Lehrstuhl für Kommunikationsnetze
Technische Universität München

Differentiated Resilience in IP-Based Multilayer Transport Networks

Achim Autenrieth

Vollständiger Abdruck der von der Fakultät für
Elektrotechnik und Informationstechnik der Technischen Universität München
zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing., Dr.-Ing. h.c. D. Schröder

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. J. Eberspächer
2. Prof. Dr. Ir. P. Demeester (Univ. Gent, Belgien)

Die Dissertation wurde am 30.9.2002 bei der Technischen Universität München
eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik
am 30.4.2003 angenommen.

ACKNOWLEDGEMENTS

This dissertation was completed during my work as researcher and teaching assistant at the Institute of Communication Networks (LKN) of the Munich University of Technology (TUM) with the guidance and support of many people.

Foremost, I want to thank my doctoral advisor Prof. Dr.-Ing. Jörg Eberspächer, who substantially helped me to successfully complete my thesis with his constant advice, support and helpful discussions during all phases of the dissertation. Jörg Eberspächer supported me with his confidence and gave me the freedom to work independently and on my own responsibility in research projects and on the thesis. Still, I could always rely on his help and guidance whenever I needed it.

Prof. Dr. Ir. Piet Demeester was project leader of the ACTS project PANEL and I am very glad that he accepted to be the second auditor of the thesis. During the PANEL project I learned from Piet that good project management and solid research as well as a friendly atmosphere and good teamwork are essential for the success of a project. Under his guidance PANEL became a very successful, fruitful and at the same time pleasurable project.

I would also like to thank all partners of the PANEL project, who became dear friends during the project duration. It is a great pleasure to meet the PANEL members again at conferences and project meetings. I also want to thank all project partners from the BMBF TransiNet and KING projects and from the Siemens IRIS project. During the project meetings I was able to broaden my knowledge horizon, and the project partners helped me see the telecommunication world from different angles. Monika Jäger and Joachim Westfahl provided me with valuable insight in the requirements and objectives of network operators. Dr. Herzog from Siemens had the confidence in me to support and promote the IRIS project.

I would like to thank Andreas Kirstädter, a former colleague at the Institute of Communication Networks, who was the supervisor of my diploma thesis. In the IRIS research cooperation with Siemens Corporate Technologies he was again my project supervisor. In many discussions he provided valuable contributions to the dissertation and he co-authored several publications.

I thank all colleagues of the LKN family who helped to complete the thesis in a friendly and creative working environment. I thank especially my friend and former roommate Andreas Iselt, who supported me in the startup phase of PANEL with his professional experience and reliable advice. The members of the research group PNNRG (Photonic Networks and Network Resilience Group) Dominic Schupke, Thomas Fischer, Claus Gruber, and Thomas Schwabe supported me with many helpful discussions and provided valuable feedback.

Several diploma thesis students, most noticeably Christian Harrer and Simon Gouyou Beauchamps performed helpful implementation and simulation work for the thesis with great enthusiasm and endurance.

Last but not least, I want to thank my family for their love and encouragement and all my friends for their support and friendship and the enjoyable time spent together.

SUMMARY

This thesis investigates the provisioning of resilience against network failures in multilayer IP-based optical networks. Failures like cable cuts or node breakdowns can have drastic impact on the communication services. Due to the ever increasing amount of data transported over a single link – more than a hundred wavelengths with a bit rate of up to 40 Gb/s each are possible on a single fiber using wavelength division multiplexing (WDM) – failures can cause tremendous loss of data, loss of revenue, and loss of reputation for the network operator.

Therefore the network has to be resilient against failures. It must be able to detect the failure and recover affected services very fast, ideally without the services realizing the outage and disconnecting. Due to the complexity of the transport *network architectures* sophisticated resilience mechanisms are needed. These may operate in multiple network technologies (or layers). The network technologies Multiprotocol Label Switching (MPLS), Asynchronous Transfer Mode (ATM), Synchronous Digital Hierarchy (SDH), and Optical Transport Networks (OTN) offer such resilience mechanisms and are considered for this work.

In this thesis a comprehensive and systematic *resilience framework* is defined to investigate and evaluate existing and novel resilience strategies. The framework consists of a definition of network survivability performance metrics and network operators' objectives, a definition of considered failure scenarios, and the definition of required failure detection functions and notification mechanisms. The generic characteristics of recovery models like protection switching and restoration are defined. Their various options in terms of network topology, resource sharing, recovery level, and recovery scope are specified. The framework is extended to cover multiple failure scenarios and multilayer recovery strategies.

The provisioning of protection flexibility, service granularity and resilience manageability are important objectives of network resilience mechanisms in addition to the optimization of performance metrics like resource efficiency and recovery time. A major contribution of this thesis is the development of a novel architecture for the flexible provisioning of *differentiated resilience* in quality-of-service-enabled IP networks. Services or flows can be assigned different levels of resilience depending on their resilience requirements. This is achieved by an extension of the traditional QoS signaling to include resilience requirements of the services. The architecture is called *Resilience-Differentiated QoS (RD-QoS)*. Four resilience classes are defined and can be mapped to appropriate recovery mechanisms with different recovery time scales. The resilience mechanisms are provided by MPLS or by lower layer recovery mechanisms. A traffic engineering process is defined for the RD-QoS architecture and a recovery time analysis model is specified for the available recovery mechanisms. Within a case study the resource efficiency and recovery time of the RD-QoS architecture is evaluated for different networks and a set of selected recovery mechanisms. The case study shows significant network capacity savings, which can be achieved by assigning each service its required level of resilience.

Finally, the thesis evaluates the multilayer resilience strategies identified in the recovery framework. The multilayer recovery options specify in which layer affected connections are recovered for a specific failure scenario. If recovery mechanisms are activated in multiple layers, the recovery actions must be coordinated. With a multilayer network simulation environment, the different strategies are investigated in detail, and a further case study is performed. Then, the multilayer recovery framework is extended to take into account the differentiated resilience requirements. Such a *differentiated multilayer resilience* approach considers the resilience requirements of the IP services and the recovery mechanisms available in different layers to select an optimal multilayer recovery strategy. The different options of this approach are discussed and their performance is evaluated in this thesis.

KURZFASSUNG

Diese Arbeit untersucht die Bereitstellung von Ausfallsicherheit gegen Netzfehler in mehrschichtigen optischen IP Transportnetzen. Bedingt durch die stetig wachsenden Übertragungskapazitäten - heutzutage sind bereits weit über hundert Wellenlängen mit Bitraten bis zu jeweils 40 Gb/s auf einer einzigen Glasfaserleitung möglich – haben Fehler wie Kabelbrüche oder Knotenausfalle drastische Auswirkungen auf Telekommunikationsdienste und können hohe Datenverluste, Umsatzeinbußen und nicht zuletzt einen Verlust an Ansehen der Netzbetreiber verursachen.

Daher müssen heutige Transportnetze gegen verschiedenste Netzfehler belastbar sein, die Fehlerauswirkungen auffangen können und in einen fehlerfreien Zustand zurückbringen (engl.: '*resilience*'). Die Netzelemente müssen eigenständig die Fehler erkennen, an andere Netzelemente und an das Netzmanagement signalisieren, sowie in möglichst kurzer Zeit die betroffenen Verbindungen wiederherstellen. Dabei können die Ausfallsicherheitsmechanismen (*engl.: 'resilience mechanisms'*) in unterschiedlichen Netztechnologien, sogenannten Netzschichten, arbeiten. Die in dieser Arbeit betrachteten *Transportnetztechnologien* sind Asynchroner Transfer Modus (ATM), Synchrone Digitale Hierarchie (SDH) und Optische Transportnetze (OTN) sowie Netze, die auf der TCP/IP Protokollfamilie mit Multiprotocol Label Switching (MPLS) basieren, und damit verbindungsorientierte Eigenschaften für die IP-Schicht realisiert.

Um vorhandene und neuartige Ausfallsicherheitsverfahren bzw. Abfederungsmechanismen (engl.: '*resilience mechanisms*') systematisch klassifizieren und bewerten zu können, wird in dieser Arbeit ein umfassendes *Rahmenwerk für Ausfallsicherheit in Transportnetzen* definiert. Dazu werden die für die verschiedenen Netztechnologien entwickelten und teilweise standardisierten Mechanismen in ein generisches, d.h. von der jeweiligen Netztechnologie unabhängiges Rahmenwerk eingebunden. In dem Rahmenwerk werden Performanzparameter und Zielvorgaben von Netzbetreibern sowie betrachtete Fehlerszenarien definiert. Ebenso werden Mechanismen zur schnellen und zuverlässigen Fehlererkennung, die für eine hohe Ausfallsicherheit notwendig sind, definiert. Die Wiederherstellungsverfahren werden in bezug auf ihre Haupteigenschaften und Optionen wie die unterstützte Netztopologie, Ressourcenverwendung, Wiederherstellungsebene und –ausdehnung klassifiziert. Nach der generischen Betrachtung der Wiederherstellungsverfahren wird auf die charakteristischen Eigenschaften der Netzschichten eingegangen und die sich daraus ergebenden Vor- und Nachteile erörtert. Das Rahmenwerk definiert außerdem Konzepte zur Behandlung von Mehrfachfehlern und erläutert Anforderungen und Strategien zur Koordination von Ausfallsicherheitsverfahren in mehreren Netzschichten.

Die Bereitstellung einer flexiblen Ausfallsicherheit, feinen Granularität und einfachen Verwaltbarkeit ist eine wichtige Eigenschaft von Ausfallsicherheitsverfahren. Die Leistungsfähigkeit der Verfahren drückt sich in einer hohen Kapazitätseffizienz und einer kurzen Wiederherstellungszeit aus. Ein wesentlicher Beitrag dieser Arbeit ist die Entwicklung einer *Architektur zur flexiblen Bereitstellung von differenzierter Ausfallsicherheit in QoS-unterstützenden IP Transportnetzen*. Dies wird durch die Erweiterung etablierter QoS-Signalisierungsarchitekturen um die Ausfallsicherheitsanforderungen von IP Diensten erreicht. Die Architektur wird *Resilience-Differentiated*

Quality of Service (RD-QoS) genannt, zu deutsch 'Dienstgüte mit differenzierter Ausfallsicherheit'. Vier Ausfallsicherheitsklassen werden definiert, die auf entsprechende Wiederherstellungsverfahren abgebildet werden können. Die Verfahren werden von der MPLS-Schicht oder von optischen Netzschichten zur Verfügung gestellt. Für die RD-QoS Architektur wurde ein Verkehrsplanungsprozess entwickelt, um die Kapazitätseffizienz der Architektur bewerten zu können. Außerdem wurden Modelle zur Analyse der Wiederherstellungszeiten verschiedener im Rahmenwerk definierter Ausfallsicherheitsmechanismen aufgestellt. In einer Fallstudie werden die Kapazitätseffizienz sowie die Wiederherstellungszeiten der RD-QoS Architektur für verschiedene Netzszenarien und einer Auswahl von Ausfallsicherheitsmechanismen analysiert und bewertet. Die Fallstudie zeigt signifikante Netzkapazitätseinsparungen, die sich mit der RD-QoS Architektur durch die Verwendung differenzierter Ausfallsicherheit erzielen lassen.

Schließlich werden die im Rahmenwerk definierten mehrschichtigen Ausfallsicherheits-verfahren untersucht. Die Strategie für *Ausfallsicherheit in mehrschichtigen Netzen* definiert, in welcher Schicht betroffene Verbindungen bei einem bestimmten Fehlerszenario wiederhergestellt werden. Wenn Wiederherstellungsvorgänge in mehreren Schichten auftreten können, müssen die verschiedenen Verfahren koordiniert werden. Die vorgestellten Strategien wurden in einer Simulationsumgebung für mehrschichtige Netze mit einem hohen Detaillierungsgrad der Netzkomponenten-modelle untersucht. Eine Fallstudie wurde durchgeführt und die Verfahren anhand der Ergebnisse bewertet. Schließlich wird das Rahmenwerk für mehrschichtige Ausfallsicherheit erweitert, um den Ansatz der differenzierten Ausfallsicherheit zu integrieren. Die *differenzierte, mehrschichtige Ausfallsicherheit* betrachtet die Ausfallsicherheitsanforderungen der IP Dienste, um eine optimale Strategie zur Fehlerwiederherstellung in mehrschichtigen Netzen auszuwählen. Dabei wurden verschiedene Optionen für mehrschichtige Ausfallsicherheitsstrategien in Betracht gezogen und ihre Eignung für unterschiedliche Netzszenarien bewertet.

TABLE OF CONTENTS

x

LIST OF USED ACRONYMS

A..... Adaptation
A..... Availability
AAL ..... ATM Adaptation Layer
ACTS ..... Advanced Communications Technologies and Services
ADM ..... Add/Drop Multiplexer
AF ..... Assured Forwarding
AIS ..... Alarm Indication Signal
ANSI ..... American National Standardization Institute
AP ..... Access Point
APS ..... Automatic Protection Switching
AS ..... Autonomous System
ASON..... Automatically Switched Optical Network
ASTN ..... Automatically Switched Transport Network
ATM..... Asynchronous Transfer Mode
AU..... Administrative Unit
AUG..... Administrative Unit Group
BA ..... Behavior Aggregate
BGP..... Border Gateway Protocol
B-ISDN ..... Broadband Integrated Services Digital Network
BLSR..... Bidirectional Line Switched Ring
BMBF ..... Bundesministerium für Bildung und Forschung
CAC ..... Call/Connection Admission Control
CI..... Connection Identifier
CIDR..... Classless Inter-Domain Routing
CP..... Connection Point
CRC ..... Cyclic Redundancy Check
D..... Downtime
DiffServ ..... Differentiated Services
DMR ..... Differentiated Multilayer Resilience
DPM..... Defects Per Million
DS ..... Digital Section
DSCP ..... Differentiated Services Code Point
DXC ..... Digital Cross-Connect
EF..... Expedited Forwarding
E-LSR ..... Egress Label Switched Router
EMF ..... Equipment Management Function
ETSI..... European Telecommunication Standards Institute
FEC ..... Forwarding Equivalence Class
FIFO..... First-In, First-out
FIT..... Failures In Time
FUR..... Router upstream of a failed link
FTP..... File Transfer Protocol
FDR..... Router downstream of a failed link

GMPLS ..... Generalized Multiprotocol Label Switching
GUI ..... Graphical User Interface
HDLC..... High-level Data Link Control
HEC ..... Header Error Control
HOP ..... Higher Order Path
IETF ..... Internet Engineering Task Force
ILP..... Integer Linear Programming
I-LSR..... Ingress Label Switched Router
IntServ..... Integrated Services
IP ..... Internet Protocol
IRIS ..... Internet Resilience & IP Survivability
ISDN ..... Integrated Services Digital Network
ISP..... Internet Service Provider
ITU-T ..... International Telecommunication Union – Telecommunication Standardization Sector
KING..... Key components for Internet Next Generation
LC ..... Link Connection
LMP ..... Link Management Protocol
LOL..... Loss of Light
LOP ..... Lower Order Path
LOS ..... Loss of Signal
LSP..... Label Switched Path
LSP..... Label Switched Path
LSR ..... Label Switched Router
MDT..... Mean Down Time
MPLS ..... Multiprotocol Label Switching
MS..... Multiplex Section
MS-SPRing ..... Multiplex Section Shared Protection Ring
MS-DPRing ..... Multiplex Section Dedicated Protection Ring
MSOH..... Multiplex Section Overhead
MSP ..... Mutiplex Section Protection
MTBF..... Mean Time Between Failures
MTTR ..... Mean Time To Repair
MTTV ..... Mean Time To Recovery
MUT..... Mean Up Time
NE ..... Network Element
NMS..... Network Management System
NSP ..... Network Service Provider
OADM ..... Optical Add/Drop Multiplexer
OAM ..... Operation, Administration and Maintenance
OCh..... Optical Channel
ODU..... Optical Channel Transport Unit
OIF ..... Optical Internetworking Forum
OMS..... Optical Multiplex Section
OPU ..... Optical Channel Payload Unit
OSPF..... Open Shortest Path First
OTN ..... Optical Transport Network

OTS ..... Optical Transmission Section
OTU ..... Optical Channel Transport Unit
OXC ..... Optical Cross-Connect
PANEL..... Protection Across Network Layers, ACTS Project AC205
PHB..... Per-Hop Behavior
POS ..... Packet over SONET
PPP..... Point-To-Point Protocol
QoS ..... Quality of Service
R..... Reliability
RC ..... Resilience Class
RDI..... Remote Defect Indication
RD-QoS ..... Resilience Differentiated Quality of Service
RM ..... Resource Manager
RS..... Regenerator Section
RSOH..... Regenerator Section Overhead
RSVP ..... Resource Reservation Protocol
SDH ..... Synchronous Digital Hierarchy
SDL..... Simple Data Link
SDL..... Specification and Description Language
SDT..... SDL Design Tool
SELANE ..... Simulation Environment for Layered Networks
SHF ..... Self-Healing Function
SNC..... Subnetwork Connection
SNCP ..... Subnetwork Connection Protection
SONET..... Synchronous Optical Network
STM ..... Synchronous Transport Modules
TCP ..... Termination Connection Point
TCP ..... Transmission Control Protocol
TE..... Traffic Engineering
TOS..... Type of Service
TP..... Transmission Path
TT..... Trail Termination
TU ..... Tributary Unit
TUG ..... Tributary Unit Group
U..... Unavailability
UPSR ..... Unidirectional Path Switched Ring
VC..... Virtual Channel
VC..... Virtual Container
VCI..... Virtual Channel Identifier
VP ..... Virtual Path
VPI..... Virtual Path Identifier
VWP..... Virtual Wavelength Path
WDM ..... Wavelength Division Multiplex
WP ..... Wavelength Path
WWW ..... World Wide Web

1 INTRODUCTION

1.1 Motivation

Network Evolution

In the recent years two factors dominated the development of the transport network infrastructure. The first factor is the advance in optical transmission and optical network technology, which made its way from research labs and test fields to operating networks. With Wavelength Division Multiplex (WDM) techniques, more than a hundred wavelengths can be transported over a single fiber, with a typical bit rate of 10 Gb/s each, in future even 40 Gb/s per wavelength. Optical network components like lasers, amplifiers, optical switches, and optical cross-connects emerged, paving the way for the deployment of purely optical transport networks.

The second trend is the continuing explosive growth of IP data traffic. According to K. G. Coffman and A. M. Odlyzko the Internet traffic approximately doubles every year (between 70% and 150% growth per year) and is becoming the dominant traffic for the global telecommunication network [Coffman-2002]. While the revenue of network operators is still largely derived from classical telecommunication services [Coffman-2002], the design of the network has to take the high data traffic volumes into account.

Internet Services

Also the type of traffic transported in IP networks changed. It used to be mainly connectionless best-effort traffic. Now mission-critical and high-priority business services and real-time, connection-oriented services are transported over the Internet. The real-time connection-oriented character of these services demanded the development of quality of service (QoS) architectures for the Internet. The two main QoS architectures standardized by the Internet Engineering Task Force (IETF) are the flow-based Integrated Services architecture with the Resource Reservation Protocol (RSVP) as signaling protocol and the Differentiated Services model, which is based on traffic aggregation and hop-by-hop traffic shaping.

Internet Resilience

Providing reliable and fault-tolerant network infrastructures is a key factor for the development of the information society [Eberspächer-2000]. The economic importance of the Internet, the increasing complexity of the network technologies and the huge amount of traffic transported over a single network element require sophisticated survivability mechanisms against failures like fiber cuts or node breakdowns. Network survivability has become a key research issue for IP-based transport networks (e.g., in [Draft-Awduche] network survivability is identified as a key requirement of traffic engineered IP networks).

Moreover, network survivability mechanisms are available in multiple network layers, and resilience strategies should make benefit of recovery at multiple layers and at the same time prevent negative interference between these mechanisms.

In circuit switched transport networks, resilience is traditionally offered as a two-state option: either no resilience has been provided for a connection, or a connection is 100% protected against a given set of failures, like all single link or node failures.

However, recovery at the Multiprotocol Label Switching (MPLS) layer allows differentiating between customers and applications requiring services with a high level of resilience and those requiring low-priority best-effort services, which could tolerate an extended period of degraded quality of service or even service outage. A resilience-differentiated approach can protect only that part of traffic, which requires a high level of service availability. This allows a cost-effective network design and traffic engineering.

The objective of this work is to investigate existing and to develop new resilience concepts. The focus is on differentiated resilience in IP-based multilayer transport networks.

1.2 Overview of the Thesis

Network Fundamentals

The thesis presents a study of the flexible provisioning of end-to-end resilience in IP over optical transport networks. To have a clear understanding of the considered transport network architectures and to define a clear terminology to be used, in the second chapter an introduction to the architecture of multilayer transport networks is given. Specifically, the architectures of Asynchronous Transfer Mode (ATM), Synchronous Digital Hierarchy (SDH), and Optical Transport Networks (OTN) as well as TCP/IP networks and Multiprotocol Label Switching (MPLS) are introduced. Additionally, networking concepts such as quality of service and traffic engineering are covered.

Network Resilience Framework

In the third chapter, an integrated multilayer resilience framework is defined. The framework covers the failure detection and signaling, the recovery mechanisms and recovery options for service restoration. Similar resilience concepts exist for individual network technologies. However, the terminology used for the different layers is often different. Therefore it is important to define a generic, architecture-independent framework with a precise and consistent terminology for all considered technologies. Nevertheless, the layer-specific capabilities are also regarded with a short introduction to the state of the art of recovery mechanisms in individual layers. The main evaluation methods used in this thesis, the spare resource usage and the recovery time analysis, are also defined as part of the resilience framework. Finally, the framework is extended to include the recovery of multiple failure scenarios as well as multilayer recovery concepts.

Resilience Differentiated Quality of Service (RD-QoS)

The novel 'Resilience-Differentiated Quality of Service' architecture (RD-QoS), which was developed in this thesis, is presented in detail in the fourth chapter. Its deployment in Quality of Service (QoS) architectures, like Differentiated Services (DiffServ) or Integrated Services (IntServ) with RSVP as signaling protocol is shown. At the border of MPLS domains, the resilience attribute defined in the RD-QoS architecture is mapped to appropriate MPLS recovery mechanisms.

For an evaluation of the RD-QoS architecture, an RD-QoS traffic-engineering (RD-QoS-TE) process is defined and a case study for different network scenarios is performed. The results are discussed in detail and show large network capacity savings achievable using the RD-QoS concept. In addition, the recovery times of the resilience classes are analyzed, and the recovery ratio over time is evaluated.

Multilayer Resilience Evaluation

In the fifth chapter, the multilayer recovery concepts are discussed and evaluated for different network scenarios. The case studies are performed on ATM over SDH network scenarios, for which the different multilayer interworking options are evaluated. Finally, the RD-QoS concept is evaluated for its applicability in multilayer IP over optical transport networks.

The thesis concludes with an overall summary and conclusion, and an outlook to future differentiated resilience strategies in IP-based multilayer transport networks is given.

2 NETWORK ARCHITECTURES AND MULTILAYER NETWORKING

2.1 Introduction

In this chapter an overview of the transport network technologies considered in this thesis is given. Specifically, the Optical Transport Network (OTN), Synchronous Digital Hierarchy (SDH), Asynchronous Transfer Mode (ATM), the TCP/IP protocol family, and Multiprotocol Label Switching (MPLS) are covered. For each technology the functional architecture and general networking concepts are summarized and the signal format is described. Finally, layering scenarios of multilayer IP over optical networks are discussed.

2.2 Transport Network Architectures

Today's transport networks are based on multiple transport technologies such as ATM, SDH or OTN, which operate in multiple layers in a client-server relationship. In the recommendation G.805 [ITU-T G.805] the International Telecommunication Union ITU-T defined a functional architecture for transport networks in a technology independent way. The European Telecommunications Standards Institute (ETSI) defined a related standard [ETSI 300 417].

In the next section the main architectural principles of [ITU-T G.805] are summarized. Based on this generic model the functional architecture of Asynchronous Transfer Mode (ATM), Synchronous Digital Hierarchy (SDH), and Optical Transport Networks (OTN) are described in the following sections. Finally, the network architecture of TCP/IP based networks is presented and layering scenarios for IP over optical networks are discussed.

2.2.1 Generic Functional Architecture

The architecture of a transport network is based on a layering and partitioning concept, where the layers interact in a client-server relation. The standards [ITU-T G.805] and [ETSI 300 417] describe a generic layering and partitioning methodology. A good introduction to the functional modeling and a summary of the standardization effort with a focus on optical networks is given in [McGuire-1998].

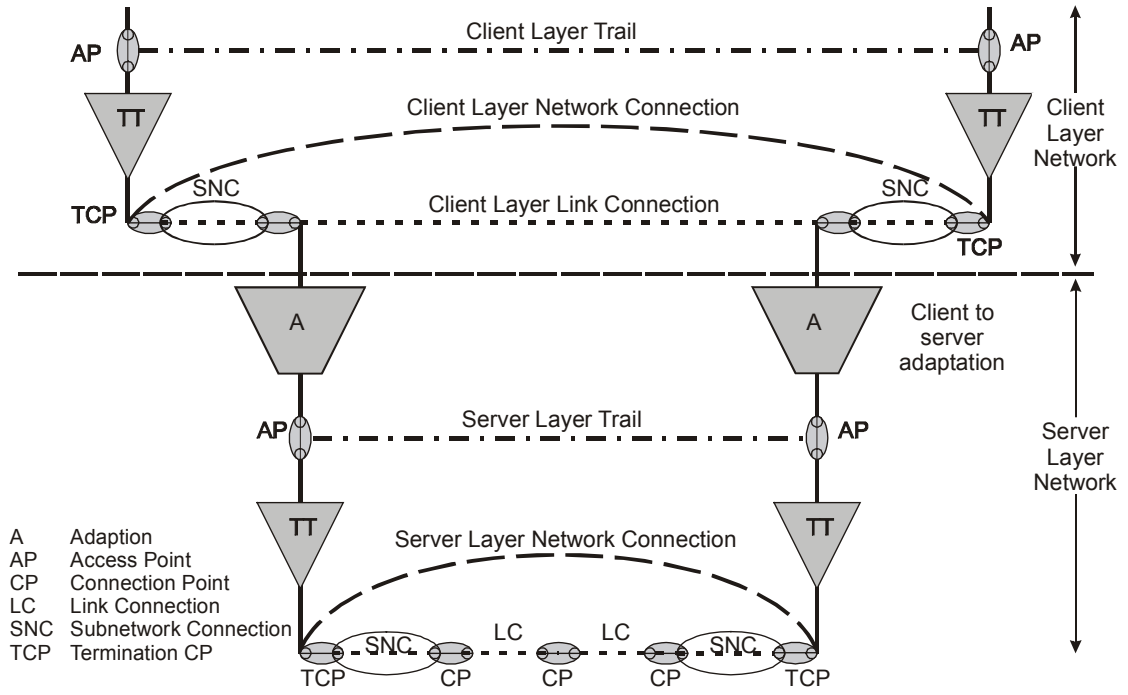


Figure 2.1: Functional model [ITU-T G.805]

The architectural components of the functional model can be seen in Figure 2.1. Four different transport entities provide transparent transport of information between reference points.

- A **link connection** transports information transparently across a link. A link connection consists of a pair of adaptation points and a trail in the server layer network.
- The **subnetwork connection** transports information transparently across a subnetwork. A subnetwork connection is a concatenation of *subnetwork connections* and *link connections*.
- The **network connection** transports information transparently across a layer network. It is formed from a concatenation of *link connections* and/or *subnetwork connections* between terminating connection points.
- The **trail** transports monitored adapted information of the client layer between access points.

In each layer transport processing functions are required to describe a transmission network. In [ETSI 300 417] transport processing functions are termed 'atomic functions'.

- The **adaptation function** represents the conversion process between a server and a client layer. The signal adaptation includes scrambling, encoding, and framing. In addition to the information adaptation it is used for multiplexing, demultiplexing and inverse multiplexing.
- The **termination function** performs the signal integrity supervision of the layer (monitoring). This is done by adding monitoring information such as cyclic

redundancy check (CRC) code at the source and removing and analyzing this information at the termination function sink. The monitoring information can be used to detect bit errors at the trail termination sink. Additionally, address information (trail trace identifier) and error signals (e.g. Remote Defect Indicator (RDI) or Alarm Indication Signal (AIS)) are monitored to detect misconnections and signal failures.

- In [ETSI 300 417] a third atomic function is defined: the **connection function**. The connection function provides flexibility within a layer. This provides a network element with routing, grooming, protection, and restoration functionality. In a network element, the connection function is realized by the switching matrix and may either be a space or time switch. In the functional architecture the connection function is always modeled as a space switch.

Figure 2.2 shows the atomic functions in a layer [ETSI 300 417].

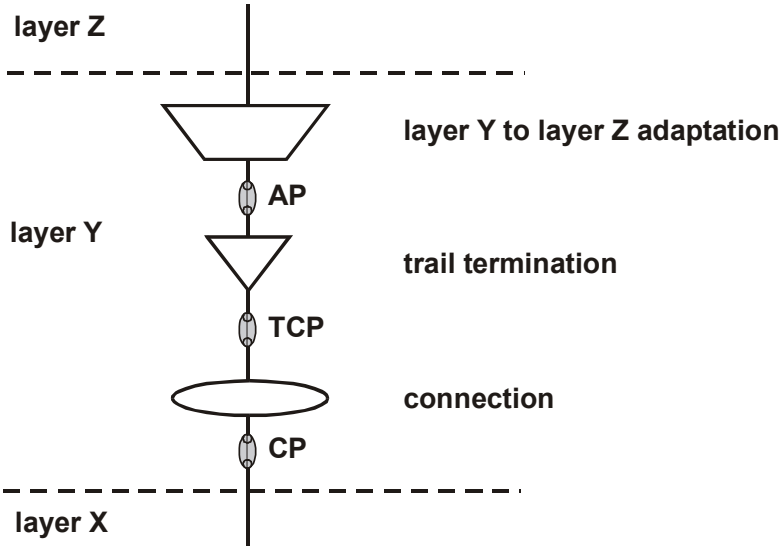


Figure 2.2: Atomic functions in a layer

A set of atomic functions can be grouped together to form a compound function. A network element can be described using a collection of atomic functions and compound functions.

Two classes of layer networks are defined in [ITU-T G.805] – path layer networks and transmission media layer networks. The path layer networks provide transmission (transfer and switching) capabilities to support various types of client services independent of underlying transmission media layer networks. Transmission media layer networks are supported by trails and link connections and may be dependent of the underlying physical media used for transmission (e.g., optical fiber or radio). Figure 2.3 shows the path layer (gray) and transmission media layer (white) networks defined for OTN, SDH and ATM.

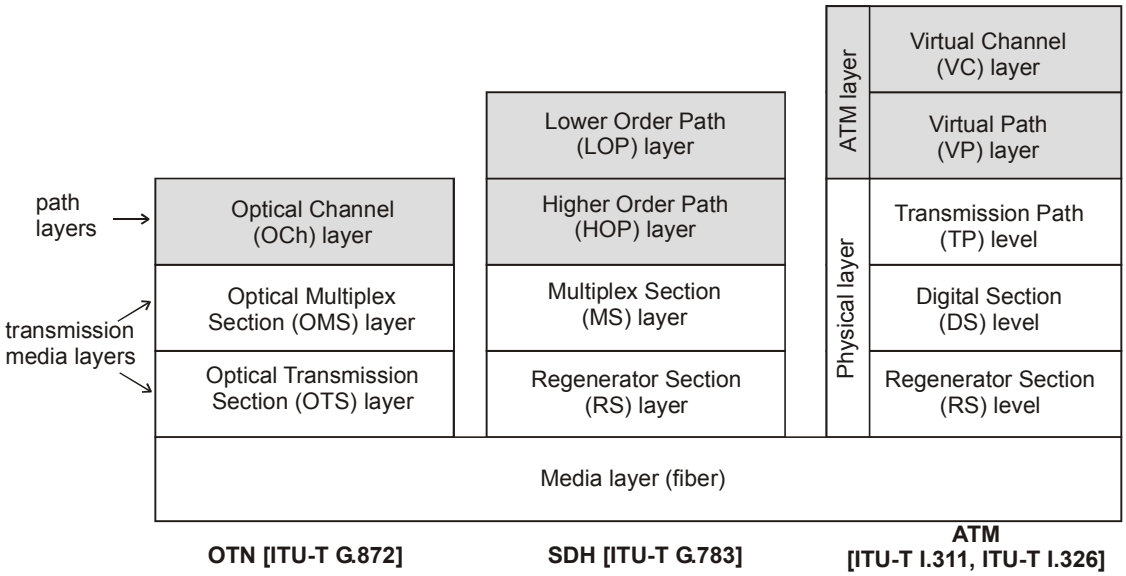


Figure 2.3: OTN, SDH and ATM network layers

In addition to the layering concepts the reference standards define a partitioning concept to represent the organizational structure within a layer. The partitioning concept is based on a recursive decomposition of a network layer in subnetworks, and a subnetwork in smaller subnetworks and link connections. The network partitions may reflect administrative structures (such as multiple operators interoperating to provide end-to-end connectivity) or organizational structures used by a single operator for administrative purposes.

2.2.2 Asynchronous Transfer Mode (ATM)

2.2.2.1 B-ISDN Reference Model and General Networking Aspects

The ITU-T defined in the recommendation [ITU-T I.321] a reference model for the Broadband Integrated Services Digital Network (B-ISDN). The Asynchronous Transfer Mode (ATM) was defined as transmission technology for B-ISDN. Figure 2.4 shows the reference model and the associated transport network layers. The reference model is divided in the physical layer, the ATM layer, an ATM adaptation layer and higher layers. For each layer a control plane and management plane is defined. The management plane is further divided in layer and plane management functions.

Recommendation [ITU-T I.311] further divides the ATM and physical layer in sub-layers, so-called levels. The ATM transport layer is subdivided in the virtual channel (VC) and virtual path (VP) level. The physical layer is subdivided in the transmission path, digital section and regenerator section level. The lowest layer, which is not shown in the figure, is the media layer.

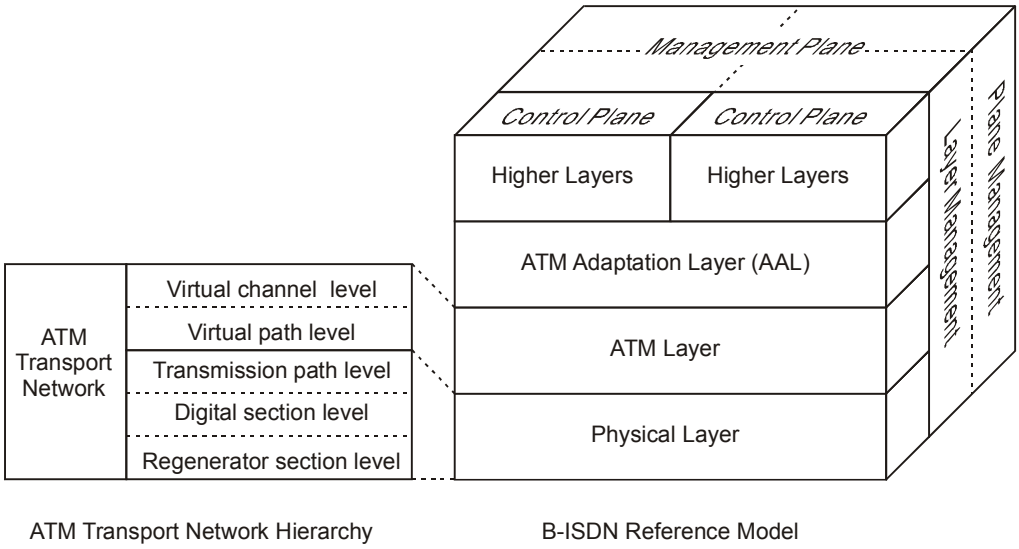


Figure 2.4: B-ISDN reference model [ITU-T I.321]

Figure 2.5 illustrates the relationship between virtual channels and virtual paths [ITU-T I.311].

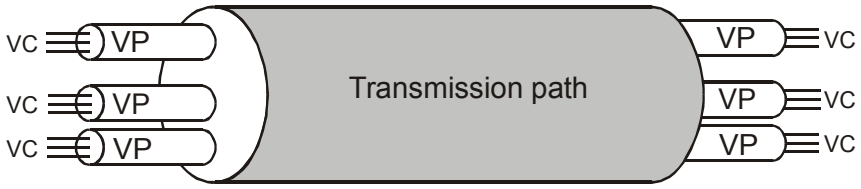


Figure 2.5: Relationship between virtual channel, virtual path and transmission path

2.2.2.2 ATM Cell Format

The ATM transport mechanism is based on a low-delay, connection oriented packet switching technique based on an asynchronous time division multiplexing. The user data is transmitted in small, 53 byte fixed-size packets with a 48 byte data field. These fixed-size packets are called ATM cells. The cell size was defined as a compromise between 32 byte cells for fast processing of real-time data required e.g. for voice communication and 64 byte proposed for a resource efficient transmission of data with low cell overhead. The basic ATM transmission bit rates are 155 Mb/s and 622 Mb/s.

The ATM cell consists of a 5 byte cell header with the control information and the 48 byte data field (see Figure 2.6). The data field (or information field) contains user data (payload) as well as control data for Operation, Administration and Maintenance (OAM) purposes. The user data is segmented to 48 byte parts. The control information in the cell header is used for the switching and transmission of the cell through the network.

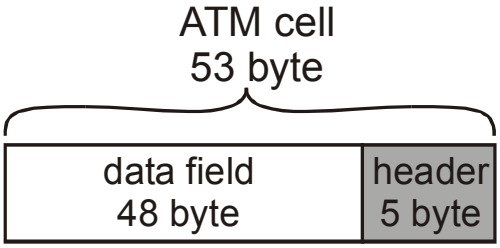


Figure 2.6: ATM cell structure

8	7	6	5	4	3	2	1	bit / byte
Generic Flow Control (GFC)				Virtual Path Identifier (VPI)				1
Virtual Path Identifier (VPI)				Virtual Channel Identifier (VCI)				2
Virtual Channel Identifier (VCI)								3
Virtual Channel Identifier (VCI)				Payload Type (PT)			Cell Loss Priority (CLP)	4
Header Error Control (HEC)								5

Figure 2.7: ATM cell header (User-Network-Interface) [ITU-T I.361]

Figure 2.7 shows the structure of an ATM cell header. Only the VPI and VCI control field will be described in detail. The definition of the other control fields can be found in [ITU-T I.361]. The cells belonging to an ATM connection are identified by the ATM Cell Identifier (CI), which consists of the control fields VPI and VCI. The CI is defined at connection setup, and cells are switched through the network based on the CI value. The VPI defines the virtual path and VCI value defines the virtual channel a cell belongs to. The relation between VPIs and VCIs is illustrated in Figure 2.5.

2.2.2.3 ATM Network Elements

The ATM network elements (NE) are defined in [ITU-T I.731] and [ITU-T I.732]. Depending on the level the ATM connection is switched through the network – VP level or VC level – the used network elements are VP or VC crossconnects (or switches), respectively. If the network element supports connection signaling, the equipment is a VP/VC switch, otherwise a VP/VC cross-connect. ATM multiplexers have only a restricted connectivity, i.e. the equipment has multiple user interfaces (tributary ports), but only a single network interface. Figure 2.8 illustrates the difference between a VP switch/crossconnect and a VC switch/crossconnect [ITU-T I.311].

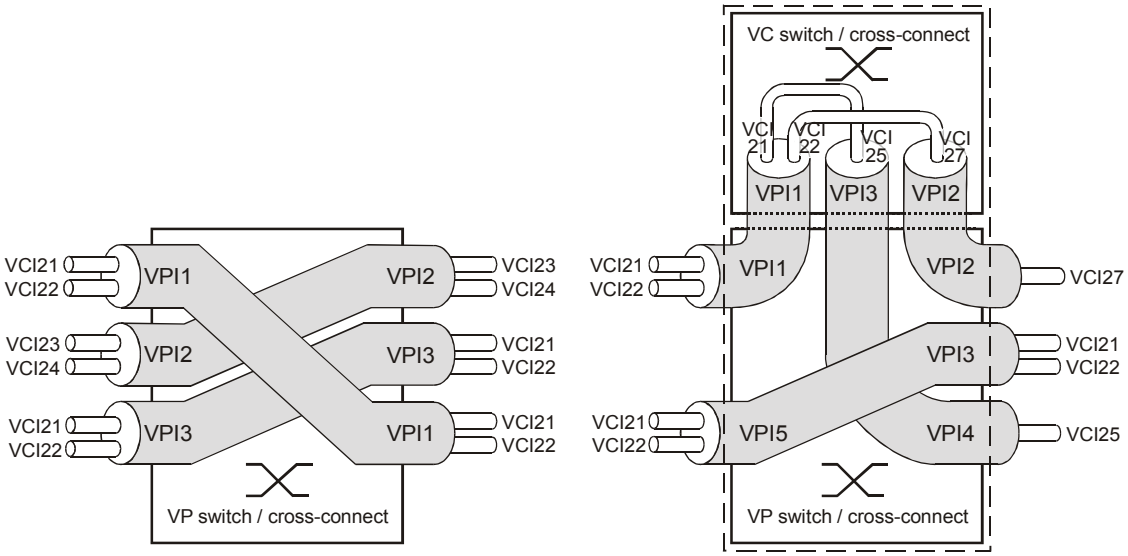


Figure 2.8: VP and VP/VC switches / crossconnects

2.2.3 Synchronous Digital Hierarchy (SDH)

2.2.3.1 SDH Functional Architecture

In Figure 2.3 the SDH network layers have already been introduced. Figure 2.9 shows again the client/server relation of the SDH layers together with a representation of the transport sections. The functional architectur e of SDH is defined in [ITU-T G.803]. The physical interface is usually an optical fiber. Alternative physical interfaces for radio and satellite links, and an electrical interface for low transmission bit rates are also defined.

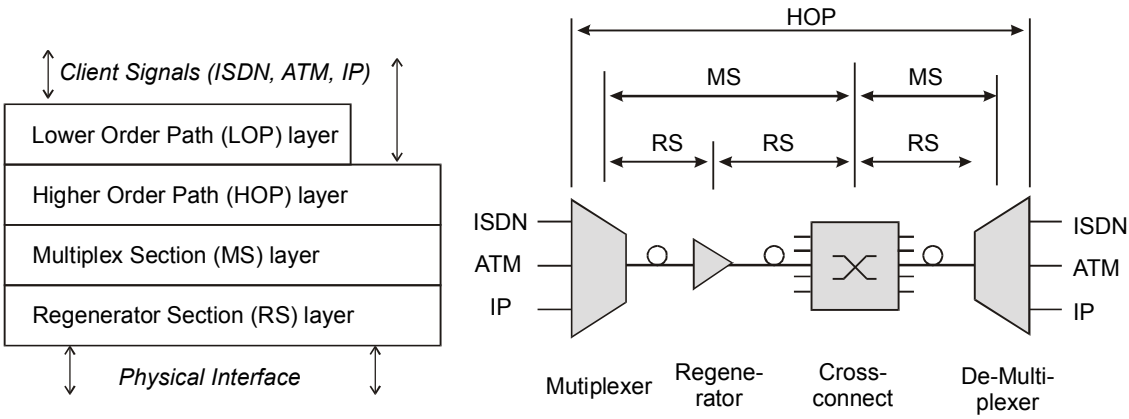


Figure 2.9: SDH layer network

The SDH network layers are the regenerator section, multiplex section, higher and lower order path layers. The right part of Figure 2.9 shows the extension of the transport sections on a sample network cutout. The SDH frame structure used for the information transport and the SDH network elements contained in the figure will be described in the following sections.

2.2.3.2 SDH Frame Structure

The SDH frame structure is based on the hierarchical multiplexing of signals with different bit rates into higher layer transport units. A transport unit is called 'virtual container' (VC). Figure 2.10 shows the mapping of client signals with their bit rates into SDH virtual containers. The virtual containers are multiplexed into synchronous transport modules (STM) [ITU-T G.707]. The tributary units (TU) and auxiliary units (AU) perform pointer processing. Multiple tributary and administrative units can be multiplexed into tributary and administrative unit groups (TUG, AUG), respectively.

The STM-N frame structure is shown in the next figure. The basic transmission frame is the STM-1 frame with 270 columns and 9 rows. The first 9 octet columns are reserved for the section overhead and a pointer to the location of the administrative unit within the frame. The section overhead contains a regenerator section overhead (RSOH) and the multiplex section overhead (MSOH). For higher order STM-N signals (STM-4, STM-16, STM-256), a byte-wise interleaving of N frame structures is used.

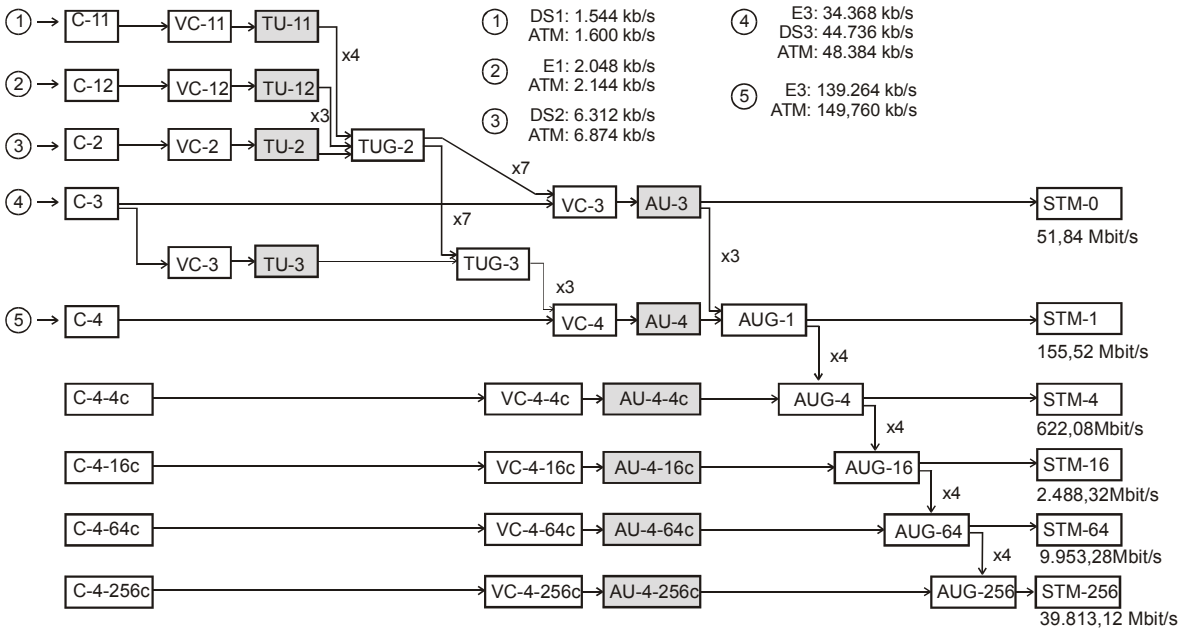


Figure 2.10: SDH multiplexing structure

The STM-0 frame has only 90 columns and 9 rows. It is primarily used for compatibility with the North American SONET (Synchronous Optical Network) standard, which is defined by the American National Standards Institute (ANSI). The SONET standard is similar to SDH, but it uses some other client signal mappings and a different multiplexing structure. In this thesis the focus is on the European SDH standard.

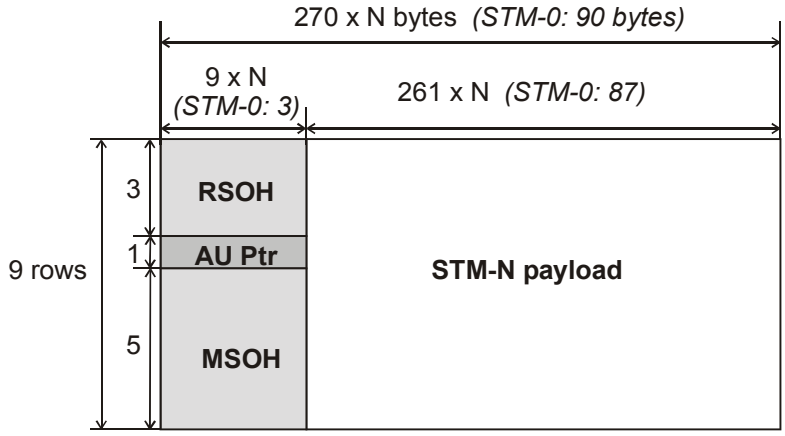


Figure 2.11: SDH frame structure

2.2.3.3 SDH Network Elements

There are four generic SDH network elements: SDH multiplexers, regenerators, add/drop multiplexers and crossconnects (Figure 2.12). The SDH multiplexer combines different client signals and lower level SDH signals in STM-N signals. The regenerator refreshes the signal, which is attenuated by the signal transmission over the physical media. The regenerator renews both signal timing and amplitude while processing and regenerating the RS overhead.

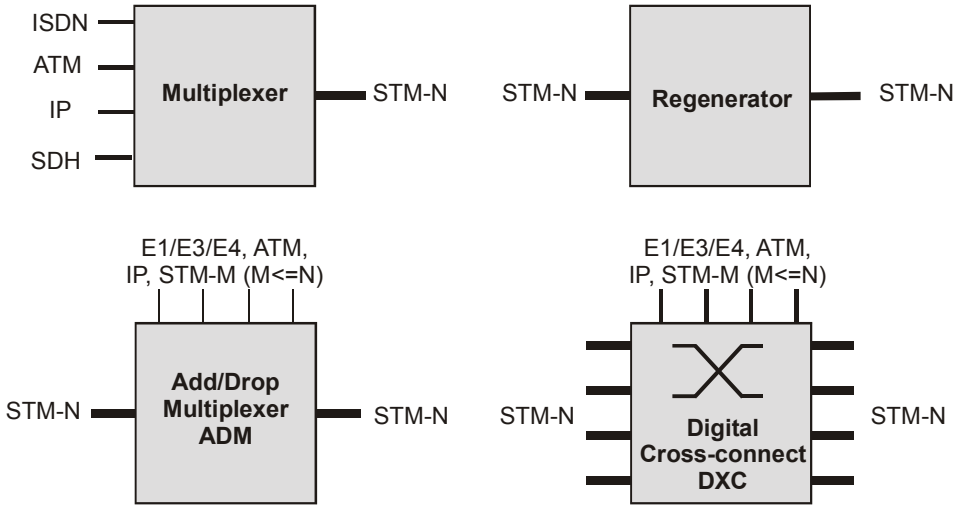


Figure 2.12: SDH network elements

The add/drop multiplexer (ADM) and the digital crossconnects (DXC) provide flexibility in the SDH network. In an ADM lower bit rate synchronous signals can be extracted from the STM-N frame and dropped to the tributary ports. New client signals can be added and multiplexed in the STM-N frame. ADMs are mainly used in SDH ring networks.

Digital crossconnects add more flexibility. VC connections can be switched between all ports. Depending on the type of equipment, the switching can either be done at the HO

level, or at LO level. The client signals can also be extracted to and included from tributary ports.

2.2.4 Optical Transport Network (OTN)

In recent years advances in optical technologies paved the way to the development of optical transport networks. With Wavelength Division Multiplexing (WDM) technology optical signals are transported with different wavelengths (colors) over a single fiber. Today, more than 100 wavelengths with a bit rate of 10 or 40 Gb/s can be transported on a single fiber. Optical Amplifiers (OA) like the erbium doped fiber amplifier (EDFA) and the combination of different optical fibers to compensate for dispersion effects allows the transmission of optical signals (wavelengths) over several thousand kilometers without electrical signal regeneration [Knudsen-2001]. Finally, more complex network elements like optical add/drop multiplexers or optical crossconnects are being developed.

2.2.4.1 OTN Functional Architecture

The main functionality provided by an OTN is the transparent transport of optical client signals and optical channel networking and protection. The optical signals transported by the OTN are individual wavelengths using the wavelength division multiplexing (WDM) technique. The architecture of optical transport networks is defined in ITU-T Recommendation G.872 [ITU-T G.872] using the modeling methodology described in [ITU-T G.805]. The functional architecture contains the description of the OTN transport layer networks, the client/server relations, optical signal transmission, multiplexing, routing, supervision, performance management, and network survivability.

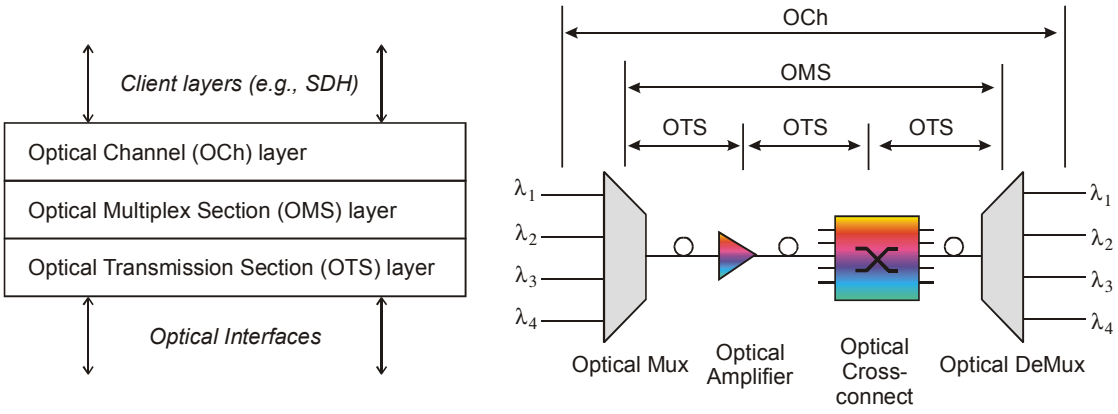


Figure 2.13: Optical layer network

Figure 2.13 shows the three layers defined for the optical transport network and their corresponding physical segments: the optical channel (OCh), optical multiplex section (OMS) and optical transmission section (OTS) layer.

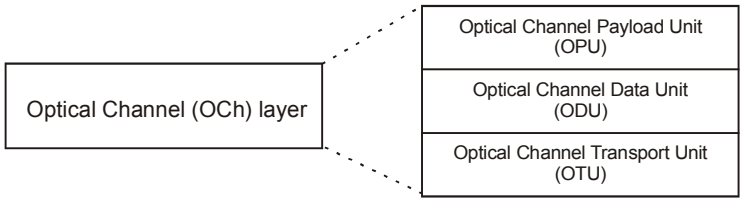


Figure 2.14: Optical channel sublayers

2.2.4.2 OTN Frame Structure

The optical signal interfaces for the OTN are defined in the recommendation [ITU-T G.709]. The optical channel is divided in three sublayers: the Optical Channel Payload Unit (OPU), the Optical Channel Data Unit (ODU), and the Optical Channel Transport Unit (OTU) (Figure 2.14). Each sublayer has a frame overhead. Figure 2.15 shows the octet-based frame structure of the OTU [ITU-T G.709] and its main elements. The frame structure is defined for three bit rates – OTUk, with k=1,2,3 – which corresponds to 2.5 Gbit /s, 10 Gbit /s and 40 Gbit /s, respectively.

The client signals (SDH, ATM or IP payload) are transported in the Optical Channel Payload Unit (OPU). In addition to the overhead bytes for the sublayers, the optical channel transport unit contains a frame alignment overhead and an OTU forward error control field. The detailed description of the fields is given in [ITU-T G.709].

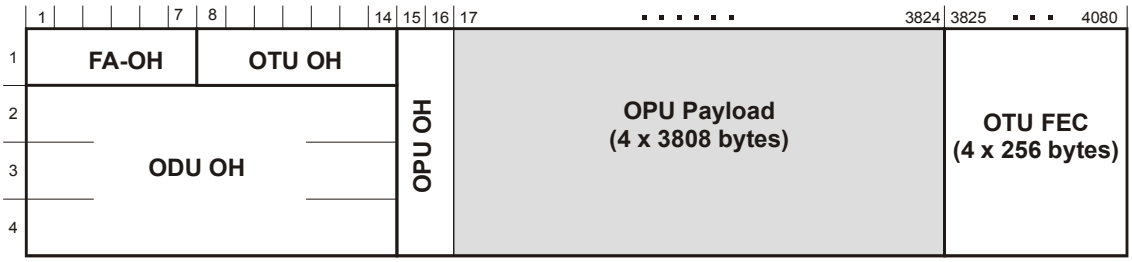


Figure 2.15: Optical channel frame structure

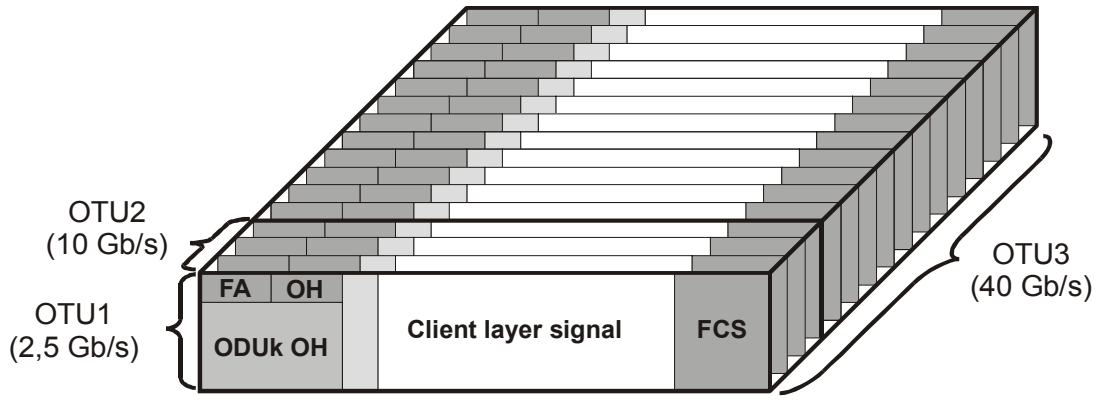


Figure 2.16: OPU, ODU, OTU signal mapping and multiplexing

Figure 2.16 illustrates the mapping of the client signal into the optical channel transport unit and the structure of the OTUk frame.

2.2.4.3 OTN Network Elements and Networking Concepts

The difference between optical networking and e.g. optical SDH signals transported over fibers is that in the former case flexibility is provided in the optical domain without opto-electrical signal conversion. The optical network elements offering this flexibility are optical add/drop multiplexers (OADM) and optical crossconnects (OXC). The complex network elements are composed of different optical components such as lasers, receivers, filters, splitters, couplers, amplifiers, switches and wavelength converters. A good introduction and overview of the optical components can be found in [Ramaswami-2002].

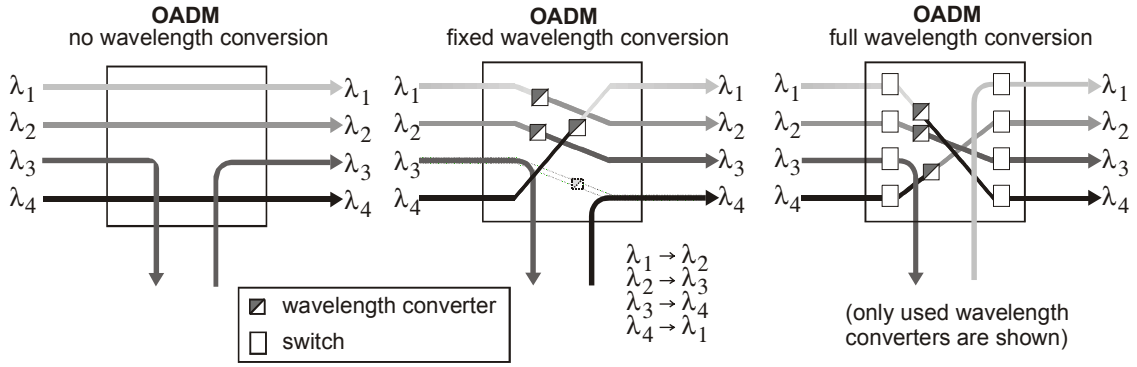


Figure 2.17: Optical add/drop multiplexer

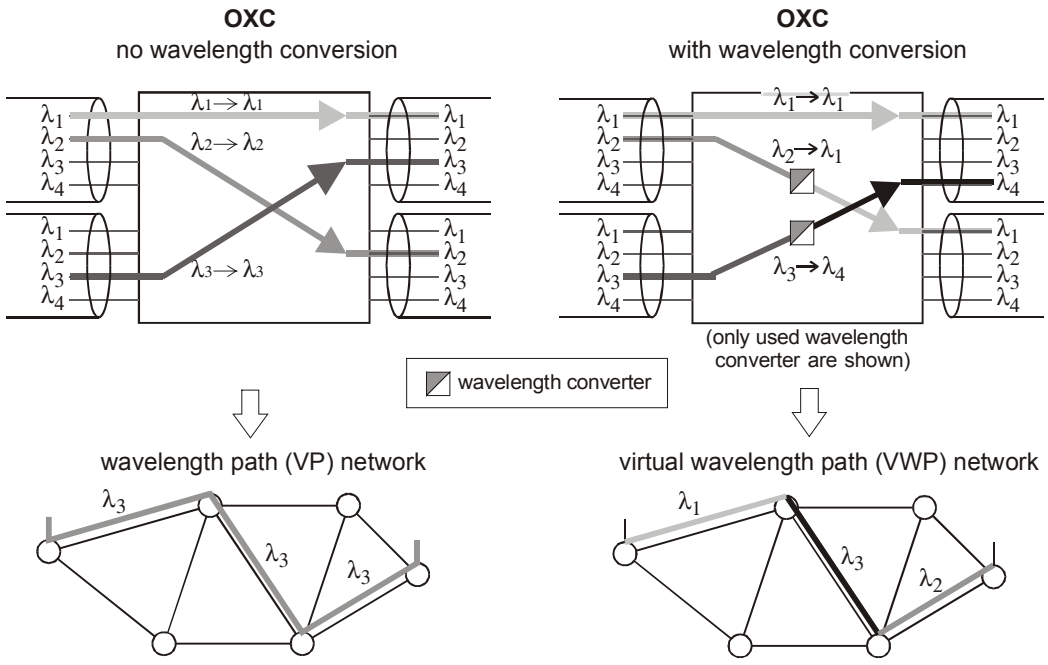


Figure 2.18: Optical crossconnects without and with full wavelength conversion

A main differentiation criterion for different types of optical network elements is the degree of wavelength conversion flexibility provided by the network element. Depending on the presence of wavelength conversion in the network, wavelength path (VP) networks and virtual wavelength path (VWP) networks can be distinguished (Figure 2.17 and Figure 2.18). In the former case an optical connection has the same color on the entire path. In the latter case, the wavelength (color) of the connection may change at every node by wavelength conversion.

In [Ramaswami-2002] optical network elements and their practical realization with optical components are presented in greater detail.

2.2.5 Simplified Network Model

The atomic functions and the decomposition in individual transport layer networks within a single technology are important for the specification and development of network elements to ensure the correct behavior of a network and interoperability between network operators. However, for the study of networking concepts often a simplified model of a network is sufficient and helpful.

The simplified model is based on the terminology used in [ITU-T I.311]. Additionally, the simplified model only shows the path layers of the different network architectures. Furthermore, only trails (or connections) and link connections (or links) are considered as transport entities. Link connections (or logical links, or links) connect network elements. The connectivity function provides flexibility within a network element. Trails (or connections) are set up by a concatenation of links between two trail termination functions. An arbitrary sequence of link connections and connections functions is called segment. Adaptation functions offer either the transfer of server layer services (native demand), or they map to a client layer link. Table 2.1 shows the relationship between the terminology used in [ITU-T I.311] and in [ITU-T I.326, ITU-T G.805].

Tranport Entities [ITU-T G.805]	**Simplified Network Model [ITU-T I.311]**
Trail	Connection
Network connection	–
Link connection	Link
Tandem connection	Segment [ITU-T I.610]

Table 2.1: Terminology used for transport entities in simplified network model

Figure 2.19 shows this simplified network model for an ATM VP over SDH HOP over OTN OCh layer (or ATM over SDH over OTN).

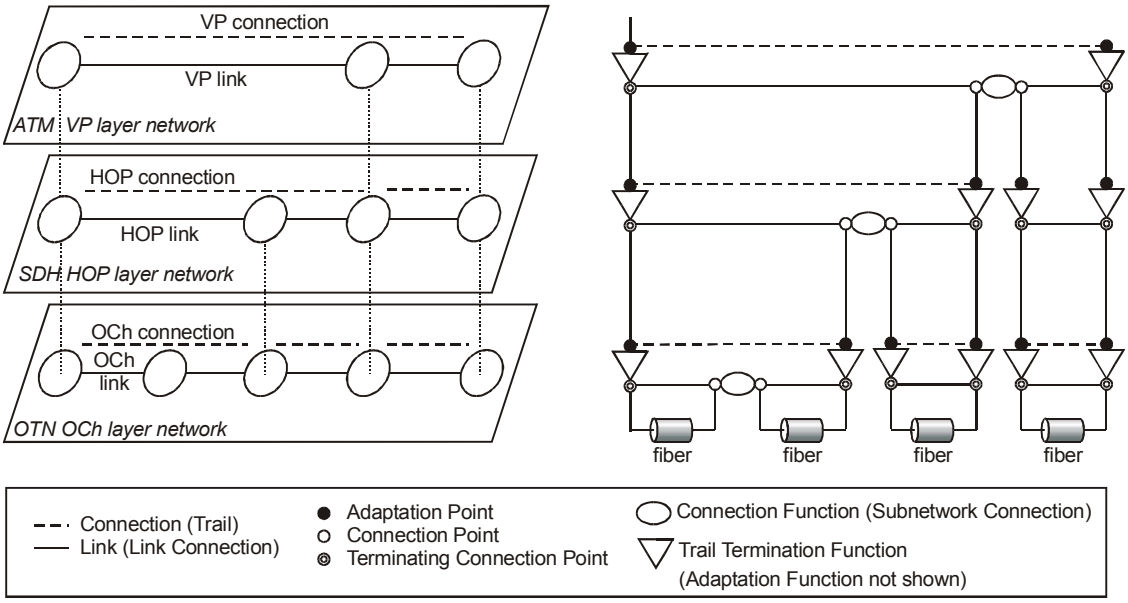


Figure 2.19: Simplified network model

2.3 The Internet and TCP/IP-Based Networks

In contrast to the ATM, SDH and OTN technologies described above, the Internet and networks based on the TCP/IP protocol family in general use an unreliable, connectionless packet switching technology. The protocols are defined and standardized by the Internet Engineering Task Force (IETF). In this section the architecture of TCP/IP networks is shortly introduced, and the used protocols are described.

2.3.1 TCP/IP Network Architecture

Like the transport networks already introduced, the TCP/IP architecture is decomposed into layers with a client/server relationship. For TCP/IP networks, five layers are defined: application layer, transport layer, network layer, data link layer and physical layer (Figure 2.20).

The application layer contains various applications such as file transfer, email, and telnet. The transport layer is providing the end-to-end connectivity between hosts. The Transmission Control Protocol (TCP) uses a reliable data transport [RFC0793], while the User Datagram Protocol provides only an unreliable (unmonitored) transmission. The Real-Time Transport Protocol (RTP) is used for time sensitive data transmission.

The packets are transported in the network layer using the Internet Protocol (IP). IP provides a best-effort connectionless packet transmission service. Additional routing protocols in the network layer provide the routing and connectivity information.

Since TCP/IP was developed to be independent of the physical media, the data link layer and the physical layer are not specified by the IETF. Various transport network technologies such as satellite and radio links, wireless networks, frame relay, ATM,

SDH and nowadays OTN can be used with additional mapping, framing and encapsulation procedures.

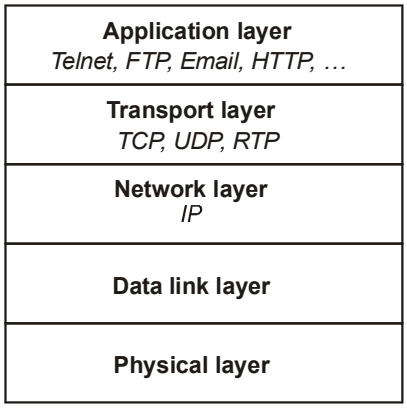


Figure 2.20: TCP/IP network layers

The Internet is the most prominent TCP/IP based network. The success of the Internet is based on the simple concepts used in TCP/IP protocols. The standardized interface between the application layer and the network layer allows a fast and easy application development independent of standardization bodies. The IP layer can be served by many different data link and physical layers with a wide range of bit rates – from modem links with few kb/s up to optical fiber links with many Gb/s on a single wavelength.

2.3.2 IP Packet Format

There are two versions of the IP protocol currently defined and in use: IP version 4 (IPv4), which is described in [RFC0791], and IP version 6 (IPv6), which is described in [RFC2460].

The IP protocol descriptions in the RFCs contain the packet format used for transmission of the client data, and routing and forwarding functions based on the IP addresses. The packet formats of both versions are compared in Figure 2.21.

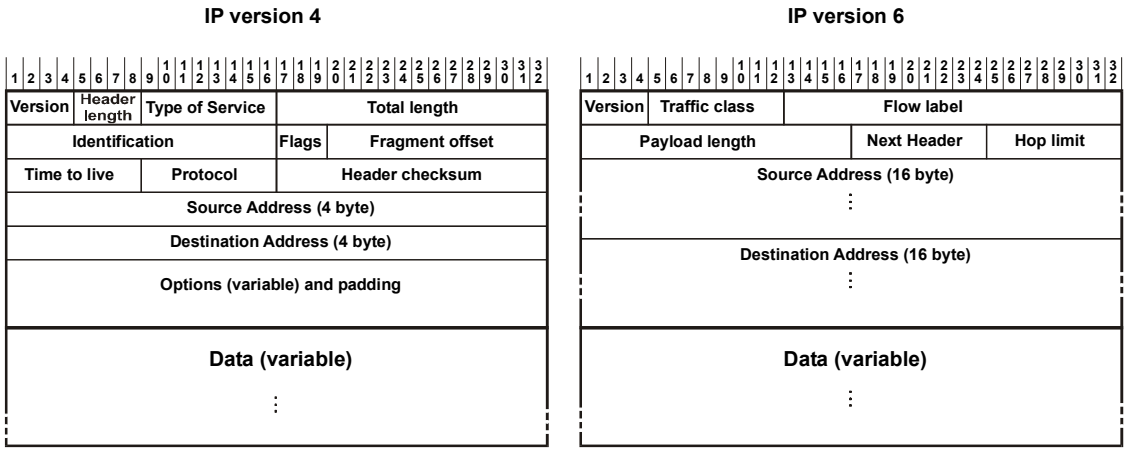


Figure 2.21: IPv4 and IPv6 packet format

The description of the header fields can be found in [RFC0791] for IPv4 and in [RFC2373] for IPv6. A major difference is the size of the address fields. While IPv4 uses 4 byte addresses, the source and destination addresses in IPv6 have a length of 16 bytes.

The IPv4 address field is divided in a host address part and a network address part. There are three unicast classes defined: Class A networks with 7 bits for the network address and 24 bit for the user address. Class B networks use 14 bits for the network address and 16 bit for the host address. Finally, class C networks use 21 bit for the network address and 8 bits for the host address. The excess bits in the 32 bit address field are used for the identification of the network class.

Since this classification proved very inflexible and made routing very difficult when the Internet grew in size, the Classless Inter-Domain Routing (CIDR) [RFC1518, RFC1519], was developed. CIDR replaces the rigid classification in three network classes with a more flexible address prefix, which defines the size of the network address. More information on the current behavior of IP addresses can be found in [RFC2101].

2.3.3 IP Routing

The Internet is not a homogenous network under a single administrative domain. It consists of many cooperating interconnected networks, so-called autonomous systems. An autonomous system itself is a group of networks under a single administrative domain. Backbone (core) IP networks have national or international extend, and they provide transit services over high bit rate links. Regional and local networks provide access services to users, corporate and campus networks. They mainly transport local traffic rather than transit traffic. The link bit rates are usually smaller compared to backbone networks. Figure 2.22 gives a high level logical view of a typical Internet architecture with two access/regional networks (AS100, AS200), and one backbone network (AS300). A good overview of the Internet architecture and its interaction with the public telephone network can be found in [ANSI TR55].

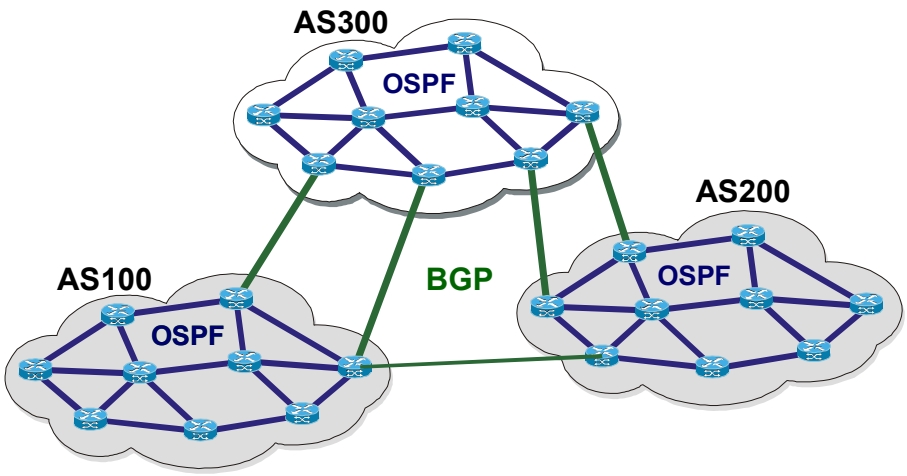


Figure 2.22: High level logical view of a typical Internet architecture

The IP routing is based on the IP destination address contained in the IP packet. The routing is a connectionless hop-by-hop forwarding of the packets. When a packet arrives at a router, the router looks for the destination address in its routing table and sends the packet to the next hop according to the information found by the routing table lookup.

The routing protocols either operate within an AS (intra-domain routing) or between ASs (inter-domain routing).

The preferred Interior Gateway Protocol (IGP) currently in use in the Internet is OSPF (Open Shortest Path First) [RFC1583]. OSPF is a link-state routing protocol. Every router periodically sends hello messages to neighboring nodes (typically every 10 seconds). The hello packets are used to automatically detect adjacent routers and to monitor the links to them. Changes in the topology are broadcasted to all routers using Link State Advertisements (LSA). With the topology information contained in the LSAs a router knows the topology and link costs of the whole AS and can calculate the routing table. In case of OSPF this is done by calculating the shortest path to all destinations using the Dijkstra algorithm.

The de-facto Internet standard for inter-domain routing is BGP (Border Gateway Protocol) version 4 [RFC1771]. BGP uses TCP as reliable transport mechanism and supports Classless Inter-Domain Routing (CIDR). BGP is based on an incremental path vector protocol that is similar to a distance vector protocol. At router initialization TCP connections are set up to the neighbors and the routing tables are synchronized. Topology update information is sent to neighbors using route announcements and withdrawals. The routing message contains information which address prefixes can be reached over an AS, the complete AS path (sequence of passed ASs) and the next hop address. The complete AS path is included in the routing information to suppress loops. The route selection is based on a longest prefix match, shortest AS path and "pre-configured policy information". Figure 2.23 illustrates the BGP routing table with the next hop entry and the AS path attribute.

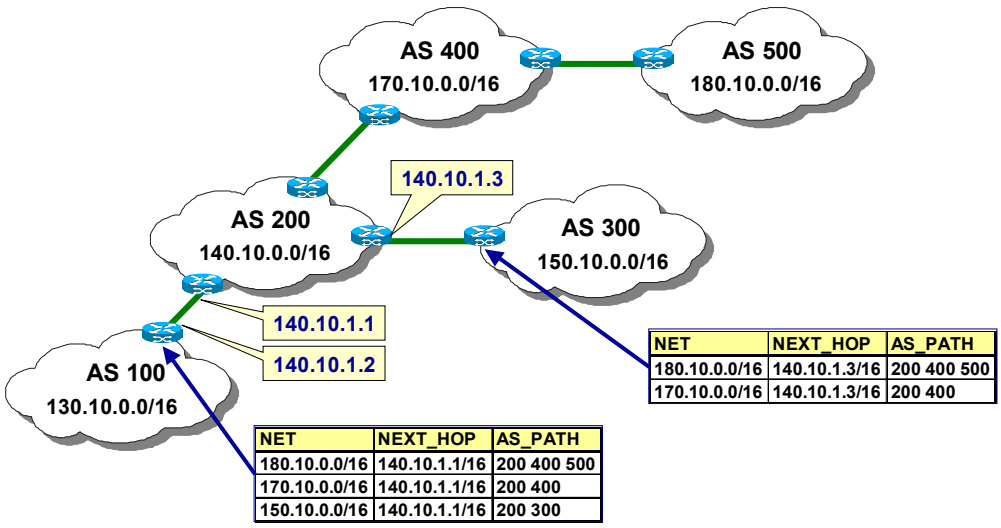


Figure 2.23: BGP routing table example

2.3.4 Support for Quality of Service in IP

The Internet was developed as a robust network with good reachability of individual hosts and reliable services even with unreliable network elements (like routers or links). The only service quality offered was 'best effort', where all packets are treated equally and no guarantee is given how, when, and if packets of a traffic flow arrive at their destination.

With the tremendous growth of the Internet, additional services with new requirements are offered. Real-time services like video-conferencing or IP telephony have a connection-oriented character and require high quality of service (QoS) in terms of delay, jitter, throughput, and packet loss.

To support the quality of service requirements of real-time, connection-oriented services, two QoS models were defined by the Internet Engineering Task Force (IETF): the Integrated Services (IntServ) architecture [RFC1633] with the Resource Reservation Protocol (RSVP) as a signaling protocol [RFC2205], and the Differentiated Services (DiffServ) architecture [RFC2475]. Multiprotocol Label Switching (MPLS), Traffic Engineering and Constraint Based Routing are also supporting QoS and QoS architectures and will be described in subsequent sections.

QoS model	**Integrated Services**	**Differentiated Services**
QoS method	Resource reservation	Traffic prioritization
Granularity	Traffic Flow	Traffic Aggregate

Table 2.2: Integrated Services and Differentiated Services classification

XiPeng Xiao gives a comprehensive survey of the different QoS strategies in the Internet in [Xiao-1999] and defines a more detailed framework in [Xiao-2000-b]. However, as an exhaustive description of all the functionalities of IntServ/RSVP and DiffServ is not needed here, only those characteristics of the QoS models that are required for the concepts presented in this thesis are summarized in the following sections. Table 2.2 shows a classification of the two QoS models based on the QoS method and the traffic granularity.

2.3.4.1 Integrated Services and RSVP

The Integrated Services model is based on a **per flow resource reservation**. Using the Resource Reservation Protocol RSVP the QoS requirements of the services are signaled through the network for individual flows and the required network resources are reserved.

In IntServ, two traffic classes are defined. Guaranteed Services [RFC2212] correspond to a constant bit rate virtual circuit with fixed delay bonds and reserved bandwidth. Controlled Load [RFC2212] services define an average delay, but no fixed limit on the delay of individual packets is given.

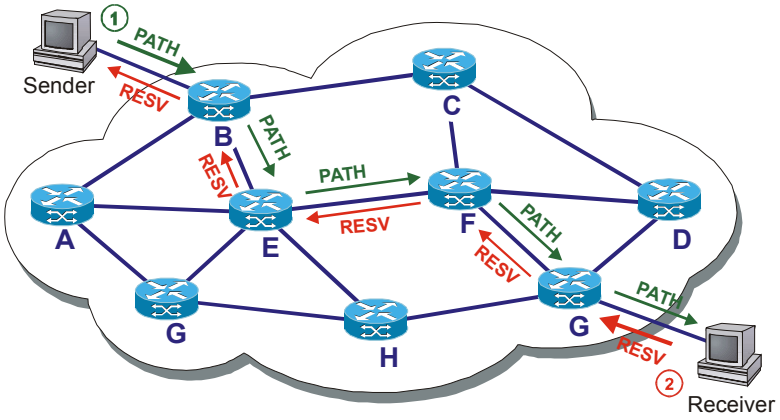


Figure 2.24: RSVP PATH and RESV signaling

Figure 2.24 illustrates the signaling sequence to set up an IntServ flow. The sender initially sends a RSVP PATH message containing the traffic flow specification (TSpec) with upper and lower bounds for delay, jitter, throughput and packet loss. The receiver responds with a reservation message (RESV) containing a request specification (RSpec) with the required traffic class – Guaranteed or Controlled Load service. The RESV message is sent along the reverse route of the PATH message. In every intermediate router the required resources are explicitly reserved using a soft state protocol. Refresh messages are periodically sent to update the flow state. The reserved resources are released if no refresh message is received within a certain period or if the flow is explicitly torn down.

The requirement to maintain states for every end-to-end flow imposed serious scalability problems on IntServ. Therefore, IntServ is mainly used in corporate and access networks but cannot be used in the Internet backbone. However, RSVP is a versatile signaling protocol, which was re-used for the setup of MPLS Label Switched Paths (LSPs).

2.3.4.2 Differentiated Services

The Differentiated Services (DS) architecture realizes IP QoS by the **prioritization** of different services on a hop-by-hop basis. Packets are classified and conditioned at the network boundary and assigned to a behavior aggregate. The behavior aggregate is identified by bit-patterns in the DS field in the IP header, so called DS code points (DSCP). The DS field is located in the IPv4 TOS octet or IPv6 Traffic Class octet. Figure 2.25 compares the TOS byte definition with the DS field.

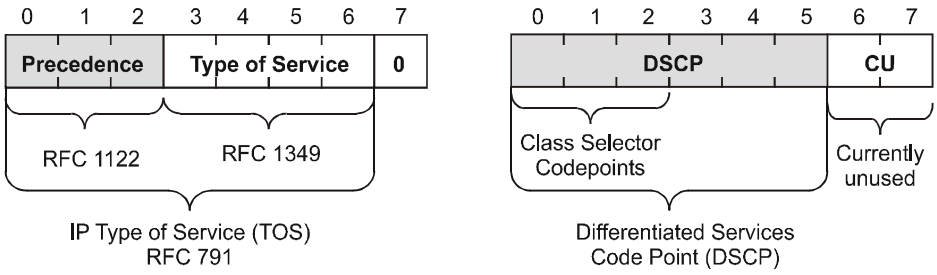


Figure 2.25: DiffServ DS field

The packet classification at the network boundary (Figure 2.26) involves processing multiple fields of the packet header, namely source address and destination address, protocol ID and source and destination port numbers. The packet is assigned to a corresponding behavior aggregate identified by a specific DSCP value (marking). The traffic conditioning includes metering, shaping and dropping. Metering means measuring the packet flow statistics like flow rate and burst size. Packet shaping tries to smooth out bursts by adding delay to the packets in buffers. If the packet stream is out of profile, the router may drop one or more packets.

In the DiffServ architecture, only the edge router performs the complex packet classification and traffic conditioning (metering, marking, shaping, dropping). At the core routers no policing occurs and the router only forward packets according to their Per-Hop Behavior (PHB) defined by the DSCP value. An Expedited Forwarding (EF) PHB [RFC2598] as well as a group of Assured Forwarding (AF) PHBs [RFC2597] are already defined with corresponding code points.

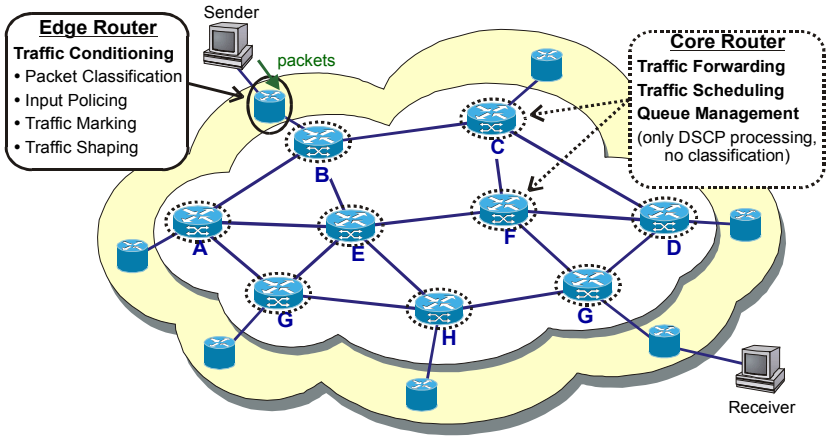


Figure 2.26: DiffServ packet classification, metering, marking and shaping

2.4 Multiprotocol Label Switching (MPLS)

Multiprotocol Label Switching (MPLS) integrates layer 3 routing and layer 2 switching functionalities [RFC3031]. MPLS is rapidly becoming a key technology for the use in core networks. MPLS introduces connection-oriented characteristics into IP by replacing the routing of IP packets (based on the IP header information) with a switching based on a short four-byte label. The technology is independent from the layer 2 technology used, and several implementation proposals have been made, e.g. for ATM, Frame Relay, and SDH/SONET. MPLS was designed to provide an elegant solution to present shortcomings of IP routing in the area of traffic engineering, QoS, virtual private networks (VPN), and resilience. In Figure 2.27 the main components of MPLS are illustrated. They are described in the following section.

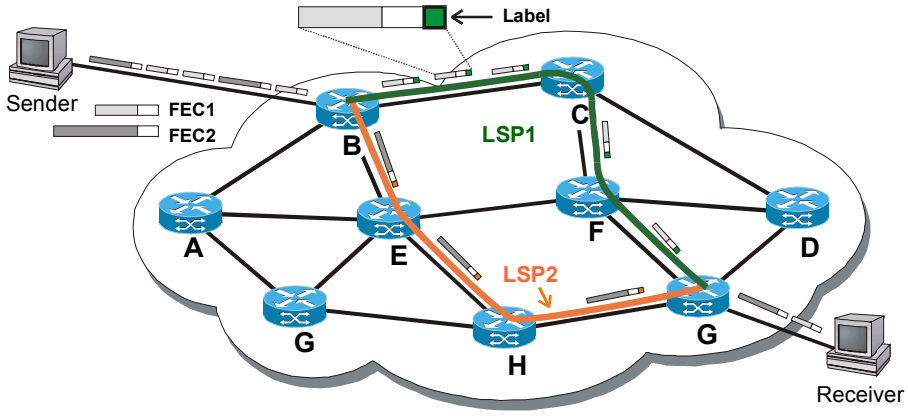


Figure 2.27: MPLS overview

2.4.1 Labels

The MPLS label is a short, 4-byte identifier added to the IP header at the Ingress Label Edge Router (I-LER) according to its Forwarding Equivalence Class (FEC) [RFC3031].

The label field contains the 20-bit label value (Figure 2.28). The 3-bit Exp field is for experimental use. Multiple labels can be added to the IP header as a last-in, first-out label stack [RFC3031]. The S-bit identifies the last entry in the label stack (bottom-of-stack). The TTL field is used to encode a time-to-live value. A more detailed description of the label encoding exceeds the scope of this thesis. For a more detailed description of the label fields and the processing of these fields refer to [RFC3032]. The label encoding (label format) is depending on the underlying transport technology and defined in [RFC3032].

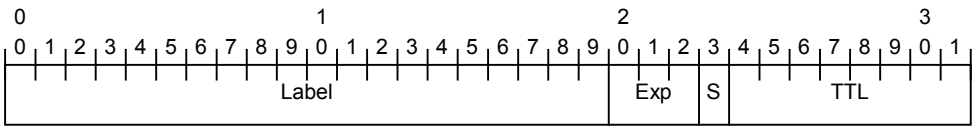


Figure 2.28: Format of the label field

The path that an IP packet follows through the network being defined by a label sequence is called a Label Switched Path (LSP) [RFC3031]. A Label Switched Router (LSR) uses a forwarding table to switch incoming packets according to their label and incoming interface to an outgoing label and interface. Each hop assigns a new label when forwarding the packet to the output port. This is called label swapping. With the concept of label stacking, tunneling and nesting of LSPs is possible. The Egress Label Switched Router (E-LSR) removes the last label from the label stack.

The binding of a label to a particular FEC may be done either data-driven or control-driven. ISPs will probably prefer a control-driven label binding, since it is more scalable and allows the deployment of traffic engineering methods in the IP network.

The label assignment may be based on forwarding criteria such as the destination address, traffic engineering and QoS requirements, or to realize multicast and virtual private network (VPN) services.

2.4.2 Signaling Protocols

To setup an LSP a signaling protocol is needed that coordinates the label distribution and (explicitly) routes the LSP. Additional (and optional) functions are the capacity reservation, the re-assignment of resources and the pre-emption of existing LSPs. Important protocol requirements are loop prevention and fault detection. The MPLS architecture doesn't mandate or even recommend a specific signaling protocol. Different signaling protocols are possible for different scenarios. The signaling can also be done "piggyback" via IP routing protocols like OSPF and BGP.

The most common signaling protocols used for MPLS are the Label Distribution Protocol LDP [RFC3036] with its extensions for Constraint-based Routing (CR-LDP) [RFC3212], and the Resource Reservation Protocol (RSVP) [RFC2205] with its traffic engineering extension (RSVP-TE) [RFC3209].

The setup is done either on a hop-by-hop basis, where each intermediate LSR defines the outgoing label and the output port based on the FEC for itself, or the LSP is set up at the source node using explicit routing.

An important feature for the setup of alternative paths is constraint-based routing (CR). CR takes parameters such as link characteristics (bandwidth, delay, etc.), hop count, and QoS into account in addition to a single cost metric like traditional routing mechanisms such as OSPF. The LSPs that are established with CR are termed CR–LSPs, where the constraints could be explicit hops or QoS requirements. Explicit hops dictate which path is to be taken. QoS requirements dictate which links and queuing or scheduling mechanisms are to be employed for the flow.

When using CR, it is possible that a longer (in terms of cost) but less loaded path is selected. However, while CR allows increased network utilization, it adds more complexity to routing calculations, as the path selected must satisfy the QoS requirements of the LSP. CR can be used in conjunction with MPLS to set up LSPs. The IETF has defined a CR–LDP component to facilitate constraint-based routes.

CR-LDP and RSVP-TE are very similar in their functionality. Which signaling protocol will be used is mainly dependent of the preference and proficiency of the equipment manufacturer.

2.4.3 MPLS Traffic Engineering

Traffic engineering is a process that enhances overall network utilization by attempting to create a uniform or optimized distribution of traffic throughout the network [RFC3272]. An important result of this process is the avoidance of congestion on any one path. It is important to note that traffic engineering does not necessarily select the shortest path between two devices. It is possible that, for two packet data flows, the packets may traverse completely different paths even though their originating node and

the final destination node are the same. This way, the less-exposed or less-used network segments can be used and differentiated services can be provided.

In MPLS, traffic engineering is inherently provided using explicitly routed paths [Awduche-1999, Awduche-2001]. The LSPs are created independently, specifying different paths that are based on user-defined policies. However, this may require extensive operator intervention. RSVP-TE and CR–LDP are two possible approaches to supply dynamic traffic engineering and QoS in MPLS. [RFC2702] defines requirements for traffic engineering over MPLS.

2.5 Layering Scenarios for IP over Optical Networks

In section 2.2, the functional architecture and network models of ATM, SDH and OTN have been introduced, and section 2.3 and 2.4 summarize the TCP/IP network architecture including QoS and MPLS. In this section, multilayer scenarios to support IP over optical networks are discussed.

The integration of IP over optical networks is a key research issue and several publications discuss integration strategies for IP over optical network (e.g. [Ghani-2000, P918-D1, P918-D2, Metz-2000]. Figure 2.29 shows possible layering scenarios. Only those layers are depicted which will be considered in the thesis.

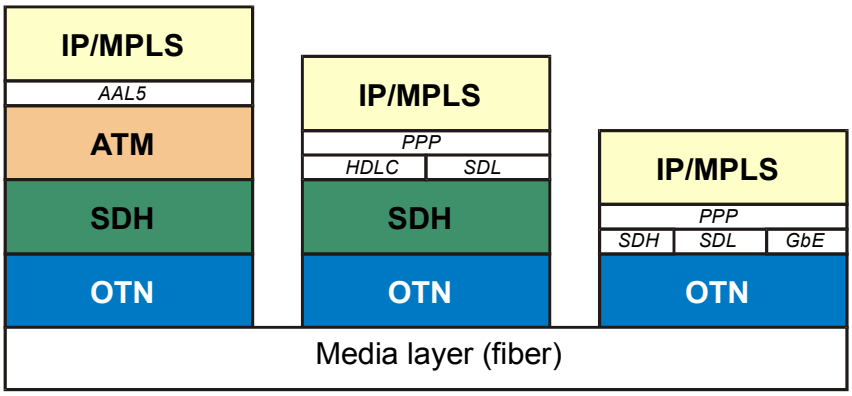


Figure 2.29: IP over optical network layering model

In the following paragraphs, the different layering scenarios will be shortly characterized, and their benefits and problems discussed.

2.5.1 IP over ATM over SDH over OTN/WDM

In the full protocol stack of IP over ATM over SDH over OTN the IP packets are segmented into ATM cells using the ATM Adaptation Layer (AAL). The ATM cells are in turn inserted into SDH frames. The SDH frames are then transported over the optical transport networks using the frame format presented in section 2.2.4.2.

An alternative IP over ATM implementation is to use MPLS over ATM, where the ATM virtual channels are set up by the MPLS protocol and the ATM VCI represents the label.

ATM allows to setup virtual channels with different capacities within virtual paths connecting to routers. This supports QoS by assigning a fixed capacity to individual customers, and of course it directly enables virtual private networks.

A drawback however is that in addition to the cell overhead the mapping of variable length IP packets in fixed size ATM cells imposes an additional fragmentation overhead, a so-called cell tax. Another problem in a layering scenario with multiple network technologies is the complex network management. It is difficult to integrate network management systems for different network technologies, and the networks are often operated by different network operators. The network cannot easily adapt to short-term changes in the network demands. Moreover, the configuration of the ATM virtual channels is complex and time-consuming, since the number of VCs grows with the square of the number of routers.

2.5.2 IP over SDH over WDM

In the second layer scenario IP is transported directly over SDH without an intermediate ATM layer. This reduces the management overhead and avoids the ATM cell tax. This scenario is also referred to as Packet over SONET (POS). For IP over SDH layering, PPP encapsulation and HDLC framing is used [RFC1661, RFC1662]. PPP (point-to-point protocol) is a standardized protocol to setup point-to-point IP links and to transport IP packets over different types of media, ranging from analogue phone lines to SDH. PPP requires framing to indicate the beginning and ending of the encapsulation. HDLC (High level Data Link Control) provides framing by adding a starting and delimiting flag and additional header fields to the PPP packet (Figure 2.30). The bit sequence of the HDLC flag must be escaped in the entire frame between the two flags.

Figure 2.30: PPP in HDLC-like framing [RFC1662]

As an alternative to HDLC the Simple Data Link (SDL) protocol can be used for framing. SDL replaces the start and end flag with a simple header containing only a packet length field and a header CRC (cyclic redundancy check), followed by the information field and an optional packet CRC.

Figure 2.31: SDL framing [RFC1662]

2.5.3 IP over OTN

To further reduce the management overhead, omitting the SDH layer yields an IP over WDM layer scenario. Different framing methods are possible for IP over WDM:

- Slim SDH framing, where only the SDH frame but no networking functionalities of SDH are used
- Pure SDL framing directly over an optical layer
- Gigabit Ethernet framing.

Depending on the control plane integration of the IP and OTN network and on the routing approaches, three different interconnection models can be distinguished: the overlay, peer and augmented model. Table 2.3 compares the three models.

Model	Control plane	Routing
Overlay model	Separate	Separate routing instances where no information is shared between domains
Augmented model	Separate	Separate routing instances with exchange of full reachability information
Peer model	Integrated	Integrated routing (single routing domain)

Table 2.3: Overlay, augmented and peer model comparison

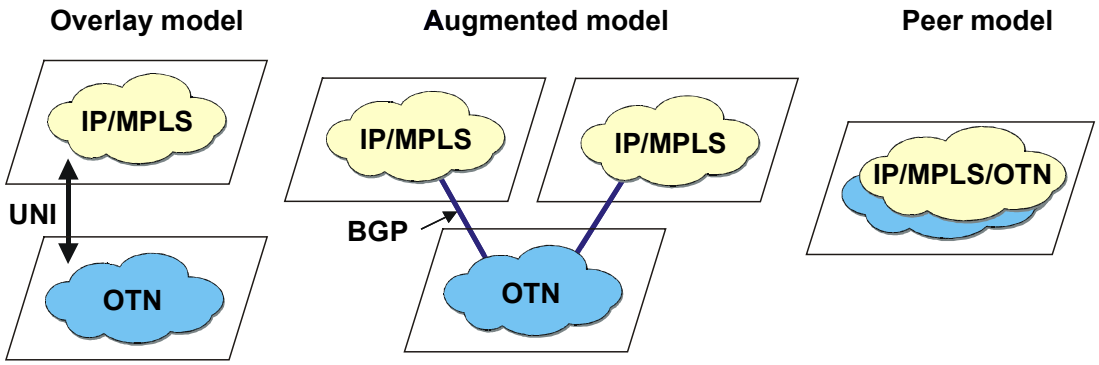


Figure 2.32: Overlay model, augmented model and peer model

In the overlay model (Figure 2.32, left side) the IP and the OTN layers are independent networks with a client/server relationship. The OTN layer offers transport functionality through optical lightpaths (OCh) to the client IP routers. Each layer uses its own routing algorithm and connection setup signaling protocols. The control signaling between both layers is done using a User-Network-Interface (UNI). This corresponds to a classical transport network layering and allows an easy integration of other client services (ATM, SDH). Two example architectures for the peer model are the Automatically Switched Optical Network (ASON) [ITU-T G.808] or, more general, the Automatically Switched Transport Network (ASTN) [ITU-T G.807].

In the peer model the IP network has full topological view of the optical network and a single routing instance is used in both networks (Figure 2.32, right side). The peer

model uses an integrated control plane for both layers. This model reaches the tightest integration of the IP and OTN network. The network operation is very difficult if different service providers own the IP network and the optical network. The Generalized Multiprotocol Label Switching (GMPLS) architecture [Draft-Mannie] is an example for the peer model.

The augmented model (Figure 2.32, middle) is a trade-off between the overlay model and the peer model where both networks have separate control planes and routing instances, but full topological information is exchanged between the networks. Thus the optical network is seen as an intermediate autonomous system between two IP domains.

2.6 Summary

In this chapter an introduction has been given to the transport network architectures, which are considered in this work. The objective of Chapter 2 is to have a common understanding of the technologies and networking concepts involved and to have a precise definition und understanding of the used terminology. A major attention was given to the functional modeling of the network architectures, and to the layering of the multiple network technologies. In addition, the general characteristics and signal formats of the involved technologies were specified. The adoption of networking concepts like Quality of Service and Traffic Engineering in IP transport networks was discussed. Finally, the layering scenarios and the control architecture of IP over optical networks currently discussed in the IETF and in standardization bodies were presented.

After the introduction to the architecture of the networking technologies considered in this thesis, in the next chapter the focus is put on network resilience concepts. Since many publications already exist on the recovery mechanisms in various layers, the objective of this chapter is to define a generic framework that integrates the resilience concepts of various technologies, and includes recovery strategies for multiple failures as well as multilayer recovery interworking aspects.

3 INTEGRATED MULTILAYER RESILIENCE FRAMEWORK

3.1 Introduction

The amount of traffic carried on single network elements like fibers or cross-connects is steadily increasing. Already the failure of a single element can cause a tremendous damage for the customers and loss of revenue for network operators. The damage in the reputation can even threaten the competitiveness of network operators in the current dynamic telecommunication environment. Moreover, natural disasters and terrorist attacks threaten the integrity of the network infrastructure. Therefore, highly resilient networks, which detect and recover network failures in a distributed manner, gained an increased interest.

In recent years, several recovery mechanisms for different transport technologies were developed. A common characteristic of those recovery mechanisms is that they only take a single transport technology into consideration. Current transport networks however are composed of multiple transport technologies working together in a client/server relation. Recovery mechanisms in different layers may interfere with each other disturbing their correct operation. Redundant protection, which is the protection of higher layer recovery resources in a lower layer, leads to a high network cost. The ACTS project PANEL investigated and evaluated advanced multilayer survivability concepts and interworking strategies making use of existing single layer recovery concepts to achieve reduced network costs and higher network survivability [*Demeester-1997*].

For the development and evaluation of the multilayer recovery strategies and interworking mechanisms an integrated multilayer resilience framework must be defined. According to [RFC2702] resilience is defined as *'...the behavior of a traffic trunk under fault conditions. That is, when a fault occurs along the path through which the traffic trunk traverses. The following basic problems need to be addressed under such circumstances (1) fault detection, (2) failure notification, (3) recovery and service restoration.*'

The integrated multilayer resilience framework has to take additional aspects into account. First of all, the operator requirements and objectives for resilience and the performance metrics to evaluate the resilience must be defined. A second step is to define faults and failures in a network and to specify considered failure scenarios. Then, the fault detection, failure notification and service restoration concepts and options are classified. These resilience techniques must be extended to multilayer scenarios, and the interworking of the different layers must be examined.

Summarizing, the multilayer resilience framework defined in this chapter includes

- Definition of network survivability performance parameters
- Definition of network operators requirements and objectives
- Definition of network failures and considered failure scenarios
- Failure detection and signaling techniques, including recovery trigger definitions
- Classification of single layer recovery mechanisms and options

- Extension to cover multiple failures
- Extension to multilayer resilience strategies

In the following sections the resilience framework used in this thesis to evaluate existing and novel recovery strategies is defined. The resilience framework and the used definitions are mainly based on ITU-T, ANSI-T1 and IETF standardization documents [ITU-T X.641, ANSI TR68, RFC2702, Draft-Sharma, Draft-Owens]. The multilayer resilience framework is based on the studies and results of the PANEL project [*Demeester-1997*].

3.2 Resilience Requirements and Performance Metrics

3.2.1 Requirements and Objectives

In this section the resilience requirements for multi-service IP over optical networks are discussed. First, the requirements are defined from an operator's perspective, user's perspective and services' perspective. Then, high-level resilience requirements and objectives are defined. A good overview of network survivability requirements with a focus on MPLS recovery is contained in [Draft-Owens]. Additionally, this IETF draft also considers recovery mechanisms that are present in multiple transport network layers (ATM, SDH/SONET, and Optical).

The resilience requirements of transport networks can be regarded from three different perspectives: the operator's, service's, and the user's perspective [*Transinet-D1*].

3.2.1.1 Operator's Perspective

Fast and Reliable Failure Detection

Failures in a network can be detected by a variety of mechanisms. A main requirement for networks with a high availability is fast and reliable failure detection. If a failure occurs, it is necessary to detect, notify and localize the failure to trigger the required recovery actions such as protection switching or rerouting [Draft-Willis]. The mechanisms to detect failures in a network will be described in Section 3.4.

Multi-Layer Recovery

With recovery functionality being available at multiple layers, faults could be detected and appropriate recovery action be triggered at the most suitable layer for obtaining best recovery performance. In reality, however, so-called race-conditions are encountered [Gerstel-2000-a] and uncoordinated recovery in multiple layers may lead to multiple service hits and unnecessary recovery operations [Autenrieth-1998-c].

Failure Coverage

Most of today's transport networks are designed to deal with a single failure at a time, but not with multiple failures. However, multiple failures become more probable as the complexity of the equipment and the size of networks increase [Gerstel-2000-a,

Schupke-2001-a]. Therefore, survivability mechanisms should be able to cope with multiple failures. This can be achieved by partitioning the network into smaller protection domains and using improved mesh protection and restoration mechanisms. Re-computing alternative paths and multilayer recovery are possible strategies to counter multiple failures in a network [Schupke-2001-a].

Cost and Resource Efficiency

Equipment redundancy and spare resources account for a large share in the total network cost. Spare capacity requirements can be more than 200% of the working capacity in case of ring protection mechanisms [Johnson-1996, Grover-2000]. Therefore the careful planning and optimization of the network survivability is a prime requirement of network operators to increase the spare capacity efficiency and to reduce network cost.

Recovery Speed

The timing requirements for the failure recovery and the completion of the service restoration largely depend on the affected application and services. The recovery should be fast to minimize the impact on the individual services.

Protection Selectivity

To optimize network resources and to take the different recovery requirements of services into account, protection selectivity should be provided. Traditionally only two classes of resilience are offered: fully protected traffic and unprotected traffic. More differentiated approaches use multiple protection classes. [Gerstel-2000-a] defines a set of 5 protection classes for traffic, which

- must be protected by the server layer (e.g. unprotected client layers)
- must not be protected (e.g. traffic protected in client layers)
- is indifferent to protection (e.g. IP traffic since resilience mechanisms wouldn't interfere)
- has best-effort protection
- has low priority (using spare capacity under normal conditions and may be preempted by resilience mechanisms)

In [*Autenrieth-2001-a*] a set of four resilience classes is defined primarily based on the recovery time requirements of the services.

Scalability

Protection schemes need to be designed in a way to allow the operator to scale efficiently from initial small systems to large-scale services. As a result, recovery schemes should be designed without inherent network size limits. Of course, the cost should also scale with the size of the network. The initial cost for the startup phase with limited network size should be low with reasonable additional costs for network growth.

Monitorability

When committing to certain performance figures in service level agreements (SLAs), network operators need to monitor the agreed performance metrics [Gerstel-2000-a] to

trigger protection mechanisms in case of irregularities, and to verify the fulfillment of their part of the contract.

The monitoring is well defined and easily feasible in traditional fixed-bitrate or hierarchical-bitrate systems. However, bitrate-transparent services like dark wavelength or fiber can constitute severe troubles for network monitoring.

3.2.1.2 Services' Perspective

Recovery Speed, Guaranteed Rerouting Time

Client services impose requirements on the recovery speed and thus the rerouting time of failed transport services. Table 3.1 shows the impact of the restoration time on various services (based on [Kawamura-1998, ANSI TR68]).

Restoration time	**Service outage impact**
0 to < 50 msec	Service "hit," reframing required
50 msec to < 200 msec	Potential voiceband disconnect (< 5%) Effect cell rerouting process
200 msec to < 2 sec	May drop voiceband calls depending on channel bank vintage
2 sec to < 10 sec	Call-dropping (all circuit switched service) Potential packet (X.25) disconnect Potential data session timeouts
10 sec to < 5 min	Packet (X.25) disconnects, data session timeout
5 min to < 30 min	Network congestion, minor social/business impact
> 30 min	Major social/business impact

Table 3.1: Restoration time impact on customers [Kawamura-1998, ANSI TR68]

Note that TCP total timeout has values of some minutes, e.g. 2 Min. (Solaris) or 9 Min. (default) [Stevens-1994].

QoS Awareness

QoS Awareness allows variable service availability requirements for different traffic types [T'Joens-2000]. Multimedia (voice, video, etc) and selected transaction oriented services require fast recovery with very low outage times, thus leading to a high availability, while adaptive traffic such as bulk data transfers tolerate outage times in the range of seconds to minutes.

Multi-Service Support, Service Granularity

The convergence of voice and data networks led to more complex networks with a variety of services. Thus the network has to cope with the different requirements and characteristics of multiple services. Also the resilience mechanisms must support the requirements of these individual services with a high service granularity.

Recovery Granularity

Recovery mechanisms at different network layers and sublayers have different recovery granularities ranging from individual MPLS LSPs with bit rates in the order of Mb/s to whole optical fibers with many wavelengths and bit rates in the order of n · 10 Gb/s.

3.2.1.3 Users Perspective

Service Availability (Connectivity)

Requirements for the completion of a protection switch largely depend upon the applications [Gerstel-2000-b]. On the other hand, users rather favor satisfying connectivity on a guaranteed level.

Service Quality

In service level agreements (SLAs), users and network operators agree upon a set of performance figures. As the performance in case of failures (e.g. disruption, survivability) is critical to specific types of services, the protection classes mentioned above should also be part of the agreement.

3.2.2 Definition of Resilience Performance Parameters

The terms to express the resilience and network survivability performance requirements defined in this section are resilience, survivability, availability, mean time between failures, and mean time to recovery. The definitions of these performance metrics are taken from standardization documents.

The main focus is put on the requirements related to network survivability performance, although it is influenced by the quality of service requirements. The basic concepts of the services are similar to the definitions for network performance like mean time between interruptions, mean interruption duration, and network accessibility.

When not specifically stated otherwise, the resilience metrics are taken from ITU recommendations [ITU-T E.800, ITU-T X.641] and from [Cisco-1999].

3.2.2.1 Resilience

In [ITU-T X.641] the resilience characteristic is defined as the ability to recover from errors. The resilience characteristic is quantified as a probability.

In [RFC2702] a similar but more detailed definition is given: The resilience attribute determines the behavior of a traffic trunk under fault conditions. That is, when a fault occurs along the path through which the traffic trunk traverses. The following basic problems need to be addressed under such circumstances: (1) fault detection, (2) failure notification, and (3) recovery and service restoration.

The latter definition of resilience will be used throughout the thesis.

3.2.2.2 Network Survivability

Network survivability is the ability of a network to maintain or restore an acceptable level of performance during network failures by applying various recovery techniques [ANSI TR24]. The 'Network Survivability Performance' is an assessment, how well the network is fulfilling its performance under abnormal conditions [ANSI TR68]. The network survivability is analyzed and outage indexes are calculated for various transport network technologies in [ANSI TR68].

3.2.2.3 Mean Time Between Failures MTBF

The Mean Time Between Failures (MTBF) is the average expected time between successive failures of a mature item, assuming the item goes through repeated periods of failure and repair. The MTBF is the reciprocal value of the failure rate λ:

$$MTBF = 1/\lambda$$

A useful interpretation of MTBF is that within the period of MTBF 63% of the product/system population is expected to have failed at least once. For example, target MTBF values of an individual circuit card range between 100,000 hours and 200,000 hours. The larger the MTBF, the better is the product. For long MTBF values and comparatively short repair times, the MTBF can be approximated by the time to failure, instead of the time between failures. As a reminder, the time to failure is defined as the duration of the operating time of an item, from the instant of time it enters an operational state, until the next failure.

3.2.2.4 Mean Time To Repair MTTR

The Mean Time To Repair (MTTR) is the average expected time interval in which an item is in a down state due to a failure. The MTTR is also called Mean Down Time (MDT). The repair is the event when an item regains the ability to perform a required function after a fault due to physical repair actions, e.g. the splicing of a broken cable, or the replacement of a defect module. Thus, MTTR includes the failure detection time, fault diagnosis, and fault isolation, trouble ticketing, repair team deployment, the actual repair, and performance test.

The MTBF and MTTR values of network equipment are important to analyze the availability of a network by calculating the system availability based on the network element availability (see 3.2.2.6). Secondly, the availability values are used for event driven simulators to randomly generate failure and repair events based on the network elements' MTBF and MTTR values.

The Mean Up Time (MUT) or Mean Time To Fail (MTTF) is the expected interval during which an item is in an up state. Thus, the MTBF is equal to the sum of MUT and MDT (or MTTR and MTTF). The Figure 3.1 illustrates the relation between the MTBF, MTTR, MTTF, MDT, and MUT.

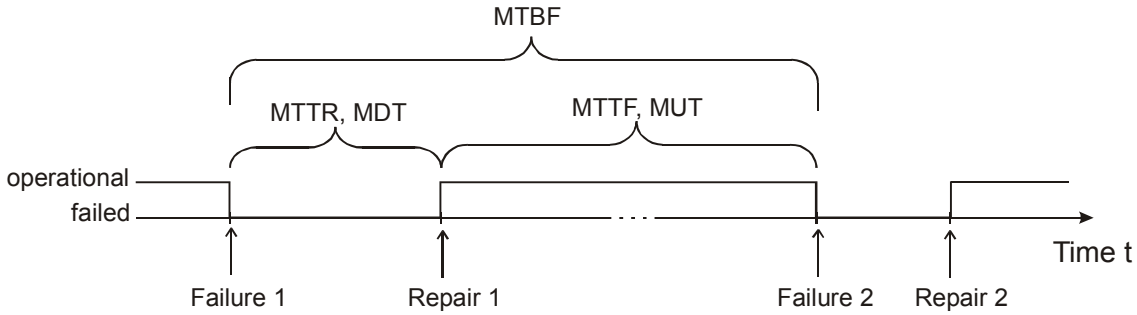


Figure 3.1: Relation between MTBF, MTTR, MTTF, MDT, and MUT

3.2.2.5 Mean Time To RecoVery MTTV

The recovery is the event when the item regains the ability to perform a required function after a fault due to protection or restoration actions. MTTV is usually one or more orders of magnitude lower than MTTR, thus improving the service availability also by orders of magnitude.

3.2.2.6 Availability A

Availability is the probability that an item will operate when needed, or, for mature communications equipment at a steady state, the average fraction of connection time that the item is expected to be in operating condition. For a communications system that can have partial as well as total system outages, availability is typically expressed as connection availability, as below [Cisco-1999]:

$$\text{Availability} = 1 - \frac{\text{Total connection outage time}}{\text{Total in-service connection time}}$$

$$= 1 - \frac{\text{No. of connections affected in outage i} \cdot \text{Duration of outage i}}{\text{No. of connections in service} \cdot \text{Operating time}}$$

In [ITU-T X.641] the availability characteristic is defined as the proportion of agreed service time that satisfactory service is available. “Agreed service time” means the aggregate time over which it is agreed between service-users and service-provider that service is to be provided. The availability metric is quantified as a probability.

Relationship between Availability, MTBF and MTTR

Knowing the mean up-time and the mean down-time, the availability can be calculated by the following relationship:

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

The MTBF and the MTTR are easier to measure for an operator. Measured over a long period of time the MDT can be approximated by the MTTR, since the *time to restore* is identical to the *down time* if the down state is caused by a failure and not by a preventive maintenance action. Similarly, when the down state is caused by a failure and not by a preventive maintenance action the *time between failures* is identical to the sum of the

Up Time and the *Down Time*. Thus, over a long measurement period the sum of MUT and MDT can be approximated by the MTBF. The availability can under these conditions be expressed as

$$A = 1 - \frac{MTTR}{MTBF}$$

3.2.2.7 Unavailability U

Unavailability is simply defined as

$$U = 1 - A$$

3.2.2.8 Downtime D

Availability and Unavailability are expressed as probability values between 0 and 1. Typical values of Availability are between 98% and 99.999%. However, these values are difficult to comprehend. The downtime D is expressed in minutes per year and is defined as the expected time that a product does not operate (is down) per unit of in-service time. The downtime is a different way to express availability. The relation between availability and downtime is indicated in the following equation:

$$D \text{ (in minutes per year)} = (1 - A) \cdot 525600$$

The number 525600 is calculated by 365 days/year · 24 *hours/day* · 60 minutes/hour. Downtime is a more convenient and intuitively comprehensible variable to evaluate compared to availability. Typical values of downtime range from 50 minutes (~ A= 99.99%) to 5 minutes (~ A= 99.999%) or even to half a minute (~ A= 99.9999%) (see Table 3.2).

Availability	Downtime D per year
99%	5256 minutes (3.65 days)
99.9%	525.6 minutes (8.76 hours)
99.99%	52.56 minutes
99.999%	5.256 minutes

Table 3.2: Relationship between availability and downtime

3.2.2.9 Defects Per Million DPM

Defects per million (DPM) is defined as the number of lost calls per million of processed calls. DPM is particularly useful for measuring the availability of switched virtual circuit (SVC) services in a multiservice switch, where connections are constantly established, sustained, and torn down.

For permanent switched (PVC) services, DPM sometimes is defined as the number of defective (outage) connection minutes per million connection minutes in service. Thus, for a mature item, availability is related to DPM by the following equation:

$$\text{DPM (for PVC)} = (1 - A) \cdot 10^6$$

Thus, an availability of 99.999% relates to a DPM of 10.

3.2.2.10 Failures In Time FIT

Failures In Time (FIT) is a measure of failure rate and is defined as the numbers of failures per 10^9 operating hours. FIT is typically used to describe the reliability of components and circuit cards. For $\lambda = 10$ FIT, the MTBF can be calculated to MTBF = $1/\lambda = 100$ Million hours.

3.2.2.11 IP Performance Parameters

The ITU-T Recommendation I.380 [ITU-T I.380] defines IP service availability performance parameters. The IP service availability is defined as a threshold function of the IP packet loss ratio IPLR:

$$\text{Outage criterion:} \quad IPLR > c_l$$

$$\text{Threshold:} \quad c_l = 0.75,$$

In other words, the IP service is in an unavailability state, if the IP packet loss ratio exceeds a threshold value of 0.75, otherwise the IP service is categorized as available. Values of 0.9 or 0.99 have also been suggested for c_l. The IP packet loss ratio is defined as the ratio of total lost IP packet outcomes to total transmitted IP packets in a population of interest [ITU-T I.380]. The minimum time interval during which the availability function is to be evaluated is T_{AV}, which is provisionally defined as 5 minutes.

The IP service unavailability, here abbreviated with U_{IP}, is defined as the percentage of total scheduled IP service time that is categorized as unavailable using the IP service availability function.

The IP service availability, here abbreviated with A_{IP}, is defined as the total scheduled IP service time that is categorized as available.

$$A_{IP} = 1 - U_{IP}$$

3.2.2.12 Further definitions

Further definitions and a discussion of high level availability objectives can be found in [Cisco-1999, ITU-T E.800, ITU-T X.641, Iselt-1999].

3.3 Network Failures

Today's economy and society heavily relies on the communication services provided by telecommunication networks. Failures in this communication infrastructure may affect thousands of customers and destroy a huge amount of data. In contrast to common

belief, failures in the network are unfortunately a relatively frequent occurrence. The networks therefore must be made resilient against these failures.

Table 3.3 illustrates this by giving the number of outages reported to the US Federal Communication Commission (FCC). All outages must be reported that last longer than 30 minutes and affect more than 30.000 customers or special services like emergency calls (911) and air traffic control. Based on the number of outages, an upper bound for the MTBF can be calculated (since not all outages are covered by these statistics).

Period	1996	1997	1998	1999	2000
# failures	219	222	217	230	225
MTBF (days)	1.7	1.6	1.7	1.6	1.6
MTTR (min.)	> 30	> 30	> 30	> 30	>30

Table 3.3: Outages affecting more than 30.000 customers in USA or special services

3.3.1 Common Failure Types

In [T'Joens-2000] three major fault causes are defined: equipment failures, cable failures and human error. Equipment failures include hardware failures related to nodes (port failure, node failure, site failure, software failures), while cable failures include all failures related to links (fiber break, cable break, duct break).

Equipment failures of single modules are quite often, but the network nodes are usually well protected by equipment protection schemes or network recovery mechanisms. Complete node failures are less often, but affect large volumes of traffic.

The most common failure type in large transport networks are fiber breaks and cable cuts. According to [T'Joens-2000], long distance carriers experience between 1.2 and 1.9 cable cuts per 1000 km of cable per year, while local exchange carriers encounter about 3 cable cuts per 1000 km cable per year.

The network perceives failures due to human error as either equipment or cable failures. If a node is unreachable due to configuration error, the node is seen as down by the network. If a cable is connected to the wrong module, the link is seen as broken.

Additionally, failures may be categorized in soft failures with a slow quality of service degradation (e.g., due to ageing of the components) or hard failures due to fire or physical damage (e.g., cable break due to road works or node failures).

The main focus of this thesis is on hard failures of network elements (equipment and cables).

3.3.2 Multiple Failures

Today's transport networks primarily assure survivability of single failures at a time. As size, integration and complexity of these networks increase, multiple failures become

more probable [*Schupke-2001-b*]. Resilience schemes may be unable to survive certain combinations of simultaneous failures in the network. According to [Gerstel-2000-b], previously unusual, multiple failure types become significant caused by

- simultaneous resource failures
- hardware failures
- software bugs
- operator errors

Unfortunately, in recent history natural disasters and terrorist attacks have become an additional source of catastrophic failure scenarios.

In [*Schupke-2001-b*] it is shown, that the mean time of multiple failures per year may be in the range of 10 to 100 hours per year for a MTTR of 48 hours depending on the fiber duct MTBF. This translates to a probability of 99.88% to 98.858%, respectively. If the resilience mechanism of the network cannot cope with multiple failures, target values of 99.999% ('five nines') availability may not be met.

3.4 Failure Detection, Notification and Signaling

3.4.1 Failure Detection

A key requirement for performing recovery actions is to detect fast and reliably failures in the network. If a failure occurs, it is necessary to detect, notify and localize the failure to trigger the required recovery actions such as protection switching or rerouting [Draft-Willis]. The failures taken into consideration are a variety of hard network resource failures. The most frequent failures are cable breaks due to construction works or node failures due to power loss or fire. Other failures may be caused by maintenance work, e.g. unplugging a cable by mistake. A common problem in optical networks is the breakdown of a laser, which results in a Loss-of-Light (LOL) failure.

Failures in a network can be detected by a variety of mechanisms. The mechanisms can be distinguished in hardware failure detection by monitoring the signal quality, and software failure detection by inserting control messages in the signal flow.

In the following, some failure detection methods are discussed.

- Loss of Signal (LOS)

  The failure of an electrical link in most cases is first detected by the line card (port). To trigger a consequent recovery action the detected failure must be reported (notified) to the node's control plane. Upon receiving such a failure notification, the node can start a rerouting process or trigger the switching of the affected connections to a pre-configured alternative route.

- Loss of Light (LOL)

The failure of an optical link may be due to a laser failure or a fiber break. As for the case of the electrical LOS the failure must be notified to the node's control plane to trigger the necessary recovery actions.

- Operation, Administration and Maintenance (OAM) Flows

  In SDH and ATM OAM flows F1 to F5 are defined to monitor the availability of transmission sections. In SDH special bytes are reserved in the section overhead of the STM frame. In ATM OAM cells with special VPI and VCI values realize the OAM flows.

- Link Management Protocol (LMP)

  In the context of GMPLS a Link Management Protocol was defined to discover and monitor links. Among other functions, the protocol is able to detect link failures using a bi-directional out-of-band control signaling.

- Hello and KeepAlive signals

  As in traditional routing protocols such as OSPF or BGP4, Hello and KeepAlive messages are defined for MPLS signaling protocols to monitor the state of the adjacent nodes and the interconnecting links. RSVP uses a Hello message, while LDP uses a KeepAlive message. The loss of multiple (at least three) hello messages is required to reliably detect a failure. Because the time between these signals should be relatively long to minimize signaling load, the time to detect a failure using such signaling mechanisms is generally an order of magnitude longer compared to hardware or lower layer detection methods.

  An advantage of such signaling failure detection methods, however, is their ability to detect software and protocol failures, which cannot be perceived by the hardware lower layer.

3.4.2 Notification and Signaling

After a failure is detected at a node, it must be notified to other network elements to take appropriate actions.

- LSP error signaling and notification

  An important role for MPLS recovery plays the failure signaling and notification of LSP error. Failures are reported to the Ingress LSR when an already established LSP fails.

  While the LSP failure notification is not as fast as hardware failure detection, it can be directly used to trigger recovery actions.

- GMPLS Notify message

  In GMPLS the Notify message extends the LSP error signaling. The Notify message can be sent to any node responsible for the recovery of a failed LSP, and the message may contain additional information, e.g. about multiple failed LSPs.

- ATM OAM

Two alarm signals are defined in ATM. The Alarm Indication Signal (AIS) is sent in forward direction after the detection of a failure. The Remote Defect Indication (RDI) is sent in the backward direction from the destination node to the source node.

- MPLS-OAM

  A new approach to solve MPLS failure notification and signaling is proposed in the Internet drafts.

  In [Draft-Willis] the motivation and high level requirements for a user plane OAM (Operation, Administration and Maintenance) functionality in an MPLS network is defined, while [Draft-Harrison] defines the requirements and mechanisms to provide OAM functionality for MPLS networks.

  The main concept is to introduce a Connectivity Verification (CV) message to monitor the integrity of links and nodes and to trigger appropriate recovery actions if a failure is detected. The CV is sent periodically (nominal 1 per second) from LSP source to LSP sink [Draft-Harrison].

Additional signals are a Forward Detect Identifier "FDI" and a Backward Defect Identifier "BDI", which carry the defect type and location to the downstream and upstream node respectively [Draft-Harrison]. The document also defines the appropriate actions related to the server and client layers of the MPLS layer.

3.5 Generic Recovery Mechanisms and Options

In [ANSI TR68] recovery mechanisms are further classified depending on which layer they operate. Four layers are defined in this context: physical layer, system layer, logical layer and service layer. The *physical layer* includes the physical components and structures of the network, i.e. the ducts, cables, fibers, node sites (houses) and network elements. Survivability techniques for the physical layer are geographical diversity, redundancy (e.g., redundant power supply) and protection against physical damage (like fire). The *system layer* represents the network transmission systems, and terminating and full-rate interface equipment. Typical system layer components are STM-N transmission channels, Add/Drop Multiplexer and Terminal Multiplexer. The *logical layer* includes lower layer transmission systems (e.g., VC-12) and their interface equipment. The logical and the system layer can be combined to the transport layer. The *service* layer contains user service network such as voice and public and private data. The type of traffic transported in the service layer is telephone calls, data packets and cells. A typical survivability mechanism in the service layer is dynamic rerouting.

In the thesis the focus is on the transport layer, that is the system and logical layer. In the next sections the recovery mechanisms for the transport layer are classified.

3.5.1 Overview

Several categorization schemes exist to classify network survivability mechanisms. The most common classification is to divide recovery mechanisms into protection switching and restoration mechanisms. Protection switching mechanisms use predefined

alternative paths, while for restoration mechanisms alternative paths are calculated on demand after the detection of a failure. A detailed definition and explanation of the recovery mechanisms follows the following paragraphs.

For ATM a slightly different classification is defined in [ITU-T I.311]. ATM recovery mechanisms are classified into protection switching, rerouting and self-healing mechanisms. Rerouting mechanisms are restoration mechanisms with centralized control, while distributed restoration mechanisms are called self-healing.

To use an unambiguous naming scheme in the recovery framework the recovery mechanisms are divided in protection switching, (distributed) restoration, reconfiguration (centralized restoration), and rerouting (at the service level). Figure 3.2 summarizes all options of the recovery framework. The recovery mechanisms and options are described in the following sections. In this work the focus is set on protection switching and restoration mechanisms.

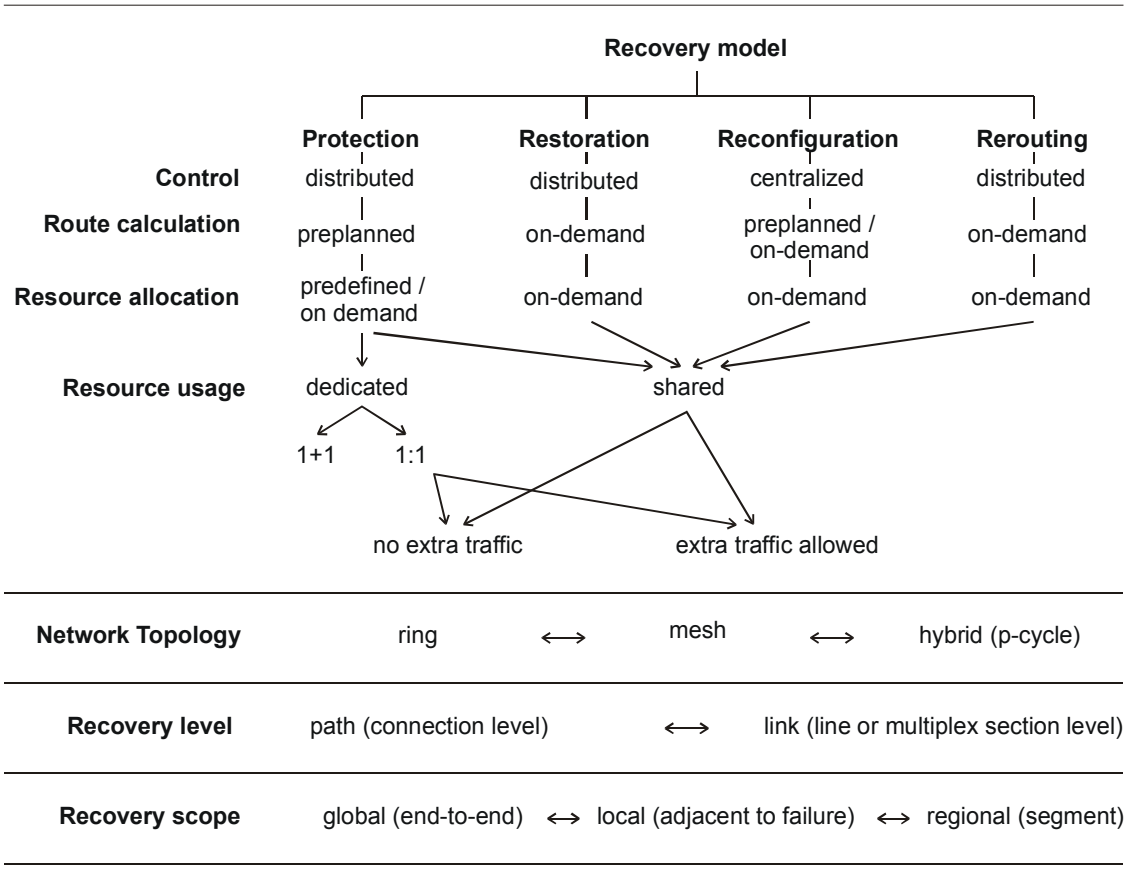


Figure 3.2: Recovery framework

3.5.2 Recovery Model

3.5.2.1 Protection Switching

In the case of protection switching, an alternative connection is pre-established and pre-reserved (pre-provisioned). Therefore, protection switching realizes the shortest disruption of the traffic, since no routing and resource allocation is required after failure detection. In the SDH standardization the maximum allowed switching time of protection switching mechanisms is defined to be 50ms [ITU-T G.841].

Depending on the recovery scope, the alternative connection is either switched at the source and target network element (global protection or *path protection*), or locally at the network element adjacent to the failure (local protection or *link protection*). Many publications deal with survivable network design using protection switching mechanisms [Wu-1992, Wu-1995]. A good overview of protection mechanisms is given in [Ramamurthy-1999-I]. Below the generic characteristics of protection mechanisms are given independently of a specific network technology.

Dedicated Protection

In case of dedicated protection, the protection resources are used dedicatedly to the corresponding working connections. There are two dedicated protection schemes: 1+1 (one plus one) and 1:1 (one for one) protection. In Figure 3.3 both dedicated protection schemes are compared. The primary path is called working or active path (a). The secondary, alternative path is called protection or backup path (b).

In 1+1 protection, the traffic is simultaneously transported over the working and protection path. In case of the failure, the target node only has to select the incoming traffic from the alternative path. With this combination, hitless recovery is possible. In [Iselt-1999] several protocols for hitless switching are analyzed.

In case of 1:1 dedicated protection, the traffic is switched to the backup path 'b' only after a failure is detected on the active path 'a'. Under normal conditions, the backup resources can be used for the transport of low-priority preemptive traffic, so-called extra traffic.

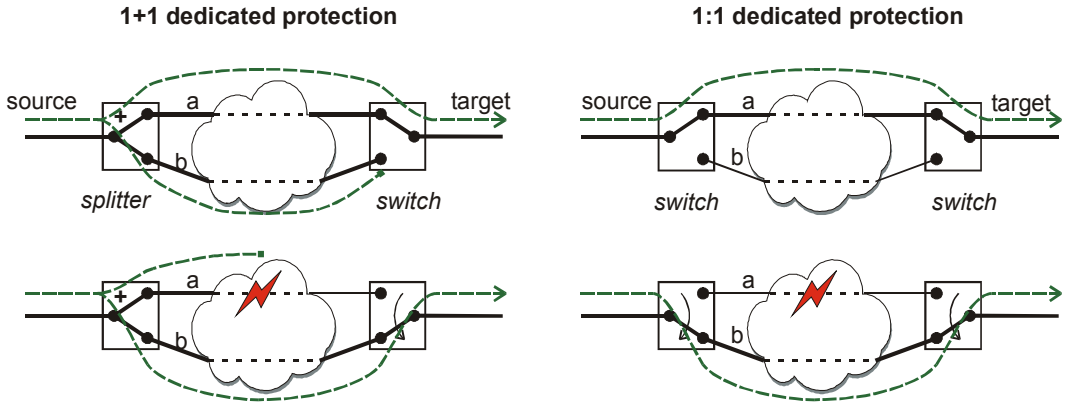


Figure 3.3: 1+1 and 1:1 protection switching

Shared Protection

With shared protection the spare resources are not dedicated for the recovery of a specific connection, but can be shared by multiple connections for different failure scenarios. Figure 3.4 illustrates the concept of shared protection.

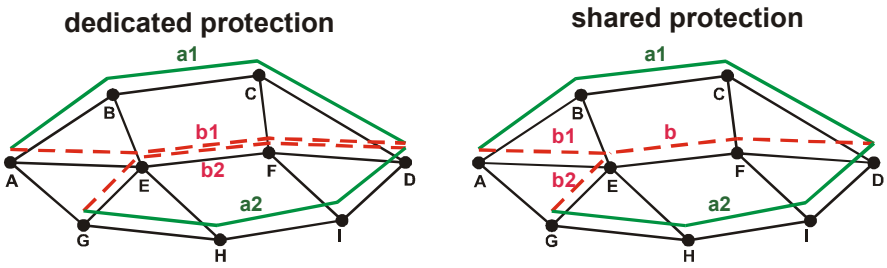


Figure 3.4: Dedicated and shared protection

On the link E-F the sum of the capacity of the two working connections A-B-C-D and G-H-I-D has to be reserved for the dedicated protection. In case of shared protection, only the larger capacity of A-B-C-D or G-H-I-D has to be reserved. In connections with equal capacity C are used, dedicated protection requires $2 \cdot C$ spare resources on link E-F and $6 \cdot C$ spare resources altogether for the protection of the two connections. With shared protection, the required spare resources are $1 \cdot C$ on the link E-F, and $4 \cdot C$ for the full connections.

Because of the sharing of the spare resource, shared protection has better resource efficiency than dedicated protection. On the other hand, it requires a more complex signaling mechanism for the activation of the alternative connection. Shared protection mechanisms are common for ring topologies, where the spare resources are provided by additional fibers used only for protection traffic and extra traffic.

3.5.2.2 Restoration

In the case of restoration, an alternative path is calculated and established on-demand after the detection of a failure. Since the calculation of alternative routes and the signaling and resource reservation of a new connection are time-consuming, restoration mechanisms are considerably slower than protection mechanisms. However, the restoration is also more resource efficient, since the spare resources can be used for the recovery of different working connections, provided these don't share the same working resources.

The recovery path is established using distributed restoration schemes after detecting the failure. There are several restoration mechanisms published, like the Selfhealing Network (SHN) [Grover-1987], FITNESS [Yang-1988], or RREACT [Chow-1993]. A good introduction to the characteristics of restoration mechanisms is given in [Ramamurthy-1999-II]. In general, restoration mechanisms search for a suitable backup path using distributed flooding mechanisms. Depending on the scope of the recovery mechanism, local or global, the node upstream of the failure or the source nodes of affected connections broadcast reservation messages on all outgoing links with enough spare capacity. When a broadcast message reaches the destination node, this node

responds with an acknowledgement message. Either the restoration is complete, when the acknowledgement message reaches the source node (2-phase algorithm), or the source node has to send a confirm message downstream to the destination node (3-phase algorithm).

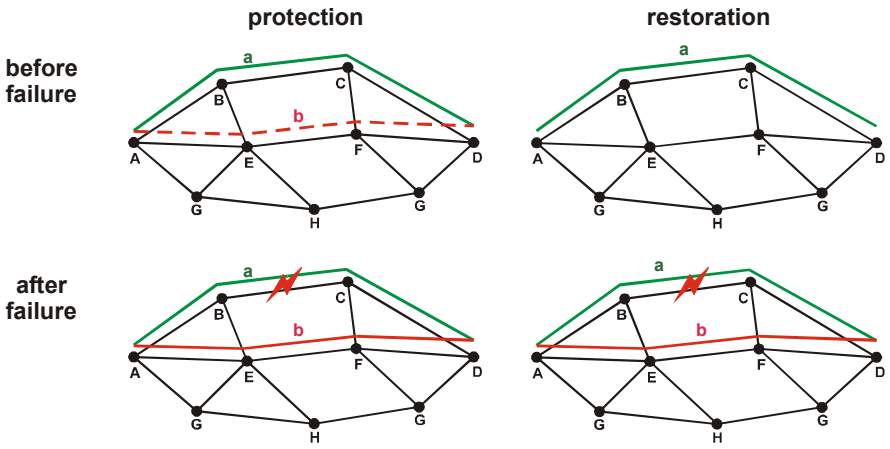


Figure 3.5: Protection and restoration

As an alternative to a flooding procedure, the upstream switching node can use a constraints based routing mechanism to calculate the full restoration route. A pre-requisite for this approach is that the nodes have a full view on the network including the available capacity on the links. Such a mechanism can be used for example in MPLS networks.

Since the calculation of new routes and the signaling and the resource reservation of a recovery path are time-consuming, restoration is considerable slower than protection mechanisms. However, restoration concepts are also more cost efficient, since the spare resources in the network can be shared for different failure scenarios.

3.5.2.3 Reconfiguration

The restoration mechanisms cannot achieve optimal resource efficiency, since they use distributed mechanisms, which haven't a full view of the network.

With centralized failure reconfiguration optimal resource efficiency can be achieved. The central network management has a full view of the current state of the network at the time of a failure. The state of the network includes the network topology and dimensioning, but also the full routing of working connections and the working and spare resources on each link. With this information, the centralized recovery can compute an optimized reconfiguration of the network.

The drawback of the centralized recovery is of course the slow recovery from failures in the time scale of several minutes. This is due to the fact that a node controller has a delay time to collect alarms and failures within the node and to filter the root failure. The filtered alarm messages must then be sent over a signaling network to the network management system (NMS). The NMS may receive alarm messages from many nodes,

and has to collect, filter and correlate these alarm messages to find the root cause of the failure. Only then the NMS can compute alternative routes for affected connections.

3.5.2.4 Rerouting

The rerouting mechanisms work at the service level. Rerouting, or dynamic rerouting, tries to reconnect a connection after it failed. In IP networks rerouting is an inherent capability of the hop-by-hop routing of IP packets. If a router detects the failure of a neighbor router or attached link, it re-computes its routing tables and forwards all subsequent packets on the alternative route. In addition, the changed network topology is notified to all adjacent nodes. In case of OSPF, this is done with so-called Link State Advertisements (LSAs).

3.5.3 Recovery Topology

The recovery mechanisms can operate in different network topologies. Ring networks are well suited for protection mechanisms, since they are the simplest form of a two-connected network. Recovery mechanisms for ring networks are described in [Eberspächer-1998]. One ring direction is used for the working traffic between source and destination, while the protection path is routed in opposite directions. In addition to two-fiber ring systems also four-fiber ring recovery mechanisms are possible. Protection mechanisms can also be used in mesh networks. Additionally, mesh network topologies also support restoration mechanisms. In [Grover-1998,Grover-2000], W.D. Grover introduced the concept of p-cycles, which is based on protection cycles (overlay ring structure) working in a mesh network, utilizing the advantages of both, ring and mesh topologies. In [*Schupke-2002*] the p-cycle concept is applied to WDM networks. Figure 3.6 illustrates the three recovery topology alternatives.

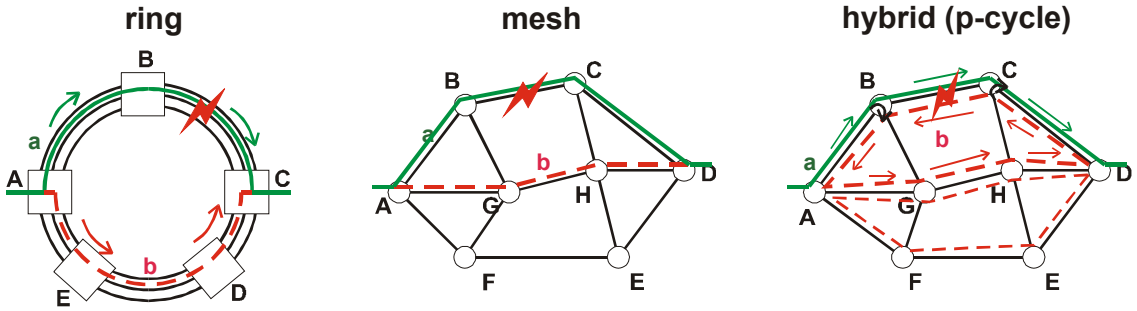


Figure 3.6: Recovery topology

3.5.4 Recovery Level and Recovery Scope

The *recovery level* refers to the vertical sublayer a mechanism is working at, the path layer or the multiplex section or line layer (also called link layer). The terms recovery level (path and link) and recovery scope (global, local, and regional or segment) are often used ambiguously.

The *recovery scope* refers to the horizontal, i.e. the geographical extension of the recovery mechanisms. In case of a local recovery scope the recovery switching is done in the nodes adjacent to the failure, while in global recovery the switching is done at the connection end-points (e.g. mesh restoration).

With *path protection*, each connection (e.g., a OCh carrying a lightpath) is switched at its endpoints. With *link protection* the multiplexed signal transmitted over a link (e.g. OMS with tens of wavelength) is switched at the same time (at multiplex section or line level). Therefore, *path protection* refers at the same time to the recovery level (connection) and to the recovery scope (global).

The commonly used terms *path restoration* and *link restoration* refer both to a recovery at *path level*; path restoration switches each affected connection at the connection endpoints, while in case of link restoration each affected connection is restored at the nodes adjacent to the failure. Therefore, the commonly used term link restoration in fact specifies a local restoration at path level.

To avoid ambiguity and to put emphasis on the difference between link protection and link restoration, in this work the terms *global, regional and local restoration* are used for a distributed, on-demand recovery of affected connections.

An intermediate recovery scope is the segment recovery. In Figure 3.7 a special case is shown where the protection-switching node (PSN) B is upstream of the failure and the protection merge node (PMN) D is the endpoint of the connection.

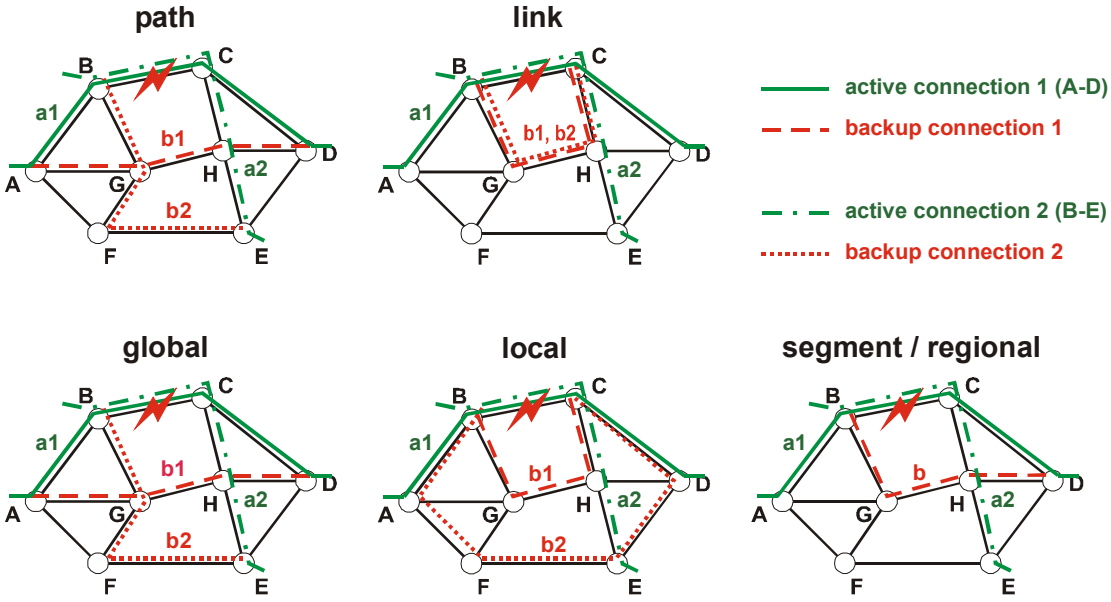


Figure 3.7: Recovery scope

3.5.5 Recovery Switching Operation Modes

3.5.5.1 Switching type: Unidirectional versus bidirectional switching

Two operation modes for recovery mechanisms exist for the recovery switching in case of failures affecting only one transmission direction. Such failures occur for example due to transmission laser outages.

With unidirectional switching, only the affected direction of the failed traffic is recovered in case of unidirectional failures. In case of bidirectional switching, both, the affected and the unaffected direction of traffic affected by a unidirectional failure are recovered. For bidirectional switching, a protection switching protocol is required to control the switching operation. For ATM and SDH networks, this protection switching control protocol is called Automatic Protection Switching (APS) protocol. In case of unidirectional switching, only the sink node controls the switching operation, so no switching protocol is required. Therefore, unidirectional switching is less complex to implement and can operate faster. Unidirectional and bidirectional switching are also termed single ended vs. dual ended switching, respectively. Figure 3.8 illustrates the two switching modes.

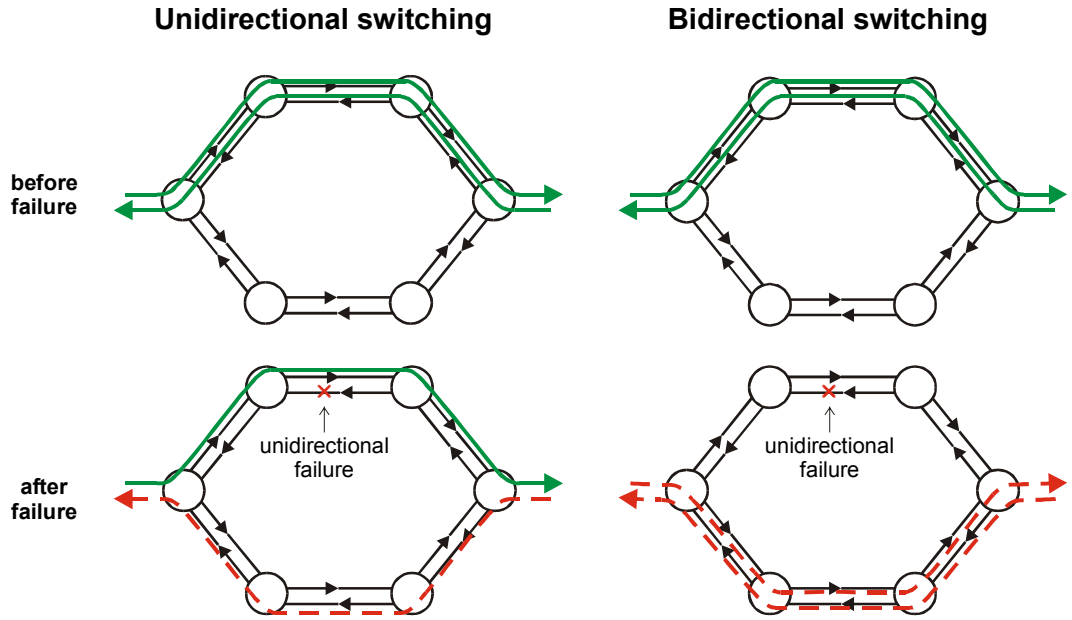


Figure 3.8: Unidirectional and bidirectional switching

3.5.5.2 Revertive / non-revertive operation

The revertive and non-revertive operations relates to the behavior of recovery mechanisms, after the network impairment (e.g., cable cut) is physically repaired.

In revertive operation, the recovered traffic is switched back to the original path automatically after the failure is repaired. In non-revertive operation, manual reaction is required to switch back the recovered traffic to the original path.

3.5.5.3 Extra-traffic supported

If in failure-free operation no traffic is transported over the spare resources, i.e. in case of 1:1 dedicated protection, shared protection and restoration, the spare resources may be used for low-priority, preemptive traffic. Supporting extra-traffic requires a sophisticated and more complex recovery protocol than recovery mechanisms without extra traffic support. The recovery protocol must take care that the extra traffic is preempted before the affected working traffic is switched to the spare resources. For ATM and SDH protection switching mechanisms, this preemption is performed by the APS protocol.

Since failures are a rare event in a network, most of the time the spare resources are not used. Therefore it is economically recommendable to use these network resources, in spite of the additional complexity of the recovery control protocol.

3.6 State of the Art of Recovery Mechanisms

A large number of recovery mechanisms for the different network technologies are standardized or published in conferences and journals. For conciseness reasons, an exhaustive description of all recovery mechanisms cannot be given here. Instead an overview of some selected mechanisms is given. The characteristics of the individual mechanisms are highlighted and the advantages and drawbacks of recovery mechanisms in the corresponding technology are discussed in general.

3.6.1 ATM Recovery Mechanisms

ATM recovery mechanisms with distributed control are classified and evaluated in [Edmaier-1996]. A more recent survey on architectures for ATM network survivability is contained in [Kawamura-1998]. Since 1999, ATM protection switching mechanisms are standardized by the ITU in [ITU-T I.630]. A special case for 1+1 protection switching for a cell-based physical layer, which allows hitless switching, is standardized in [ITU-T I.480]. In this section the main characteristics of ATM protection switching and restoration mechanisms are presented, and the failure detection and signaling methods specified.

OAM Functions

The defect and failure detection and notification and the activation of recovery mechanisms in the ATM layer are realized using specific OAM (Operation, Administration, and Maintenance) cells [ITU-T I.610]. According to the five layers of the B-ISDN reference model, five hierarchical OAM flows F1 top F5 are defined are defined (see Table 3.4).

OAM level	Network level	Network layer
F5	Virtual channel level	ATM layer
F4	Virtual path level	
F3	Transmission path level	Physical layer
F2	Digital section level	
F1	Regenerator section level	

Table 3.4: OAM levels

The bi-directional F4 and F5 management flows use in-band OAM cells, which have pre-assigned VCI (Virtual Connection Identifier) and PTI (payload type identifier) values. In case of SDH-based transmission, the F1, F2, and F3 flows are transported in overhead bytes. In addition to the vertical hierarchy, F4 and F5 flows can either cover an entire network connection (end-to-end flow) or a part of a connection (segment flow). Table 3.5 lists the ATM OAM cells defined in [ITU-T I.610].

OAM cell	Type	Function
CC	Continuity Check	Failure detection
LB	Loopback	Failure localization
AIS	Alarm Indication Signal	Failure notification
RDI	Remote Defect Indication	Failure notification
APS	Automatic Protection Switching	Recovery protocol

Table 3.5: OAM cell types

Protection Switching Mechanisms

ITU-T recommendation [ITU-T I.630] defines mechanisms for 1+1 and 1:1 bi-directional protection switching as well as 1+1 unidirectional protection switching. The description of the mechanisms includes the trigger mechanism, a hold-off time mechanism to delay recovery actions, and the protection switching control protocol. For the protection mechanisms a revertive and non-revertive operation is defined. The 1:1 operation is supported with and without extra traffic. The protection switching can be performed on either individual VP/VC connections, or on VP/VC protection groups. The VC/VP protection group alleviates the problem, that a large number of VP/VC connections can be affected by a single failure.

Figure 3.9 shows the temporal model to evaluate ATM protection switching performance [ITU-T I.630]. The model is based on a general temporal model for restoration times defined in [ITU-T M.495].

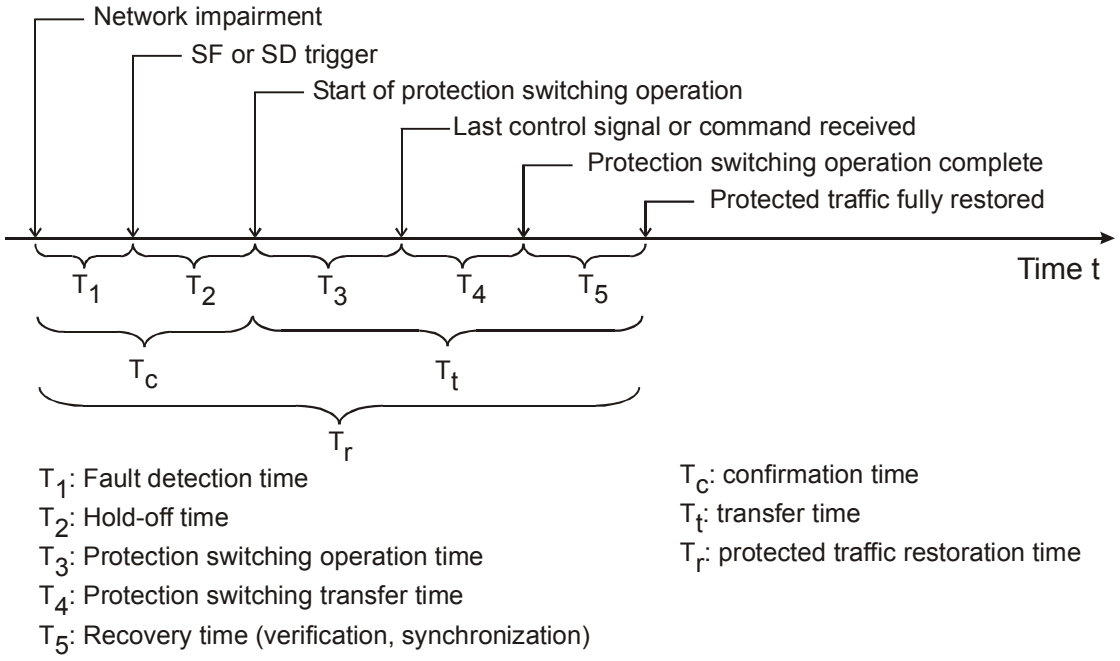


Figure 3.9: Protection switching timing model [ITU-T I.630]

In recommendation [ITU-T I.480] a 1+1 protection switching mechanisms is defined for cell-based physical layers (as opposed to SDH/SONET-based physical layers), which provides a hitless protection switching (without signal disruption) between two sources and sinks of F3 OAM flows. The hitless switching is realized by duplication and synchronization of the two signal flows using sequence numbers in the OAM F3 cell flow. At the sink node one signal is selected and forwarded. Hitless protection switching mechanisms and their synchronization algorithms are also studied in [Iselt-1999].

Restoration Mechanisms

So far no distributed restoration mechanism is standardized for ATM. However, a number of restoration mechanisms is studied and published in literature. The predominant property of ATM restoration mechanisms is the ability to pre-establish backup paths using zero capacity VPs (e.g, [Kawamura-1994, Anderson-1994]). Restoration schemes that use pre-defined backup VPs are called preplanned (Backup-VP) Self-Healing Network (SHN) schemes [Kawamura-1998]. In [Kawamura-1995-b] the preplanned SHN schemes are compared to SHN restoration schemes based on flooding mechanisms.

3.6.2 SDH and SONET Recovery Mechanisms

With the resilience framework defined in the previous chapters, SDH (and SONET) recovery mechanisms are categorized in protection and restoration mechanisms, which will be briefly discussed in the following.

3.6.2.1 Restoration Mechanisms

W. D. Grover first introduced the Self-Healing Network (SHN), which realizes distributed restoration mechanism for DXC-based networks [Grover-1987]. Since then a number of publications deal with algorithms for path or link restoration for SDH (or for circuit-switched networks in general) and evaluate the spare-capacity utilization of different approaches [Chow-1993, Herzberg-1995, Iraschko-1996, Yang-1988]. For example, in [Iraschko-1996] the authors show that path restoration achieves 19% lower spare capacity utilization than link restoration. The algorithms to find suitable restoration paths were introduced in Section 3.5.2.2.

3.6.2.2 Protection Mechanisms

SDH protection mechanisms are standardized in [ITU-T G.841]. Detailed descriptions of the mechanisms can be found in [Wu-1992, Ramaswami-2002, Wu-1995, PANEL-D2a]. Table 3.6 gives an overview of the most common protection mechanisms defined by the ITU-T and by ANSI-T1. The recommended target for the restoration time of SDH/SONET protection switching mechanisms is 50ms [ITU-T G.841].

ITU-T	ANSI T1	Level	Topology
MSP (1+1, 1:1, 1:N)	Line protection	Span (MS)	Point-to-point
2-fiber/4-fiber MS-SPRing	BLSR	Link (MS)	Ring
SNCP/mesh	Path protection	Path (HOP, LOP)	Mesh
SNCP/ring	UPSR	Path (HOP, LOP)	Ring

Abbreviations

APS	Automatic Protection Switching
BLSR	Bidirectional Line Switched Ring
HOP	Higher Order Path
LOP	Lower Order Path
MS	Multiplex Section
MS-DPRing	Multiplex Section Dedicated Protection Ring
MS-SPRing	Multiplex Section Shared Protection Ring
MSP	Multiplex Section Protection
UPSR	Unidirectional Path Switched Ring
ULSR	Unidirectional Line Switched Ring
SNCP	Subnetwork Connection Protection

Table 3.6: SDH / SONET Protection Mechanisms

The table classifies the SDH protection mechanisms depending on the recovery level (span, link, and path) and recovery topology (ring, mesh).

Linear Multiplex Section Protection (1+1, 1:1, 1:N APS)

The Linear Multiplex Section Protection (MSP) mechanism is called line protection in ANSI-T1 terms. In contrary to the 1+1 and 1:1 protection switching introduced in Section 3.5.2.1, SDH linear MSP is applicable to spans (point-to-point networks) only.

In case of 1+1 MSP, one working fiber is protected with one dedicated protection fiber and the traffic is transmitted over both fibers simultaneously. In normal operation mode, the sink node forwards the signal from the primary fiber. In case of a failure or signal degrade the sink node can immediately switch to the alternative fiber. The 1+1 MSP has a very low complexity and achieves very fast recovery times. On the other hand, the mechanism requires 100% of spare resources. 1+1 MSP is generally operated in a unidirectional switching mode.

In case of 1:1 MSP, one protection fiber is dedicated to one working fiber, but the traffic is transmitted on the protection fiber only in case of a failure. This allows the transmission of extra traffic on the protection fiber, but requires a more complicated recovery switching protocol. 1:1 MSP is generally operated in a bidirectional switching mode.

In case of 1:N MSP, N working fibers share a single protection fiber. The APS protocol controls the recovery switching and takes care, that the protection fiber is not used for the recovery of more than one working fibers at the same time.

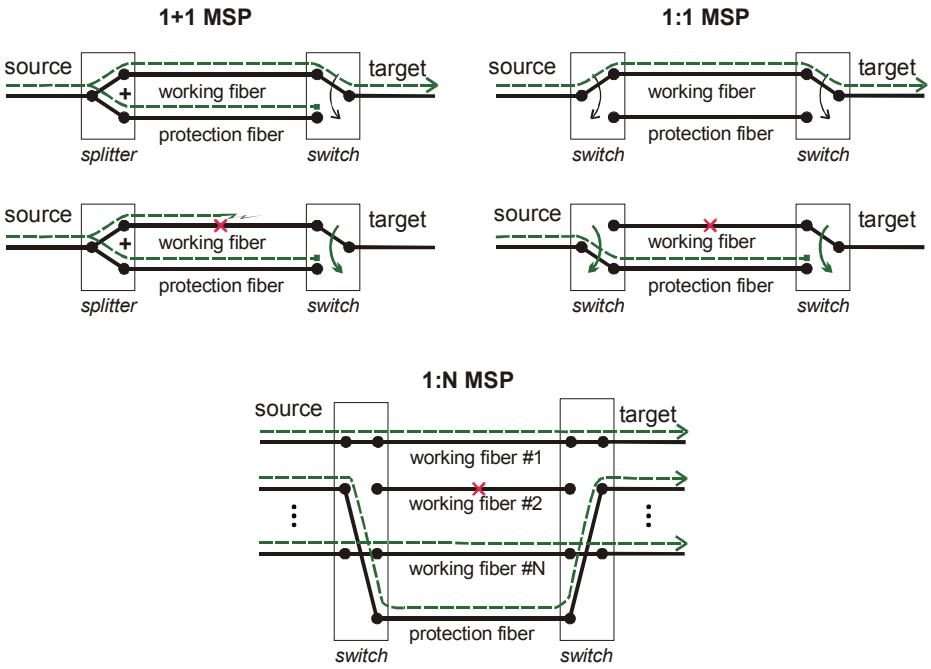


Figure 3.10: SDH Linear Multiplex Section Protection

SDH/SONET Self-Healing Rings

SDH and SONET networks with ring structures are currently the most commonly deployed network infrastructure of telecommunication operators. One reason for this is

that recovery mechanisms were standardized and available for SDH/SONET Add/Drop-Multiplexers.

The SDH Multiplex Section Shared Protection Ring (MS-SPRing) and the SONET Bidirectional Line Switched Ring (BLSR) architecture is a shared protection scheme and operates at link level. The MS-SPRing architecture uses a bidirectional ring with two or four fibers where working traffic is transmitted over both directions of the ring – clockwise and counter-clockwise (Figure 3.11). The protection capacity of the ring is shared by all traffic on the ring. The shared protection mechanism allows the transmission of extra traffic when no failure is present. For the control of the bidirectional protection switching and preemption of extra traffic, an APS protocol is defined. The MS-SPRing APS signals are transported in the K1/K2 bytes of the SDH/SONET overhead.

In the two-fiber configuration half the capacity of each fiber is used for working traffic and the other 50% for protection traffic (Figure 3.11). This configuration only supports bidirectional ring switching.

In the four-fiber configuration there is for each direction one fiber for working traffic and one fiber for protection traffic (Figure 3.12). The four-fiber configuration supports bidirectional ring switching in case of cable and node failures and additionally span switching in case of unidirectional failures.

A detailed description of the MS-SPRing architecture and the MS-SPRing APS control protocol is given in [Wu-1992] and [Ramaswami-2002].

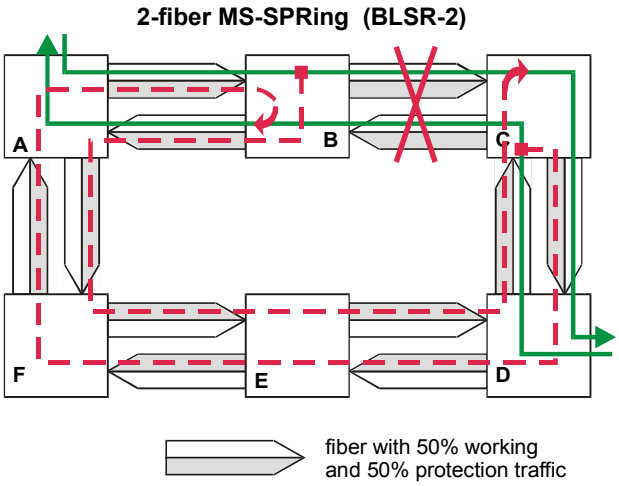


Figure 3.11: Two-fiber MS-SPRing

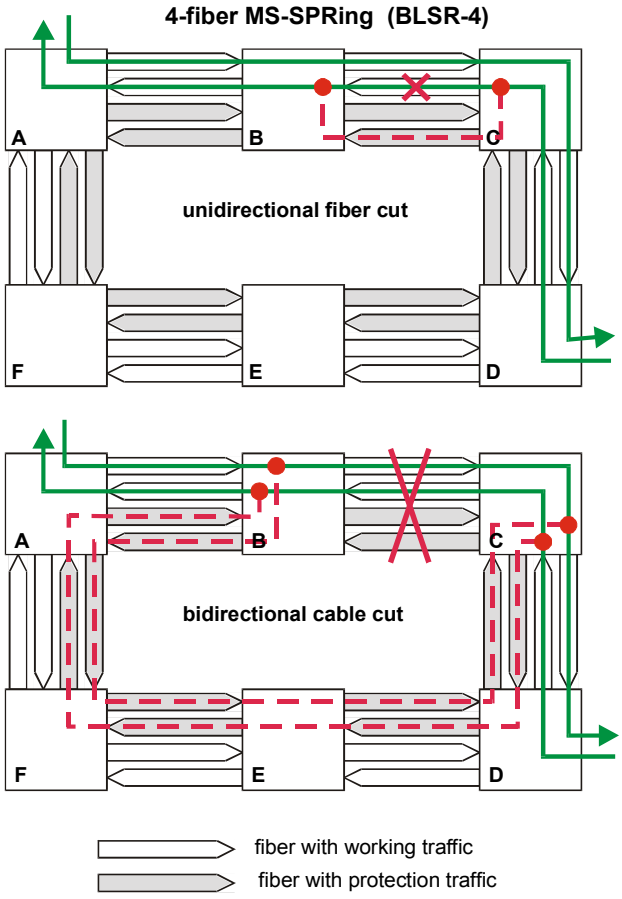


Figure 3.12: Four-fiber MS-SPRing

Subnetwork Connection Protection (SNCP)

The Subnetwork Connection Protection architecture supports ring and mesh network structures. In SONET standards the corresponding recovery architectures are called Path Protection for mesh networks and Unidirectional Path Switched Ring (UPSR) for ring networks.

In ring networks working traffic is sent unidirectional over the ring. That means, upstream and downstream traffic are traversing different portions of the ring (Figure 3.13). After the detection of a failure, the failure information is notified to the source and sink nodes of the affected connections using an APS protocol. The affected connections are switched at the end nodes using unidirectional protection switching. In the 1:1 operation, extra traffic is supported and must be preempted in case of failures.

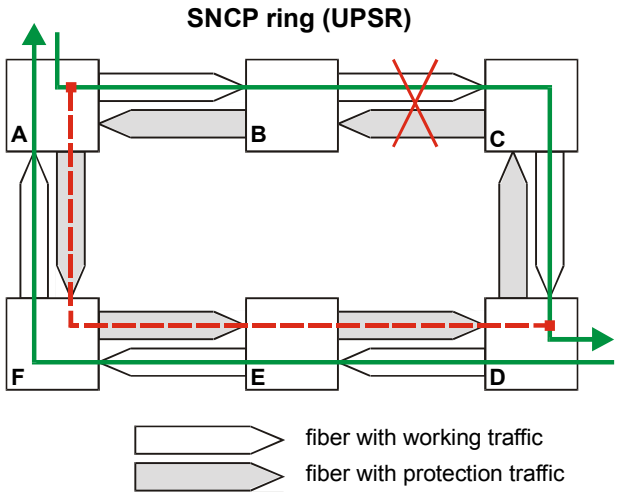


Figure 3.13: SNCP ring (UPSR)

3.6.3 OTN Recovery Mechanisms

3.6.3.1 General considerations

By using optical layer protection, significant cost savings can be realized compared to the protection of client layers of the optical network [Ramaswami-2002]. An additional advantage of optical layer protection is the bulk protection of physical failures instead of protecting many individual client signals. The drawback is of course that the smallest recovery granularity of the optical layer is one lightpath. Moreover, protection routes in the optical layer have length limitations due to optical attenuation and dispersion budgets.

Optical layer recovery alone cannot recover failure of client layer equipment and client connections terminating in failed optical nodes. If optical layer recovery is used in addition to client layer recovery, an additional degree of resilience can be achieved. Single failures in the optical layer can be recovered by the optical layer alone. Secondary failures and client layer failures can be recovered in the client layer. However, this requires careful attention to interworking schemes. A detailed framework for multiple failure recovery is defined in Section 3.8 and the multilayer recovery framework is defined in Section 3.9.

3.6.3.2 OTN Protection Switching Mechanisms

Optical layer recovery mechanisms operate either at the Optical Multiplex Section (OMS) or at the Optical Channel (OCh). The first corresponds to link recovery, the second to path recovery mechanisms. Table 3.7 gives an overview of optical protection mechanisms defined in [Ramaswami-2002]).

ITU-T	Type	Level	Topology
OMSP (1+1, 1:1)	Dedicated, shared	Span (OMS)	Point-to-point
OMS-DPRing	Dedicated	Link (OMS)	Ring
OMS-SPRing	Shared	Link (OMS)	Ring
OCh-DPRing	Dedicated	Path (OCh)	Ring
OCh-SPRing	Shared	Path (OCh)	Ring
OCh-SNCP	Dedicated	Path (OCh)	Mesh

Abbreviations

OMS-DPRing	Optical Multiplex Section Dedicated Protection Ring
OMS-SPRing	Optical Multiplex Section Shared Protection Ring
OMSP	Optical Multiplex Section Protection
OCh-SNCP	Optical Channel Subnetwork Connection Protection

Table 3.7: OTN Protection Mechanisms

The optical layer protection mechanisms are similar to their SDH and SONET counterparts. However, their implementation is substantially different. In addition to [Ramaswami-2002] a good discussion of implementation aspects of optical layer protection mechanisms is given in [Gerstel-2000-b]. A difference, which must be considered for the implementation, is the costs of optical transmission equipment, which grows which the number of wavelengths to be multiplexed and terminated. In addition, optical paths are length constraint due to physical link budgets. This is especially important for link based ring mechanisms, where the protection path is traversing the full ring except for the failed ring. A connection, which traverses half the ring under normal conditions, can be almost three times as long after link protection switching. If no wavelengths conversion is possible, the wavelength continuity must be complied with also for the protection path.

3.6.3.3 OTN Restoration Mechanisms

Using distributed control architectures defined for Automatically Switched Optical Networks (ASON) and for the Generalized Multiprotocol Label Switching (GMPLS) architectures fast optical mesh restoration is feasible.

In the ASON architecture, an end-to-end path restoration mechanism can be implemented. Since ASON supports automatic end-to-end path provisioning, this mechanism can be used for an end-to-end path restoration with 100ms restoration times [FASHION-D1]. Upon detection of a failure in the end nodes of optical lightpaths, the control plane can automatically set up an alternative connection. Since the distributed control plane has information about the full network, the restoration can use a constraint based routing instead of flooding mechanisms.

Optical restoration mechanisms for the GMPLS architecture are proposed in the IPO (IP-over-Optical) and CCAMP (Common Control and Management Protocol) workgroups of the IETF.

3.6.4 MPLS Recovery Mechanisms

MPLS recovery is a key research issue in the IETF. Several IETF drafts and a framework proposal [Draft-Sharma] are discussed in the MPLS working group and present different recovery mechanisms.

According to [Draft-Sharma] benefits from the MPLS Recovery are:

- Finer recovery granularity (compared to Layer-1 recovery)
- Protection selectivity based on service requirements becomes possible
- Efficient and flexible resource usage (e.g., recovery path may have reduced performance requirements)
- Allows end-to-end protection of IP services
- Uses lower layer alarm signals (contrary to current IP rerouting)

Several functions are required to provide resilience in a MPLS network:

- Fast and reliable failure detection
- Recovery framework
  - Selection of recovery options
- Resilience provisioning and signaling
  - Traffic engineering aspects
  - Resilience-constrained LSP setup
  - Protection selectivity support

Using the concept of the LSP the provisioning of resilience similar to classical link restoration or protection switching mechanisms is possible. Since MPLS LSPs are generally unidirectional the recovery mechanisms work unidirectional, too. The main options and parameters for MPLS recovery mechanisms defined in [Draft-Sharma] are the recovery model (protection switching, restoration, rerouting), the path setup (pre-established, pre-qualified, established-on-demand), the resource allocation (pre-reserved, reserved-on-demand), and the resource usage (dedicated-resource or extra-traffic-allowed). Some sample recovery mechanisms and their applicable recovery options will be discussed and illustrated with some network examples in the following sections.

3.6.4.1 Protection Switching

In the case of protection switching, the alternative LSP is pre-established and pre-reserved (pre-provisioned). That is why protection switching realizes the shortest disruption of the traffic. Depending on the recovery scope, the LSP is either switched at the ingress and egress LSR (path protection), or locally at the LSRs adjacent to the failure (local protection).

Link Protection

A protection switching scheme where recovery LSPs are pre-established for each link is often called MPLS Fast Reroute. Several different proposals are currently discussed in the IETF. The advantage of such a fast rerouting scheme is that no end-to-end failure

notification and signaling is required for the protection switching. A node detecting a physical failure at its port may immediately switch the affected traffic to the recovery path. To reduce the number of recovery LSPs a node has to configure, a single recovery LSP could be configured to protect several LSPs running over the link and belonging to the same FEC.

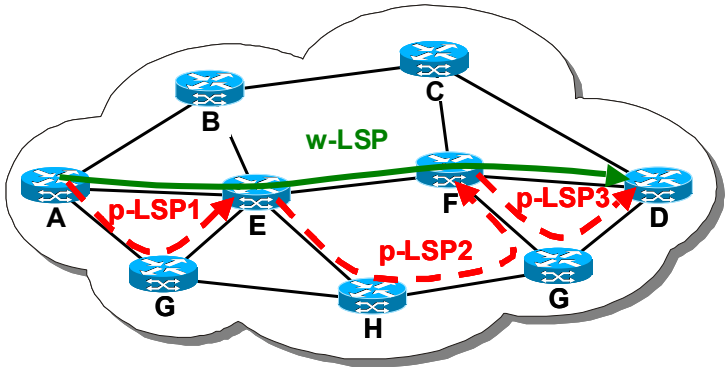


Figure 3.14: MPLS link protection (fast reroute)

Another method to setup an alternative label switched path to handle fast rerouting is proposed by Haskin [Draft-Haskin]. The mechanism is similar to a classical SDH MS-SPRing mechanism. Figure 3.15 illustrates the mechanism.

For each LSP an alternative recovery LSP is set up as indicated from the last-hop switch in reverse direction to the source of the working LSP and along a node-disjoint path to the destination switch.

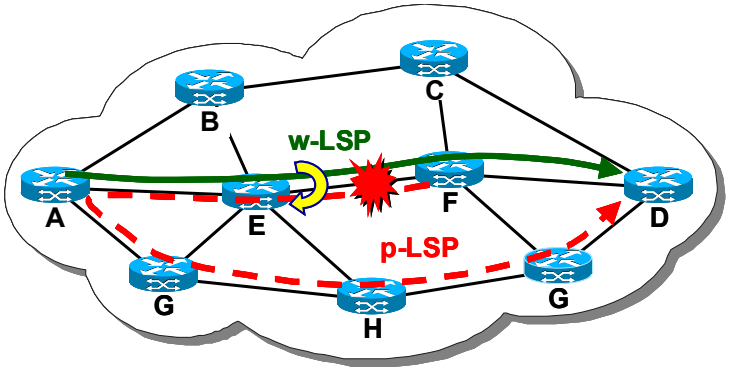


Figure 3.15: MPLS segment protection (by Haskin)

When a failure is detected (1), the adjacent upstream node immediately switches the working LSP to the recovery LSP (2).

The advantage of this approach is, that only a single protection-LSP must be set up, and the rerouting may still be triggered based on a local decision in the node directly upstream of the failure. Thus no recovery signaling is needed.

Path Protection

Protection switching schemes with global scope are called path protection. For each protected LSP a protection LSP is established either between the ingress and egress LSR (Figure 3.16), or between designated recovery-switching points (so-called segment protection). The switching LSR must be notified that an LSP failed, in order to switch the LSP to the protection LSP. The MPLS signaling protocols CR-LDP and RSVP-TE are extended to support such failure notification.

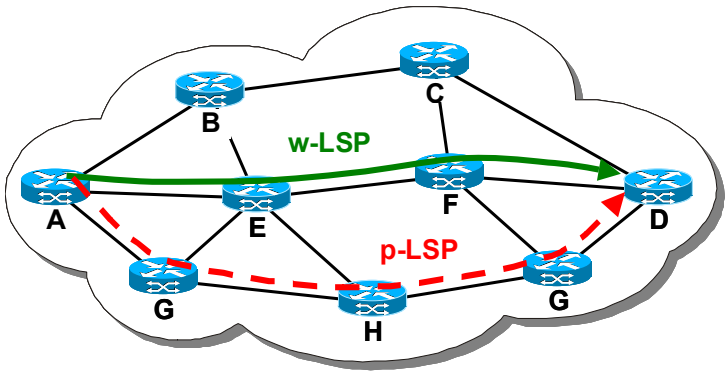


Figure 3.16: MPLS path protection

Resource Allocation and Usage

Several options are possible for the resource usage of the recovery path [Draft-Sharma].

In 1+1 ("one plus one") operation a copy of the working traffic is always transported over the recovery path. To recover from a failure the egress LSP must only select the incoming traffic from the protection LSP instead of the working LSP. No signaling is required in this case.

In 1:1 ("one for one") operation, the working traffic is only switched to the recovery LSP if a failure occurred on the working LSP. Depending on the selected resource usage, dedicated or shared, the recovery LSP may be used only to recover a single working LSP, or it may be used to recover different LSPs with the same LSP end points (see also Figure 3.16). In the second case, the working LSPs follow disjoint routes through the network. Otherwise a single failure could disrupt both working paths, and there wouldn't be sufficient protection resources to recover both paths.

If a 1:1 resource allocation is used the recovery LSP may additionally carry low-priority, pre-emptible traffic - so-called extra-traffic - when no failure is present in the network. This extra traffic must be dropped if the LSP is needed for the recovery of a failed LSP.

3.6.4.2 Restoration (MPLS Rerouting)

When deploying an MPLS rerouting scheme, recovery LSPs are established-on-demand after the detection of a failure. In contrary to classical IP rerouting, MPLS may utilize a fast hardware detection to decrease the recovery time needed to restore the affected traffic.

In analogy to protection switching, the recovery can be done locally around the failed link or node, or globally starting at the ingress and egress LSP. Figure 3.17 illustrates global restoration, Figure 3.18 regional restoration, and Figure 3.19 local restoration in MPLS.

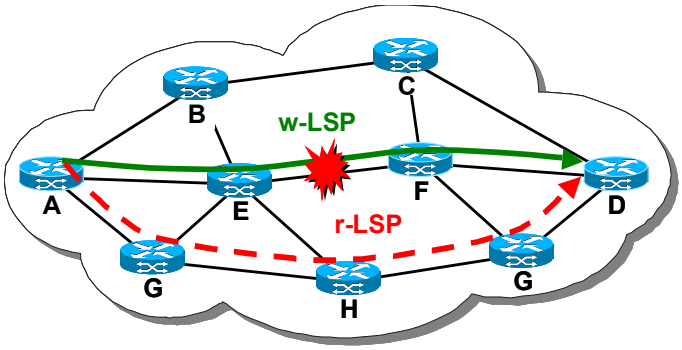


Figure 3.17: MPLS global restoration

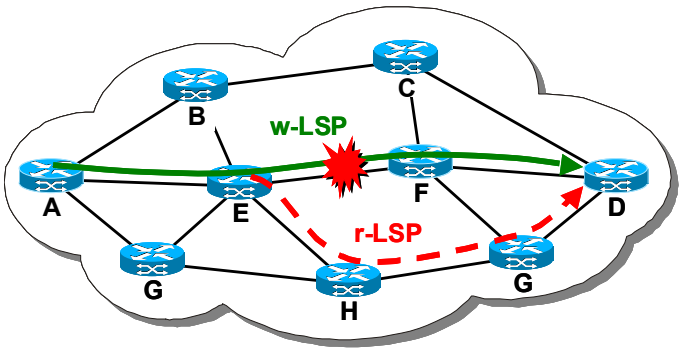


Figure 3.18: MPLS regional restoration

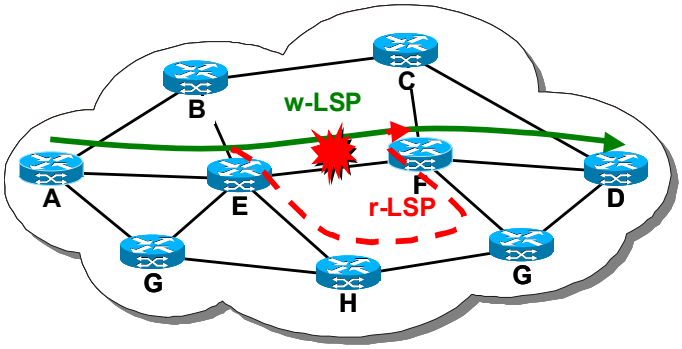


Figure 3.19: MPLS local restoration

The recovery path is established using constraint-based routing and signaling protocols after detecting the failure. Since the calculation of new routes and the signaling and resource reservation of a new LSP are time-consuming, MPLS rerouting is considerable slower than protection mechanisms.

3.7 Performance Evaluation of Resilience Concepts

Different evaluation methods and performance parameters can be used for the evaluation of resilience strategies.

- Terminal pair availability analysis
- Network availability analysis
- Spare capacity planning and resource efficiency evaluation
- Recovery time analysis
- Recovery time performance simulation
- Protocol simulation
- Protocol complexity evaluation

The performance evaluation methods mainly used in this work are the resource efficiency evaluation, recovery time analysis and protocol complexity evaluation. For the evaluation of the differentiated resilience architecture, resource efficiency based on traffic engineering methods and the recovery time are evaluated using graph theory analysis. For multilayer recovery analysis a protocol simulation using a highly detailed modeling of the simulated network elements was performed, which at the same time also allows a discrete-time event-oriented performance simulation to evaluate the recovery time of multilayer interworking strategies.

In the following subsections, the methods for spare capacity planning, resource efficiency evaluation and recovery time analysis are briefly introduced.

3.7.1 Spare Capacity Planning

The provisioning of spare resources for network survivability directly translates to increased network costs. Depending on the used recovery model and options, the required normalized spare resource costs may be between 150% for path restoration methods [Johnson-1996] and 200 – 300% for protection against single link or node failures [Grover-2000]. If survivability guarantees against multiple failures are required, the spare resource requirements may be even higher.

An introduction to network planning including spare capacity dimensioning is contained in [Eberspächer-1998]. The planning of spare resources has been studied extensively in literature (e.g, [Iraschko-1996, Caenegem-1997-a, Caenegem-1997-b, Herzberg-1994, Herzberg-1995, Herzberg-1997, Grover-1991]). The publications covering the planning of spare resources usually start with a greenfield scenario and optimize the network capacity using graph theory and optimization techniques like Integer Linear Programming (ILP) or heuristics. The result of the planning process is a capacitated network with resources for working and protection/restoration traffic.

3.7.2 Resource Efficiency Evaluation

In the recovery efficiency evaluation used in this thesis a traffic engineering approach is used instead of a network planning approach: instead of optimizing the network capacity

a capacitated network is used as starting point. The backup resources required for a given resilience strategy are evaluated using graph theory procedures and heuristics for the routing of the primary and secondary paths. In contrast to the spare capacity planning, only the resources dedicated to the recovery of working traffic are considered, but not free network capacity. In chapter 4 this traffic engineering approach is explained in more detail for the evaluation of the RD-QoS architecture.

3.7.3 Recovery Time Analysis

In [ITU-T M.495] a generic temporal model for service restoration is defined. Based on this model a restoration temporal model specific for ATM protection switching is defined in [ITU-T I.630]. A similar model for the recovery time cycle in MPLS networks is defined in [Draft-Sharma]. Figure 3.20 illustrates this recovery cycle model.

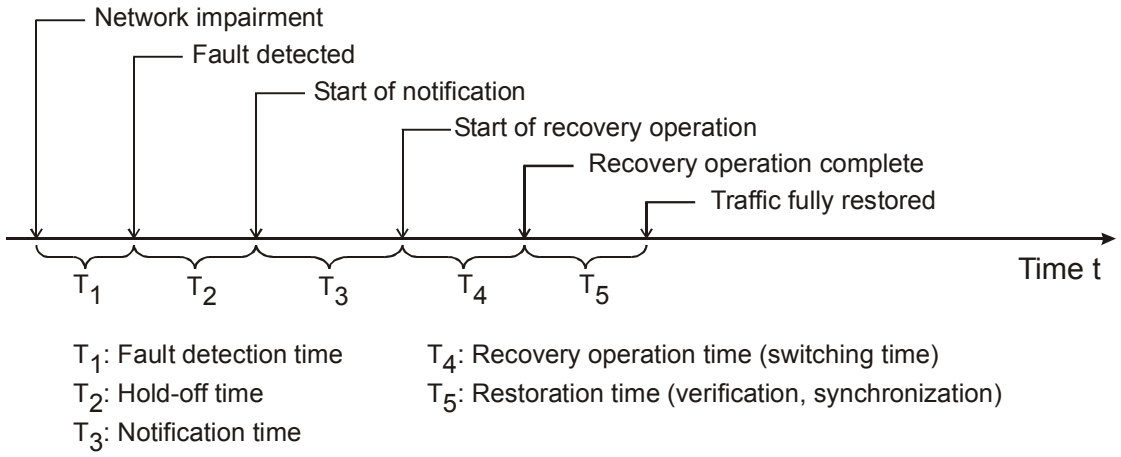


Figure 3.20: Recovery cycle model (based on [Draft-Sharma, ITU-T I.630])

The recovery time analysis model used in this work is based on the above recovery cycle model and on the analysis model for the restoration times published in [Ramamurthy-1999-II]. However, the latter model was extended in some parts. In the following the recovery timing model and the extensions to the work of Ramamurthy is explained.

The following list summarizes the extensions considered in this thesis compared to [Ramamurthy-1999-II].

- Adaptation to MPLS recovery mechanisms.
- Relation to restoration time model and MPLS recovery cycle [Draft-Sharma].
- Queuing and sequential processing of control messages for multiple recovered connections. In [Ramamurthy-1999-II] it is assumed that the message queuing time is included in a fixed message processing time.
- A route calculation time for restoration schemes using constraint based explicit routing instead of a flooding mechanism.

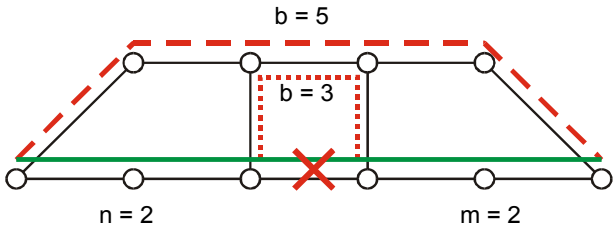
- Use of real link lengths for calculation of propagation delay. [Ramamurthy-1999-II] assumes fixed link propagation delays of 400µs corresponding to a link length of 80 km.
- Additional switching time in destination node for restoration mechanisms.
- Analysis of multiple affected connections (ranging from a few to several thousand connections).

The following notation and parameter values are used for the timing model.

- *n:* Number of nodes from the node upstream of the failure (FUR) to the source node of the flow (I-LSR).
- *m:* Number of nodes from the label merge router (LMR) to the target node of the flow (E-LSR).
- *b:* Number of links on the backup route of a failed connection from the recovery switching node to the recovery merging node.
- *f:* Flow/LSP index.
- T_1: Failure detection time. Depending on the used failure detection mechanisms, highly different values are possible for the failure detection time. In the case studies this value is assumed to be 20 ms.
- T_2: Hold-off time is assumed to be 0 ms
- T_3: Notification time is calculated
- T_4: Recovery operation time is calculated
- T_5: Traffic restoration time is calculated
- *D:* Message processing time in the node is 10µs as in [Ramamurthy-1999-II]. However, in contrast to [Ramamurthy-1999-II] in this thesis the queuing and processing of multiple control messages is modeled with a simple sequential processing of individual flows. That is, the processing of the second flow is finished after $2 \cdot D$, and the processing of the flow f is finished after $f \cdot D$.
- *P:* Propagation delay on a link is 5µs per km, corresponding to 250µs for a link length of 50km
- L_i: Length of link i in km
- *C:* Time to configure, test and setup a forwarding table (switching time). Values between 1 ms and 10 ms are considered for C in the case studies. In contrast to [Ramamurthy-1999-II] it is assumed, that the switching is in parallel to the message processing. A LSR needn't wait for the completion of the switching before the message (e.g., backup LSP setup) can be forwarded.
- *R:* Time to calculate new constraint based backup route. This time is depending on the size of the network and on the number of constraints. A value of 2 ms is assumed for the route calculation. For the constraint based routing, the route must be calculated for each individual LSP.

The link propagation time is calculated for the traversed edges using the physical link length of the considered network scenarios. In the model, the following notation is used for the link propagation:

$\sum_{i=1}^{n} (L_i P)$ Propagation time over the set of edges E_n. The time is composed of the sum of the link lengths of all traversed edges multiplied with the propagation delay.



Recovery model	n	b	m
Path protection / restoration (dashed line)	2	5	0
Link protection / restoration (dotted line)	0	3	2
Segment protection by Haskin (not shown)	0	7	0
Regional restoration (not shown)	0	4	0

Figure 3.21: Relation between restoration time parameters n, m and b

Figure 3.21 shows the relation between the parameters n, m and b. In the following sections the timing models for selected recovery mechanisms are defined with a focus on MPLS networks.

3.7.3.1 Protection Switching

Path Protection

After the detection of a failure the upstream node sends a failure detection message to the I-LSRs of all affected flows traversing the failed link. The source nodes switch the failed LSPs to the preconfigured protection LSP by reconfiguring the label forwarding table. No additional signaling is required since MPLS LSPs are unidirectional and the protection LSP is predefined. The intermediate LSRs needn't perform any functions and in the label merge router (LMR) the protection and the working traffic are merged automatically by the preconfigured label forwarding table. In MPLS, the labels of the backup paths can already be preconfigured in the label forwarding table, since the capacity of the LSP is not exclusively reserved for the flow. Instead, if the backup path is not used, the capacity is available automatically to other traffic (including low priority, extra traffic). Preemption of the low priority traffic on the backup route is performed implicitly in MPLS by utilizing a lower holding priority for the extra traffic [RFC3212]. The restoration process is finished when the traffic sent from the I-LSR over the backup LSP reaches the E-LSR. Thus the total restoration time for a flow *f* using a path protection scheme in MPLS is:

$$T_r(f) = T_1 + T_3 + T_4 + T_5$$

$$T_r(f) = T_1 + [\, f \cdot D + n \cdot D + \sum_{i=1}^{n} (L_i \cdot P) \,] + C + \sum_{i=1}^{b} (L_i \cdot P)$$

T_1 is the fixed failure detection time in the router upstream of the failed link (FUR). The notification time is composed of a sequential message processing in the FUR and the message propagation and processing until the notification message reaches the I-LSR. In the I-LSR the LSP forwarding table is reconfigured.

Link Protection

The link protection timing model is similar to the path protection case. The main difference is that the signaling to the I-LSP is not applicable. Upon detection of the failure, the FUR must identify the failed LSPs and can then immediately switch the failed flows to the backup LSP. The restoration process is finished, when the traffic sent from the FUR over the backup route reaches again the E-LSR.

$$T_r(f) = T_1 + [\, f \cdot D \,] + C + \sum_{i=1}^{b+m} (L_i \cdot P)$$

Segment Protection

The segment protection timing model is almost identical to the link protection case, only the backup route is different. For the segment protection mechanisms defined by Haskin the backup route is usually much longer than the backup route in case of link protection.

$$T_r(f) = T_1 + [\, f \cdot D \,] + C + \sum_{i=1}^{b} (L_i \cdot P)$$

3.7.3.2 Restoration

The restoration time model is given for the global, local and regional MPLS recovery mechanisms described in Section 3.6.4.2. In contrast to [Ramamurthy-1999-II], where flooding based restoration mechanisms are assumed, restoration mechanisms with constraint based explicit routing in the LSRs are assumed here.

Global Restoration

The restoration time is composed of the failure detection time T_1, the notification time T_3, the switching time T_4, and the restoration completion time T_5. As with path protection, the notification time is composed of a sequential message processing in the FUR and the message propagation and processing until the notification message reaches the I-LSR. The switching requires a sequential constraint based route calculation. After the route calculation is finished, a backup LSP setup message is sent to the E-LSR using explicit routing. When the signal is sent back, the LSP forwarding table must be

updated on every intermediate node. Since the message can be forwarded before the forwarding table update is completed, only one switching time (at the I-LSR) is added to the restoration time. Finally, the traffic must propagate from the I-LSR to the E-LSR.

$$T_r(f) = T_1 + [f \cdot D + n \cdot (P + D)] + [f \cdot R + 2 \cdot b \cdot D + 2 \cdot \sum_{i=1}^{b} (L_i \cdot P) + C] + \sum_{i=1}^{b} (L_i \cdot P)$$

Local Restoration

With local restoration no failure notification is required. The FUR looks up the affected LSPs, computes the backup path for each failed LSP, and sends a backup LSP setup message to the router downstream of the failed link (FDR). When an acknowledgement message arrives back at the FUR the LSP can be switched to the alternative route.

$$T_r(f) = T_1 + [f \cdot D] + [f \cdot R + 2 \cdot b \cdot D + 2 \cdot \sum_{i=1}^{b} (L_i \cdot P) + C] + \sum_{i=1}^{b+m} (L_i \cdot P)$$

Regional Restoration

With the exception of the traffic restoration on the backup path the regional restoration timing model is the same as in the link protection case.

$$T_r^f = T_1 + [f \cdot D] + [f \cdot R + 2 \cdot b \cdot D + 2 \cdot \sum_{i=1}^{b} (L_i \cdot P) + C] + \sum_{i=1}^{b} (L_i \cdot P)$$

3.7.3.3 Recovery Time Analysis

The restoration time is calculated for a given set of expected failures, e.g. all single link failures. The expected time to restore is calculated as an average of the restoration time of all flows and all expected failures. Alternatively, the restoration ratio over time can be calculated, which indicates the ratio of recovered flows to the number of affected and recoverable flow over the time. The restoration ratio can be calculated for a single failure, or averaged over a given set of failures.

This restoration time analysis is used for the performance evaluation of the RD-QoS architecture presented in Chapter 4.

3.8 Multiple Failure Recovery Framework

In Section 3.3 an overview of possible failure of a network is given, and in [*Schupke-2001-a*] network-wide significance of multiple failures as well as the vulnerability of different recovery mechanisms in a network is shown. In this section a framework for the recovery of multiple failures is defined.

As already stated, recovery mechanisms are commonly classified into protection, restoration and reconfiguration according to two criteria: the time of the backup route calculation and the type of switching control instance.

It should be noted that protection schemes are more vulnerable to multiple failures than restoration schemes since they use predefined recovery paths. If the primary failure in a network affects the protection path of a connection, the connection cannot be recovered if it is affected by a second failure. Secondly, if the original connection was affected by the first failure and was recovered using the protection path, an additional failure on the protection path cannot be recovered.

Restoration schemes on the other hand are inherently more robust against complex multiple failure scenarios. The alternative path is either computed centrally or searched using a distributed flooding mechanism. Therefore, provided that enough spare capacity is available in the network, alternative paths can be found.

The classification of recovery mechanisms takes only the behavior for single failures into consideration. In Figure 3.22 a framework is defined to classify recovery of multiple failures.

Multiple Failure Recovery Approach

Horizontal approach	Vertical approach
• Network partitioning • Pre-computed recovery (before first failure) • Re-computed recovery (after first failure) • Re-Restoration (After secondary failures)	• Recovery at higher layer (Escalation strategy with hold-off time) • Recovery at lower layer (OMS-Restoration) • Central reconfiguration (by NMS with access to several layers)

Figure 3.22: Multiple failure recovery framework

Two approaches are distinguished: a horizontal approach, where the secondary failures are recovered at the same layer (but possibly with another recovery scheme) and a vertical approach, where secondary failures are recovered at another layer.

A second criterion is, at which point in time the alternative route for secondary failure is computed (before primary failure, after primary failure, after secondary failure). The different options of the multiple failure recovery framework are discussed in the next sections.

3.8.1 Horizontal Approach

Network Partitioning

A horizontal partitioning of the network in protected domains can be used to minimize the probability of multiple failures. Examples for network partitioning are multiple interconnected rings and p-Cycles.

A main consideration to handle multiple failures is the computation of the route for recovery from a secondary failure in the presence of a primary failure. Three different options can be distinguished:

Pre-Computed Recovery (Before First Failure)

In case of pre-computed multiple failure recovery the alternative paths of secondary failures are calculated already before a primary failure occurs. This option is useful for primary protection or centralized restoration mechanisms.

Practical approaches for this scenario are protection tables with multiple entries, where a node must select a protection entry, which is still available after a failure. In the planning process it must be considered, that the protection routes should be maximally disjoint to increase the probability that there is still a protection path available for an arbitrary failure scenario.

A second approach is to compute for each node a second protection table for every single link and/or single node failure. After the first failure, all nodes load the corresponding secondary protection table.

Re-Computed Recovery (After First Failure)

In this case the alternative paths for working and backup paths affected by a primary failure are calculated after the primary failure occurrence.

After the occurrence of a first failure, the nodes or the network management system compute for all affected working paths new protection paths and for all affected protection paths new alternative paths. The figure below illustrates this. In the left part of the figure the working path W-1 using the route F-G-H-D is affected by a failure and is switched over to the protection path P-1 using F-E-D. Since this connection is now vulnerable against secondary failures, a backup path P'-1 is recomputed for this connection, using for example F-G-B-C-H-D.

In the right part of the figure the situation is shown where the backup path is affected by a failure (P-2). In this case the working path W-2 is also unprotected against secondary failures. Therefore the protection path P'-2 is re-computed.

This multiple failure recovery option corresponds to a protection mechanism in the single failure scenario. The network is protected against multiple failures after a relatively short re-computation phase.

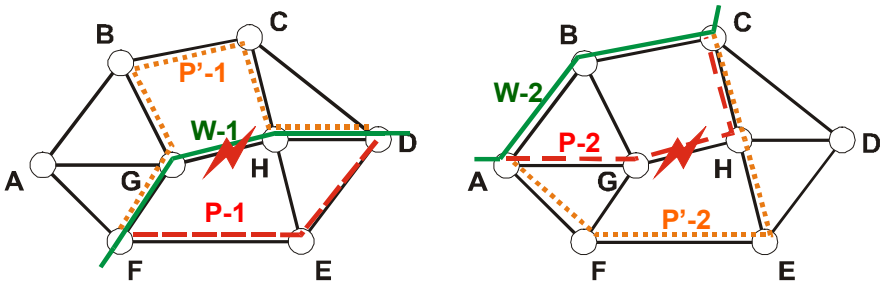


Figure 3.23: Two scenarios for protection path re-computations

Re-Restoration (After Secondary Failures)

Corresponding to single failure restoration, the alternative route computation for secondary failures could be done on-demand after the occurrence of a secondary failure. Please remark, that in this case the primary recovery mechanism can also be a protection scheme.

3.8.2 Vertical Approach

In today's networks, different transport technologies can participate in the recovery of failure recovery. However, the layers would by default not coordinate themselves when attempting recovery from outages. On the one hand, server layer(s) are left unaffected by client layer failures for the most part. On the other hand, due to dependency on lower-layer services, client layers are in most cases struck by failures on the layer(s) below, which inherently causes multiple simultaneous failures.

Recovery at Higher Layer (Escalation Strategy with Hold-Off Time)

Recovery in such systems with vertical separation may converge to a steady state relatively slow or not at all. A simple, but certainly sub-optimal solution is to introduce escalation tactics [Manchester-1999]. A server layer is given the opportunity to solve the problem during a certain period of time. When the hold-off timer expires, control is passed to its client layer if the measures had no effect so far, and so on. Further escalation or interworking strategies published in [*Demeester-1999*] are described in section 3.9.

As an alternative, explicit failure messaging across layers could be introduced. Besides standardized failure types/messages, this would also include failure localization procedures. Based on knowledge about the capabilities of the given layers, recovery could be explicitly delegated or executed, respectively, by a generalized control plane. This is currently pursued by the GMPLS standardization effort, where the various transport technologies are logically merged to a uniform infrastructure layer [Draft-Lang] rather than attempting vertical messaging across layers. The standardization organization ANSI and the Optical Interworking Forum (OIF) also work on layer coordination [Gerstel-2000-b].

Recovery at Lower Layer (e.g. OMS-Restoration)

A second option for a vertical multiple failure recovery approach works on two granularities, namely on path (optical channel or SDH/SONET signal) and line (WDM signal) layers, respectively. While the first occurring failure is quickly overcome on the fine-granular higher layer, another failure is repaired at the layer below, using optical crossconnects (OXCs). That way, there is no vulnerability anymore in the period between occurrence and repair of the first failure [Gerstel-2000-b].

A drawback is, that with the mechanism described in [Gerstel-2000-b] the switching of the lower layer recovery mechanism causes an additional short-term disruption of the already recovered connections.

Central Reconfiguration (by NMS with Access to Several Layers)

Such systems typically perform a kind of resource optimization on the network, trying to accommodate all demands with the current topological restrictions in mind. Therefore, a centralized entity re-computes and updates routing in the network. The NMS may have the advantage to access several layers for service restoration.

3.8.3 Multiple Failure Spare Capacity Planning

To be robust against multiple failures, enough spare capacity must be present in the network. This should be taken into regard already in the network planning process by including multiple failures into the anticipated failure scenarios.

Instead of calculating the required spare capacity to restore the traffic only for single link or node failures, multiple failure scenarios should already be taken into account. Such additional multiple failure scenarios could include all double link and node failures as well as simultaneous single link and node failures.

3.9 Multilayer Recovery Framework

It is generally possible to reach the survivability requirements of network services with resilience mechanisms present at different layers, e.g. SDH, ATM, OTN and MPLS [Draft-Owens]. Moreover, these resilience mechanisms may even be in operation in multiple layers at the same time.

The ACTS project PANEL [*Demeester-1997*] investigated the interworking of multilayer recovery mechanisms, and a multilayer recovery framework was defined and published in [*Demeester-1999*]. The multilayer framework, concepts and results of the PANEL project influenced a white paper on multi-layer survivability published by Lucent Technologies [Meijen-1999], which was presented as ANSI-T1 contribution T1A1.2/2000-008 in January 2000. This finally resulted in multi-layer survivability definitions in the ANSI-T1 *Technical Report on Enhanced Network Survivability Performance* [ANSI TR68].

The multilayer recovery framework considers the recovery mechanisms present in single layers, the planning of spare resources in multiple layers, the multilayer recovery strategy and interworking mechanisms.

3.9.1 Multilayer Resilience Considerations

Before selecting a multilayer resilience strategy, the benefits and drawbacks of resilience at the individual layers must be analyzed and some pitfalls with multilayer recovery are illustrated.

3.9.1.1 Recovery at Optical Layer Only

A main advantage of the optical layer is the ability to detect failures very fast (<10ms). The optical layer can either protect individual lightpaths or all wavelengths in a fiber.

Since the cost of router ports is higher than the cost of optical ports, the recovery at the optical layer is more cost efficient [Ramaswami-2002].

On the other hand, protection at the optical layer cannot restore client layer failures. Moreover, in case of node failures, client traffic terminating at that node cannot be recovered. Figure 3.24 illustrates this scenario: The client layer connection uses two links which are mapped to corresponding connections in the server layer. The server layer connections are protected in the server layer against link failures (A-C is protected by A-B-C and C-D is protected by C-E-D). However, in the case of node failure C, the client node is isolated and cannot be recovered by the server layer.

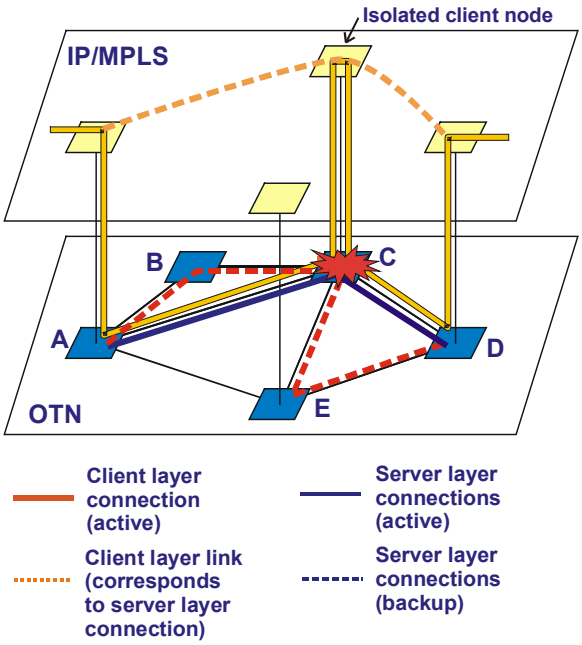


Figure 3.24: Isolated client node scenario

3.9.1.2 Recovery at IP/MPLS Layer Only

While the recovery at lower layers generally has advantages in the time scale of the recovery operation due to bulk recovery of physical failures, the recovery at the IP/MPLS layer allows higher recovery flexibility and QoS granularity. In addition, if services are protected with a differentiated level of resilience IP/MPLS recovery can be more resource efficient [Autenrieth-2002-a], [Autenrieth-2002-b].

Moreover, failures in the client layer and the above mentioned failure scenario with the isolated client node can only be recovered with client layer recovery.

3.9.2 Multilayer Recovery Strategies

If failure recovery is possible at multiple layers, it should be coordinated, which layer is responsible for the recovery of a specific failure scenario. In the multilayer recovery

framework published in [*Demeester-1998-a*, *Demeester-1999*] a recovery at highest layer scheme and a recovery at lowest layer scheme is defined.

If recovery is possible in multiple layers for a specific failure scenario, a multilayer interworking mechanism must be defined to coordinate the recovery mechanisms in both layers.

3.9.2.1 Highest Layer Recovery

With *recovery at highest layer*, the traffic is recovered at the layer where the traffic is injected. In case of an IP/MPLS over OTN network, this means that the IP/MPLS traffic will be recovered by MPLS recovery mechanisms and native OTN traffic is recovered using the OTN recovery, independent of the type of failure. Figure 3.25 illustrates this concept. In any of the 3 failures X_1, X_2 and X_3 the working MPLS label switched path (LSP) will be recovered at the MPLS label switched routers (LSR) A' and D'. The OTN working lightpath s_w is in this example only affected by the cable break (X_1) and the OTN node failure (X_2). In both cases the OTN connection is recovered by the OTN protection path (s_p).

This recovery approach allows to provision different resilience classes at a high granularity. On the other hand, in the case of physical failures like cable cuts, a large number of connections must be recovered individually. This leads to a slower recovery performance.

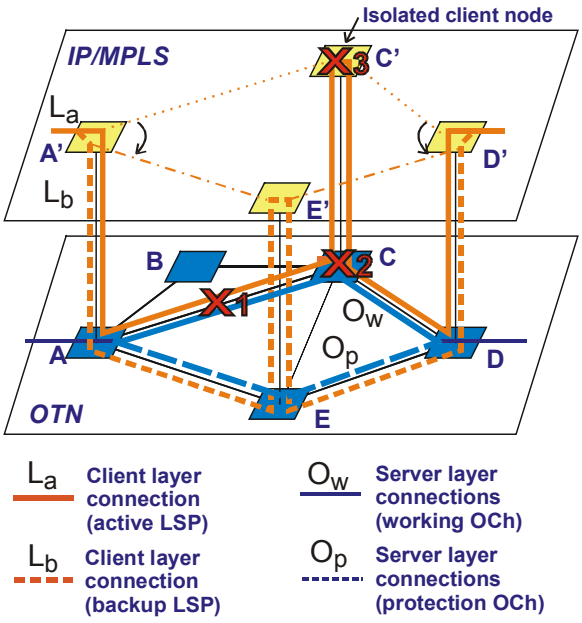


Figure 3.25: Highest/lowest layer recovery scenario

With *recovery at the highest layer* the single layer recovery mechanisms don't need to be coordinated, since the mechanisms in the different layers work independently of each other. However, in the planning phase of the network it must be taken care that active and backup client connections are not protected within the server layer.

3.9.2.2 Lowest Layer Recovery

When using *recovery at the lowest layer* disrupted traffic is restored at the layer closest to the failure. This concept can also be illustrated using Figure 3.25. In case of a cable break (X_1) for example, OTN and IP/MPLS traffic would be recovered using a OTN recovery mechanism, while in case of an MPLS LSR breakdown (X_3), the affected traffic would be recovered using an MPLS recovery mechanism.

In case of an OTN node failure (X_2), however, the MPLS LSP C_a transiting the IP/MPLS node (collocated to the failed OTN node C) cannot be recovered by the OTN layer (isolated client node). This means that IP traffic transiting the failed node has to be recovered using a secondary IP/MPLS layer recovery.

If a network has to be 100% protected against all single link and single node failures, spare resources in the client layer are only needed for multi-hop client connections. Due to fault propagation the client layer detects failures caused by server layer or physical layer failures. It must be stressed that the client layer in some cases cannot distinguish whether the defect detection is due to a client layer failure or due to a server layer failure. Therefore, the recovery mechanisms must be coordinated to prevent a contention between the recovery mechanisms. Possible interworking strategies will be presented next.

3.9.3 Multilayer Recovery Interworking

3.9.3.1 Uncoordinated Recovery

In the *uncoordinated recovery* no interworking mechanism is used to coordinate the recovery mechanisms. The client layer switches an affected connection to a backup path as soon as a defect is detected. However, if the failure occurred in the server layer, the server layer also starts the recovery. Due to a time delay between the detection of failures and the completion of the recovery mechanisms, it may happen that failures, which were already recovered by the server layer, are again switched to another recovery path in the client layer. This second switching causes a secondary temporary disruption of already recovered connections and client layer spare resources will be unnecessarily occupied. If low priority traffic was carried as extra traffic on these spare resources, this traffic is unnecessarily pre-emptied, causing additional loss of (low priority) traffic.

3.9.3.2 Hold-Off Time

The drawbacks of an uncoordinated recovery prove the need for an interworking strategy for the coordination of the recovery mechanisms. A simple and robust interworking strategy is the adoption of a *hold-off time* to delay the activation of a higher layer recovery mechanism. The hold-off time has to be dimensioned large enough to make sure that the lower layer recovery has enough time to finish the recovery process. If after the elapse of the hold-off time the upper layer connections are still affected by the failure, the higher layer recovery will be activated.

The advantage of this solution is its simplicity and robustness. The obvious drawback is that in some failure scenarios the higher layer recovery is delayed even if the lower layer recovery is not able to recover the failure.

3.9.3.3 Recovery Token

To improve the recovery performance in situations where the lower layer recovery fails, a *recovery token* to trigger the higher layer recovery may be employed. As soon as the server layer realizes, that it cannot restore the affected connections anymore, an explicit signal is sent to the client layer activating the client layer recovery. An alternative approach to a recovery token is based on the monitoring of the protection path of the server layer recovery. The basic idea is that the server layer triggers the client layer recovery if the server layer recovery failed. This may be the case for multiple link failures if the protection path is affected by a prior failure. A second scenario where the recovery token will be submitted is in the case of a node failure terminating a server layer connection, which carries client demand. Connections being terminated at the failed node cannot be recovered. For both scenarios, the far-end node detects defects on both, the working and protection path of the affected connections. A prerequisite for this defect detection is to include a waiting time for the detection of additional failures before triggering the recovery process. Since the protection path may be longer and may contain more intermediate nodes than the working path, it takes longer for the defects to propagate. The terminating node must delay the activation of the recovery mechanisms for this waiting time, which is in the order of a few milliseconds. If after the elapse of the waiting time no defects are detected on the protection path, the server layer recovery is triggered. Otherwise a recovery token is issued to the client node activating the client layer recovery.

3.9.3.4 Integrated Multilayer Recovery Approach

If an integrated control plane for the IP over optical network is used (peer model), an integrated multilayer recovery may be envisaged. Depending on the failure scenario and network state, the control plane may decide if affected MPLS LSPs are recovered by setting up an alternative lightpath in the optical layer or with a backup LSP in the MPLS layer. The integrated approach is the most flexible recovery strategies, but it is also the most complex approach. Moreover, an integrated approach is only feasible for the IP over OTN peer model.

3.9.4 Multilayer Spare Capacity Design

In network scenarios with resilience mechanisms deployed in multiple networks, special care has to be taken to achieve an efficient spare capacity design. In [Gryseels-1998-a,Gryseels-1998-b] the '*protection selectivity*' and '*common pool of spare resources'* concepts were introduced to reduce spare capacity.

Protection selectivity refers to the ability of a network to protect only selected connections, and to configure other connections as unprotected [*Demeester-1998-b*]. In the multilayer context this ability is used in case of highest layer recovery, where the protection is provided in the client layer. The server layer connection carrying the client

layer links must then be unprotected. However, native server layer connections, which may use the same physical fibers, may be protected.

The second concept is the common pool of spare resources, where the protection resources of the client layer are carried as preemptive extra traffic in the server layer. If the server layer uses a shared protection model, the extra traffic may be transported using the shared protection resources.

3.10 Summary

In this chapter a comprehensive resilience framework was defined. The framework includes network operator requirements, network survivability performance parameters, failure detection and signaling methods and single layer recovery mechanisms and options. In addition, extensions for multiple failure recovery as well as for multilayer resilience were defined.

The different resilience mechanisms like protection and restoration have all their specific advantages and disadvantages. The benefit of a resilience strategy depends on the specific network scenario and how much weight a network operator puts on a specific performance metric. For example, a network operator may put more emphasis on the resource efficiency of a resilience mechanism than on its recovery speed, or he requires a very fast recovery mechanism to meet his customer contracts, which may result in higher costs.

Faced with a variety of recovery mechanisms with additional options, how can a network operator define the optimal resilience strategy? He has to take into consideration the customer demands, the network topology, and the resilience mechanisms in different layers, and trade off the cost, complexity and recovery speed.

Protection flexibility and simple resilience provisioning can help to solve that problem. A network operator can offer different levels of resilience for different services to his customer. The possibility for service differentiation is a strength of the IP layer. In fact, protection flexibility and resilience differentiation is defined as a requirement for traffic engineered MPLS networks [RFC2702].

In the next chapter, resilience differentiation is investigated for IP-based networks. The chapter starts with a definition of the resilience requirements of IP services. Then, related work on differentiated resilience subject is presented. The RD-QoS architecture for the provisioning of differentiated resilience in IP-based network, which was developed in this thesis, is defined and presented in detail. The architecture is evaluated in terms of recovery capacity and recovery time for selected network and traffic scenarios.

4 RD-QoS: Resilience Differentiated Quality of Service

4.1 Introduction

Motivation and Related Work

The provisioning of QoS and resilience is a key requirement for today's and future IP-based networks. The current research centers on MPLS as a common platform to support both, QoS and resilience in IP-based networks.

A large effort was put into the development of QoS architectures like the Integrated Services or Differentiated Services architecture. The interworking of MPLS and QoS architectures already allows an assignment of the Forward Equivalence Class (FECs) of a flow according to its QoS class.

As mentioned in the previous chapter, the strength of MPLS resilience is protection flexibility and service granularity. This strength is employed to its most, if services are assigned a differentiated level of resilience. Thus, a well-designed MPLS resilience strategy should take the resilience requirements of individual flows into account.

Unfortunately, so far resilience is seen as a network property and not as a service property. However, the current trend in the Internet is clearly towards a service driven transport architecture. Therefore, the resilience requirements of IP services should be taken into consideration for the service provisioning just like classical QoS requirements such as bandwidth and end-to-end delay.

A resilience-differentiated approach can be compared to Quality of Service provisioning. It is economically unfeasible to provide a high level of service quality to all services by simply over-provisioning the network capacity. By introducing service quality parameters individual services can have a differentiated quality of service. This results in better resource efficiency. It additionally increases a network operator's service portfolio. The differentiated services can be offered to the customers at different costs reflecting the network resource usage.

The over-provisioning of the network can be compared to a full protection of all services independently of their real resilience requirements. This also results in unnecessary high network costs. In the same way as the Differentiated Services model a differentiated resilience model can provide services with a customized level of resilience. Network operators can offer the resilience as a value added service. Moreover, critical, highest available services like government or emergency services can be protected with an extra level of protection against disasters or terrorist attacks.

Surprisingly, only little research work is done on differentiated resilience. In 1997, Yahara and Kawamura published a virtual path self-healing scheme based on multi-reliability ATM network concept [Yahara-1997]. This concept is based on a restoration order control function to restore high reliability VPs with a high priority. Additionally,

low-reliability VPs can be bumped (preempted) to capture capacity needed to restore high-reliability VPs.

Andrea Fumagalli et al. propose a concept for "Differentiated Reliability in Multi-Layer Optical Networks" [Fumagalli-2001]. In this concept, MPLS flows are protected by a lower layer optical recovery depending on the reliability degree required by the application. The different reliability degrees (classes) are offered by a common protection mechanism but with different failure coverage. Connections with a lower reliability degree end up unprotected against some failure scenarios. Which failures are covered is determined by an optimization scheme using the target reliability metric and the MTTF and MTBF values of the link.

A drawback of this approach is that in case of a failure of an unprotected link the affected connections may experience very long outage times, if no spare resources are available allowing a centralized recovery. In such cases, the outage time may be in the range of hours or days depending how long it takes to physically repair the failed element. Another problem is the very coarse granularity of the optical network. The finest granularity at which transport services are offered by the optical transport network is an optical channel with a capacity of 2.5Gb/s.

The differentiated resilience architecture proposed in this thesis differentiates the services regarding their recovery time requirements. As proposed in [*Autenrieth-2000*] the resilience requirements of IP services are included in the QoS signaling as an additional resilience attribute. In case of failures, affected services are restored according to this resilience attribute using appropriate MPLS recovery mechanisms.

Outline of the Chapter

In the following sections the resilience requirements of IP services are discussed. The differentiated resilience architecture enabling the provisioning of tailored levels of resilience is introduced. After an overview of this so-called 'Resilience-Differentiated QoS' (RD-QoS) architecture, the components of this architecture are discussed in detail. After a specification of resilience classes, the resource management is discussed and a traffic engineering process defined. The resilience requirements are signaled in existing QoS architectures using an additional resilience attribute. At the border to an MPLS domain, the resilience requirements are mapped to appropriate recovery mechanisms. Using a software environment, the proposed RD-QoS architecture is evaluated regarding its resource efficiency and recovery time performance compared to network scenarios with unprotected or fully protected MPLS traffic.

4.2 Resilience Requirements of IP Services

To support the Quality of Service requirements of real-time, connection-oriented services, two QoS models were defined by the IETF: the Integrated Services (IntServ) architecture with RSVP as a reservation protocol, and the Differentiated Services (DiffServ) architecture.

Even though the availability of a service is an important attribute of the service quality, no resilience attribute is currently defined for these QoS architectures. Network survivability is treated independently from the QoS architectures.

4.2.1 QoS and Resilience Requirements of IP Services

A first approach to QoS awareness could be to provide survivability mechanisms for all services which require high QoS, and no survivability for services with low or no QoS requirements. However, some IP applications (e.g., most e-commerce applications) require high service availability but have low QoS requirements. On the other hand, some QoS services may well tolerate low network resilience, e.g. low cost voice-over-IP services. Non-interactive real-time services like streaming media tolerate longer outages than interactive services.

At a closer look, it becomes obvious that resilience requirements of single applications are independent of their "classical" quality-of-service requirements (bandwidth, delay, delay jitter). The table below illustrates the resilience and traditional QoS requirements of some services.

		Resilience requirements	
		high	**low**
Service requires traditional QoS	**high**	mission-critical VoIP and multimedia services	standard VoIP and multimedia services
	low	database transactions, mission-critical control terminals, e-commerce services	e-mail, FTP, standard WWW

Table 4.1: Resilience and QoS requirements of sample IP services

Services requiring high traditional QoS and high resilience

- Mission critical voice-over-IP

  Sessions covering financial matters and discussions between business executives don't go along well with interruptions. They require both high quality-of-service and resilience. Therefore, business customers won't base their voice communication infrastructure on IP networks without a service level agreement assuring high service availability.

- Mission critical multimedia communication

  Real-time multimedia services, e.g. a business videoconference are severely affected even by short network outages. These applications require both QoS and resilience.

Services requiring high traditional QoS and low resilience

- Standard VoIP and multimedia-over-IP applications

  Quality-of-service may be required depending on the user preferences, but resilience might not be necessary as far as it goes along with additional cost and network failures are expected to have very low probabilities. These services can be defined as low-priority QoS traffic.

Services requiring low traditional QoS and high resilience

- Remote database transactions

  Delay jitter and bandwidth fluctuations are less critical for remote database transactions but the database is commonly locked during a transaction in order to avoid consistency problems. If due to a link or node failure the current transaction is interrupted, other transactions will stay locked-out until a time-out occurs. Thus the number of possible transactions per time interval is reduced, as will be the turnover in a commercial application. It has to be mentioned that any failure along the complete path between the database and the current customer will lead to this problem; it is not limited to failures on the access link to the network of the database host.

- Critical control terminals

  Remote control terminals of mission critical processes are often connected to the controlled system over an IP network. The transported data has low QoS requirements, but a service outage over an extended period of time may result in a breakdown of the process and possible safety problems.

- E-commerce applications often require only the exchange of a few data packets with no QoS requirements at all. Service outages however directly result in loss of revenue and loss of reputation. E-commerce customers will therefore request high network and service availability guarantees from ISPs.

  This is equally true for Web-based stock exchange as well as Application Service Provisioning (ASPs), which relies on the flexible provisioning of services with high resilience and low traditional QoS.

Services requiring both low traditional QoS and low resilience

- E-mail, FTP or WWW

  The classical best-effort services like e-mail, FTP or WWW have no QoS or resilience requirements and tolerate service degradation. In case of failures such services may well tolerate decreased service quality to maintain the QoS and availability requirements of high-priority services. It must be noted that such low-priority traffic may be discarded even if not directly affected by a network failure to allow resources previously used for the low-priority traffic be used for the alternate route of the high-priority traffic.

4.2.2 Extended QoS Signaling

The observation of independence of QoS and resilience leads to the concept of postulating an extended quality-of-service: the combination of the commonly discussed quality-of-service in terms of bandwidth and delay together with the resilience requirements of the application. Thus the correct way for signaling the resilience requirements is to include the corresponding signaling into the quality-of-service signaling between the application and the network. Corresponding to the different quality-of-service approaches (IntServ, DiffServ) this could either be done per flow or per packet.

The architecture proposed in this thesis supports resilience requirements and QoS requirements of IP services in an integrated way. This '*Resilience Differentiated QoS'* (RD-QoS) architecture is presented in the next sections.

4.3 RD-QoS Architecture – Concepts, Network Model and Components

In [*Autenrieth-2001-a*, *Autenrieth-2001-b*] the Resilience-Differentiated QoS (RD-QoS) architecture is defined, which integrates the signaling of resilience requirements with the traditional QoS signaling. The applications signal their resilience requirements in addition to their QoS requirements to the network edge. The network takes the resilience requirements into consideration for both resource management and traffic handling. At the border of MPLS domains the resilience requirements can then be directly mapped to the appropriate MPLS recovery options. In this section the RD-QoS architecture is presented. After an overview of the architecture, a service classification is proposed, and the interworking with existing QoS architectures is discussed. Then a mapping of the resilience classes to MPLS recovery mechanisms is defined. Finally, the RD-QoS architecture is evaluated in terms of resource efficiency and recovery time performance.

4.3.1 Overview

Figure 4.1 shows the RD-QoS network model and contains the basic RD-QoS components. The network is divided into QoS-enabled access networks and a traffic engineered MPLS core network

In the access networks, the QoS signaling includes a resilience attribute of the services. Packets are marked at the network boundary according this resilience attribute and assigned to a certain forwarding equivalence class (FEC). This resilience attribute is taken into consideration for the resource management in the access network.

At the border to MPLS domains, the resilience attribute of the QoS service is mapped to a FEC with appropriate recovery mechanisms. When a failure occurs in the MPLS network, affected flows can be recovered using the assigned recovery mechanism.

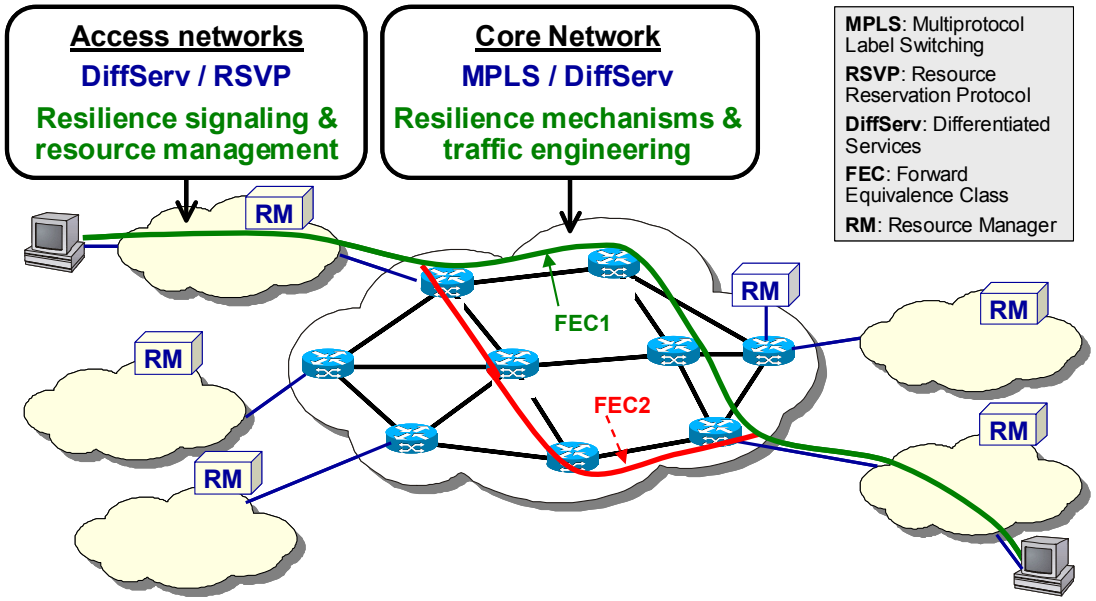


Figure 4.1: RD-QoS network model

In order to support the RD-QoS Architecture, the network must handle several key requirements and provide necessary functional components.

Signaling

Depending on the QoS architecture, the signaling may be along a full end-to-end route or from the application to the network boundary. In either case the signaling includes a resilience attribute identifying the resilience requirements of the service. The resilience signaling is discussed for the IntServ/RSVP architecture and for the DiffServ architecture in a later section.

Resource Management

The resource management performs acceptance control and traffic shaping functions depending on the resilience attribute. That means, the resource manager must take care that the required QoS level can be maintained in case of a network failure with a minimum of service outage time. This requires a careful capacity and resource management which reserves enough spare resources to allow service continuity for a given set of expected failures. The capacity management may either reserve dedicated resources on two physically disjoint paths through the network, or keep a shared pool of spare resources, which can be shared by multiple services in the event of failures.

Packet Classification

To be able to enforce a specific QoS policy, packets must be classified into a specific flow or service class. The packets classification in the RD-QoS architectures is extended in the way that the resilience attribute of the services is identified. The packet classification is done depending on the QoS architecture.

Packet Marking

Packets belonging to a specific resilience flow or service class will be marked accordingly. Depending on the QoS architecture the marking may be done using the flow label or the IP TOS-byte.

In the event of a network failure, packets belonging to a specific resilience class may be remarked to assure correct forwarding by routers which didn't detect the network failure or which don't support RD-QoS.

Traffic Conditioning

Under normal conditions (i.e. without any network failure present), traffic is handled by the QoS architecture according to the negotiated service level agreement without the RD-QoS extension.

In the event of a failure however, the traffic conditioning in the access networks may already take the resilience requirement of the service class into consideration. Packets with no resilience requirements may be dropped to free network resources for services with resilience requirements. This includes also packets of services, which were not directly affected by the network failure. Depending on the negotiated level or resilience, packets may be re-marked, and their scheduling and drop precedence redefined.

However, QoS architectures like DiffServ and IntServ/RSVP are not capable of recovering traffic flows affected be network failures in short time. Fast traffic recovery with resilience mechanisms is only possible in the MPLS-based core network.

Resilience Provisioning in the MPLS Domain

The MPLS domain must support the extended QoS architecture. The flows with resilience requirements are mapped to LSPs with appropriate resilience mechanisms. Of course, the MPLS domain must support fast and reliable failure detection and must be able to recover the affected flows after a network failure in the MPLS domain.

4.3.2 RD-QoS Resilience Classes

To reflect the resilience requirements of the services a set of four resilience classes - primarily distinguished by their recovery time requirements - is defined in [*Autenrieth-2001-a*]. Table 4.2 lists these four resilience classes.

Resilience Class 1: High resilience requirements

Services of Resilience Class 1 require recovery times of 10 to 100ms.The resilience scheme proposed for these services is protection switching. Both 1+1 and 1:1 protection is possible. For a 1+1 protection, packets must be forwarded on a working and an alternative path simultaneously. In the case of a failure on the working path, the downstream side simply selects packets from the alternative path. In the case of 1:1 protection the packets are forwarded on a predefined alternative path only when a network failure occurs. The protection resources may be used for low-priority, preemptive traffic as long as no failures are present in the working path. This requires a recovery signaling to handle unidirectional failures.

Service Class	RC1	RC2	RC3	RC4
Resilience requirements	High	Medium	Low	None
Recovery time	10-100 ms	100ms - 1s	1s - 10s	n.a.
Resilience scheme	Protection	Restoration	Rerouting	Pre-emption
Recovery path setup	pre-established	on-demand immediate	on-demand delayed	none
Resource allocation	pre-reserved	on-demand (assured)	on-demand (if available)	none
QoS after recovery	equivalent	may be tempo-rarily reduced	may have reduced QoS	none

Table 4.2: Proposed service classes and corresponding resilience options

Resilience Class 2: Medium resilience requirements

For medium resilience requirement, restoration techniques (or fast rerouting) may be used where the recovery path is setup after failure detection. In this case spare resources are inherently shared for the recovery of different working paths. On service setup, the resource management has to assure that enough spare resources are available for a given set of expected failures. In case of a network failure, packets are forwarded after a fast rerouting and reservation of spare resources

Resilience Class 3: Low resilience requirements

For services with low resilience requirements, recovery resources are not considered during traffic engineering processes (neither exclusively nor shared). In case of a failure, packets may be forwarded after a rerouting and reservation phase, if enough resources are available. This implies that the services may experience reduced QoS after the recovery.

Resilience Class 4: No resilience requirements

In case of a network failure in the administrative domain, packets with no resilience requirements may be discarded / dropped. This may happen even if the traffic is not directly affected by the network failure but rather by a re-routing of other traffic having higher resilience requirements. This corresponds to low-priority, preemptive traffic in telecommunication networks.

The resilience classes define the basic resilience behavior of the service. For more efficient resource management, additional resilience attributes may be defined. These attributes could specify if the service tolerates a reduced Quality of Service in the event of a network failure. The drawback of additional resilience attributes is that the signaling and resource management is more complex. This complexity must be weighted against achievable capacity savings.

4.3.3 RD-QoS Traffic Engineering

The classical QoS traffic engineering (TE) process for MPLS networks (see e.g. [Aukia-2000, Xiao-2000-a]) has to be extended to take the resilience differentiation into account. This RD-QoS TE process must be performed using offline routing, since a

global knowledge of the used resources and the routing of the demands is required for the determination of the resources needed for the recovery of RC2 working demands.

The RD-QoS TE process can be combined with an online routing approach. The used resources on each link for the restoration of RC2 demands are calculated offline at a network management system (NMS). In addition, the NMS calculates the available resources for RC2 demands for all ingress-egress node pair. These values are notified to the nodes. When a new RC2 service request arrives at an ingress node, this node checks if in addition to the working resources enough spare resources are available to the egress node.

The RD-QoS TE process is executed for each QoS class. Throughout the thesis and in the case studies bandwidth-guaranteed LSPs are assumed. Other traffic metrics beside the bandwidth (such as delay, delay jitter, etc.) must be mapped to an effective bandwidth requirement for the LSP.

For the RD-QoS TE process the used resources for the resilience classes on each link must be calculated. Figure 4.2 shows the resource partition on a link for a single QoS class. Resources are reserved for the active paths of RC1 and RC2 and for RC3. The demand of RC4 can share the resources of the backup paths of RC1 and RC2. In the case of a failure, the RC4 LSPs are preempted, making the resources available for the recovery of RC1 and RC2 LSPs.

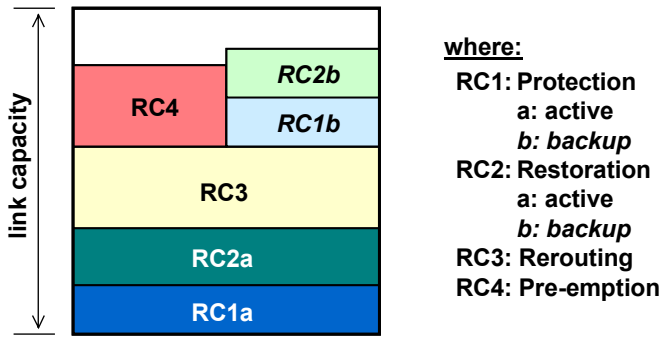


Figure 4.2: Link capacity management

The total capacity usage on each link is the maximum of

$(RC1a+RC2a+RC3+RC4)$ and $(RC1a+RC2a+RC3+RC1b+RC2b)$.

For the calculation of the resource usage all demands are routed on the network according to their resilience class. RC1 demands are assigned the highest setup priority and they are routed first. RC2, RC3 and RC4 demands are routed in turn. Within the resilience classes, the demands of different node pairs are sorted descending by their capacity. The demands of the node pair with the highest capacity are routed first. This is a simple heuristic to improve the routing.

The proposed recovery scheme for RC1 demands is protection switching. The proposed recovery scheme used for RC2 demands is a shared restoration scheme where the recovery path is calculated after the detection of a failure. The restoration mechanisms are recommended, since the recovery time requirement of RC2 (see Table 4.2) allows on-demand path calculation.

RC3 and RC4 demands can share the remaining capacity of the link, and RC4 demands can additionally use the backup resources allocated for RC1 and RC2 demands. In case of failures, RC4 demand can be preempted.

4.3.4 RD-QoS Signaling

The proposed way for signaling the resilience requirements of IP services is to include the corresponding signaling into the quality-of-service signaling between the application and the network. Corresponding to the different quality-of-service approaches (IntServ, DiffServ) this could either be done on a per-flow or on a per-packet basis.

4.3.4.1 RD-QoS Signaling in Integrated Services / RSVP Networks

In the IntServ architecture [RFC1633] resources are reserved by the RSVP [RFC2205] for every QoS flow on every router along a path from sender to receiver. In the RD-QoS architecture, the signaling is extended to include resilience requirements of the services. Additionally, spare resources are reserved in the network to provide some degree of survivability against network failures (see Figure 4.3).

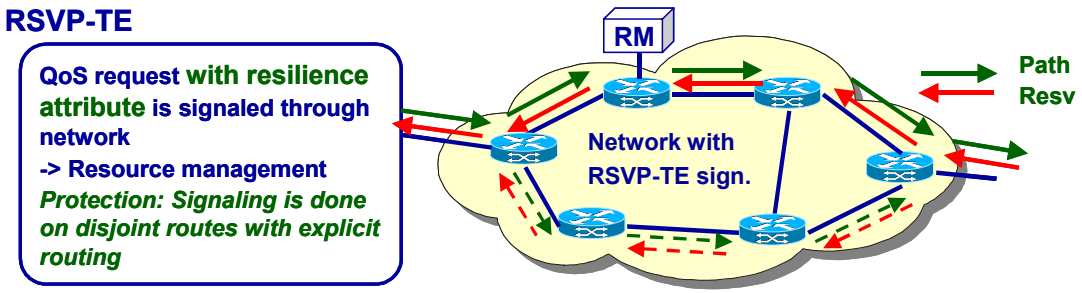


Figure 4.3: IntServ architecture with RD-QoS support

Extension to Signaling Protocol

The RSVP message formats are extended in the sense that the end user's terminal is able to signal a resilience requirement to the network in addition to the bandwidth requirement.

The proposed way to do this is to include the resilience requirement in the Rspec [RFC2210] of RSVP. The three IntServ classes – Guaranteed, Controlled Load and Best-Effort – are combined with a two-bit resilience attribute identifying the resilience class of the service.

Setup of Flow

When a RD-QoS flow is set up, the network may additionally reserve an alternative and disjoint explicit route for the flow with resilience requirements. In case of a link or node failure, the network switches the flow to the alternative route.

Note: If no purely QoS-oriented RSVP path had been identified in advance the network has to consider the path the packets would take under non-failure circumstances before

the disjoint path can be identified. This could be achieved by observing the flow of RSVP messages.

In [Beauchamps-2001] the RSVP-TE architecture was extended to support the signaling of RD-QoS with alternative paths over multiple domains.

4.3.4.2 RD-QoS Signaling in Differentiated Services Networks

The Differentiated Services (DS) architecture [RFC2475] realizes service quality by the prioritization of different services on a hop-by-hop basis. Packets are classified and conditioned at the network boundary and assigned to a behavior aggregate (see Figure 4.4). The behavior aggregate is identified by bit-patterns in the DS field, so called DS codepoints (DSCP). The DS-Field is located in the IPv4 TOS octet or IPpv6 Traffic Class octet. A specific DSCP selects a corresponding Per-Hop-Behavior (PHB) for the packet. An Expedited Forwarding (EF) PHB [RFC2598] as well as a group of Assured Forwarding (AF) PHBs [RFC2597] are already defined in RFCs with corresponding codepoints. The possible DSCPs are defined in [RFC2474].

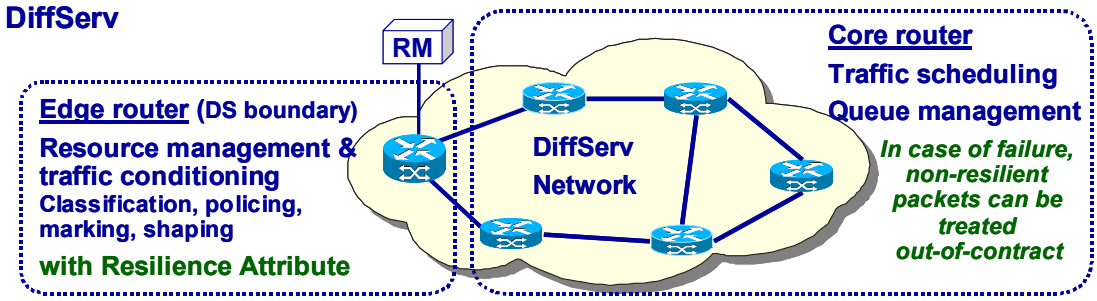


Figure 4.4: DiffServ architecture with RD-QoS support

In the following subsections only those aspects of the DiffServ architecture are discussed, which are affected by the RD-QoS Architecture.

Signaling of Resilience Requirements

The packets with resilience requirements are marked according to their resilience attribute by the application or the edge device when they enter the DiffServ network. The Resource Management takes the resilience requirements of the services into account. The required resources are reserved according to the estimated or negotiated (by service level agreements) amount of traffic having resilience requirements. In case of a network failure services with low resilience requirements are treated as out of contract and may be dropped to free resources required for services with high resilience requirements.

DSCPs for Resilience PHB

The marking of the packets is done using DSCP values. Several different options are possible for the definition of the DSCP. It is possible to define specific PHBs for behavior aggregates (BA) requiring resilience. These BAs may be independent from the already defined behavior aggregates or extend them.

To allow the provisioning of end-to-end resilience over multiple administrative domains a standardized definition of the resilience behavior is needed. This may be done be using the proposed set of four resilience classes.

The bit patterns for resilience DSCPs may either be taken from the DSCP standardized pool or the pool for local and experimental use. It must however be taken care, that a correct mapping of the resilience attributes over the domain borders is assured.

In the following, three alternatives for the coding of the DSCPs are discussed.

Use of DSCP-Bit 5 as a Resilience Identifier

The DSCP-Bit 5 distinguishes, whether a codepoint belongs to Pool 1 (Standards Actions) or to either Pool 2 or 3 (experimental or local use) [RFC2474].

The DSCP Bit 5 could be used as a resilience marker to distinguish between services with negotiated resilience guarantees and services without (Table 4.3).

Codepoint Space	**Traffic Conditioning**
xxxxx0	Without resilience constraints
xxxxx1	With resilience constraints

Table 4.3: DSCP Bit 5 as resilience identifier

The advantage of this option is that no specific bit-patterns are used to define resilience requirements. A single bit distinguishes between services requiring resilience and those without resilience requirements.

Additionally, DSCPs for resilience are only located within experimental or local use pool. This allows a correct traffic conditioning of the service classes in domains not supporting resilience differentiation.

The disadvantage for this option is that only two resilience classes can be distinguished. This disadvantage is circumvented using one of the following two DSCP coding options.

Special PHBs for resilience behavior aggregates

It is also possible to define specific PHBs for behavior aggregates (BA) requiring resilience. These BAs are independent from and parallel to other behavior aggregates.

Such resilience PHBs may be defined locally in DiffServ domains. Therefore, the DSCP for these BA should be taken from the pool 2 or pool 3.

This approach allows the highest flexibility, since no resilience PHBs must be standardized. It allows individual ISPs to offer resilience guarantees to their customers independent of neighboring domains.

The drawback of this approach is that end-to-end resilience over multiple domains is very difficult to achieve. For this objective, universally defined resilience PHBs are preferable.

Resilience PHB for existing behavior aggregate definitions

A second option is to define resilience DSCP and resilience PHBs for already defined behavior aggregates. So far, 14 PHBs are defined: the EF PHB, the AF PHB with 4 traffic priority classes with 3 drop preferences each, and a Default PHB. For these 14 defined PHBs, resilience PHBs could be defined with specific codepoints. These codepoints could be taken from the pool of recommended codepoints or from experimental and local use codepoints.

The Resilience PHBs have the same traffic conditioning characteristics as their corresponding standardized PHBs, but define additionally the traffic conditioning characteristics in case of network failures.

Traffic Conditioning

In case of link or node failures, the traffic conditioner considers flows for which no resilience requirements were signaled to be out-of-profile. These flows are either directly dropped at the boundary router, or they are assigned a very low drop precedence.

For flows with resilience requirements the required spare resources were reserved in advance. This maintains the same level of service quality in the presence of network failures. Depending on the required level of resilience, the packets may be remarked and the traffic profile modified. This influences the way the traffic is shaped and which packets are dropped.

4.3.4.3 Provisioning of RD-QoS over Multiple Domains

One objective of the RD-QoS architecture is the end-to-end provisioning of resilient services over multiple domains. For this aim, a mapping between resilience classes of different domains is needed. If both domains use the same underlying QoS architecture (e.g., DiffServ), the mapping is simple for standard resilience classes with well-defined characteristics. If both domains use different resilience classes and / or different QoS architectures, the mapping of resilience classes requires careful planning.

The European Telecommunication Standardization Institute ETSI is defining an architecture for the Telephony and Internet Protocol Harmonization Over Networks (TIPHON), where interworking issues between multiple operators are considered, however without the notion of resilience in the signaling. Following the primary design concept of RD-QoS, that is the integration of the resilience signaling and provisioning in the existing QoS and Traffic Engineering environment, the resilience should also be included in inter-domain QoS and Traffic Engineering interworking of multiple operators.

However, a detailed investigation of the provisioning of RD-QoS over multiple administrative domains is outside the scope of this work.

4.3.5 RD-QoS Recovery

With Multiprotocol Label Switching (MPLS) a Label Switched Path (LSP) is established between edge routers using MPLS signaling protocols. Since MPLS is path-

oriented and allows an explicit route definition, various recovery mechanisms are possible [Draft-Sharma]. The combination of MPLS and DiffServ [RFC3270] or RSVP [RFC3209] with the RD-QoS extension allows the provisioning of end-to-end QoS and resilience over multiple administrative domains.

The extended quality-of-service definition including resilience attributes allows the mapping of RD-QoS classes to MPLS LSPs with different protection levels according to the resilience requirements. Thus an integrated approach for the provisioning of end-to-end QoS and resilience can be accomplished.

For example, in the case of RD-QoS with DiffServ, the behavior aggregate (BA) of services with high resilience requirements can be assigned to a LSP with a defined protection path. A detailed mapping of RD-QoS classes to MPLS recovery schemes is proposed in the following section.

4.3.5.1 Interworking of RD-QoS with MPLS Recovery

The extended Quality-of-Service definition allows the direct mapping of RD-QoS classes to MPLS LSPs with different protection levels and recovery options according to the negotiated resilience requirements.

Resilience Class 1

According to the RD-QoS definition of resilience classes, the FECs of services with high resilience requirements (RC1) should be assigned to an LSP with a predefined protection path. While the recovery scope (path protection or fast reroute) and the actual recovery mechanism is left to the network operator's discretion it is strongly recommended to allow extra-traffic on the protection LSP. This allows working LSPs of RC4 to use the backup capacity of RC1.

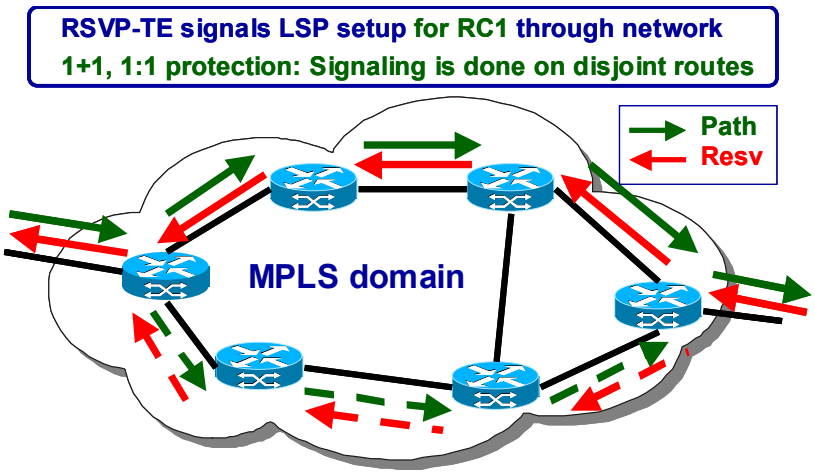


Figure 4.5: RSVP-TE protection signaling

When an LSP with high resilience requirements (RC1) is established the MPLS network (additionally) signals an alternative and disjoint explicit route using constraint- based routing extensions of the signaling protocols. In Figure 4.5, path protection signaling is shown for RSVP-TE. So far, the RSVP-TE protocol is not capable of such kind of

alternate path signaling. In [Beauchamps-2001] an extension of the RSVP-TE protocol was proposed and implemented, to provide RD-QoS in MPLS networks with RSVP-TE signaling.

After the detection of a link or node failure the network drops low priority traffic (if present) and switches the LSP to the alternative route.

Resilience Class 2

For service classes with medium resilience requirements (RC2) an LSP with a MPLS rerouting scheme is proposed. At LSP setup, only the active LSP is signaled through the network. However, the resource management must reserve enough spare resources that in the event of a failure an alternative path can be found with the required QoS.

After the failure detection the alternate path is established. To meet the required recovery time fast failure detection within a few milliseconds is required. This can be achieved using hardware failure detection and a fast Hello, KeepAlive or OAM signaling.

Resilience Class 3

For lower resilience classes (RC3) no MPLS recovery is configured and no additional resources or alternative paths are reserved.

After a failure, the network tries to recover the affected traffic only when the recovery of RC1 and RC2 is completed. This recovery may be done by the IP layer or also by MPLS. In the latter case a hold-off time is proposed to give RC1 and RC2 enough time to complete the recovery. Thus it is assured that the setup of alternative paths for RC3 doesn't occupy spare resources needed for the recovery of RC2 LSPs.

After the elapse of the hold-off time, MPLS signaling could try to establish an LSP, which may even have reduced QoS requirements.

Resilience Class 4

Low-priority LSPs with no resilience requirements can be transported as extra traffic using the protection and spare resources of higher resilience classes (RC1 to RC3) when no failures are present in the MPLS domain.

To free network resources, which are needed for services with resilience requirements, flows of RC4 may be dropped. This will happen when not enough spare resources are available for the recovery of RC2 and RC3 flows or when the RC4 flows are transported over the protection LSPs of RC1.

4.4 Implementation for RD-QoS Evaluation

4.4.1 Introduction

A software environment was implemented for the evaluation of the RD-QoS architecture. The objective of the program is to evaluate the backup capacity efficiency and the recovery time performance of RD-QoS for different networks. Therefore, the

RD-QoS TE process and a recovery time calculation are implemented in the program. In addition, the calculation times were measured as an indication for the complexity of the traffic engineering process.

The program is implemented in C++ using the LEDA library [LEDA] for graph algorithms. Additionally, the NPL library is used, an extension of LEDA for the handling of multiple network layers (demand layer, fiber layer, physical layer), which was developed at the Institute of Communication Networks. The routing mechanisms used for the RD-QoS evaluation were the DIJKSTRA algorithm in a standard version [LEDA] and in a modified version [Bhandari-1999] as well as a DisjointPath algorithm published by Bhandari [Bhandari-1999].

To evaluate the RD-QoS concepts on various networks with different demand matrices, the network topology, demand matrix and the evaluation parameters are read in from configuration files.

Generally, the RD-QoS evaluation uses capacitated networks and routes the active and backup flows through this network. Depending on the evaluated recovery model, either a sum capacity or a maximum capacity reservation was performed for the allocation of the backup resources.

4.4.1.1 Flow Diagram

Figure 4.6 shows the flow diagram of the RD-QoS evaluation program. The input files contain the physical and demand graphs as well as general program parameters. After reading in the configuration files, the network graphs are processed and additional graphs and objects are generated. The used graphs are the physical duct graph, the physical fiber graph and the demand graph. The demand between each node pairs is split into individual flows, which can be individually routed. In the second step these demand flows are routed on the physical topology and assigned to links with sufficient free capacity according to the configured routing strategy. Then, depending on the recovery strategy, the recovery routes are calculated and required backup resources are reserved. The used resources are stored in a traffic-engineering object for each fiber.

When all active and backup demand flows are routed, the recovery time is calculated for all possible link failures.

Finally, the routes, link capacity allocations and recovery times are written to files as well as to the screen with a configurable detail level, and the results are optionally visualized using a graphical display provided by LEDA.

4.4.1.2 Program Structure

Figure 4.7 illustrates the graphs and main objects and their relations modeled in the RD-QoS evaluation program.

The demand graph is read in from the configuration files. The demand between node pairs is divided into multiple flows, which can be individually routed. The demand edge class contains attributes for the resilience class of the demand, its capacity, the number of flows, and the capacity of each flow.

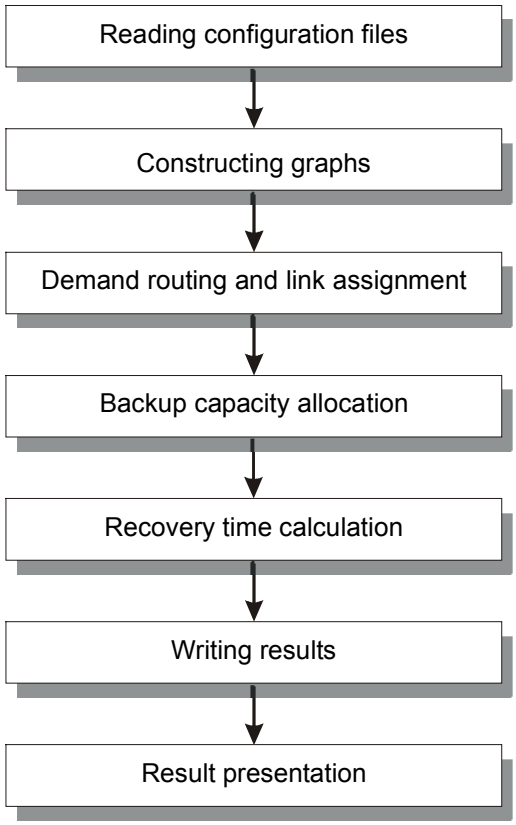


Figure 4.6: Flow diagram of RD-QoS evaluation program

The flows are stored in an array of the demand edge. In the demand graph configuration file, the demand can either be configured with a specific resilience class, or if the resilience class is unspecified, the demand is split into resilience classes according to a specified ratio. The flow capacity can also be configured in the configuration files. Thus, the number and size of flows as well as the ratio of resilience classes can be flexibly configured.

The resilience class attribute in the flow class is of course the same as the related demand edge, but it is stored redundantly to speed up reading access of the value. For each flow, the active and backup route is stored as a list of traversed fibers as well as ducts. Finally, the flow contains an identification number, which defines its position in the flow array of the related demand.

The physical duct topology is also read in from the configuration files. Each duct stores as attributes its length and the number and capacity of fibers in the duct.

During the *graph construction* phase, the fiber graph is generated, and the relations between the fiber and duct graphs are stored. Each fiber contains a list of active and backup flows traversing it. These lists are edited in the phase. During the *demand routing and link assignment and* the *backup capacity allocation* phases the allocated capacities are stored for each resilience class the RD-QoS TE class.

The classes for graph edges and nodes are derived from NPL base classes. Table 4.4 lists the main RD-QoS classes. The demand routing and backup capacity allocation is presented in the next sections.

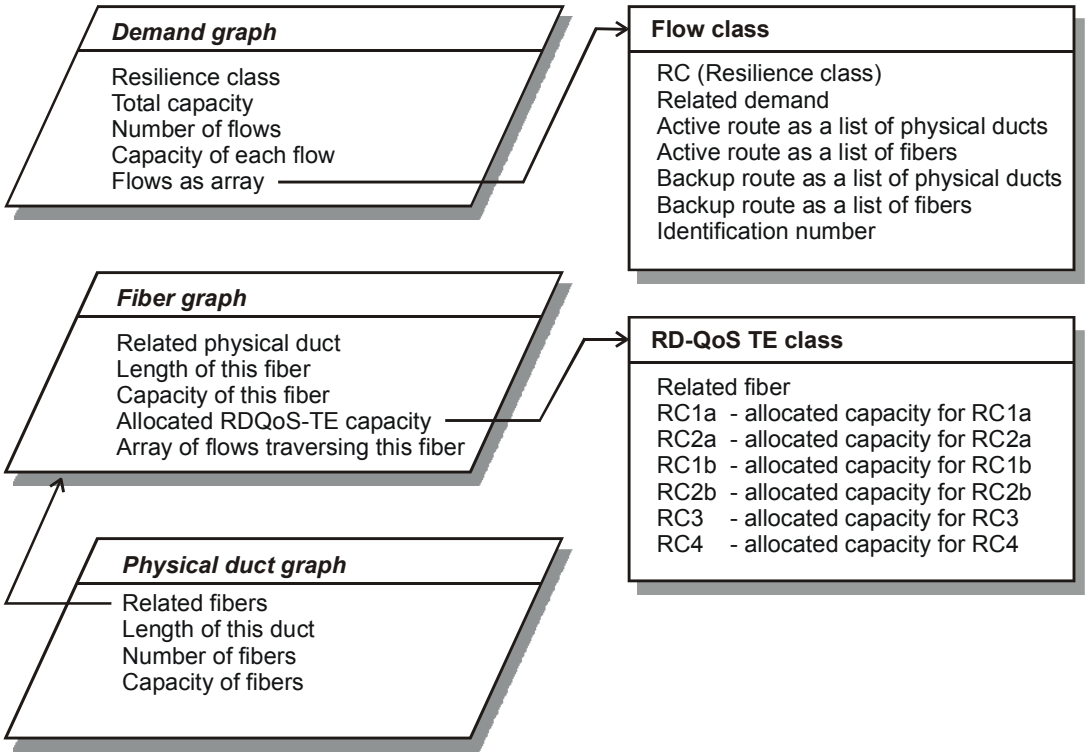


Figure 4.7: Graphs and objects of the RD-QoS evaluation program

Class Name	Description
Main	Main program flow control, reading in of configuration parameters, result output and presentation
RDQOSgeneral	Global definitions, included in every class
RDQOScontrol	Container for graphs, configuration and evaluation parameters
RDQOSconstruction	Graph construction, routing and link assignment and capacity allocation
RDQOSdemand_edge	Edge class of the demand graph
RDQOSdemand_node	Node class of the demand graph
RDQOSfiber_edge	Edge class of the fiber graph
RDQOSfiber_node	Node class of the fiber graph
RDQOSphysical_edge	Edge class of the physical duct graph
RDQOSphysical_node	Node class of the physical duct graph
RDQOSflow	Flow class for splitting demand pairs into multiple flows
RDQOSte	Resource allocation class for each fiber
DisjointPaths	Calculation of k-disjoint paths

Table 4.4: Overview of main classes of the RD-QoS program

4.4.1.3 Recovery Mechanisms

Figure 4.8 illustrates the six MPLS recovery mechanisms implemented and used for the evaluation of RD-QoS. The recovery mechanisms are described in Section 3.6.4.

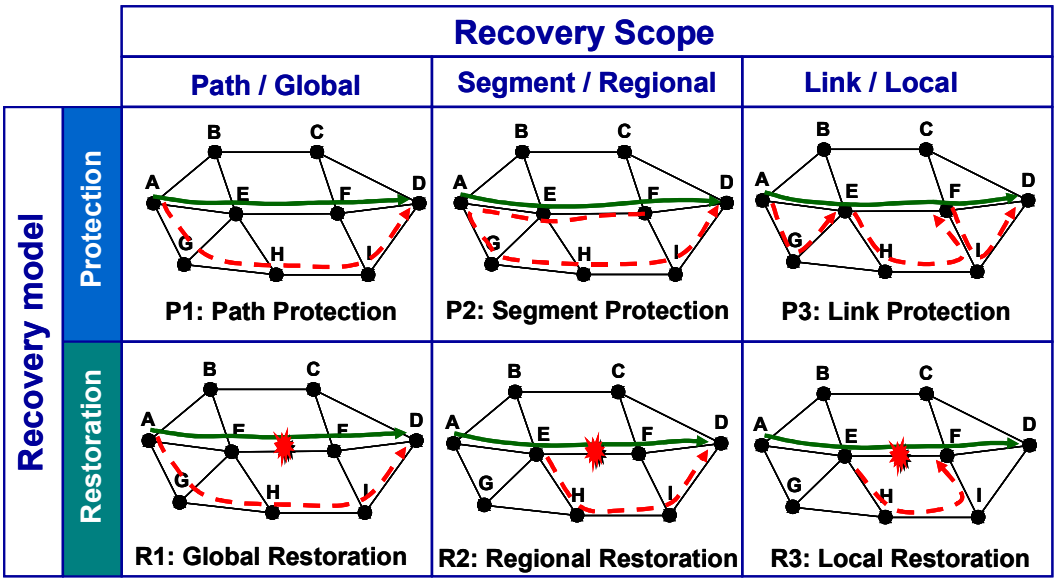


Figure 4.8: Recovery mechanisms

4.4.1.4 RD-QoS TE Process Implementation

The routing of RC1 demands is first done on the physical graph topology. In a second step the demand is assigned to a specific fiber in the fiber graph, and the resources are reserved. RC1 demands with path protection mechanisms and segment protection mechanism (P1 and P2) use a k-shortest path routing mechanism, such as the Disjoint Path routing defined in [Bhandari-1999] (see Figure 4.8). The backup resources are reserved according to the protection mechanism. For RC1 demands with a link protection scheme (fast reroute) a local shortest path routing is performed for each link of the primary path. The resilience class 1 uses a 1:1 dedicated protection with extra traffic allowed. In other words, the RC1 spare resources are dedicatedly reserved for the demand, but can be used for low-priority, preemptive traffic (RC4).

The RC2 demands are all routed using a DIJKSTRA shortest path algorithm. When all demands are routed, the restoration paths for all demands affected by a link failure are calculated. Depending on the restoration scheme, the backup path is calculated from the ingress node to the egress node (R1), from the node upstream of the failure to the egress node (R2), or from the two nodes directly adjacent to the failure (R3).

The required backup resources to recover all demands affected by a link failure are saved for every link. The total required backup resources on each link are the maximum of the required backup resources on that link for any single link failure.

The demands of RC3 are routed using a common DIJKSTRA algorithm. No spare resources are calculated.

The demands of RC4 are also routed using DIJKSTRA, however, the spare resources reserved for the backup paths of RC1 and RC2 may be reused to route the demands of RC4.

4.4.1.5 Recovery Time Calculation

The recovery time analysis model defined in Section 3.7.3 is implemented in the RD-QoS evaluation program. After the completion of the traffic engineering process, the active and backup routes of all flows are stored in the flow object.

The recovery time of the RC1 and RC2 flows is calculated for all single link failures. First of all, all flows affected by a link failure must be determined. Next, the recovery time is calculated based on the active and backup route and the number of affected flows. The recovery ratio is calculated as the quotient of the number of recovered flows to the total number of affected flows. Finally, the mean recovery ratio over all link failures is calculated.

4.4.2 Network Scenarios

To evaluate the RD-QoS two network topologies with several demand scenarios were used. The network scenarios and demand patterns will be described in the next sections.

Northern Italian Network (PANEL)

The first network topology is the North-Italian network published in [*Demeester-1999*] with 16 nodes and 36 links (Figure 4.9). The network was defined by the PANEL project. In the demand matrix the minimum demand between a pair of nodes was 1 Gb/s; the maximum demand was 16 Gb/s. If not stated otherwise, the routing was done on demand units with a capacity of 100 Mb/s.

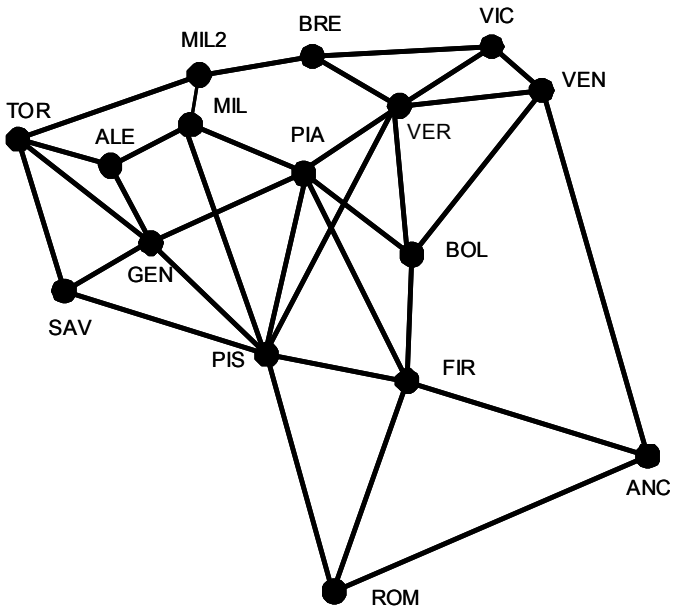


Figure 4.9: Northern Italian network (PANEL)

COST 239 Network

The second network is defined by the COST 239 project [COST239]. The pan-European network consists of 11 nodes and 25 links (see Figure 4.10). The demand matrix given in [COST239] was scaled by a factor of four – the minimum demand between a pair of nodes was thus 10 Gb/s, the maximum demand was 110 Gb/s. The routing was done on demand units with a capacity of 1 Gb/s.

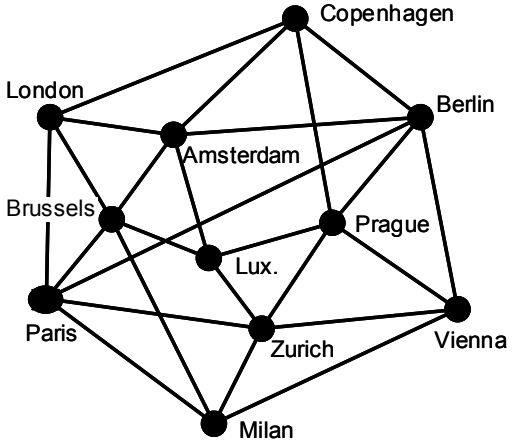


Figure 4.10: The COST239 network

For the physical network, each link was assumed to consist of eight fibers with 40 Gb/s each for each direction.

4.5 Discussion of Results

4.5.1 Resource Usage

To evaluate the backup resource efficiency of the RD-QoS architecture, the RD-QoS TE process was performed on the two presented network scenarios. For the scenarios with multiple resilience classes, a ratio of 10% RC1, 20% RC2, 40% RC3 and 30% RC4 traffic was assumed. The TE process was executed for three RC1 recovery mechanisms and three RC2 recovery mechanisms.

For comparison, the case study was performed for additional scenarios with no reserved spare resources (corresponding to 100% RC3 demands), with full restoration (100% RC2) and full protection (100% RC1). The two latter cases were done using the three different recovery mechanisms each.

The results obtained in the two case studies are published in [*Autenrieth-2002-a*] and [*Autenrieth-2002-b*]. In the following the results are discussed for the PANEL network.

PANEL Network

The scenarios are numbered from A to P in Figure 4.11 and in Table 4.5. The bars in the diagram show the total used resources per resilience class in Gb/s. The double bars of

the scenarios B to J are drawn as in Figure 4.2 The used resources are the sum of the reserved capacity for all demands on all links.

The most obvious result is that with a flexible service-differentiated resilience provisioning, the total resource usage can be reduced drastically. The required resources for the RD-QoS scenarios B-J are only slightly larger than the resource requirements without any survivability requirements (A). Compared with the fully-protected or fully-restorable scenarios (K-P), resource savings of 34% to 65% can be achieved.

The RC4 resources use the spare resources of the resilience classes RC1 and RC2. Since only those services are protected which require resilience, a gain of over 50% can be achieved depending on the recovery mechanisms used. This capacity gain may well justify the additional complexity of the TE process.

As can be seen throughout the scenarios, the 1:1 shared-path protection scheme generated by Path Protection performs better than the fast rerouting scheme proposed by Haskin (in terms of capacity requirements). However, it must be remarked that this resource gain is partially offset by additional signaling complexity and recovery delay. The worst resource efficiency can be seen with a link-protection scheme. This is because long recovery paths can be shared by more working connections than locally-isolated recovery paths. A similar behavior can be seen for the restoration mechanisms. Again, the best results can be obtained with a global restoration scheme, followed by the local-to-egress restoration scheme. The purely local restoration scheme needs the most resources for the same reason as indicated above. Again, the resource efficiency must be traded off against more complex failure notification and recovery signaling.

Regarding the complexity of the TE process, it is interesting to note that the calculation of a single scenario took less than 20 seconds on an Intel Pentium III machine with 600MHz.

COST 239 Network

The results obtained for the cost network are shown in Figure 4.12 and in Table 4.6. The results are comparable to the PANEL case study. Similar results were also obtained for various test networks with different topologies and demand patterns.

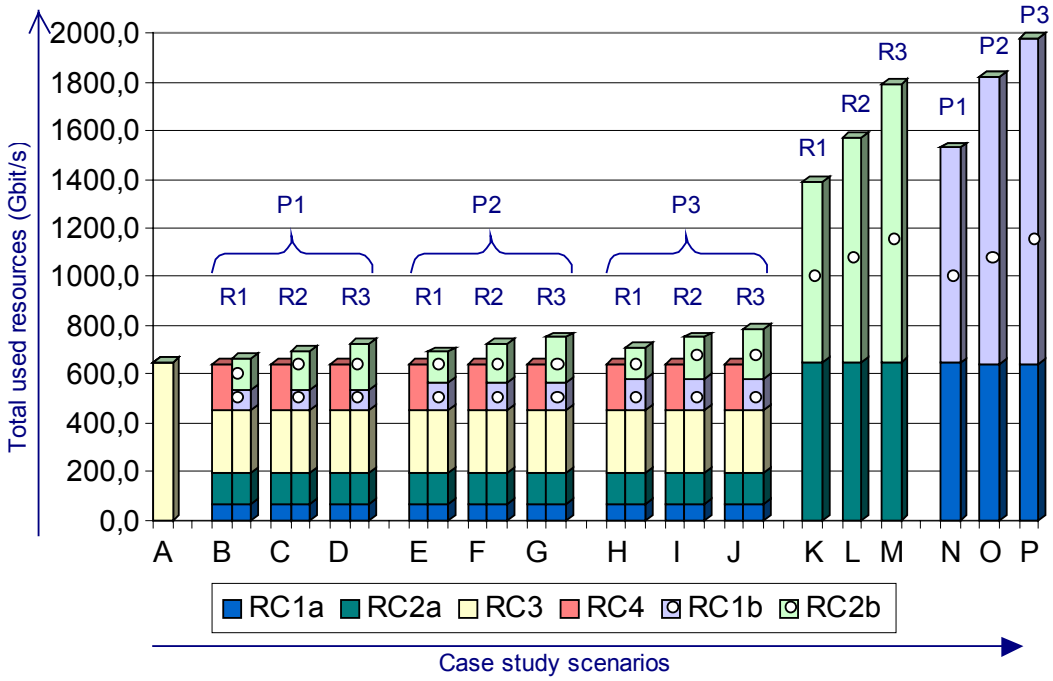


Figure 4.11: PANEL RD-QoS resource usage

	Recovery Options		Used resources per Resilience Class						Total
	RC1	*RC2*	*RC1a*	*RC2a*	*RC3*	*RC4*	*RC1b*	*RC2b*	
A	-	-	0,0	0,0	646,0	0,0	0,0	0,0	646,0
B	Path Protection	Global Rest.	64,2	128,2	256,4	192,3	87,6	127,8	714,4
C	Path Protection	Regional Rest.	64,2	128,2	256,4	192,3	87,6	159,5	741,0
D	Path Protection	Local Rest.	64,2	128,2	256,4	192,3	87,6	186,6	776,7
E	Segment Prot.	Global Rest.	64,2	128,2	256,4	192,3	117,0	124,7	744,7
F	Segment Prot.	Regional Rest.	64,2	128,2	256,4	192,3	117,0	159,8	773,8
G	Segment Prot.	Local Rest.	64,2	128,2	256,4	192,3	117,0	185,2	806,2
H	Link Protection	Global Rest.	64,1	128,2	256,4	192,3	133,3	125,4	772,4
I	Link Protection	Regional Rest.	64,1	128,2	256,4	192,3	133,3	168,3	807,7
J	Link Protection	Local Rest.	64,1	128,2	256,4	192,3	133,3	204,5	858,6
K	-	Global Rest.	0,0	645,5	0,0	0,0	0,0	744,1	1389,6
L	-	Regional Rest.	0,0	645,5	0,0	0,0	0,0	928,2	1573,7
M	-	Local Rest.	0,0	645,5	0,0	0,0	0,0	1147,1	1792,6
N	Path Protection	-	643,7	0,0	0,0	0,0	886,6	0,0	1530,3
O	Segment Prot.	-	642,8	0,0	0,0	0,0	1180,2	0,0	1823,0
P	Link Protection	-	642,2	0,0	0,0	0,0	1339,5	0,0	1981,7

Table 4.5: RD-QoS Resource usage in the PANEL network

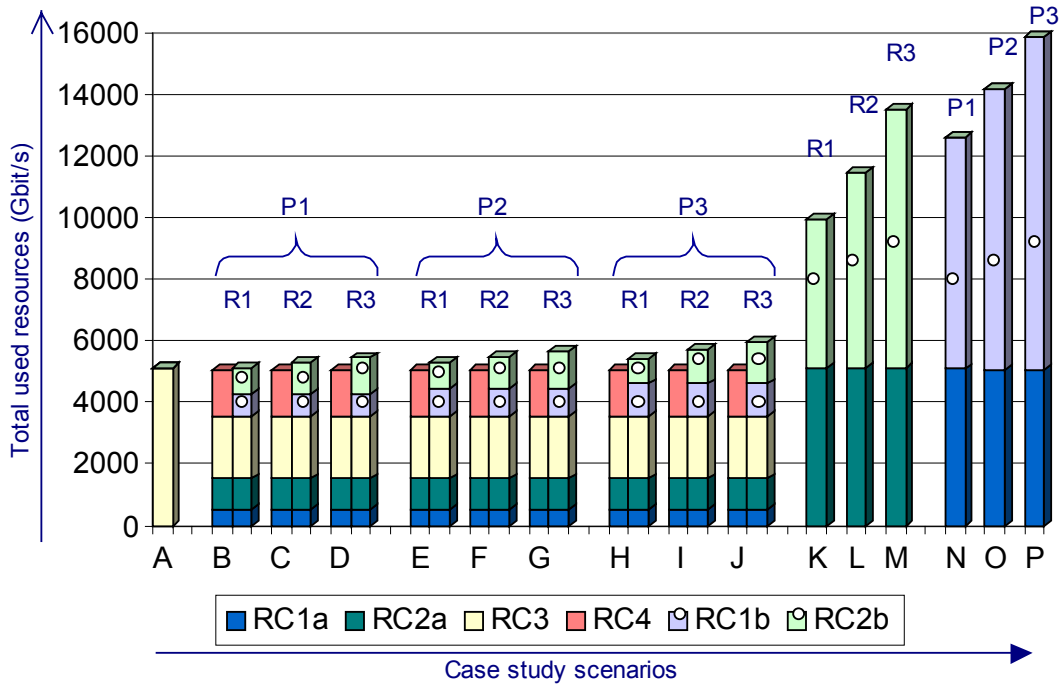


Figure 4.12: COST 239 RD-QoS resource usage

	Recovery Options		Used resources per Resilience Class						Total
	RC1	*RC2*	*RC1a*	*RC2a*	*RC3*	*RC4*	*RC1b*	*RC2b*	
A	-	-	0	0	5126	0	0	0	5126
B	Path Protection	Global Reroute	507	1014	2028	1521	750	811	5464
C	Path Protection	Regional Rest.	507	1014	2028	1521	750	1028	5712
D	Path Protection	Local Reroute	507	1014	2028	1521	750	1160	5949
E	Segment Prot.	Global Reroute	507	1014	2028	1521	909	831	5668
F	Segment Prot.	Regional Rest.	507	1014	2028	1521	909	1041	5880
G	Segment Prot.	Local Reroute	507	1014	2028	1521	909	1205	6080
H	Link Protection	Global Reroute	507	1014	2028	1521	1056	805	5926
I	Link Protection	Regional Rest.	507	1014	2028	1521	1056	1107	6209
J	Link Protection	Local Reroute	507	1014	2028	1521	1056	1350	6531
K	-	Global Reroute	0	5121	0	0	0	4861	9982
L	-	Regional Rest.	0	5121	0	0	0	6371	11492
M	-	Local Reroute	0	5121	0	0	0	8429	13550
N	Path Protection	-	5089	0	0	0	7540	0	12629
O	Segment Prot.	-	5081	0	0	0	9141	0	14222
P	Link Protection	-	5070	0	0	0	10849	0	15919

Table 4.6: COST 239 RD-QoS resource usage

4.5.2 Recovery Time Analysis

The recovery time analysis (RTA) was performed on the PANEL network for different numbers of flows. The graph in Figure 4.13 shows the mean RD-QoS recovery ratio averaged over all link failures as a function of the time for the three protection switching

mechanism. The traffic mix was 10% RC1, 20% RC2, 40%, RC3 and 30% RC4. As can be seen, the recovery of all flows is finished between 31 and 39 ms. The smallest recovery time is achieved by link protection, followed by path protection and segment protection.

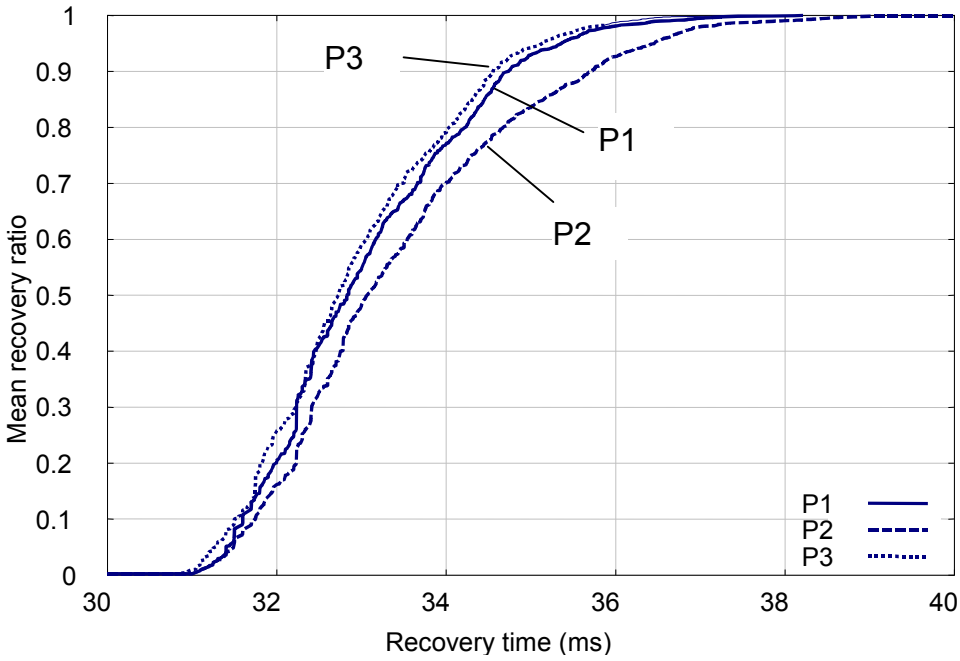


Figure 4.13: Protection switching RTA

MPLS can achieve very fast protection switching times, since only unidirectional protection switching is used and no protection switching signaling is required. Moreover, the label forwarding tables in intermediate nodes can be preconfigured.

This also explains the relatively small differences between the three protection switching models. The segment protection using the protection switching mechanisms proposed by Haskin is the slowest, since this mechanism results in the longest backup routes.

In Figure 4.14 the mean recovery ratio for the three restoration mechanisms is given in addition to the protection switching recovery ratios shown in Figure 4.13. The MPLS restoration mechanisms are much slower than protection switching mechanisms, since the explicit routes are calculated for each LSP sequentially, and the backup LSP must be setup using a label distribution protocol. Local and regional restoration is slower than global restoration, since it has a larger computation time: all affected LSPs are rerouted at the same LSR. With global restoration, the computational load is distributed between multiple I-LSRs, which have to reroute a smaller set of LSPs, only. Since the route computation time is large in comparison to the signaling time, the mean recovery time for regional restoration and link restoration are almost equal.

The maximum restoration times computed in this case study are 164 ms for the global restoration, and 242 ms for regional and local restoration.

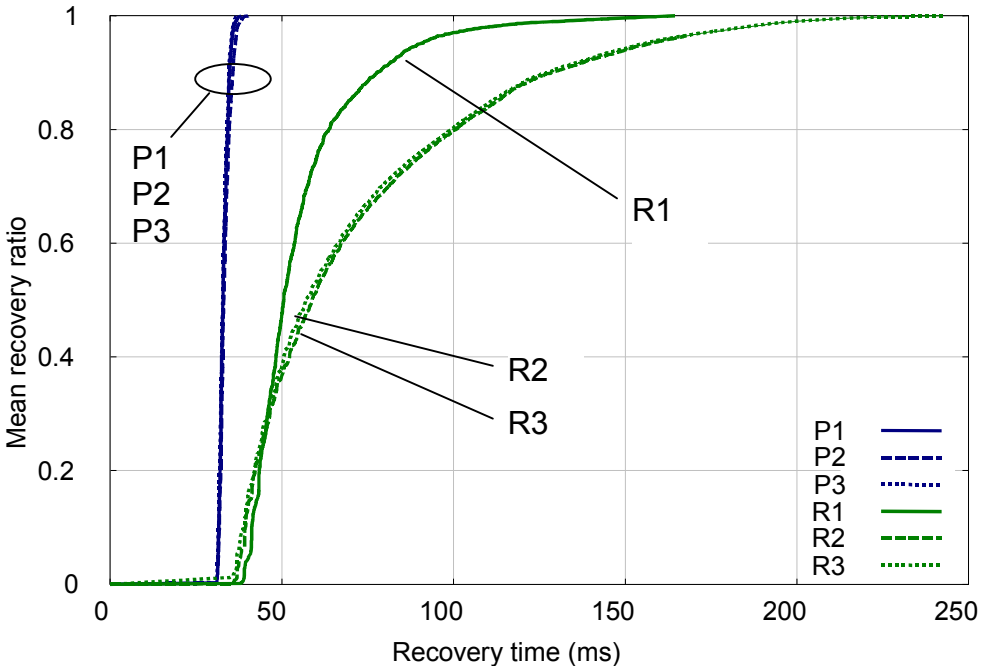


Figure 4.14: Protection and restoration RTA

In Figure 4.15 the graphs resulting from a traffic mix with 10% RC1 and 10% RC2 are compared with a scenario with 100% RC2 traffic. The latter case corresponds to a 100% restorable network using restoration mechanisms. In contrast to the previous figures, the x-axis is now in logarithmic scale.

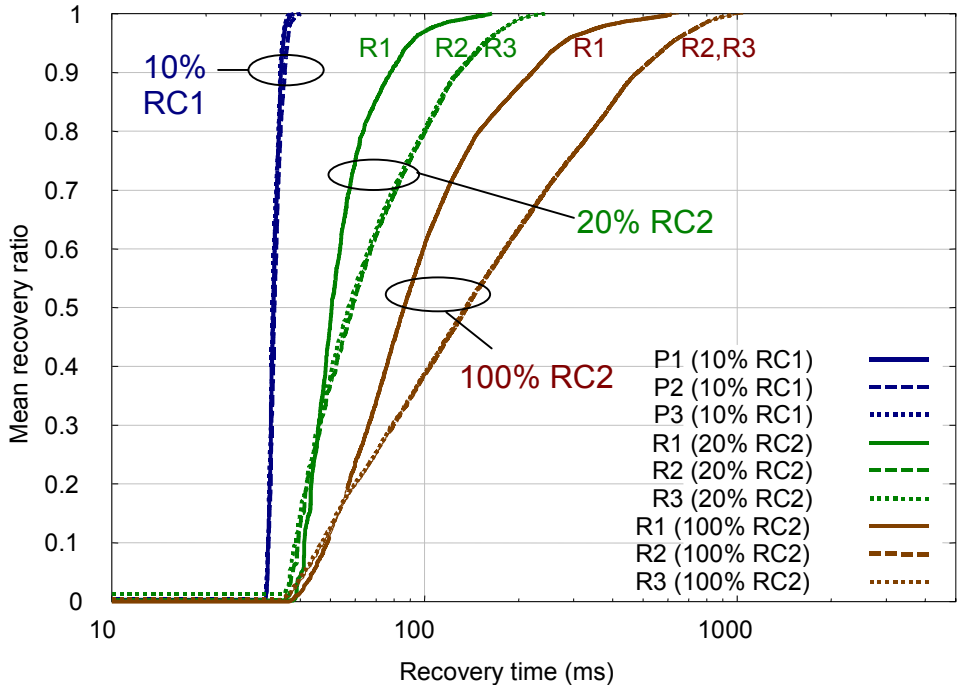


Figure 4.15: RTA comparison with 100% RC2

Since with a traffic mix of 100% RC2 all flows must be restored, the resulting recovery time is significantly larger than with a differentiated traffic mix.

An immediate advantage of the differentiated resilience approach is that the total number of restorable flows is reduced, thus reducing the computational load on the routers and the resulting recovery times.

Table 4.7 summarizes the mean and maximum recovery times calculated for the nine graphs shown in Figure 4.15 (all timing values are given in ms).

Traffic mix	10% RC1, 20% RC2						100% RC2		
Rec. mechanism	P1	P2	P3	R1	R2	R3	R1	R2	R3
MTTV (ms)	33.1	33.4	32.9	55.9	73.0	71.9	114.1	208.9	207.8
MaxTTV (ms)	38,2	40,2	37,7	164,5	242,4	242,4	650,3	1044,0	1044,0

Table 4.7: Mean and maximum recovery times

4.5.3 Used Resource Versus Maximum Recovery Time

It is interesting to compare the performance of the different recovery mechanisms in terms of recovery time and resource usage directly. In Figure 4.16 the sums of active and backup resources of RC1 and RC2 are plotted against the maximum recovery time for the six recovery mechanisms.

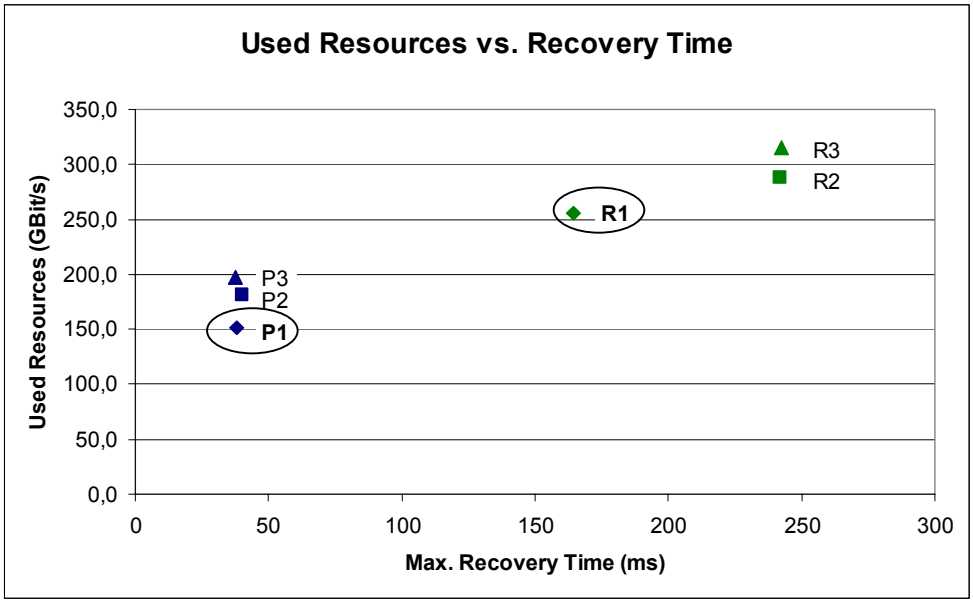


Figure 4.16: Used resources versus recovery time

The three protection switching mechanisms differ only marginally in the maximum recovery time, but the used resources of P2 and P3 are 19,4% and 30,0% higher than those of P1. The resource usage of the restoration mechanisms R2 and R3 are 12,4%

and 23,0% higher than the resource usage of R1. Moreover, the maximum recovery time of R1 is much better, as well. The maximum recovery times of R2 and R3 are both 47,4% higher than R1.

Summarizing the global restoration and path protection mechanisms (encircled and printed in bold in Figure 4.16) are performing much better than their regional and local counterparts regarding the resource usage and recovery time. Therefore, these mechanisms should be preferred when selecting a resilience strategy for the RD-QoS architecture.

4.5.4 Influence of Number of Flows

The previous case studies were performed with a flow size of 0.1 Gb/s. In case of the PANEL network, this results in 10 to 160 flows per demand pair, or a total number of 3881 flows. In this section the dependence of the recovery time from the number of flows is evaluated. For this case study, the total demand stays fixed, but the flow size is varied from 10 Mb/s to 100 Mb/s. Only path protection and global restoration mechanisms were used for this case study.

In Figure 4.17 the mean recovery ratio of 10% RC1, 20% RC2 and 100% RC2 is shown for three different values of the flow size, resulting in nine graphs. While the mean recovery time of the protection switching case is growing very slowly, the recovery time of restoration mechanisms is growing much faster with the number of flows.

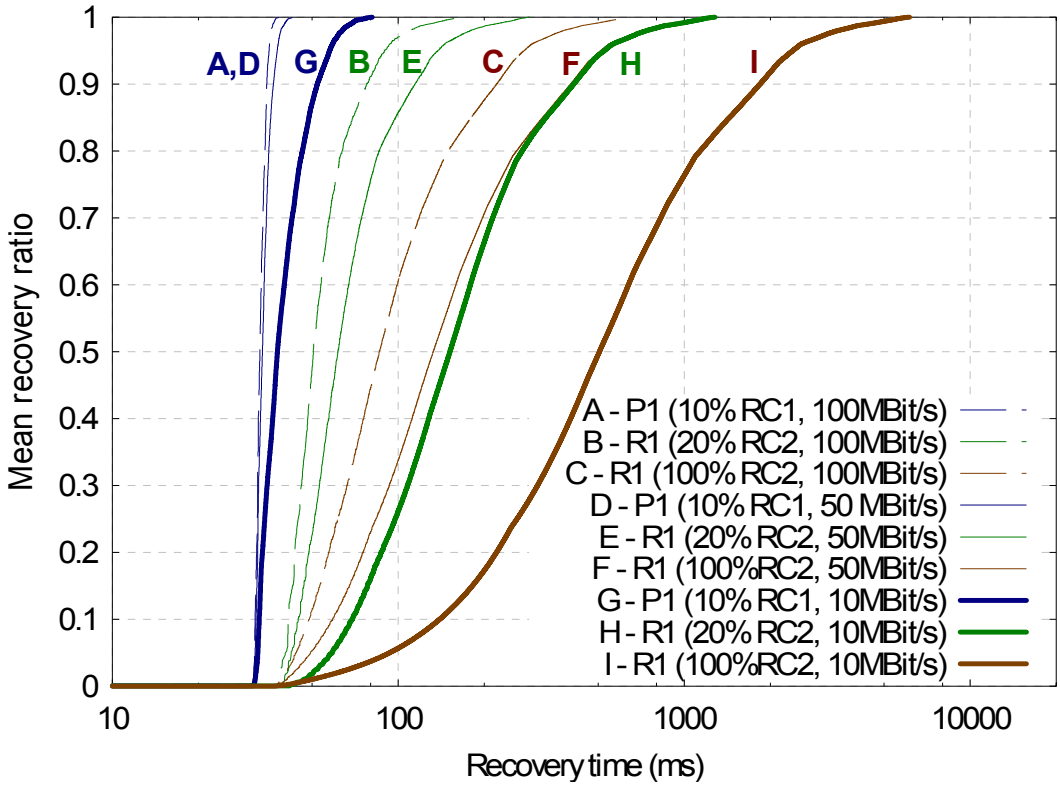


Figure 4.17: Mean recovery ratio for different numbers of flows

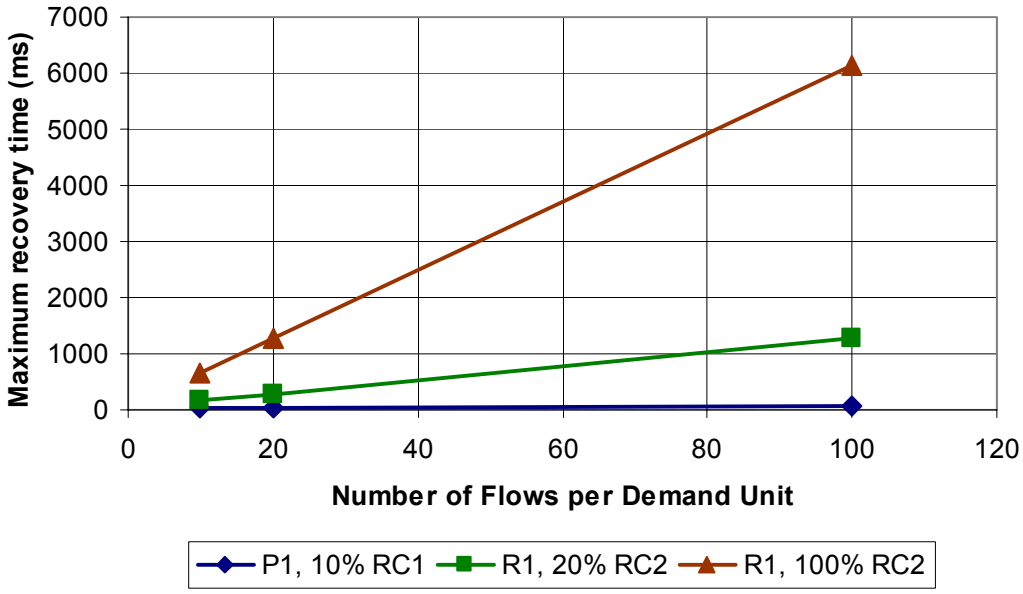


Figure 4.18: Maximum recovery time for different numbers of flows

Figure 4.18 illustrates the dependence of the recovery time on the number of flows. Only the maximum recovery time is depicted for the three traffic cases. The number of flows per demand unit is calculated by the minimum demand between a pair of nodes, which is one Gb/s, divided by the size of a flow. For the size of the flows, values of 100 Mb/s, 50 Mb/s and 10 Mb/s are considered, resulting in 10, 20, and 100 flows per demand unit.

It can be seen from the graph, that even with the fastest restoration mechanism, the recovery time target of one second for RC2 flows cannot be reached if the number of flows increases. For the RD-QoS traffic mix with 20% RC2 traffic this threshold is exceeded for 80 flows per demand unit, while with 100% restoration traffic the threshold is already exceeded for 15 flows per demand unit.

4.5.5 Summary of Results

MPLS is a promising architecture for the resilience provisioning in IP-based networks. The RD-QoS architecture additionally allows the signaling of resilience requirements, thus offering a customized level of resilience. The benefits of resilience provisioning with RD-QoS and MPLS must however be compared to alternatives such as classical IP rerouting, pure MPLS protection and restoration and lower layer recovery (e.g. SDH Automatic Protection Switching or MS-SPRing).

Resource Efficiency

MPLS Recovery with RD-QoS allows the flexible provisioning of differentiated resilience to service classes. Since the services are protected with exactly the required degree of resilience, high resource efficiency can be achieved. The resource efficiency of RD-QoS is very high in comparison to pure MPLS restoration or protection. Since only

a smaller portion of traffic is protected, the required backup resources are naturally lower. In addition, the backup resources can be used to offer unprotected, low-priority traffic.

A direct comparison of the resource efficiency of RD-QoS with multilayer recovery approaches was not done in the case studies. However, since the total IP traffic transported over the server layer in the RD-QoS case is not much higher than with best-effort IP traffic, a multilayer strategy protecting all IP traffic will certainly require more server layer capacity than the RD-QoS case. An extension of the RD-QoS architecture is to use the RD-QoS architecture to signal the IP resilience requirements, but to provide the actual recovery switching in the lower layer. This approach is discussed in the next chapter.

Recovery Time

As shown in the case studies, the recovery time requirements specified for the resilience classes can be reached. However, if the number of RC2 flows is high, the restoration time of RC2 services may exceed one second.

Due to the coarser protection granularity and the fast hardware failure detection, lower layer recovery strategies are generally faster than client layer recovery strategies.

Protection Granularity

The protection granularity at lower layers is very coarse. Commonly, the smallest protection unit is a VC-4 container. For optical layer protection, the smallest protection unit is one optical channel with a capacity of 2.5 Gb/s, 10 Gb/s, or 40 Gb/s.

Recovery in MPLS allows the assignment of different recovery options to individual FECs based on their destination and QoS requirements. With RD-QoS signaling, FECs may additionally be assigned based on their resilience requirements. Therefore, the protection granularity of MPLS-based recovery mechanisms is one LSP.

Failure Scope

Lower layer recovery offers fast recovery against link failures like fiber cuts and intermediate node outages. However, failures of nodes terminating the client layer connections and failures of client layer equipment cannot be recovered in the server layer. Only resilience mechanisms present in the client layer are able to restore these failures.

However, if the MPLS and the server layers are planned independently and no information about the physical routing of server layer connections carrying MPLS LSPs is given, a single link failure in the server layer may result in multiple failures in the client layer. The recovery of these multiple failures could fail. The simplest solution to this problem is to use a one to one mapping between the physical duct topology and the MPLS link topology.

Protocol Complexity

The finer the recovery granularity the more connections must be recovered in the case of failures. Multiple recovery classes increase the protocol complexity even more. In addition, the QoS architectures must also support the resilience classes. This increases the complexity of the QoS architecture implementation.

On the other hand, the integrated signaling of resilience and QoS allows the services based provisioning of resilience from the application. The alternative is to setup service resilience by the network management. This increases the operational costs of the network.

4.6 Summary

Network survivability is a key requirement of traffic engineered networks. Survivability mechanisms are available at multiple network layers, e.g. SDH/SONET, OTN and MPLS. Moreover, these resilience mechanisms may even be in operation in multiple layers at the same time. While recovery at lower layers generally has advantages in the time scale of the recovery operation, recovery at the IP or MPLS layer allows a better resource efficiency, recovery granularity and QoS granularity. A resilience-differentiated approach could protect only those traffic flows that require a high level of service availability. This results in a more cost-effective network design and traffic engineering.

Therefore it is reasonable for an ISP to provide the required network survivability using only resilience mechanisms in the IP layer. That way, also the network operation and management complexity could be reduced, since all traffic-engineering aspects (including resilience) are managed in the IP layer only. ISPs can offer unprotected and protected services (the latter at higher cost) with a single administrative platform, including user authentication and billing. This is a major advantage since it reduces the operational cost of the network and increases service flexibility. Customers who accept lower network resilience may be offered lower-cost network services. Customers demanding high network resilience are charged according to the level of resilience.

In this section an extension of the Quality of Service signaling to include resilience requirements of IP services was presented. RD-QoS defines an architecture for the flexible provisioning of differentiated resilience to service classes. Since the services are protected with exactly the required degree of resilience, high resource efficiency can be achieved. The immediate advantage for an ISP is, that the resilience can be treated as a value-adding service, which can be charged for.

After a detailed description of the RD-QoS architecture, a traffic engineering process was defined. The backup resource usage was evaluated in a software evaluation showing the significant capacity saving achievable with this approach. Additionally the recovery timing for RC1 and RC2 services was evaluated. This showed, that path protection and global restoration mechanisms have the shortest recovery times in addition to the better resource efficiency.

The current trend is clearly towards a service-driven transport architecture. The resilience requirements should therefore be included in the QoS signaling just like the bandwidth and end-to-end delay requirements.

5 MULTILAYER RESILIENCE EVALUATION

5.1 Introduction

For the simulation and evaluation of the multilayer recovery strategies and interworking mechanisms introduced in Section 3.9 an integrated multilayer simulation environment has been developed, allowing the detailed modeling of the network elements and recovery protocols. A comprehensive description of the multilayer simulation environment and an evaluation of the obtained performance simulation results are contained in project deliverables and reports [*PANEL-D4*, *PANEL-D5*, *PANEL-D6*, *PANEL-FR*, *Autenrieth-1998-a*] as well as in several publications [*Autenrieth-1998-b*, *Autenrieth-1998-c*, *Demeester-1999*].

The simulation tool, its network model and its network element model will be presented in the next section. Then, performance results of a case study of two multilayer recovery strategies for an SDH based ATM network are presented and discussed.

5.2 Multilayer Simulation and Evaluation Environment

The network simulator SELANE (Simulation Environment for Layered Networks), developed in the PANEL project [*PANEL-D6*, *PANEL-FR*], is implemented using the 'Specification and Description Language' (SDL) [ITU-T Z.100], a formal description language especially suited for the modeling of communication protocols. The use of SDL allows a detailed specification of the signaling protocols and their timing aspects for protocol simulations. Additionally, performance simulations for the evaluation of different protocols are possible.

Another important aspect for the decision for the formal description language SDL is the fact that different parts of the simulation environment have been developed separately at two research institutes. Due to the hierarchical structure and the graphic representation of SDL, the integration of the different modules was possible without problems and in a very short time.

5.2.1 Network Model

The approach chosen for PANEL is to model the whole network within a single SDL system. The different physical network components (e.g., ATM VP cross-connects) are implemented as SDL block types defined at system level (see Figure 5.1). The individual network elements of the simulated network are dynamically instantiated from their corresponding block type. The concept of dynamic instantiated SDL block types used for the implementation is published in [*Iselt-1997*]. SDL processes within the cable block are used to model the cables and fibers interconnecting these network elements. At simulation start-up, the required number of link processes is instantiated dynamically depending on the network topology.

The logical topology of the network is represented by variables stored in the individual network elements. This distributed concept allows a very flexible configuration of the simulation environment. The physical and logical network topology can be imported from a planning file to configure the instantiated network components. Thus, multiple networks with different technologies and topologies can be simulated without recompiling the system.

The Central Manager block indicated in the figure fulfils some functionality of the network management system (NMS) and additional simulation related functions. Examples for such functions are the configuration of the network based on the imported planning information, the measurement of performance data, and the statistical calculation and injection of random network element failures. Another function of the Central Manager is the communication with a graphical user interface (GUI) connected to the simulator. The GUI allows an interactive setting of simulation parameters like timing values, the graphical injection of failures, and the control and visualization of the simulation itself. The simulation flow can be manually controlled step-by-step or by setting the speed of the signal flow. The level of detail of the visualized messages can also be defined. The methodology to control the SDL simulation from a graphical user interface is published in [*Kellerer-1998*, *Kellerer-2000*].

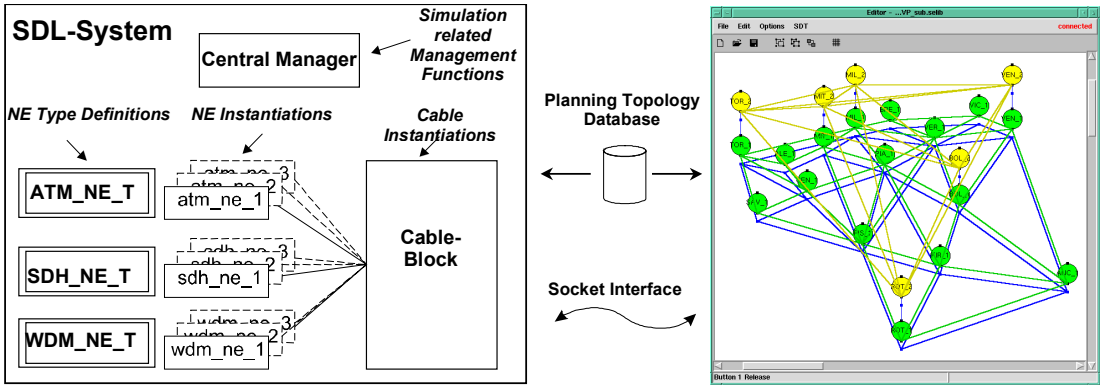


Figure 5.1: System level of the SDL specification and graphical user interface (GUI)

5.2.2 Network Element Model

For network simulations the management, control, and transfer functionality of a network element has to be modeled. Figure 5.2 shows the layering model, the atomic functions within a layer, and the functional architecture of network elements used in the simulation environment.

The layering concept used for the specification of the network elements is based on the recommendations from ITU-T and ETSI [ITU-T G.805, ETSI 300 417]. The left part of Figure 5.2 shows the relevant layers and their client/server relations for the ATM, SDH and WDM technology.

The basic functionality in the transfer layer, i.e. the adaptation, trail termination, and connection atomic functions, is common to every layer and is specified as block and process types. The trail termination, connection, and connectivity functions are grouped

together to simplify the implementation of the model and to increase the simulation performance. These generic block types are instantiated for every layer present in a specific network element.

The right side of Figure 5.2 shows as an example the functional architecture of an ATM VP cross-connect. The management plane consists of a block representing the equipment management function (EMF) and the coordination function (CoF). For each layer present in the transfer function, a corresponding block in the layer management plane exists. All layer specific functions like the detection of a VP alarm indication signal (AIS) are implemented in the layer management blocks.

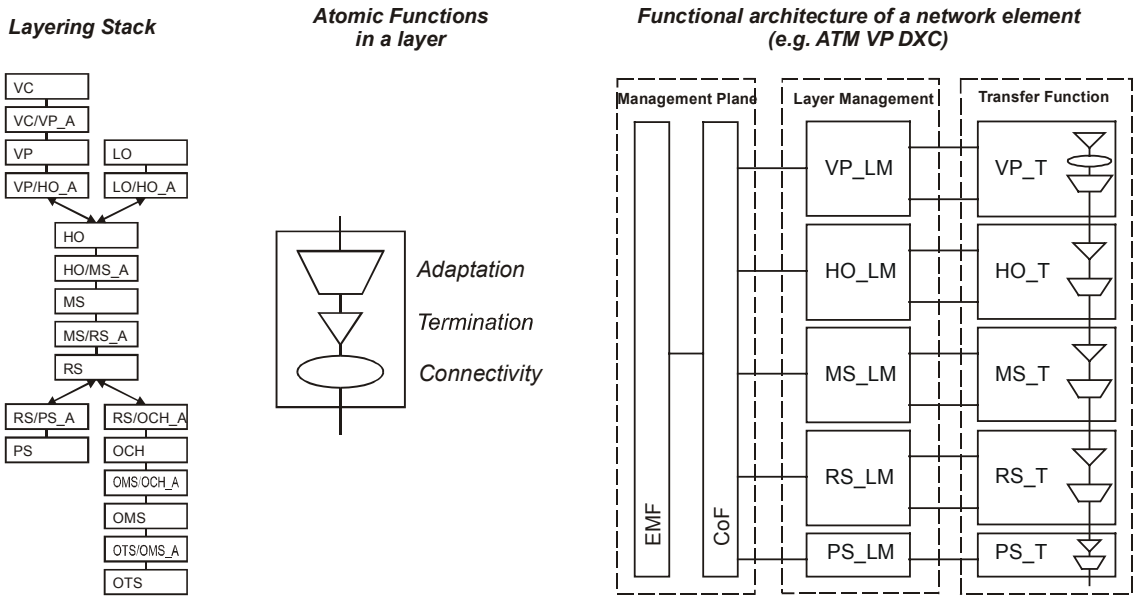


Figure 5.2: Layering stack, atomic functions and functional architecture of a network element

5.2.3 Signaling

The communication between the network elements is based on the transmission of signals. For performance reasons, user data is not simulated. Only operation, administration and maintenance (OAM) signals and alarm signals are transmitted in the simulation. The correct routing of the signals through the network and within the network elements is done by routing information added to the signals. This routing information consists of the trail and link connection identifiers of the route.

5.2.4 Timing Model

To be able to perform a detailed timing analysis of the events in the network, the simulation environment has to incorporate different delays. The main delay types are discussed in the following chapters.

5.2.4.1 Link Propagation Delay

Each fiber adds a propagation delay to the messages, depending on the length of the fiber. For example, a signal transmission over 50 km of fiber has a delay of 250 µs. In SELANE each fiber is modeled as a SDL process. Upon the reception of a signal, a "propagation-delay" timer is set. When this timer expires, the message is forwarded to the next network element.

5.2.4.2 Network Element Propagation Delay

Within network equipment, transfer delays may be introduced by buffers (e.g., pointer buffer, mapping and de-mapping buffers, jitter reduction buffers), switch fabrics, frame alignment circuits, encoders and decoders, series/parallel and parallel/series converters, OAM processing and other equipment specific processes. These transfer delays depend on the hardware implementation, and thus it is difficult to assign the individual delays to specific atomic functions. In SELANE the network element propagation delay is assigned to the connection atomic function. The values for a SDH VC4 crossconnect are set to 1 µs by default, while for an ATM VP cross-connect the value is set to 5 µs. The values can be flexibly configured using the initialization file.

5.2.4.3 Processing Time Within Nodes

In the case, when recovery from failures is provided using distributed restoration mechanisms, the software protocol to process the OAM messages is implemented in the controller (EMF) of the network element. The timing model within the EMF is based on the timing model published in [Kawamura-1995-a]. However, unlike the bulk processing model (-/$D^{[N]}$/1) described there, a simpler multi-server queue (-/D/n) is used. Three processes are dealing independently of each other with the OAM signal processing (Figure 5.3).

The first process takes care of the transport of the OAM messages from the DXC to the EMF. This is realized with *n* independent single FIFO-servers (IF) ①. The value of *n* is typically 16 and the serving time is estimated to be 20 ms. Examples of signals reported to the EMF are the detection and clearance of defects, APS messages, and recovery token.

The second process handles the self-healing actions. This is implemented as a single FIFO queue with a serving time of 2 ms②. Examples of signals processed in this self-healing function (SHF) are the backup VP trigger signals (e.g. recovery token or signal fail, possibly after hold-off time) and the K1-K2 APS bytes [ITU-T I.630]. If a shared backup VP algorithm is used, the capacity needs to be captured in the call admission control (CAC) process③. Note that the self-healing process is not included in the SHF. The CAC processing time is estimated to be 30 ms. To switch over from working to protection entity (upon command of the SHF) the routing tables in the DXC need to be updated⑤. This update requires 30 ms and does not block the SHF.

The third process takes care of the injection of new OAM cells. This is being realized with 20 independent Output FIFO (OF) queues④. The serving time is estimated to be 90 ms.

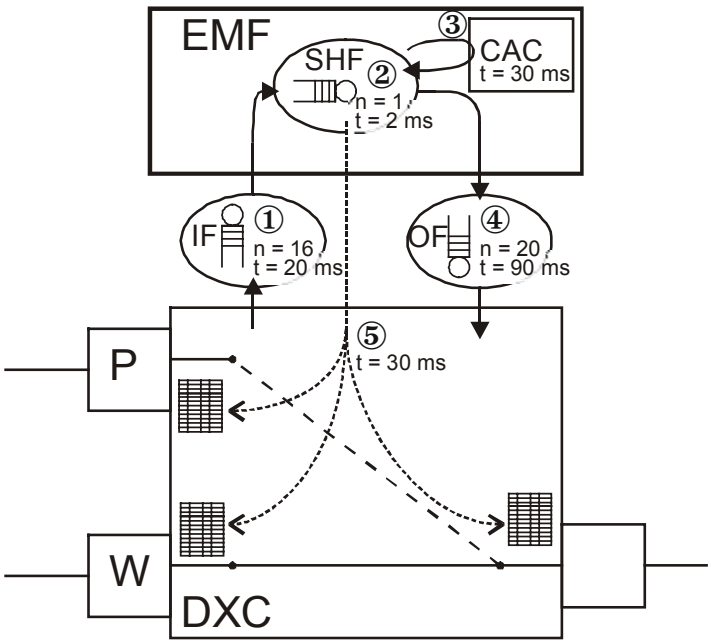


Figure 5.3: Processing model inside the VPXC

The table below compares the settings used in PANEL with the settings described in [Kawamura-1995-a].

Parameter	**Kawamura**	**PANEL**
Queue model	-/$D^{[N]}$/1	-/D/n
Number of input processes	16	16
Processing time of input process	18 ms	20 ms
Number of output processes	20	20
Processing time of output process	90 ms	90 ms
Number of self-healing processes	1	1
Processing time of self-healing process	2 ms	2 ms
Processing time of capacity allocation	-	30 ms
Processing time of routing table update	-	30 ms

Table 5.1: Comparison of the processing model of PANEL and of Kawamura

5.3 Discussion of Results

The different interworking strategies were simulated and evaluated using the simulation environment described in Section 3.9. The case studies are based on a 32-node network spanning the whole of Italy, and a simplified network with only the northern part of Italy and 16 node sites. For each network topology there was either a high or a low ATM

demand, resulting in the following four different networks (Table 5.2). Figure 5.4 shows the 16-node network with 8 equipped ATM offices as an example.

	SDH nodes	**ATM nodes**	**Physical links**	**SDH VC-4s**	**ATM 8 Mbit VPs**
Network 1	16	6	35	400	750
Network 2	16	8	35	500	2200
Network 3	32	10	70	900	1850
Network 4	32	16	70	1100	5300

Table 5.2: Parameters of network scenarios

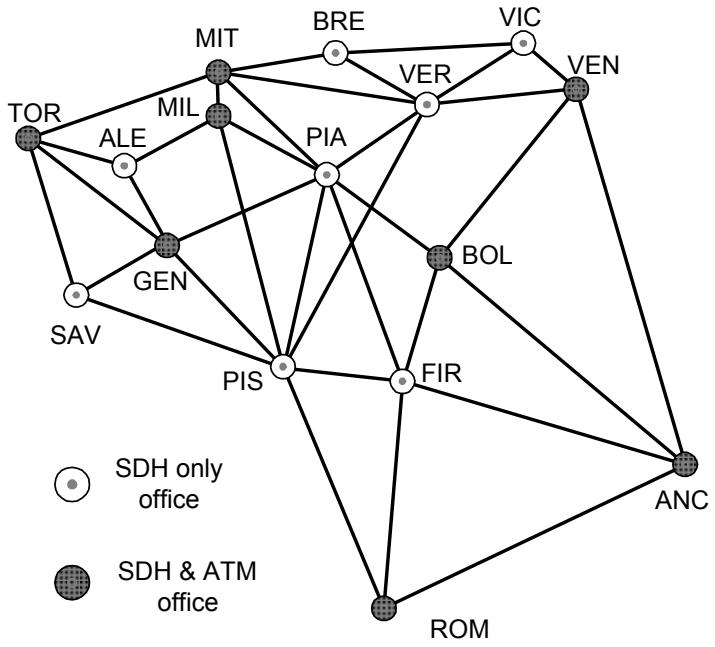


Figure 5.4: Network example: 16-node topology with 8 ATM equipped offices

The chosen single layer recovery mechanisms are SNCP 1+1 for the SDH layer [ITU-T G.841] and a dedicated backup-VP protocol in the ATM layer [ITU-T I.630].

The set of expected failures for the case study are single link failures (cable cuts) and single node failures (either ATM or SDH node failures). In the case of a SDH node failure in a joint ATM and SDH site, the collocated ATM node is also completely disconnected, and therefore all connections traversing that ATM node failed. The spare capacity was planned so that for all expected failures a recovery ratio of 100% is reached.

In addition to the set of expected failures, double link failures as the most probable unexpected failures were simulated. Obviously, for such unexpected failures the

recovery ratio is less than 100%, since the spare resources of some failed connections may be affected by the second failure.

The timing parameters used in the simulations are shown in Table 5.3:

SDH switching matrix reconfiguration time	25 ms
ATM hold-off time	100 ms
ATM routing table reconfiguration time	30 ms
Link propagation delay	250 µs/50 km

Table 5.3: Simulation timing parameters

For the processing of ATM OAM messages within the Equipment Management Function the timing model published by Kawamura in [Kawamura-1995-a] is used.

5.3.1 Uncoordinated Recovery

Figure 5.5 shows the mean recovery ratio in the ATM layer over all link failures for the uncoordinated recovery. The values are obtained by sequentially injecting a link failure in every link, running the simulation and calculating the mean recovery ratio from all results.

The steep increase of the recovery ratio at about 30 ms is due to the server layer recovery (SDH). The objective for a SNCP 1+1 recovery is to complete the recovery process after 50 ms, including the fault detection and propagation. In the simulations all affected ATM client layer connections were restored after about 40 ms by the server layer recovery. This is as expected, since all single link failures are protected by the server layer recovery in a *recovery at lowest layer* approach.

However, after about 50 ms, the mean recovery ratio decreases again, even though it already reached 100%. Before the server layer recovery completes the recovery, defects are detected in the ATM layer. This defect detection immediately triggers the ATM layer recovery. The consequence is, that despite successful SDH layer recovery the ATM layer unnecessarily reroutes the affected connections. This switching of the connections causes additional disruptions of the traffic for a short time.

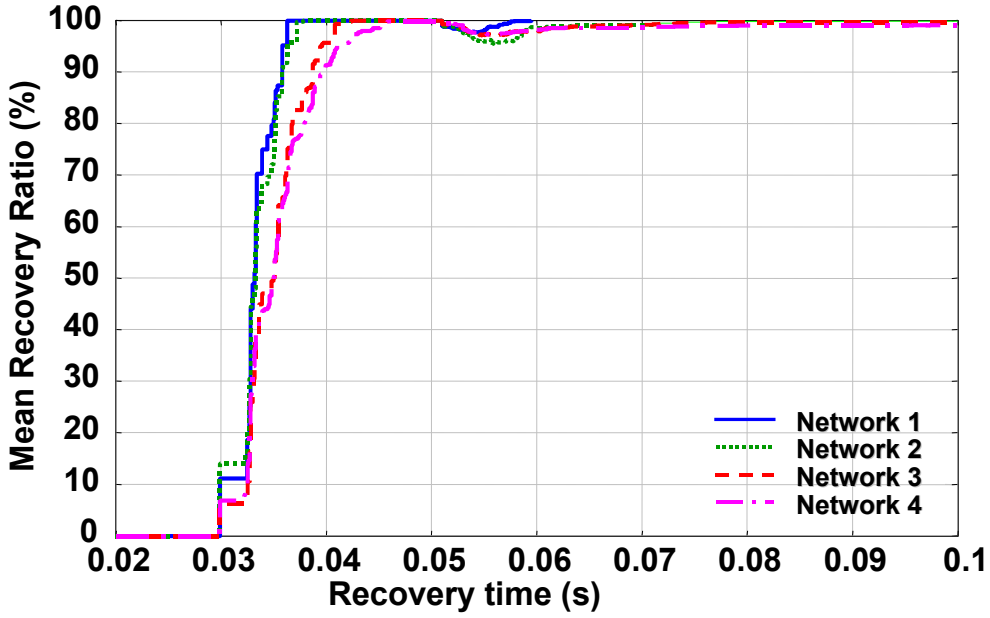


Figure 5.5: Uncoordinated recovery after link failures

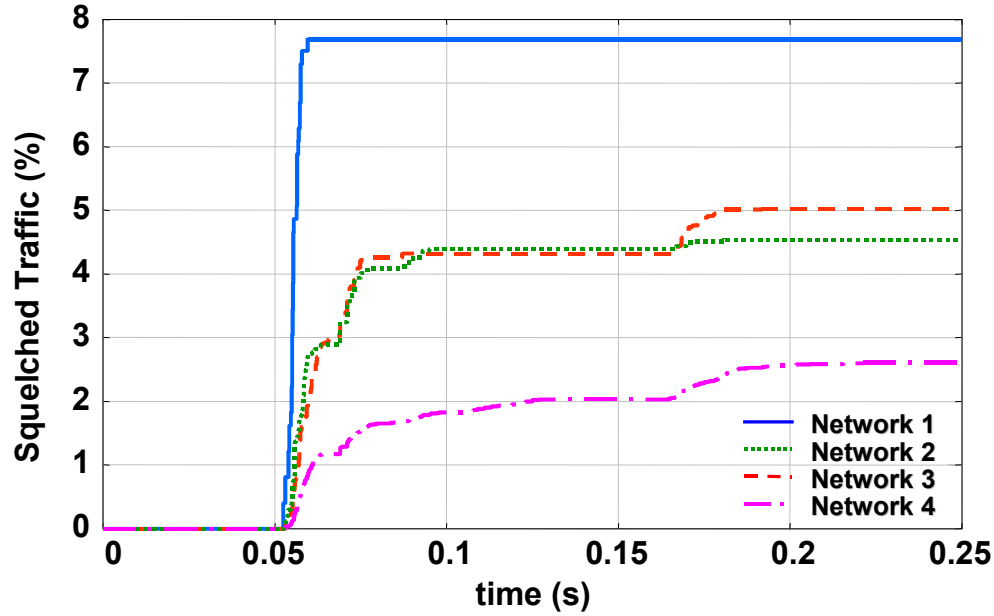


Figure 5.6: Squelched traffic with uncoordinated recovery

As mentioned before, the spare resources in a network can be used for extra traffic, which is unprotected low priority traffic. This traffic type is pre-empted (or squelched) if the spare resources are needed for the recovery of high priority traffic. With uncoordinated recovery, extra traffic will in some cases be unnecessarily squelched. The graph in Figure 5.6 shows the percentage of squelched traffic for single link failures. Since for link failures all ATM traffic should be recovered by the SDH layer recovery, the extra traffic shown in Figure 5.6 is unnecessarily pre-empted. In networks with a

higher ratio of multi-hop connections the ratio of lost extra traffic can increase significantly.

5.3.2 Recovery at Lowest versus Recovery at Highest Layer

The performance of *lowest layer recovery* with *hold-off time* interworking is now compared with a *highest layer recovery* approach in the graphs of Figure 5.7 and Figure 5.8. The graphs show the mean recovery ratios over all SDH node failures. Similar to the previous graphs, the values were obtained by simulating every node failure one after the other and calculating the mean recovery ratio from all results.

The lower graphs in both figures show the recovery ratios in the SDH layer. As can be seen, in the SDH layer the recovery performance is almost the same for both layer assignment strategies. For the recovery at the highest layer approach (bottom right graph), only SDH connections carrying native traffic are taken into consideration, since only these connections are recovered in the SDH layer. Since the SDH demand is the same for the low ATM demand networks (Network 1 and 3) and high ATM demand networks (Network 2 and 4), only two graphs appear in the figures. With recovery at the lowest layer approach (bottom left graph), SDH connections carrying ATM traffic are recovered in the SDH layer also. The main influence on the recovery time in SDH comes from the time to reconfigure the SDH switching matrix, which is assumed to be 25 ms. In addition, a fixed time for the detection of a defect (less than 1 ms) and for the fault propagation to the terminating node of multi-hop connections is needed. With the given network and timing parameters, the SDH recovery was always finished in less than 40 ms. Thus, a requirement to achieve recovery times of below 50 ms SDH equipment must be able to reconfigure the switching matrix in less than 35 ms.

The most noticeable characteristic of the ATM recovery ratio (top part of Figure 5.7) is the two-step curve in case of a lowest layer recovery (upper left). This is because the majority of the failed ATM connections are recovered by the SDH layer recovery. In accordance with the recovery time in the SDH layer, this server layer recovery is completed after approx. 40 ms and reaches an average recovery ratio of about 74% for the smallest network and 95% for the largest network. However, ATM connections transiting the failed SDH site cannot be recovered in the SDH layer, if the serving VC-4 connection is terminated at the failed node. These multi-hop connections are recovered using the dedicated backup VP. The ATM recovery is triggered after the hold-off time (100 ms) elapses and needs 30 ms for the setting of the (pre-configured) translation table. Additionally, a few milliseconds signaling time are needed for each failed connection due to a FIFO queuing in the controller executing the recovery protocol. With recovery at the lowest layer, all ATM connections were recovered after 300 ms for the 16-nodes networks and after 600 ms for the 32-nodes networks.

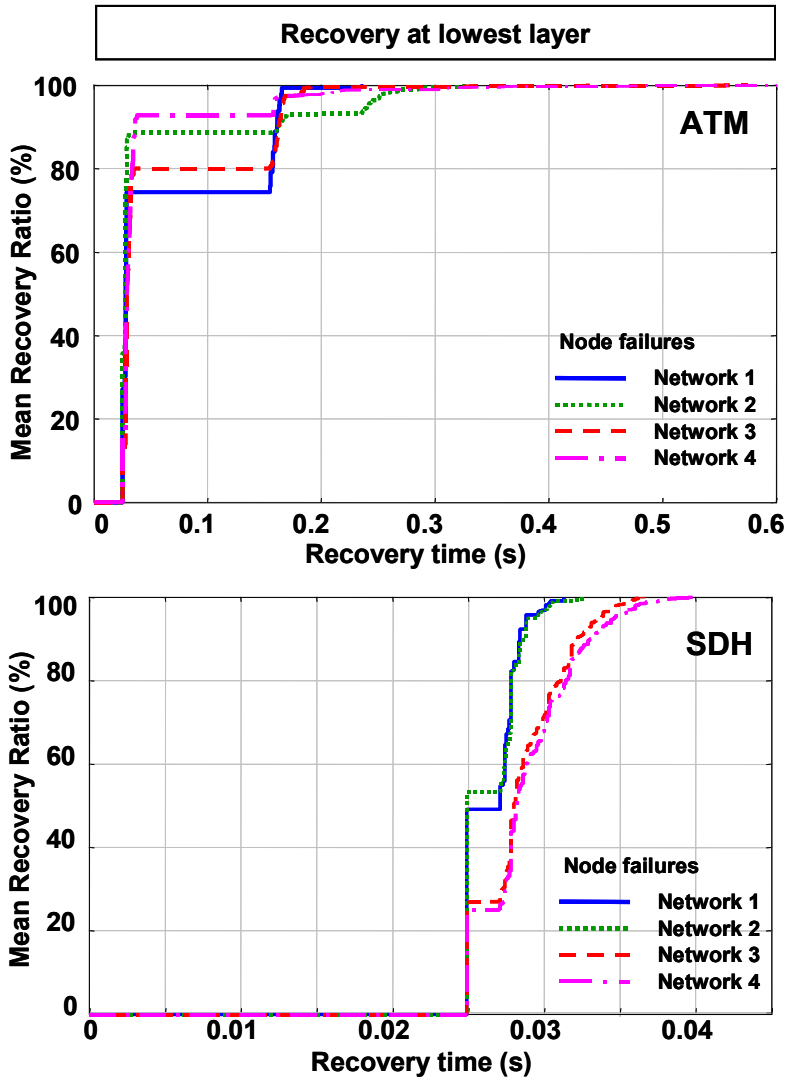


Figure 5.7: Mean Recovery Ratios with recovery at lowest layer for all node failures

In case of recovery at the highest layer (Figure 5.8), all affected ATM connections are recovered in the ATM layer. Even though no hold-off time is needed as interworking mechanism, the recovery is slower in comparison to the first case, since a much higher number of ATM connections has to be individually recovered using the backup VP. It is interesting to see, that the main influence for the ATM recovery ratio is not the number of nodes, but the number of ATM connections that have to be recovered. With the networks having low ATM demand, all affected connections were recovered after 1.3 s, while for the networks having large ATM demand the recovery took up to 3.5 seconds.

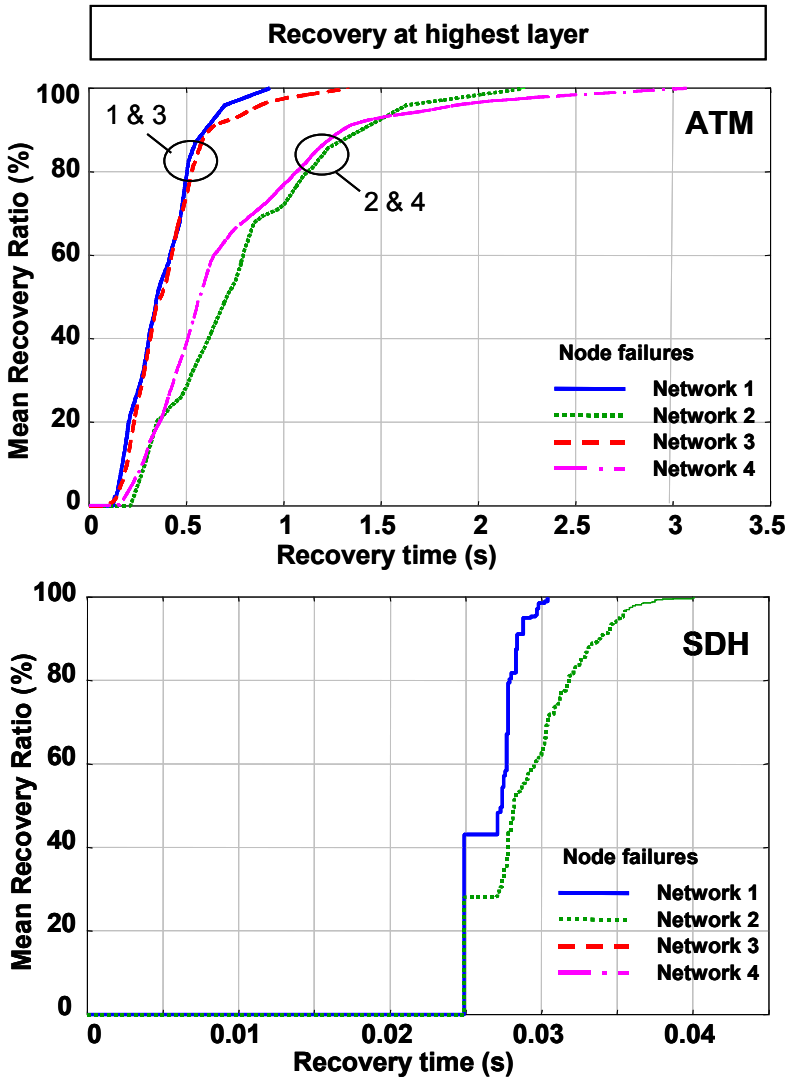


Figure 5.8: Mean Recovery Ratios with recovery at highest layer for all node failures

5.3.3 Hold-Off Time Versus Recovery Token Interworking

The recovery token mechanism introduced in Section 3.9.3.3 aims to reduce the time delay caused by the hold-off time. The ATM recovery will be triggered using the recovery token, if the server layer recovery fails.

In Figure 5.9 the performance of the recovery token is compared to a hold-off time interworking mechanisms. The graphs show the mean recovery ratios for both interworking mechanisms in case of node failures. The hold-off time graphs are identical with the graphs in Figure 5.7. With the recovery token mechanism, as soon as the SDH layer recovery failed, a Recovery Token is emitted to the VP-Layer and propagated to the Node terminating the affected connections. At the terminating ATM-Node, the Recovery Token is put in the FIFO-Queue for the EMF signaling processing. When the Recovery Token is received at the Equipment Management Function, the

Backup-VP process is triggered and the affected connections rerouted. The rerouting is performed by activating a pre-configured routing table in the ports of affected connections.

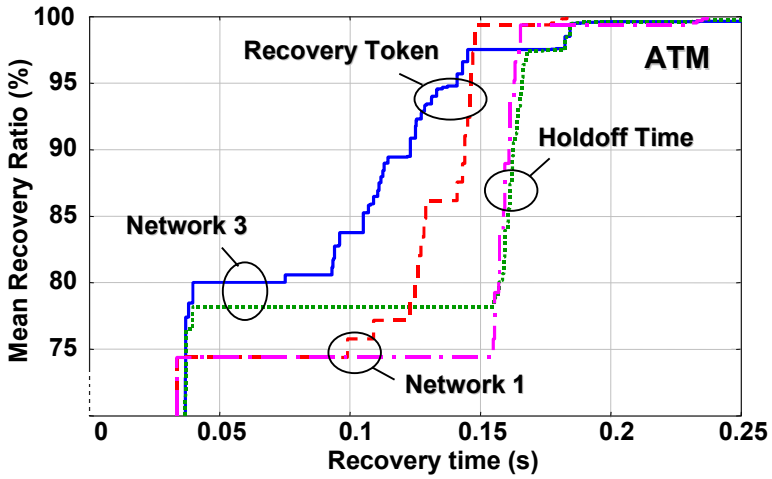


Figure 5.9: Comparison of Recovery Token and Hold-off Time interworking strategies

As can be seen from the graphs, the recovery in the ATM layer starts about 50 ms earlier than the recovery using a hold-off time. The gradient of the recovery ratio is however smaller than in the hold-off time graph. The reason for the slower recovery speed is that also defect detection and alarm signals are reported to the EMF. These signals are still in the FIFO-queue when the Recovery Trigger is received, thus delaying the processing of the Recovery Token. However, the recovery ratio of the Recovery Token is always greater or even to the hold-off time case. In addition, using a priority based FIFO -buffer for the EMF, the processing of the Recovery Token signals could be accelerated. Then the 50 ms gain can be achieved for all recovered connections.

5.3.4 Summary of Results

The integrated simulation environment based on SDL presented in this paper allows a detailed protocol simulation and performance evaluation of multilayer recovery strategies. Using this simulation environment, a case study comparing two multilayer recovery strategies has been carried out.

One outcome of this case study is that *uncoordinated recovery* in two layers leads to unexpected and undesirable effects. A second switching action in the client layer can temporarily disrupt traffic already recovered in the server layer. The unneeded rerouting in the client layer will waste spare resources, which could otherwise be used for the recovery of unexpected failures, e.g. multiple failures. If extra traffic is carried in the client layer spare resources, this traffic may be unnecessarily squelched.

An advantage of the multilayer recovery approach *recovery at highest layer* is that different reliability grades can be assigned at a fine granularity. However, the recovery time performance of the interworking strategy *recovery at the lowest layer* is

significantly higher than the *recovery at the highest layer*, if the SDH recovery is faster than the ATM recovery. This is, because single cable cuts, which are the most probable failures in a network, can be solved at the SDH layer. With the fast recovery of a single SDH connection a very large number of ATM connections will be recovered jointly. Only for node failures additional ATM spare resources protecting multi-hop VP connections will be needed with lowest layer recovery. The additional gain due to the ATM recovery is therefore only significant, if the average VP hop count is significantly larger than one.

The *hold-off time* is a simple and robust interworking mechanism to coordinate the recovery mechanisms in multiple layers. To reduce the delay caused by the hold-off time, a simple *recovery token* mechanism was introduced. It was shown, that the recovery time can be improved by about 50 ms in case of a hold-off time of 100 ms.

The simulation results and the evaluations of multilayer recovery strategies were used to define guidelines for the optimized selection of a recovery strategy for different multi-layer network architectures [*Demeester-1999*].

5.4 Differentiated Multilayer Resilience (DMR)

In the previous chapters, an overview of recovery mechanisms and options for single layer recovery and for multilayer recovery strategies was given. With the RD-QoS architecture, a differentiated level of resilience can be flexibly provided for IP services using MPLS recovery.

In this section, the concept of differentiated resilience is extended to multilayer recovery approaches.

The motivation for the differentiated resilience concepts supplements the provisioning of QoS in IP networks. Internet service providers must offer differentiated services to be economically competitive. The provisioning of best-effort services only proved to be a weak business model.

To provide a premium level of resilience to all services can be compared to providing a premium level of QoS to all services. For QoS this is commonly considered to be too expensive. For resilience this is the current practice. Differentiated Resilience is a feasible approach to improve the resource efficiency of MPLS recovery. This differentiated resilience approach is extended to multilayer resilience scenarios [*Autenrieth-2002-c*].

5.4.1 Multilayer Resilience Classes

For the provisioning of differentiated resilience in a multilayer network appropriate resilience classes must be defined taking multilayer coordination into consideration.

In [Gerstel-2000-a] a set of five services classes for traffic in the optical network is defined, which is referred to as multilayer resilience classes (MRC). The five classes are now briefly characterized:

- MRC1: service must be protected by the server layer (e.g., unprotected client layers or native traffic)
- MRC2: service must not be protected (e.g., traffic protected in client layers)
- MRC3: service is indifferent to protection (e.g., IP traffic, since resilience mechanisms would not interfere)
- MRC4: service has best-effort protection in the server layer
- MRC5: service has low priority (using spare capacity under normal conditions and may be preempted by resilience mechanisms)

In a static optical network, a network operator can offer these five multilayer resilience classes to its customers. In a dynamic optical network (which is considered in this work), the multilayer resilience classes must be included in the control signaling between the layers. This is either done via the User Network Interface in the Overlay Model, or the common control plane in the Peer Model.

5.4.2 DMR Approaches

5.4.2.1 Differentiated Highest Layer Recovery

In a *differentiated highest layer recovery approach*, the client layer services are protected in the client layer depending on their resilience class. In the service layer, all client traffic uses MRC2 (must not be protected in the server layer).

For native optical traffic or traffic from clients without protection capabilities the multilayer resilience classes MRC1 and MRC3-5 are used depending on the resilience class of the service.

5.4.2.2 Differentiated Lowest Layer Recovery

In a *differentiated lowest layer recovery* approach, client layer services are using the multilayer resilience classes MRC1 and MRC3-5 depending on their resilience class. In addition, to be able to cope with failure scenarios that cannot be recovered by the server layer, a client layer recovery may be triggered after a hold-off time or by a recovery token.

Spare resources in the IP/MPLS layer which will be used to protect against client failures and node isolation scenarios can be carried as MRC5 using the common pool concept proposed in [Gryseels-1998-a].

5.5 Summary

Despite current layer convergence efforts for IP over optical networks the transport networks will always consist of multiple layers. Moreover, server layer networks will transport more than one client layer technology.

Since resilience mechanisms are an integral part of all backbone transport networks, the issue of multilayer resilience was recognized in recent year and some research effort was put on this topic [*Demeester-1997*, Meijen-1999, P918-D5].

The PANEL project played a key role in the development of multilayer recovery strategies. Part of the work of this thesis was done in that project. The results of the PANEL project influenced the standardization efforts of ANSI-T1 [ANSI TR68]. The ITU-T Study Group 15 works in the study period 2001 to 2004 on the 'Specification of survivability capabilities and development of a strategy for multi-layer survivability interactions' [ITU-T SG15].

The integrated simulation and evaluation environment presented in this chapter allowed the detailed analysis of the proposed interworking strategies and helped to design the PANEL multilayer recovery guidelines published in [*Demeester-1999*].

Resilience differentiation helps reduce the network costs and allows network operators to offer resilience as an additional service quality. Especially for tightly integrated control architectures as it is envisaged for IP over optical networks the differentiated multilayer resilience approach proposed in this chapter offers a high flexibility and allows the mapping of resilience differentiated services to server layer recovery mechanisms.

6 SUMMARY AND CONCLUSION

6.1 Summary of Key Contributions of this Thesis

Our society depends on the correct and faultless operation on the global communication infrastructure. This becomes especially transparent with natural disasters and terrorist attacks in the recent history. A transport network must be able to cope with failures like cable cuts due to road works and node breaks due to power loss or fire accidents.

The growing complexity of the network technologies increases the probability for failures. Due to the large capacity of network elements a huge amount of traffic can be lost even in case of single failures. The sensitivity of services to even short outages requires a very fast recovery of network failures. The complexity of a recovery strategy, its resource efficiency and mean time to recover failed traffic must be weighted against each other.

So far, resilience was considered as an attribute of the transport network, and consequently only two levels of resilience were offered by network operators: either 100% protected traffic or unprotected traffic. Moreover, the traffic granularity for the selection of these protection levels is very coarse. The smallest granularity an optical transport network (OTN) is offering to its client layers is an optical channel with 2.5 Gb/s.

However, the current trend is clearly towards service-driven transport architectures. As Quality-of-Service (QoS) architectures are firmly established, new concepts to offer differentiated levels of resilience to IP services in multilayer transport networks must be developed. In the following the key contribution of this thesis to this research topic are summarized.

Integrated Multilayer Resilience Framework

The thesis contains a brief introduction to the network architectures and technologies considered in this thesis. The aim of this introduction is to define the functional models and terminology.

Each considered network technology includes mechanisms to provide resilience. For historic reasons, the terminology used for the description of the resilience concepts is often not consistently. In this thesis an integrated multilayer resilience framework (IMRF) is developed to define a common view and consistent terminology for resilience concepts. The multilayer resilience framework includes the definition of resilience requirements and performance parameters, and the description failure detection and notification methods. The core of the integrated multilayer resilience framework is a generic definition of recovery mechanisms focusing on protection switching and distributed restoration mechanisms. The generic, technology independent definition of recovery mechanisms is followed by an overview of the state of the art of recovery mechanisms at the considered network architecture.

The Multilayer Recovery Framework defined by the ACTS project PANEL is included in the IMRF as well as a framework to handle multiple failures in a network.

The performance evaluation methods for the resource efficiency and recovery time analysis used in this work are also defined in the IMRF.

Resilience-Differentiated Quality of Service Architecture

A key contribution of this thesis is the definition of a novel architecture to provide differentiated resilience in IP/MPLS networks. The setup of resilient services is achieved by extending existing QoS mechanisms to include the signaling of resilience requirements of IP services. In analogy to the differentiated services model, the resilience requirements of IP services are mapped to resilience classes. In this thesis four resilience classes are proposed and specified. The concept and key building blocks of the RD-QoS architecture are described in detail and their interworking with QoS architectures are discussed. Possible mappings to MPLS recovery mechanisms are proposed. An important part of the RD-QoS architecture is the definition of a traffic engineering process to manage the link resources used by the different resilience classes. The resource efficiency and the recovery time of the RD-QoS architecture are evaluated in a case study using the traffic engineering process and the recovery-timing model.

Multilayer Recovery Evaluation

The multilayer resilience strategies proposed by the PANEL project are evaluated in terms of their recovery time. For this evaluation a detailed simulation and evaluation SELANE (Simulation Environment for Layered Networks) environment is developed. The simulation environment is designed using the Specification and Description Language SDL with a detailed modeling of the network elements following the functional model of transport network architectures introduced in Chapter 2. With this simulation environment a detailed protocol simulation of the interworking strategies is possible as well as a performance evaluation of the investigated recovery mechanisms. The results of a case study to evaluate the proposed multilayer interworking strategies are presented in this thesis.

Finally, the extension of the multilayer recovery strategy to support the RD-QoS architecture and differentiated resilience concepts in general is discussed. This leads to a proposal for a Differentiated Multilayer Resilience concept.

Outlook

Differentiated resilience is a very promising concept for future multilayer transport networks. An detailed proposal with an intensive discussion and evaluation to offer differentiated resilience is presented in this thesis.

Based on the results and recommendations of this thesis, some additional topics can be studied. The resilience classes proposed in this thesis could be subdivided to allow a specification of additional resilience requirements. Especially the ability to allow a reduced QoS on the backup paths is a promising concept. In the traffic-engineering model this can be included by reserving a reduced service bandwidth on the backup link resources compared to the active links. Another issue for further studies is the multi-domain interworking of the RD-QoS architecture.

Regarding multilayer resilience, the study and development of integrated multilayer recovery mechanisms using the ASON or GMPLS control plane seems promising. To offer differentiated multilayer resilience, the resilience attributes of services must be included in the inter-layer control signaling.

With the integrated control plane in IP over optical networks, a better coordination between multiple layers is possible. Recovery actions can be triggered in different layers depending on the failure scenario, and strategies for fully integrated recovery mechanism can be conceived.

6.2 Conclusion

Differentiated Resilience

The interworking of MPLS-based core networks with QoS access network is a very promising architecture to provide differentiated resilience. A benefit of service-based differentiated resilience is that service providers can charge their customers for the level of resilience they receive. This service differentiation allows a capacity efficient network design. In the case studies it could be shown, that with only 6,6% higher total resource usage 10% of high resilience traffic and 20% of medium resilience traffic can be offered. The high resilience traffic uses protection switching mechanisms with a recovery time requirement of 100ms, and medium resilience traffic uses restoration mechanisms with several hundred milliseconds up to one second recovery time. With a recovery time analysis it could be show, that the recovery time requirements are met as long as the number of flows between a node pair is below about 1280 flows for global restoration mechanism. Since in MPLS backbones LSPs are setup for flow aggregates, this number will probably not be exceeded.

Multilayer Resilience

Resilience mechanisms working in multiple layers at the same time can improve resilience performance, since a larger failure scope can be covered. Lower layer resilience is well suited for the recovery of physical failures, while resilience in higher layers can offer higher protection selectivity and finer recovery granularity. Especially the recovery of multiple failures can be improved by using resilience mechanisms in multiple layers.

However, resilience mechanisms in the different layers must be coordinated to prevent unpredictable and adverse behavior of the recovery mechanisms. The careful planning and control of recovery mechanisms extending over multiple layers is an important issue for multilayer transport networks. The multilayer recovery interworking mechanisms proposed the ACTS project PANEL were evaluated using the simulation environment presented in this thesis. The results and guidelines of the PANEL project for multilayer recovery influenced the work of standardization bodies like ANSI-T1 and ITU-T.

INDEX OF FIGURES

INDEX OF TABLES

BIBLIOGRAPHY

The references in the first section of the bibliography, which are printed in Italics, are authored and co-authored publications of the author of this thesis. These are pre-publications ("Vorveröffentlichungen") of the thesis according to §4, section (5) of the doctorate graduation guidelines ("Promotionsrichtlinien") of the Technische Universität München.

Authored and Co-Authored Publications

[*Autenrieth-1998-a*] Autenrieth, Achim, Brianza, Carlo, Clemente, Roberto, Demeester, Piet, Gryseels, Michael, Harada, Yohnosuke, Jajszczyk, Andrej, Janukowicz, D., Kalbe, Gustav, Ohta, S., Ravera, Mauro, Rhissa, A. G., Signorelli, Giulio, Van Doorselaere, Kristof, "Resilience in a multi-layer network", CSELT Technical Reports, vol. 26, no. 6, pp. 869-82, 1998

[*Autenrieth-1998-b*] Autenrieth, A. , Van Doorselaere, K., Iselt, A., Demeester, P., Struyve, K., Vandendriessche, L., "Simulation and Evaluation of Multi-layer Broadband Networks", First International Workshop on the Design of Reliable Communication Networks (DRCN'98), Brugge, Belgium, 1998

[*Autenrieth-1998-c*] Autenrieth, Achim, Van Doorselaere, Kristof, Demeester, Piet, "Evaluation of multi-layer recovery interworking strategies", Eunice'98 - Open European Summer School, Munich, Germany, 1998

[*Autenrieth-2000*] Autenrieth, Achim, Kirstadter, Andreas, "Fault-Tolerance and Resilience Issues in IP-Based Networks", Second International Workshop on the Design of Reliable Communication Networks (DRCN 2000), Munich, Germany, 2000

[*Autenrieth-2001-a*] Autenrieth, Achim, Kirstädter, Andreas, "Resilience-Differentiated QoS – Extensions to RSVP and DiffServ to Signal End-to-End IP Resilience Requirements.", Third International Workshop on the Design of Reliable Communication Networks (DRCN2001), Budapest, Hungary, 2001

[*Autenrieth-2001-b*] Autenrieth, Achim, Kirstädter, Andreas, "Components of MPLS Recovery Supporting Differentiated Resilience Requirements", Seventh EUNICE Open European Summer School in association with IFIP Workshop on IP and ATM Traffic Management WATM'2001, Paris, France, 2001

[*Autenrieth-2002-a*] Autenrieth, Achim, Kirstädter, Andreas, "Engineering End-to-End IP Resilience Using Resilience-Differentiated QoS", IEEE Communications Magazine, vol. 40, no. 01, pp. 50-57, 2002

[*Autenrieth-2002-b*] Autenrieth, Achim, Kirstädter, Andreas, "RD-QoS - The Integrated Provisioning of Resilience and QoS in MPLS-Based Networks", IEEE International Conference on Communications (ICC 2002), New York, USA, 2002

[*Autenrieth-2002-c*] Autenrieth, Achim, "Differentiated Multilayer Resilience in IP over Optical Networks", SSGRR 2002, L'Aquila, Italy, July 29 – August 4, 2002

[*Demeester-1997*] Demeester, Piet, Gryseels, Michael, Struyve, Kris, Van Doorselaere, Kristof, Autenrieth, Achim, Brianza, Carlo, Signorelli, Giulio, Clemente, Roberto, Ravera, Mauro, Jajszczyk, Andrej, Roszkiewicz, M., Kalbe, Gustav, Harada, Yohnosuke, Yada, T., Rhissa, A, "PANEL - protection across network layers", NOC '97, Antwerp, Belgium, 1997

[*Demeester-1998-a*] Demeester, P., Gryseels, M., Van Doorselaere, K., Autenrieth, A., Brianza, C., Signorelli, G., Clemente, R., Ravera, M., Jajszcyk, A., Janukowicz, D., Kable, G., Harada, Y., Otha, S., Rhissa, A.G., "Resilience in a multi-layer network", First International Workshop on the Design of Reliable Communication Networks (DRCN'98), Brugge, Belgium, 1998

[*Demeester-1998-b*] Demeester, Piet, Gryseels, Michael, van Doorselaere, Kristof, Autenrieth, Achim, Brianza, Carlo, Signorelli, Giulio, Clemente, Roberto, Ravera, Mauro, Jajszcyk, Andrej, Geyssens, A., Harada, Y., "Network resilience strategies in SDH/WDM multilayer networks", 24th European Conference on Optical Communication (ECOC '98), Madrid, Spain, 1998

[*Demeester-1999*] Demeester, Piet, Gryseels, Michael, Autenrieth, Achim, Brianza, Carlo, Castagna, Laura, Signorelli, Giulio, Clemente, Roberto, Ravera, Mauro, Jajszczyk, A., Janukowicz, D., Van Doorselaere, Kristof, Harada, Yohnosuke, "Resilience in multilayer networks", IEEE Communications Magazine, vol. 37, no. 8, pp. 70-76, 1999

[*Iselt-1997*] Iselt, A., Autenrieth, A., "An SDL-based platform for the simulation of communication networks using dynamic block instantiations", SDL '97, Evry, France, 1997

[*Kellerer-1998*] Kellerer, W., Autenrieth, A., Iselt, A., "SDL based protocol engineering and visualization for education: ISDN Q.931 case study", FORTE/PSTV'98 - Tutorials/ECASP, International Conference on Formal Description Techniques for Distributed Systems and Communication Protocols (FORTE XI) and Protocol Specification, Testing and Verification (PSTV XVIII), Paris, France, 1998

[*Kellerer-2000*] Kellerer, Wolfgang, Autenrieth, Achim, Iselt, Andreas, "Experiences with evaluation of SDL-based protocol engineering in education", Computer Science Education, vol. 10, no. 3, pp. 225-41, 2000

[*PANEL-D4*] PANEL Deliverable D4, "Software Testbed Description", August 1997

[*PANEL-D5*] PANEL Deliverable D5, "Software Testbed Results", August 1998

[*PANEL-D6*] PANEL Deliverable D6, "Demo and Results Description", March 1999

[*PANEL-FR*] PANEL Final Report, March 2000

[*Schupke-2001-a*] Schupke, D.A., Autenrieth, Achim, Fischer, Thomas, "Survivability of Multiple Fiber Duct Failures", Third International Workshop on the Design of Reliable Communication Networks (DRCN2001), Budapest, Hungary, 2001

[*Schupke-2001-b*] Schupke, D. A., Fischer, T., Autenrieth, A., Feng, H., Kravcenko, V., Patzak, E., Saniter, J., Jäger, M., Westphal, F.-J., Fitzek, F. H. P., Woesner, H., Dolzer, K., Finsterle, L., Gauger, C., "TransiNet - Innovative Transport Networks for the Broadband Internet", ITG-Fachtagung Photonische Netze, Dresden, Germany, 2001

[*Schupke-2002*] Schupke, D.A., Gruber, C., Autenrieth, A., "Optimal Configuration of p-Cycles in WDM Networks", IEEE International Conference on Communications (ICC 2002), New York, USA, 2002

[*Transinet-D1*] TransiNet AG-a Deliverable 1, "Network Requirements based on Analysis of Services", July 2001

Other Publications

[Anderson-1994] Anderson, J., Doshi, B.T., Dravida, S., Harshavardhana, P, "Fast Restoration of ATM Networks ", IEEE Journal on Selected Areas in Communication (JSAC), vol. 12, no. 1, pp. 128 – 138, January 1994

[ANSI TR24] ANSI T1, "A Technical Report on Network Survivability Performance", Committee T1 Technical Report No. 24, 1993

[ANSI TR55] ANSI T1, "Reliability and Survivability Aspects of the Interactions Between the Internet and the Public Telecommunications Network", Committee T1 Technical Report No. 55, 1998

[ANSI TR68] ANSI T1, "A Technical Report on Enhanced Network Survivability Performance", Committee T1 Technical Report No. 68, 2001

[Aukia-2000] Aukia, P., al., et, "RATES: A server for MPLS traffic engineering", IEEE Network Magazine, vol. 14, no. 2, pp. 34-41, 2000

[Awduche-1999] Awduche, D., "MPLS and traffic engineering in IP networks", IEEE Communications Magazine, vol. 37, no. 12, pp. 42-47, 1999

[Awduche-2001] Awduche, D., Rekhter, Y., "Multiprotocol lambda switching: combining MPLS traffic engineering control with optical crossconnects", IEEE Communications Magazine, vol. 39, no. 3, pp. 111-16, 2001

[Beauchamps-2001] Simon Gouyou Beauchamps, "Implementation and Simulation of the RSVP architecture with RD-QoS support", Master Thesis, Institute of Communication Networks, Munich University of Technology, 2001

[Bhandari-1999] R. Bhandari, "Survivable Networks – Algorithms for Diverse Routing", Kluwer Academic Publishers, Boston/Dordrecht/London, 1999.

[Caenegem-1997-a] B. Van Caenegem, N. Wauters, P. Demeester, "Two Techniques for Spare Capacity Assignment in Mesh Survivable Networks", Proc. of 5th International Conference on Telecommunication Systems, Modeling and Analysis, pp 98-102, Nashville, March 20-23, 1997

[Caenegem-1997-b] B. Van Caenegem, N. Wauters, P. Demeester, "Spare Capacity Assignment for Different Restoration Strategies in Mesh Survivable Networks", Proc. of IEEE International Conference on Communications 1997 (ICC'97), pp. 288-292, Montreal, Canada, June 8-12, 1997

[Chow-1993] C. E. Chow, J. Bicknell, S. McCaughey, and S. Syed, “A Fast Distributed Network Restoration Algorithm”, IEEE IPCCC ’93, Tempe, Arizona, USA, pp. 261-267, March 1993.

[Cisco-1999] Cisco Systems, "Always-On Availability for Multiservice Carrier Networks", Public paper, http://www.cisco.com/warp/public/cc/so/neso/wan/aoav_wp.htm

[Coffman-2002] Coffman, K. G., Odlyzko, A. M., "Growth of the Internet", in Optical Fiber Telecommunications IV B: Systems and Impairments, I. P. Kaminow and T. Li, eds., Academic Press, 2002, pp. 17-56.

[COST239] Batchelor, P. et al., "Ultra high capacity optical transmission networks", Final report of action COST 239, Faculty of Electrical Engineering and Computing, University of Zagreb, ISBN 953-184-013-X, 1999

[Draft-Awduche] Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., Xiao, X., "A Framework for Internet Traffic Engineering" draft-ietf-tewg-framework-05.txt, Internet Draft, June 2001.

[Draft-Harrison] N. Harrison, et al., "OAM Functionality for MPLS Networks", Work in Progress, Internet Draft, <draft-harrison-mpls-oam-00.txt>, February 2001

[Draft-Haskin] D. Haskin, R. Krishnan, "A Method for Setting an Alternative Label Switched Paths to Handle Fast Reroute", Work in Progress, Internet Draft, <draft-haskin-mpls-fast-reroute-05.txt>, November 2000

[Draft-Lang] Lang, Jonathan P., Mitra, Krishna, Drake, John, Kompella, Kireeti, Rekhter, Yakov, Berger, Lou, Saha, Debanjan, Basak, Debashis, Sandick, Hal, Zinin, Alex, Rajagopalan, Bala, " Link Management Protocol (LMP)", Internet Draft, Work in Progress, March 2002 <draft-ietf-ccamp-lmp-03.txt>

[Draft-Owens] Owens, Ken, Sharma, Vishal, Oommen, Mathew, Hellstrand, Fiffi, "Network Survivability Considerations for Traffic Engineered IP Networks", Work in Progress, Internet Draft, <draft-owens-te-network-survivability-03.txt>, May 2002

[Draft-Willis] P. Willis, et al.,"Requirements for OAM in MPLS Networks", Work in Progress, Internet Draft, < draft-harrison-mpls-oam-req-01.txt>, November 2001

[Draft-Sharma] Sharma, V., Hellstrand, F. (Editors) "A Framework for MPLS-based Recovery," Work in Progress, Internet Draft, draft-ietf-mpls-recovery- frmwrk-04.txt, May 2002

[Draft-Mannie] Mannie, E., Papadimitriou D., et al., "Generalized Multiprotocol Label Switching Architecture", Internet Draft, Work in progress, draft-ietf-ccamp-gmpls-architecture-03.txt, August 2002.

[Eberspächer-1998] Eberspächer, Jörg, "Systemtechnische Grundlagen", in: "Handbuch für die Telekommunikation", V. Jung, H.-J. Warnecke (Eds), Springer Verlag Berlin, 1998

[Eberspächer-2000] Jörg Eberspächer, Piet Demeester (Eds.), "Proceedings of the Second International Workshop on the Design of Reliable Communication Networks (DRCN 2000)", Herbert Utz Verlag, Munich, ISBN 3-89675-928-0, 2000

[Edmaier-1996] Edmaier, Bernhard, "Pfad-Ersatzschalteverfahren mit verteilter Steuerung für ATM-Netze", PhD Thesis, Institute of Communication Network, Munich University of Technology, Munich, 1996

[ETSI 300 417] ETSI EN 300 417, "Transmission and Multiplexing (TM), Generic requirements of transport functionality of equipment", October 2001

[FASHION-D1] Eurescom Project P1012 – FASHION - Deliverable D1, "Switched Optical Networks: Architecture and Functionality", Jacques Robadey (Ed.), March 2002

[Fumagalli-2001] Fumagalli, A. and Tacca, M., "Optimal Design of Optical Ring Networks with Differentiated Reliability (DiR)", International Workshop on QoS in Multiservice IP Networks, Rome, Italy, January 2001

[Gerstel-2000-a] Gerstel, O., Ramaswami, R., "Optical Layer Survivability: A Services Perspective", IEEE Communications Magazine, vol. 38 (2000), no.3, pp.104-113.

[Gerstel-2000-b] Gerstel, O., Ramaswami, R., "Optical Layer Survivability - An Implementation Perspective", IEEE Journal on Selected Areas in Communications, vol.18, no.10, pp.1885-1899, October 2000.

[Ghani-2000] Ghani, N., Dixit, S., Wang, T. S., "On IP-over-WDM integration", IEEE Communications Magazine, vol. 38, no. 3, pp. 72-84, 2000

[Grover-1987] Grover, W.D., "The Selfhealing Network, A Fast Distributed Restoration Technique for Networks Using Digital Cross-connect Machines", Proceedings of IEEE GLOBECOM '87, Tokyo, Nov. 1987. pp.1090-1095.

[Grover-1991] Grover, W.D., Bilodeau, T.D., Venables, B.D., "Near Optimal Spare Capacity Planning in a Mesh Restorable Network," Proceedings of IEEE Globecom '91, Phoenix, Arizona, Vol. III, pp. 2007-2012, December 1991.

[Grover-1998] Grover, W.D., Stamatelakis, D., "Cycle-oriented distributed pre-configuration: ring-like speed with mesh-like capacity for self-planning network restoration," in Proc. IEEE International Conf. Commun. (ICC '98), Atlanta, June 8-11, pp. 537-543.

[Grover-2000] Grover, W.D., Stamatelakis, D., “Bridging the ring-mesh dichotomy with p-cycles”, (Invited Paper) in Proc. IEEE / VDE Design of Reliable Communication Networks (DRCN 2000), Munich, Germany, April 2000, pp. 92-104.

[Gryseels-1998-a] Gryseels, Michael, Stuyve, K, Pickavet, M., Demeester, P., "Common pool survivability for meshed SDH-based ATM networks", Int. Symp. Broadband European Networks (SYBEN'98), Zurich, Switzerland, May 1998, pp. 267-278

[Gryseels-1998-b] Gryseels, M., Clemente, R. and Demeester, P., “Common pool survivability for ATM over SDH ring networks”, Proc. 1st Intl. Workshop on the Design of Reliable Communication Networks, Brugge, 1998.

[Herzberg-1994] Herzberg, M. and Bye, S. J., "An Optimal Spare-Capacity Assignment Model for Survivable Networks with Hop Limits", Proc IEEE GLOBECOM '94, 1994

[Herzberg-1995] Herzberg, M., Bye, S. J., and Utano, A., "The Hop- Limit Approach for Spare-Capacity Assignment in Survivable Networks," IEEE/ACM Transactions on Networking, vol. 3, no. 6, December 1995, pp. 775-784.

[Herzberg-1997] Herzberg, M., Wells, D., and Herschtal, A., "Optimal Resource Allocation for Path Restoration in Mesh-Type Self-Healing Network," Proceedings of the 15th International TeletraÆc Congress - ITC 15, Washington, DC, 22-27 June 1997, pp. 351-360, Elsevier Science B. V., 1997.

[Iraschko-1996] R.R. Iraschko, M.H. MacGregor, W.D. Grover, "Optimal Capacity Placement for Path Restoration in Mesh Survivable Networks," IEEE International Communications Conference (ICC'96), Dallas, June 23 - 27, 1996. vol.3, pp. 1568-1574.

[Iselt-1999] Iselt, Andreas, "Ausfallsicherheit und unterbrechungsfreies Ersatzschalten in Kommunikationsnetzen mit Redundanzdomänen", PhD Thesis, Institute of Communication Network, Munich University of Technology, Munich, 1999

[ITU-T E.800] ITU-T Recommendation E.800, "Terms and definitions related to quality of service and network performance including Dependability", August 1994

[ITU-T G.707] ITU-T Recommendation G.707/Y.1322, "Network node interface for the synchronous digital hierarchy (SDH)", October 2000

[ITU-T G.709] ITU-T Recommendation G.709/Y.1331, "Interfaces for the optical transport network (OTN)", February 2001

[ITU-T G.803] ITU-T Recommendation G.803, "Architecture of transport networks based on the synchronous digital hierarchy (SDH)", March 2000

[ITU-T G.805] ITU-T Recommendation G.805, "Generic functional architecture of transport networks", March 2000

[ITU-T G.807] ITU-T Recommendation G.807/Y.1302, "Requirements for automatic switched transport networks (ASTN)", July 2001

[ITU-T G.808] ITU-T Recommendation G.808/Y.1304, "Architecture for the Automatically Switched Optical Network (ASON)", November 2001

[ITU-T G.841] ITU-T G.841, "Types and Characteristics of SDH Network Protection Architectures", Geneva, Switzerland, 1995

[ITU-T G.872] ITU-T Recommendation G.805, "Architecture of optical transport networks", February 1999

[ITU-T I.311] ITU-T Recommendation I.311, "B-ISDN general network aspects", August 1996

[ITU-T I.321] ITU-T Recommendation I.321, "B-ISDN protocol reference model and its application", April 1991

[ITU-T I.326] ITU-T Recommendation I.326, "Functional architecture of tansport networks based on ATM", November 1995

[ITU-T I.361] ITU-T Recommendation I.361, " B-ISDN ATM layer specification", February 1999

[ITU-T I.380] ITU-T Recommendation I.380, " Internet protocol data communication service – IP packet transfer and availability performance parameters", February 1999

[ITU-T I.480] ITU-T Recommendation I.480, " 1+1 protection switching for cell-based physical layer", March 2000

[ITU-T I.610] ITU-T Recommendation I.610, "B-SIDN operation and maintenance principles and functions", November 1995

[ITU-T I.630] ITU-T Recommendation I.630, "ATM protection switching", February 1999

[ITU-T I.731] ITU-T Recommendation I.731, " Types and general characteristics of ATM equipment", October 2000

[ITU-T I.732] ITU-T Recommendation I.732, " Functional characteristics of ATM equipment", October 2000

[ITU-T M.495] ITU-T Recommendation M.495, "Transmission restoration and transmission route diversity - terminology and general principles", November 1988

[ITU-T SG15] ITU-T Study Group 15 (Study Period 2001 - 2004), "List of Questions under Study - Question 9/15 - Transport equipment and network protection/restoration", http://www.itu.int/ITU-T/studygroups/com15/sg15-q9.html (last visited: 2002-09-23)

[ITU-T X.641] ITU-T Recommendation X.641, " Information technology - Quality of service: Framework", December 1997

[ITU-T Z.100] ITU-T Recommendation Z.100, "Specification and description language (SDL)", November 1999

[Johnson-1996] Johnson, David, "Survivability Strategies for Broadband Networks", Proc. of IEEE Globecom '96 conference, London, 1996

[Kawamura-1994] Kawamura, R., Sato, K., Tokizawa, I., " Self-healing ATM Networks Based on Virtual Path Concept", IEEE Journal on Selected Areas in Communication (JSAC), vol. 12, no. 1, pp. 120 – 127, January 1994

[Kawamura-1995-a] Kawamura, Ryutaro, Hadama, Hisaya, Tokizawa, Ikuo, "Implementation of Self-healing Function in ATM Networks Based on Virtual Path Concept", Journal of Network and System Management, September 1995

[Kawamura-1995-b] Kawamura, Ryutaro, Tokizawa, Ikuo, "Self-healing Virtual Path Architecture in ATM Networks", IEEE Communications Magazine, vol. 33, no. 9, pp. 72-79, September 1995

[Kawamura-1998] Kawamura, Ryutaro, "Architectures for ATM Network Survivability, IEEE Communication Surveys and Tutorials, Fourth Quarter 1998, Vol. 1, No. 1 (http://www.comsoc.org/pubs/surveys/4q98issue/pdf/Kawamura.pdf)

[Knudsen-2001] Knudsen, S. N., Peckham, D. W., Pedersen, M. Ø., Zhu, B., Judy, A. F., Nelson, L. E., "New dispersion-slope managed fiber pairs for ultra long haul transmission systems", National Fiber Optic Engineers Conference, 2001

[LEDA] "LEDA, A Platform for Combinatorial and Geometric Computing", Kurt Mehlhorn, Stefan Näher, Cambridge University Press, 1999

[Manchester-1999] Manchester, J., Bonenfant, P., Newton, C., "The evolution of transport network survivability", IEEE Communications Magazine, vol. 37, no. 8, pp. 44-51, 1999

[McGuire-1998] McGuire, Alan, Bonenfant, Paul, "Standards: The Blueprints for Optical Networking", IEEE Communications Magazine, vol. 36, no. 2, pp. 68-70, February 1998

[Meijen-1999] Meijen, Johan, Varma, Eve, Wu, Ren, Wang, Yufei, "Multi-Layer Survivability", Lucent Technologies White Paper, http://optical.web.lucent.com/salestools/white_paper/wp008.pdf, 1999, presented as ANSI-T1 contribution T1A1.2/2000-008, January 2000

[Metz-2000] Metz, Chris, "IP over Optical – From Packets to Photons", IEEE Internet Computing, pp. 76-82, November/December 2000

[P918-D1] Eurescom Project P918 - Integration of IP over Optical Networks: Networking and Management, "Deliverable 1 - IP over WDM, Transport and Routing", 1999

[P918-D2] Eurescom Project P918 - Integration of IP over Optical Networks: Networking and Management, "Deliverable 2 - Network scenarios for IP over optical networks", 2000

[P918-D5] Eurescom Project P918 - Integration of IP over Optical Networks: Networking and Management, "Deliverable 5 - Survivability Strategies with Interactions between Layer Networks", 2001

[PANEL-D2a] PANEL Deliverable D2a: “Overall Network Protection - Version 1”, April 1997

[Ramaswami-2002] Ramaswami, Rajiv, Sivarajan, Kumar N., "Optical Networks – A Practical Perspective, Second Edition", Morgan Kaufmann Publishers, 2002

[RFC0791] J. Postel (Ed.), "Internet Protocol", RFC 791, September 1981

[RFC0793] J. Postel, "Transmission Control Protocol", RFC 793, Date: September 1981

[RFC1518] Y. Rekhter, T. Li, "An Architecture for IP Address Allocation with CIDR", RFC 1518, September 1993

[RFC1519] V. Fuller, T. Li, J. Yu, K. Varadhan , "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy", RFC 1519, September 1993

[RFC1583] J. Moy, "OSPF Version 2", RFC 1583, March, 1994

[RFC1633] R. Braden, D. Clark and S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633, June 1994.

[RFC1661] Simpson, W. (Ed.), "Point-to-Point Protocol (PPP) ", RFC 1661, July 1994

[RFC1662] Simpson, W. (Ed.), "PPP in HDLC-like Framing", RFC 1662, July 1994

[RFC1771] Rekhter, Y., and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1771, March 1995.

[RFC2101] B. Carpenter, J. Crowcroft, Y. Rekhter, " IPv4 Address Behaviour Today", RFC 2101, February 1997

[RFC2205] R. Braden (Ed.), L. Zhang, S. Berson, S. Herzog, S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 FunctionalSpecification", RFC 2205, September 1997.

[RFC2210] J. Wroclawski, "The Use of RSVP with Integrated Services", RFC 2210, September 1997.

[RFC2212] S. Shenker, C. Partridge, R. Guerin, "Specification of Guaranteed Quality of Service", RFC2212, September 1997.

[RFC2373] R. Hinden, S. Deering, "IP Version 6 Addressing Architecture", RFC 2373, July 1998

[RFC2460] S. Deering, R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998

[RFC2474] K. Nichols, S. Blake, F. Baker, D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

[RFC2475] D. Black, S. Blake, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

[RFC2597] Heinanen, J., Baker, F., Weiss, W. and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999.

[RFC2598] V. Jacobson, K. Nichols, and K. Poduri, "An Expedited Forwarding PHB", RFC 2598, June 1999.

[RFC2702] D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999

[RFC3031] E. Rosen, A. Viswanathan, R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001

[RFC3032] E. Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., Conta, A., "MPLS Label Stack Encoding", RFC 2032, January 2001

[RFC3036] Andersson, L., Doolan, P., Feldman, N., Fredette, A. and B. Thomas, "LDP Specification", RFC 3036, January 2001.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V. and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[RFC3212] Jamoussi, B., Andersson, L., Callon, R., Dantu, R., Wu, L., Doolan, P., Worster, T., Feldman, N., Fredette, A., Girish, M.,

Gray, E., Heinanen, J., Kilty, T. and A. Malis, "Constraint-based LSP Setup Using LDP", RFC 3212, January 2002.

[RFC3270] Le Faucheur et al, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC3270, May 2002.

[RFC3272] D. Awduche, A. Chiu, A. Elwalid, I. Widjaja, X. Xiao, "Overview and Principles of Internet Traffic Engineering", RFC 3272, May 2002

[Ramamurthy-1999-I] Ramamurthy, S., Mukherjee, B., "Survivable WDM mesh networks, Part I - Protection", INFOCOM '99, Page(s): 744 -751 vol.2, 1999

[Ramamurthy-1999-II] Ramamurthy, S., Mukherjee, B., "Survivable WDM mesh networks, Part II - Restoration", ICC '99, Page(s): 2023 –2030, vol.3, 1999

[Stevens-1994] Stevens Richard, "TCP/IP Illustrated, Vol. 1 – The Protocols", Addison-Wesley Publishing Company, USA, 1994

[T'Joens-2000] Yves T'Joens, Gary Ester, Marc Vandenhoute, "Resilient Optical and SONET/SDH-based IP networks", Proc. 2nd International Workshop on the Design of Reliable Communication Networks (DRCN2000), April 2000, Munich, Germany

[Wu-1992] T. Wu, Fiber Network Service Survivabdity, Norwood, MA: Artech House, 1992.

[Wu-1995] T. Wu, “Emerging Technologies for Fiber Network Survivability”, IEEE Communications Magazine, vol. 33, no. 2, pp. 58-74, February 1995.

[Xiao-1999] XiPeng Xiao, Lionel Ni, "Internet QoS: A Big Picture", IEEE Network Magazine, March/April, pp. 8-18, 1999.

[Xiao-2000-a] XiPeng Xiao, A. Hannan, B. Bailey, S. Carter, L.M. Ni, "Traffic engineering with MPLS in the Internet", IEEE Network Magazine, March/April 2000, vol. 14, iss. 2, pp. 28-33.

[Xiao-2000-b] XiPeng Xiao, "Providing QoS in the Internet", Ph.D. thesis, Michigan State University, 2000 (http://www.cse.msu.edu/~xiaoxipe/papers/thesis/thesis.pdf)

[Yang-1988] Yang, C. H., Hasegawa, S., "FITNESS-Failure Immunization Technology for Network Services Survivability", Proc. GLOBECOM '88, 1988, vol. 3, pp. 1549 –1554

[Yahara-1997] Yahara, Tahishi, Kawamura, Ryutaro, "Virtual path self-healing scheme based on multi-reliability ATM network concept", IEEE GLOBECOM'97, November 1997