–A Primer”www.zzyzx.com/products/whitepapers/pdf/MTBF_and_availability_primer.pdf. White paper Zzyzx peripherals 2001.

[ZHA02] Hongwei Zhang, Arjan Durresi, “Differentiated Multi-Layer Survivability in IP/WDM Networks”, 8th IEEE-IFIP Network Operations and Management Symposium (NOMS 2002).

Glossary

A	Availability
AS	Autonomous System
ATM	Asyncronous Transfer Mode
BGP	or Border Gateway Protocol
CR-LDP	Constraint-Route LDP
DAP	Dynamic-Alternative Path
DCS	Digital cross-connect system
DR-PI	Dynamic-Routing with Partial-Information
DWDM	Dense wavelength-division multiplexing
EXP	Experimental field
FEC	Forwarding Equivalence Class
FIS	Fault Indication Signal
FR	Failure Rate
FSC	Fiber switch capable
GMPLS	Generalized MPLS
IETF	Internet Engineering Task Force
IP	Internet Protocol
ISIS	Intermediate System to Intermediate System
LD	Link Degraded
LDP	Label Distribution Protocol
LER	Label Edge Router
LF	Link Failure
LFP	Link Failure Probability
LIB	Label Information Base
LMP	Link Management Protocol

LoL	Loss of Light
LSC	Lambda switch capable
LSP	Label Switch Paths
LSP	Label Switch Paths
LSR	Label Switch Router
MHA	Min-Hop Algorithm
MIRA	Minimum Interface Routing Algorithm
MPLS	Multiprotocol Label Switching
MRB	Maximum Reservable Bandwidth
MTBF	Mean Time Between Failures
MTTF	Mean Time to Failure
MTTR	Mean Time to Repair
OSPF	Open Shortest Path First
OXC	Optical Cross-connect
PBR	Profile-Based Routing
PD	Path Degraded
PF	Path Failure
PML	Protection Merge LSR
PSC	Packet switch capable
PSL	Protection Switch LSR
QoS	Quality of Service
R	Reliability
RSVP	Resource Reservation Protocol
RSVP-TE	RSVP Traffic-Engineering.
SRLG	Shared Risk Link Groups
SWP	Shortest-Widest Path
TDM	Time-division Multiplexing

TSC	TDM switch capable
TTL	Time To Live
U	Unavailability
VC	Virtual Circuits
VCC	Virtual channel connection
VP	Virtual Paths
WSP	Widest-Shortest Path

APPENDIX A

Publications and projects

Throughout this thesis many publications has been published and presented. In this chapter the main publications related with this work are listed.

PUBLICATIONS

Journals and books

Eusebi Calle, Jose L Marzo, Anna Urra "Protection Performance Components in MPLS Networks" Accepted in **Computer Communications Journal,** Elsevier **2004**

J. L. Marzo, **E. Calle**, C. Scoglio, T. Anjali "QoS On-Line Routing and MPLS Multilevel Protection: a Survey" **IEEE Communication Magazine**, vol. 41(10), pp. 126-132, October **2003**

Eusebi Calle, Anna Urra, Jose L Marzo "Multiagent system for controlling GMPLS network protection" **Artificial Intelligence Research and Development**. Frontiers in Artificial Intelligence and Applications. IOS Press **2003**. Pages 256-264. ISBN 1-58603-378-6

International Conferences

Eusebi Calle, Anna Urra, Jose L. Marzo " Network Reliability and Failure Recovery Time Evaluation in GMPLS-Based Networks using Segment and Path Protection". (Submitted to) of Symposium on Performance Evaluation of Computer and Telecommunication Systems **SPECTS 2004**. San Jose. USA.

E. Calle, J.L. Marzo, A. Urra, LL. Fàbrega "Enhancing fault management performance of two-step QoS routing algorithms in GMPLS" (to be presented) International Conference on Communications IEEE **ICC 2004**, 20-24 June, 2004, Paris (France).

Eusebi Calle, Jose L Marzo, Anna Urra "Evaluating the probability and the impact of a failure in GMPLS based networks" In proceedings of of IEEE Design of Reliable Communications Networks IEEE **DRCN 2003**, Alberta (Canada) 2003, ISBN 0-7803-8118-1

Eusebi Calle, Jose L Marzo, Anna Urra, Pere Vila "Enhancing MPLS QoS routing algorithms by using the Network Protection Degree paradigm." In proceedings of IEEE Global Communications Conference **GLOBECOM 2003**, San Francisco (USA)

Eusebi Calle, J L Marzo, Anna Urra "Protection performance components in MPLS networks." In proceedings of of Symposium on Performance Evaluation of Computer and Telecommunication Systems **SPECTS 2003**, Montreal (Canada) ISBN 1-56555-269-5

J. L. Marzo, **E. Calle**, C. Scoglio, T. Anjali "Adding QoS Protection in Order to Enhance MPLS QoS Routing" In proceedings of International Conference on Communications IEEE **ICC 2003**. Anchorage, Alaska (USA). IEEE 2003, ISBN 0-7803-7804-4

Eusebi Calle, Pere Vilà, Jose L. Marzo, Santiago Cots "Arquitectura del sistema de gestión de ancho de banda y protección para entornos de redes MPLS (SGBP)" Symposium on Informatics and Telecommunications, **SIT 2002**. September 25-27, 2002. Sevilla, Spain. ISBN: 84-699-9417-4. Pages 143-154.

Eusebi Calle, Teo Jové, Pere Vilà, Josep L Marzo "A Dynamic Multilevel MPLS Protection Domain" In Proceedings of **DRCN 2001**, Budapest (Hungary), 7-10 October, 2001. Edited by Tibor Cinkler

National Conferences

Anna Urra, **Eusebi Calle**, Jose L Marzo "Sistema multi-agent per al control de la protecció en xarxes GMPLS" Accepted in Congrés Català d'Intel·ligència Artificial **CCIA 2003**, Palma (Spain)

Research Reports

Eusebi Calle "A Dynamic Multilevel MPLS Protection Domain" UdG Reserach Report IIi A 02-18-RR

José L. Marzo, **E. Calle**, Pere Vilà "QoS Protection: Formulation and experimental analysis of the MPLS case" UdG Research Report IIi A 02-17-RR

Other publications:

This is the list of publications not directly related with this work.

Yezid Donoso Meisel, Ramon Fabregat, Jose Luis Marzo, **Eusebi Calle** "Extensión de los métodos Hop-by-Hop, CR-LDP y RSVP-TE para Multicast IP sobre MPLS" **InfoUYclei 2002**. November 25 - 29, Montevideo, Uruguay. ISBN: 9974-7704-1-6

Pere Vilà, José L. Marzo, **Eusebi Calle** "Dynamic Bandwidth Management as part of an Integrated Network Management System based on Distributed Agents" In proceedings of IEEE Global Communications Conference **GLOBECOM 2002**, Taipei (Taiwan), November 17-21, 2002 IEEE 2002, ISBN 0-7803-7633-1

Yezid Donoso Meisel, Ramon Fabregat, Jose Luis Marzo, **Eusebi Calle** "Multidifusión IP sobre MPLS sin y con QoS: propuesta y análisis de rendimiento" Symposium on Informatics and Telecommunications, **SIT**

2002. September 25-27, 2002. Sevilla, Spain. ISBN: 84-699-9417-4.

P. Vilà, J.L. Marzo, **E. Calle**, L. Carrillo "Lightweight Monitoring of Label Switched Paths for Bandwidth Management" (to be presented) IEEE Symposium on Computers and Communications (**ISCC'04**), Alexandria (Egypt), June 29 - July 1, 2004

PROJECTS

This work has been supported by the following projects:

"Arquitectura multiservicio orientada a la fiabilidad y accesibilidad de redes IP" MCyT [TIC2003-05567] 2004 - 2006

"Xarxa temàtica de gestió de xarxes IP orientades a circuits virtuals (GMPLS/MPLS)" [2003/XT/00037] AGAUR 2003 - 2004

"Workshop multiprotocol label switching (MPLS)" DURSI 2002 arcs00279 28/03/2003 - 28/03/2003

"Red temática de gestión de redes IP orientadas a circuitos vistuales (MPLS)" AAEE (MCYT) TIC2002-10150-E 1/2003 - 12/2004

"Gestión Inteligente de redes orientadas a las nuevas aplicaciones telematicas con requerimientos de calidad de servicio (Girona TRECS)" CICYT 10/1999 - 9/2002 TEL99-0976