Hybrid Routing in Next Generation IP Networks: QoS Routing Mechanisms and Network Control Strategies

Antoine B. Bagula

December 2006

Doctoral thesis

TRITA-ICT/ECS AVH 06:08
ISSN 1653-6363
ISRN KTH/ICT/ECS AVH-06/08--SE

Telecommunication Systems Laboratory
Department of Electronics, Computer and Software Systems (ECS)
KTH, Royal Institute of Technology
Stockholm, Sweden

A dissertation submitted to Royal Institute of Technology (KTH) for partial fulfillment of the requirements for the Doctor of Technology degree.

Telecommunication Systems Laboratory
Department of Electronics, Computer and Software Systems (ECS)
KTH, Royal Institute of Technology
Stockholm, Sweden

Abstract

Communication networks have evolved from circuit-switched and hop-by-hop routed systems into hybrid data/optical networks using the Internet as a common backbone carrying narrow- and broad-band traffic offered by a multitude of access networks. This data/optical backbone is built around a multi-technology/multi-protocol routing architecture which runs the IP protocols in a collapsed IP stack where ATM and SONET/SDH have been replaced by the suite of Generalized Multiprotocol Label Switching (GMPLS) protocols. A further evolution referred to as "IP over Photons" or "All IP - All Optical" is expected where "redundant intermediate layers" will be eliminated to run IP directly on top of optical cross-connects (OXCs) with the expectation of achieving savings on operation expenditures (OPEX) and capital expenditures (CAPEX). "IP over Photons" has been stalled by the immaturity in the control and data plane technologies leading to complex and time-consuming manual network planning and configurations which require a group of "layer experts" to operate and maintain a hybrid data/optical network.

By making the status of each link and node of a data/optical network visible to a common control, GMPLS protocols have opened the way for automated operation and management allowing the different layers of an IP stack to be managed by a single network operator. GMPLS protocols provide the potential to make more efficient use of the IP backbone by having network management techniques such as Traffic Engineering (TE) and Network Engineering (NE), once the preserve of telecommunications, to be reinvented and deployed to effect different Quality of Service (QoS) requirements in the IP networks. NE moves bandwidth to where the traffic is offered to the network while TE moves traffic to where the bandwidth is available to achieve QoS agreements between the current and expected traffic and the available resources. However, several issues need to be resolved before TE and NE be effectively deployed in emerging and next generation IP networks. These include (1) the identification of QoS requirements of the different network layer interfaces of the emerging and next generation IP stack (2) the mapping of these QoS requirements into QoS routing mechanisms and network control strategies and (3) the deployment of these mechanisms and strategies within and beyond an Internet domain's boundaries to maximize the engineering and economic efficiency.

Building upon different frameworks and research fields, this thesis revisits the issue of Traffic and Network Engineering (TE and NE) to present and evaluate the performance of different QoS routing mechanisms and network control strategies when deployed at different network layer interfaces of a hybrid data/optical network where an IP over MPLS network is layered above an MPλS/Fiber infrastructure. These include mechanisms and strategies to be deployed at the IP/MPLS, MPLS/MPλS and MPλS/Fiber network layer interfaces. The main contributions of this thesis are threefold. First we propose and compare the performance of hybrid routing approaches to be deployed in IP/MPLS networks by combining connectionless routing mechanisms used by classical IGP protocols and the connection oriented routing approach borrowed from MPLS. Second, we present QoS routing mechanisms and network control strategies to be deployed at the MPLS/MPλS network layer interface with a

focus on contention-aware routing and inter-layer visibility to improve multi-layer optimality and resilience. Finally, we build upon fiber transmission characteristics to propose QoS routing mechanisms where the routing in the MPLS and MPλS layers is conducted by Photonic characteristics of the fiber such as the availability of the physical link and its failure risk group probability.

Acknowledgments

First my thanks go to my main advisor, Professor Björn Pehrson who made this dissertation possible and introduced me to a wider academic networking community. My special thanks go also to Professor Anthony Krzesinski whose support and contribution to this work is most valuable.

I would like to thank my colleagues of the COE group at Stellenbosch university and those of the TSLAB at KTH for their willingness to help on many occasions. Margreth Helberg, Lena Wosinska, Americo Muchanga, Björn Knutsson, Jon-Olov Vatn, Erik Eliasson, Markus Hidell, Amos Nungu, Godfrey Chikumbi, Vytautas Valencia, Khurram Jahangir, and Voravit Tanyingyong have made my time enjoyable and stimulating at TSLAB. Thanks to Marlene Botha, Hong Feng Wang, Atumbe Baruani, Gaston Mazandu, Lusilao Zodi, Tayseer Fath Elrahman and Okito Jean Andre for their contribution to this work.

A special thought goes to my father who taught me to have confidence, to do my best and run the last mile. Though no longer with us, his advice was of great help during my seemingly never ending student life. My mother had promised to assist in the defence of this thesis work. Unfortunately she passed away five months before the defence date. I am grateful for her moral support and encouragement during this thesis. I would like to thank doctors Amuli Itegwa and Bagula Moise, Brigitte Samba, Dada Bagula, Annie Bagula, Justine Bagula and Claudine Bagula for the support received from them during this thesis work.

A special thank goes to the Bakomito and Boroto families who have been supportive in many ways since our arrival in Cape Town. Their support has contributed to the completion of this thesis work.

Last but not the least, my thanks go to my wife and children. From the depths of my heart, I would like to thank my wife Yvette for her love and support during my long term student life. Yvette has tried her best to keep the children on the "right track" by playing both mom and dad roles during my frequent travels. I have signed a contract with my children for them to "deliver" at school and for me to deliver my PhD. I am delivering this PhD as the accomplishment of my part of the contract. Thanks to my champions Herman, Fortuna, Nancy, Amani and Grace for your friendship, love, trust, honesty and for making me one of the happiest fathers of the world.

This work would not be possible without the financial support of TELKOM SA and Siemens Telecommunications through the Centre of Excellence for ATM and Broadband Networks and the Department of Electronics, Computer and Software Systems of KTH.

Contents

Chapter 1

Introduction

The past decades have witnessed an intensive activity within standardization bodies such as the IETF, OIF and ITU and a deluge of proposals to extend traditional hop-by-hop IP and circuit-switched protocols and design new protocols to support Quality of Service (QoS). This was driven by several factors including (1) the evolution of the core of the Internet from a copper into a fiber network infrastructure (2) incentives to use the Internet as a data/optical backbone carrying narrow- and broad-band traffic offered by a multitude of access networks (3) the emergence of the new suite of Generalized Multiprotocol Label Switching (GMPLS) [1] protocols used to support network management such as Traffic Engineering (TE) and Network Engineering (NE) in IP networks and (4) the expectation of using these protocols to collapse traditional routing architectures from a four layer protocol stack layering IP over ATM networks above SONET over WDM networks into an "All IP-All Optical" stack running IP directly on top of cross-connects (OXCs). Figure 1.1 taken from [2] depicts one of the many visions [3, 4, 5, 6] of the evolution of the IP stack from a first generation of networks where IP over ATM networks are layered above SONET/SDH over fiber networks to a next generation IP stack where IP with GMPLS control is layered above a third generation DWDM network using Optical Label Switching (OLS) and Optical Label Switching Routers (OLSRs). However at the writing of this thesis the evolution towards "IP over Photons" is still in an early stage where several networks are operated by running IP in a traditional IP stack while a thinner IP stack is appearing in experimental research networks where an IP/MPLS is layered above an MPλS/Fiber infrastructure.

1.1 Traffic and network engineering

TE and NE are network management techniques allowing the traffic to be efficiently routed through a routed or switched network to effect QoS agreements between the offered traffic and the available resources by either using TE to move the traffic to where the network resources are available or NE to move the resources to where the traffic is offered to the network. The advent of new protocols such as MPLS, MPλS, and GMPLS have impacted network management techniques in two different but complementary ways. On one

hand, a set of single layer traffic engineering (STE) protocols such as MPLS and MPλS were invented to provide packet switching capability (PSC) in data networks and lambda switching capability (λSC) in wavelength-routed networks using different control planes and a master-slave relationship between the data and optical layer. On the other hand, GMPLS was proposed as an extension to these STE protocols to support inter-layer visibility by providing a common control plane that makes each node and link of the lower layers of a hybrid data/optical network visible to the upper layers.

1.1.1 Single layer traffic engineering (STE)

At its outset, MPLS was intended to overcome the limitations related to the *longest prefix match* deployed in traditional IP routing through *label swapping*, a packet forwarding paradigm where the route followed by the IP packets is defined by pre-computed labels. These labels are swapped into the core of the network by MPLS routers referred to as Label Switched Routers (LSRs) to define the next-hop to the destination. MPLS uses a connection-oriented routing model borrowed from the ATM *virtual connection* paradigm where the traffic is routed over bandwidth tunnels referred to as Label Switched Paths (LSPs). From a "fast forwarding" technology, MPLS has evolved into a set of protocols that offer advanced traffic engineering capabilities, Virtual Private Networks (VPNs) and multi-protocol support through logical separation between IP forwarding and routing. MPLS has experienced a wide deployment on the ISP backbone as replacement for ATM and traditional IP routing and made its way into metro-, access-networks and even some private enterprise networks. MPLS provides the capability to aggregate different traffic streams (flows) and forward these streams into one or multiple LSPs to form a single-layer grooming architecture. MPλS extends the MPLS features to wavelength routed networks to add intelligence and reconfigurability to traditional optical core networks by assigning labels to wavelengths and switching these labels to setup wavelength-switched paths referred to as λ-Switched Paths or λSP tunnels.

1.1.2 Multi layer traffic engineering (MTE)

The similarity of the LSRs to the optical cross-connect (OXCs) has led to the generalization of MPLS to cover optical networks under the *Generalized Multiprotocol Label Switching (GMPLS)* umbrella. GMPLS achieves efficient use of the high bandwidth provided by the optical technology by extending MPLS to use a single control plane that includes heterogeneous network elements supporting different routing/switching capabilities. These include IP routers that switch packets, MPLS routers and ATM switches with layer 2 switching capabilities, SDH cross-connects with TDM switching capabilities, OXCs with wavelength switching capabilities and fiber cross-connects that can switch an entire fiber. By providing a common control plane, GMPLS is expected to discharge the optical network manager from the complex and time-consuming manual network planning and configuration and effect automated management functionalities such as connection creation, connection provision, connection modification and connection deletion. GMPLS is based on a hybrid architecture that allows different logical networks to be layered above a unique physical

(fiber) network to form a *multi-layer routing architecture* and uses a path multiplexing technique referred to as *traffic grooming* where a set of logical paths located in a higher layer of the multi-layer network are multiplexed into paths located into a lower network layer to form a *multi-layer grooming architecture*.

1.2 The main issues

While TE and NE have been widely deployed in telecommunication networks, several issues need to be resolved before these network management mechanisms become wide-spread in the IP networks. These include (1) the identification of QoS requirements of the different network interface layers of an IP stack, (2) the mapping of these requirements into QoS routing mechanisms and network control strategies and (3) the deployment of these mechanisms and strategies within and beyond Internet domain's boundaries to maximize the network engineering and economic efficiency.

1.2.1 QoS requirements of different IP network layer interfaces.

Backbone communication networks are evolving from an IP architecture where IP over ATM networks are layered above SONET/SDH over DWDM/Fiber infrastructure into a thinner IP stack where IP/MPLS networks are layered over MPλS/Fiber networks with the expectation of further evolution towards the next generation "IP over Photons" architecture where IP will be layered directly on top of cross-connects. This evolution has been made possible by having the IP/ATM network interface layer replaced by an IP/MPLS interface and collapsing the SONET/SDH layer by having some of its functions such as "fast-reroute" moved into the MPLS layer and other such as the "ring topology" delegated to the DWDM layer where the SONET rings are replaced by most resilient mesh-interconnected optical cross connects (OXCs). The replacement of traditional network layer interfaces by new ones raises new QoS requirements for these interfaces. These include the design of new protocols and the redesign of network management mechanisms to be deployed at the IP/MPLS interface to replace IP over ATM mechanisms and at the MPLS/MPλS interface to replace ATM over SONET/SDH routing. Enhancements to the mechanisms previously deployed at the WDM/Fiber interface are also needed to account for the advances in switching and routing technologies provided by DWDM and fiber technology.

1.2.2 Mapping application QoS requirements into QoS strategies.

Most current generation IP protocols were designed to deliver best-effort service to the IP applications when IP transport was concerned with only data transmission. The Internet has since developed into a common transport infrastructure requiring QoS routing to meet the QoS demanded by the mixture of real-time and best-effort applications carried by a multitude of access networks. These include applications with (1) hard real-time constraints such as remote sensing, voice over IP, home automation, (2) soft real-time constraints

such as streaming video, and (3) best-effort constraints such as FTP, Secure Shell, etc. The wide-spread deployment of QoS routing by network operators will require mechanisms to (1) map the different applications into traffic flow classes, (2) identify the QoS to be provided to these flows, and (3) ensure QoS delivery for these flows. The last two steps have been extensively researched by the IP community but only few steps have been made in the quantitative evaluation of integrated systems combining the three steps and much less in the mapping of the application into traffic classes. This is according to some service provider opinion the reason for the slow deployment of QoS routing in the emerging multi-service Internet.

1.2.3 Deploying QoS within an Internet domain boundaries.

Most currently deployed routing mechanisms for IP networks are based on routing metrics (cost metrics) which optimize system-wide measures of performance such as average response time, throughput, delay , etc. discounting the diversity of QoS requirements from the mixture of narrow- and broadband applications carried by the new multi-service Internet. Managing cost to support QoS routing is a challenging problem which has been tackled by the IP community using different optimization approaches. These include (1) the use of cost metrics which reflect the current resource availability such as implemented by constraint-based routing (CBR) [11], (2) the deployment of traffic-aware routing algorithms [14, 16] proposed in the context of Multiprotocol Label Switching (MPLS) [7], and (3) the deployment of multiple metrics used either separately or combined into a mixed cost metric such as proposed in [10]. Though overcoming the limitations of the OSPF [23], CBR is poorly equipped for traffic engineering support under heavy load conditions. The proposed traffic-aware algorithms are concerned with bandwidth maximization only and incur additional complexity which does not necessarily translate into equivalent performance gains. Multiple metric routing has been suggested to be used at best as an indicator in path selection since it may result in unknown composition rule for the path cost.

1.3 Contributions and outline

It is not known how fast the collapse of IP layers will lead the evolution to "IP over Photons". However we assume that IP, MPLS, MPλS will survive the collapse of the IP stack by having these layers visible in the IP stack or moved either into an enhanced IP layer or an enhanced DWDM layer. Building upon this assumption and different frameworks and research fields, this thesis revisits the problem of Traffic and Network Engineering (TE and NE) to propose and evaluate the performance of QoS mechanisms and network control strategies which can deployed in the emerging and next generation IP networks with a specific focus on the IP/MPLS, MPLS/MPλS and MPλS/Fiber network layer interfaces. The main contributions of this thesis are threefold.

1. IP/MPLS aware routing. First we propose an STE framework to be deployed at the IP/MPLS network layer interface using an IGP+MPLS

routing approach where the performance of a network is improved by using either classical optimization methods to improve MPLS routing or genetic optimization to improve IGP routing. This framework may be used in IP infrastructures that provide different routing capabilities such as "per-application filtering" where specific end points are routed based on applications to allow different applications to be reached through different LSPs or "per-domain filtering" where specific domain prefixes are reached differently to allow the traffic destined to different websites for example to be routed over different LSPs.

2. MPLS/MPλS aware routing. Contention avoidance for bandwidth on links, inter-layer visibility and routing cooperation are three key features which can be used to improve the routing and rerouting of IP tunnels in MPLS over MPλS networks. We propose and evaluate the performance of a multilayer routing framework using a "tunnel separation/multiplexing" paradigm to avoid per-tunnel contention for bandwidth. Building upon emerging distributed router technologies, we consider the impact of "Sub-second convergence" and "inter-layer visibility" in IP networks with the expectation of improving the multi-layer resilience. A cooperative routing framework is also considered using a hybrid network management model where NE is used to complement TE.

3. MPλS/Fiber aware routing. Finally we present QoS routing mechanisms to be use in MP(L/λ)S over fiber infrastructure by deploying a layered routing approach where the routing of (L/λ)SPs is conducted by photonic characteristics of the underlying fiber. These include availability aware routing mechanisms and risk group aware control strategies. While the latter are built around the availability framework, the former are based on the "Failure Risk Group" paradigm, a survivability strategy which consider the stochastic principle that different nodes or links of a network can have similar reliability characteristics to derive QoS routing mechanisms which are more robust than classically deployed diversity routing schemes.

These contributions are depicted by Figure 1.2 which groups the different QoS routing mechanisms and network control strategies in three main parts representing the content of this thesis. In Part 1, the IP/MPLS interface is addressed using classical and genetic optimization methods to improve the performance of IP, MPLS and hybrid routed networks combining the strength and both IP and MPLS engineering. Part 2 introduces the problem of multi-layer routing and resilience by addressing the issue of contention avoidance among competing tunnels and proposing inter-layer signalling strategies which can be used to achieve inter-layer visibility. The work presented in Part 3 consists of using the photonic characteristics of the underlying fiber infrastructure to conduct the routing in the MP(L/λ)S networks. These include the fiber availability and failure risk group probability.

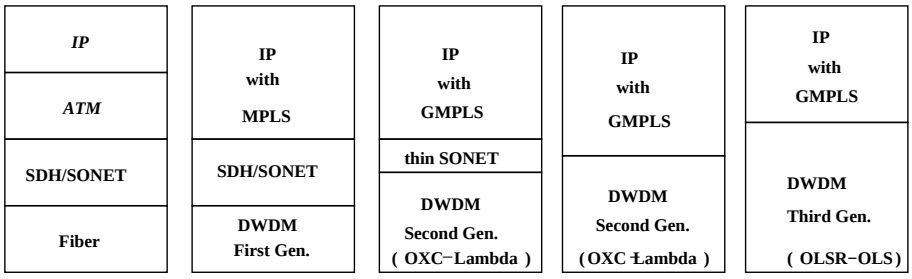


Figure 1.1: IP stack evolution

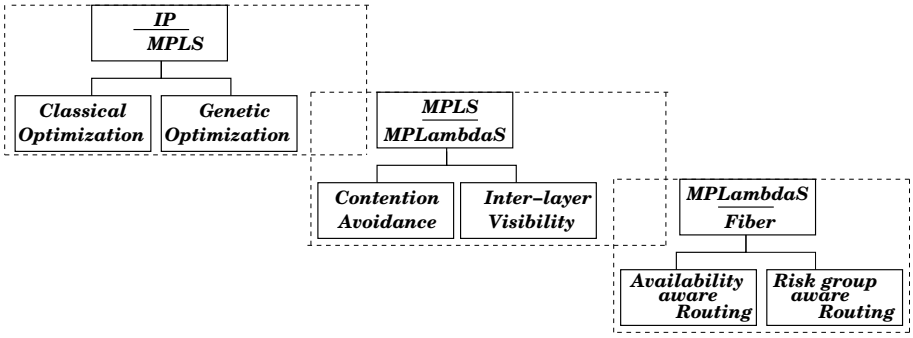


Figure 1.2: Hybrid routing

Part I

IP/MPLS aware routing

Chapter 2

MPLS routing

At its outset, Multi-protocol Label Switching (MPLS) [7] was proposed as a flexible IP forwarding mechanism using source- or flow-based routing to make IP routing at switching speed by avoiding the longest prefix-match of traditional IP networks and overcome the limitations related to the transport of IP traffic over ATM networks. In source- or flow-based routing, the complete path from a source to a destination is computed and included in the header of the packet. MPLS engineering has experienced a wide deployment in the Internet backbone as replacement for ATM switches and traditional IP routers and made its way into metro-, access-networks and even some private enterprise networks. This expansion has been driven by the potential of MPLS to build a connection-oriented network above the connectionless IP network by packet multiplexing IP and frame-relay/ATM services traditionally offered by service providers with emerging fast Ethernet services into shared channels referred to as Label Switched Paths (LSPs). Flow-based routing allows macro-bandwidth management by having the traffic be aggregated and the routes followed by the traffic to be computed on demand using a different constraint for each computed path. However basing the routing decisions on multiple constraints raises scalability issues.

This chapter presents a framework for routing these LSPs to achieve optimal network configurations maximizing bandwidth usage (optimality), minimizing the interference among competing LSPs (survivability), reducing the complexity related to using multiple constraints (scalability) and deployable by standardized routing protocols with minimum changes (migration). Unless specified explicitly, the terms channel, tunnel, flow and LSP, λSP will be used equivalently in the rest of this thesis.

2.1 A multi-constrained optimization problem

Consider a network represented by a directed graph $(\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ is a set of N nodes and $\mathcal{L}$ is a set of L links. Let C_ℓ denote the maximum reservable bandwidth of link ℓ and let $\mathcal{P}_{i,e}$ denote the set of feasible paths connecting the ingress-egress pair (i, e). Assume an on-line setting where a request $r_{i,e} = (i, e, d_{i,e})$ to setup a bandwidth-guaranteed tunnel (LSP) of $d_{i,e}$ bandwidth units between an ingress-egress pair (i, e) is received and that future

demands concerning tunnel setup requests are not known.

Let $L_p = \sum_{\ell \in p} L_\ell(n_\ell, f_\ell)$ denote the cost of path p where $L_\ell(n_\ell, f_\ell)$ is the cost of link ℓ when carrying n_ℓ tunnels and f_ℓ is the total bandwidth reserved by the tunnels traversing link ℓ.

Problem 1.1. The tunnel routing problem consists of finding the least cost path $p \in \mathcal{P}_{i,e}$ ($L_p = \min_{k \in \mathcal{P}_{i,e}} L_k$) satisfying the routing constraints

$$d_{i,e}(p) < \min_{\ell \in p}(C_\ell - f_\ell) \quad (2.1)$$

$$I_\ell(p) = \min_{k \in \mathcal{P}_{i,e}} I_\ell(k). \quad (2.2)$$

The link interference n_ℓ denotes the number of tunnels carried by the link ℓ and n_ℓ^* is a pre-assigned maximum number of tunnels that can be carried by link ℓ. $I_\ell(x) = Prob(n_\ell \rightarrow n_\ell^*)$ denotes the probability that the link interference approaches its maximum value. Eqn. (2.1) expresses the QoS routing constraints in terms of bandwidth usage maximization and Eqns. (2.2) express the interference among the competing tunnels. Note that while (2.1) represents a hard constraint to be met when routing the MPLS tunnels, (2.2) defines a soft constraint expressing the need to lead the interference n_ℓ on a link $\ell \in p$ far from its maximum value.

2.2 A single constrained optimization problem

The solution to the optimization problem above subject to two concave metrics (bandwidth and interference) is NP-hard [10]. We therefore consider a heuristic solution based on a routing scheme where the routing constraints are relaxed to find a single constraint optimization problem. This relaxation is achieved by deploying network controls in constraint-based routing [11] to guaranty that constraint (2.1) is met and using cost-based route optimization to increase the probability of meeting the routing constraint (2.2). Note that while the former is met by pruning the network, the latter is met by using a cost optimization model which maximizes the residual bandwidth $R_\ell = C_\ell - f_\ell$ and the interference difference $I_\ell = n_\ell^* - n_\ell$ or equivalently minimizes their inverses $1/R_\ell$ and $1/I_\ell$ or else minimizes the flow f_ℓ and the interference n_ℓ.

These objectives can be combined into a cost function embedding the semantics of both routing constraints (2.1) and (2.2) to express a penalty for leading the link load f_ℓ closer to its capacity C_ℓ and the link interference n_ℓ closer to its preassigned maximum value n_ℓ^*. These functions include the power-based link function

$$L_\ell(n_\ell, f_\ell) = \frac{n_\ell^\alpha}{(C_\ell - f_\ell)^{1-\alpha}} \quad (2.3)$$

or other functions derived from an additive composition rule of the two routing constraints (2.1) and (2.2). The calibration parameter $0 < \alpha < 1$ balances the impact of the two constraints on the link cost: α can be set to a high value ($\alpha \rightarrow 1$) to minimize the interference among competing tunnels or to a low value ($\alpha \rightarrow 0$) to optimize bandwidth usage following the Constraint Shortest Path First (CSPF) routing model [11]. The relaxation of the routing constraints

leads to a single constrained optimization problem expressed by

Problem 1.2. Given the network model described previously, the tunnel routing problem consists of finding the least cost path $p \in \mathcal{P}_{i,e}$ (e.g $L_p = \min_{k \in \mathcal{P}_{i,e}} L_k$) subject to the constraint

$$d_{i,e}(p) < \min_{\ell \in p}(C_\ell - f_\ell) \tag{2.4}$$

where the path cost $L_p = \sum_{\ell \in p} L_\ell$ and the link cost L_ℓ is expressed by (2.3). We showed in [32] that the optimal value of the calibration parameter α is $\alpha = 0$ under light load and $\alpha = 1$ under heavy load conditions.

2.3 The LIOA algorithm

The Least Interference Optimization Algorithm (LIOA) was proposed in [12] as a routing algorithm that uses the same procedure as the standard Constraint Shortest Path First (**CSPF**) [11] to route and reroute bandwidth-guaranteed tunnels in MPLS networks. It follows the same steps as CSPF but achieves shortest path computation using the link cost (2.3) instead of link weights which are inversely proportional to the residual link capacities: $L_\ell(CSPF) = 1/(C_\ell - f_\ell)$.

Consider a demand for $d_{i,e}$ bandwidth units between two nodes i and e. LIOA executes four steps in routing this demand

1. **Prune the network.** Eliminate all links with residual capacities less than $d_{i,e}$ to form a reduced network whose links have sufficient spare capacity to carry the demand $d_{i,e}$.

2. **Find the new least cost path.** Use Dijkstra's algorithm to find the new least cost path p from i to e in the reduced network.

3. **Route the traffic demand.** Assign the traffic demand $d_{i,e}$ to the path p.

4. **Update the link flows and interference.** For each link $\ell \in p : f_\ell := f_\ell + d_{i,e}$ and $n_\ell := n_\ell + 1$.

2.4 Related work

We presented in [12] the least interference optimization Algorithm (LIOA) to achieve on-line traffic engineering in IP networks. [12] showed that LIOA uses the cost metric (2.3) to support network optimization and protection under single link failure by balancing the number and the intensity of the flows offered to the routes. We presented in [13] an approximation of the cost metric (2.3). This approximation builds upon the stochastic property that different links of a network can have different flow carrying probabilities to express the interference among competing flows by a link flow carrying probability. [13] revealed that the approximation leads to performance improvements similar to those achieved in [12].

The work most related to ours is the Minimum Interference Routing Algorithm (MIRA) [14], a flow-based routing algorithm proposed in the context of MPLS networks to set up bandwidth-guaranteed LSPs. MIRA uses knowledge of the ingress-egress pairs to prevent the creation of bottlenecks for flows in a network by selecting a path for an LSP request that maximizes the minimum available capacity between all other ingress-egress pairs, thus reducing the rejection rate of LSP requests between a specified subset of ingress-egress pairs. However, we showed in [12] that MIRA route computations can be computationally expensive and do not necessarily translate into equivalent performance gains.

Chapter 3

IGP routing

Despite its scalability which contributed to the large expansion of the Internet, destination-based routing leads to opportunistic bandwidth sharing overloading some portions of the network while leaving some others unused. In a network carrying voice and data, an unbalanced network configuration may result in unattractive behaviour such as routing a voice-over-IP call over a high propagation delay path while a low-latency path is available or routing data traffic over high-utilised links while some portions of the network are still under-utilised. A cost-based QoS adaptation model was proposed in [38] under the *Link Weight Optimization (LWO)* label. This model is based on a routing optimization model consisting of fine-tuning the link cost metrics (also referred to as link weights) to overcome the limitations of destination-based routing. The link weight optimization model proposed in [38] has the advantages of (1) simplicity (2) capability of using diverse performance constraints and (3) compatibility with traditional IGPs. However, finding link metrics which minimize the maximum utilization is NP-hard.

*Genetic Algorithms (**GAs**)* belong to a class of *evolutionary algorithms* which can find acceptably good solutions to this problem by examining and manipulating a set of possible solutions from a set of designs. However GAs are not guaranteed to find the global solution to the problem since they may find a local optimum which does not necessarily converge to a global optimum. Memetic algorithms (MAs) are hybrid genetic algorithms which attempt to overcome this limitation by using a local search to complement classical global search.

This chapter revisits the problem of Traffic Engineering (TE) to evaluate the strength of the evolutionary algorithms when used as IP weight optimizers in destination-based routing and assess the relevance of using genetically tuned IGP weights as static costs in flow-based routing as suggested by network operator's best current practices [17].

3.1 A multi-constrained optimization problem

Consider a directed network $G(\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ is a set of N nodes and $\mathcal{L}$ is a set of L links. Let C_ℓ denote the bandwidth of link ℓ and each link ℓ has a color c_ℓ expressing some administrative constraints on that link such as the

type of traffic to be carried by that link or its loading. Assume an offline setting where the traffic $d_{i,e}$ offered to the I-E pair (i, e) is known a-priory and each request $r_{i,e} = (i, e, d_{i,e}, c_{i,e})$ demanding $d_{i,e}$ units of bandwidth between i and e is colored by $c_{i,e}$ to specify its administrative constraints. Let $P_\ell(f_\ell, n_\ell)$ denote the penalty function for link ℓ when the link carries a flow f_ℓ and n_ℓ tunnels. The IGP routing problem can be expressed as follows.

Problem 3.1. We wish to find an optimal link flow vector f and an interference vector n such that for each path p carrying a request $r_{i,e}$

$$\min_{f,n} \sum_{\ell \in \mathcal{L}} P_\ell(f_\ell, n_\ell) \tag{3.1}$$

subject to the constraints

$$f_\ell \leq C_\ell \tag{3.2}$$
$$n_\ell \leq n_\ell^* \tag{3.3}$$
$$c_\ell(p) = c_{i,e} \tag{3.4}$$

where $c_\ell(p)$ expresses the color of a link ℓ which is traversed by path p. Note that all the constraints (3.2), (3.3) and (3.4) are hard constraints which need to be satisfied in order to route the flows. It is known that if the penalty function $P_\ell(f_\ell, n_\ell)$ is convex, and if the derivatives of the penalty function are continuous and non-negative, and if all additional constraints that exist are included in $P_\ell(f_\ell, n_\ell)$ as penalty functions then the routing problem expressed by *Problem* 3.1 can be formulated as a constrained nonlinear multi-commodity flow problem. The mathematical programming literature provides general techniques for solving multi-commodity flow problems. However, the straightforward application of these techniques to the routing problem in large networks proves to be computationally intractable.

Problem 3.1. can solved using an off-line heuristic solution based on constraint based routing. This heuristic follows the same model as CSPF routing but takes advantage of the a-priory knowledge of traffic matrix to rank the demands based on their importance (priority) and route them to allow the highest priority requests to be routed first and thus receive better resources. A high level description of this algorithm is as follows

1. Sort the demands in decreasing order of importance to allow the largest demand to take the best possible path
2. For each demand in the ordered set,
   (a) Prune all the links that do not meet the constraints (3.2), (3.3) and (3.4),
   (b) Run Dijkstra's algorithm on the pruned network to find the least cost path
   (c) Adjust the used bandwidth and interference on links to reflect the current bandwidth availability and interference.
3. Use the resulting network configuration to find a set of link weights to be used in IGP routing.

3.2 An unconstrained optimization problem

Penalty function methods can be used to transform *Problem* 3.1 into an unconstrained problem which is easily solved using evolutionary algorithms. This is done by integrating both equality and inequality constraints into the objective function and introducing a penalty adjustment parameter which is updated during generations to improve the genetic process. They have been widely deployed using genetic optimization to find solutions to diverse engineering problems. We consider in this chapter the application of two penalty functions methods in IP routing namely the "adaptive" and the "co-evolutionary" also referred to as "Self-Adaptive" penalty function.

3.2.1 Adaptive Penalty

The adaptive penalty function method was initially proposed in [18]. It uses an evolutionary method where the constraints of a multi-constraint problem are integrated in the fitness function of a genetic algorithm as follows

$$fitness(x) = P(x) + \lambda(t)\big[\sum_{i=1}^{m} {g_i}^2(x) + \sum_{i=1}^{p} |h_i(x)|\big] \tag{3.5}$$

where $m = 2$ and $p = 1$

$$g_1 = f_\ell - C_\ell \tag{3.6}$$
$$g_2 = n_\ell - n_\ell^* \tag{3.7}$$
$$h_1 = c_\ell(p) - c_{i,e} \tag{3.8}$$

and $\lambda(t)$ is updated at every generation t in the following way:

$$\lambda(t+1) = \begin{cases} \frac{1}{\beta_1}\lambda(t) & \text{if } case\,1 \\ \beta_2\lambda(t) & \text{if } case\,2 \\ \lambda(t) & \text{otherwise} \end{cases} \tag{3.9}$$

Case 1 and case 2 denote situations where the best individual in the last k generations was always feasible (case 1) or was never feasible (case 2) while $\beta_1, \beta_2 > 1$ are parameters which are set to $\beta_1 \neq \beta_2$ to avoid cycling. Equation (3.9) expresses the fact that the penalty component $\lambda(t+1)$ for generation $t+1$ is decreased if all best individuals in the last k generations were feasible. It is increased if they were all infeasible. The penalty does nor change if there are some feasible and infeasible individuals tied as best in the population.

3.2.2 Co-evolutionary Penalty

Self-Adaptive penalty function or co-evolutionary penalty function method was proposed in [19] to solve problems with only inequality constraints. This method reduces to finding a fitness function of the form:

$$fitness(x) = P(x) - (coef \times w_1 + viol \times w_2) \tag{3.10}$$

where the objective function $P(x)$ takes it's value from a given set of variable encoded in a chromosome; w_1 and w_2 are two penalty factors (considered as

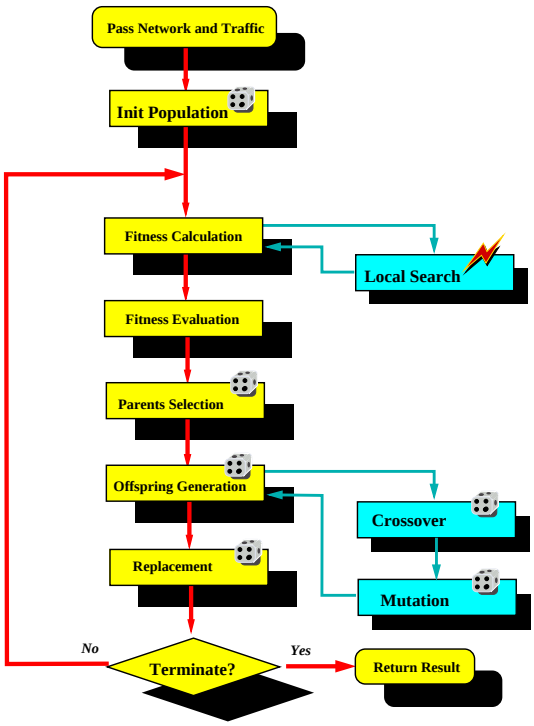


Figure 3.1: "MA algorithm"

integers); the sum of all the amounts by which the constraints are violated is given by $coef$ where:

$$coef = \sum_{i=1}^{m} {g_i}^2(x) \qquad \forall\, g_i(x) > 0 \tag{3.11}$$

The integer parameter $viol$ takes the value 0 initially, and increases by one for each constraint that has been violated.

3.3 Genetic optimization

Evolutionary algorithms use concepts from real-world genetics to evolve solutions to problems. They are based on an evolutionary paradigm where each iteration of the algorithm transforms one population of individuals into a new generation, using some pre-determined fitness measure for an individual. In applications of evolutionary algorithms, potential solutions must be represented and encoded in terms of *genome*. Each problem generally has its own genome representation, and more than one representation could be used for a given problem. The fitness measure or *fitness function* determines how good the solution represented by some genome is. The appropriate fitness function is determined by the problem and genome representation.

3.3.1 Memetic Algorithms (MAs)

*Memetic Algorithms (**MAs**)* belong to a class of evolutionary algorithms which are based on a population selection where the evolution of one generation into another relies on the three main genetic operations and a local search. These three operations depicted by Figure 3.1 are (1) *replacement* (2) *crossover* and (3) *mutation*. *Replacement* is a direct copying of a member of the current generation into the next generation. *Crossover* is the combination of two genomes from the current generation into two different genomes in the next generation. Crossover attempts to combine good solutions to find potentially better solutions. *Mutation* is the random permutation of one of the tokens in the Genome representation of a member of the current generation. By introducing new solutions at each stage of the algorithm, mutation ensures that the evolution process does not get stuck at a local optimum. There are probabilities associated with the crossover, replacement and mutation operations as illustrated by the dice in Figure 3.2. These probabilities are denoted P_c, P_r and P_m respectively, and $P_c + P_r + P_m = 1$. In general, $P_m \ll P_c$ and $P_c \approx P_r$. Candidates for the genetic operations are chosen randomly, but the selection is *fitness-proportionate* to ensure the survival of good solutions over generations. The conditions for the termination of the algorithm are problem-specific, although for practical reasons one often limits the number of iterations. The local search is used to complement the global search implemented by classical **Genetic Algorithms** to improve the genetic individuals fitness through hill-climbing and speed the genetic algorithm. It is used to map the link weights to the offered load by diverting traffic from the link with the highest utilization such as described in [31].

3.3.2 Gene Expression Programming

While **MAs** individuals are symbolic strings of fixed length referred to as "chromosomes", GEP individuals are expressed as non-linear entities of different sizes and shapes referred to as "Expression Trees" consisting of a function of + and terminals as usually expressed by the *Karva* GEP language [40]. As illustrated by Figure 3.2 (b) *Gene Expression Programming (**GEP**)* uses more operators and a different genome representation than **MA**. The *Fitness Calculation* procedure uses translated link weights patterns to compute the fitness by calling *LocalSearch*. The *LocalSearch* procedure deals with the TE computation to speed up and complement the global search procedure. The *Evaluation* procedure evaluates the fitness function and modifies the probabilities and powers which are needed by other procedures. The *Parents Selection* procedure randomly selects two parents in the population set according their probabilities and generates two offsprings which are used in other procedures. The *Transposition* procedure calls the *ISTransposition, RISTransposition, and GeneTransposition* procedures based on a given probability. These procedures achieve respectively (1) *IS transposition* with IS length of three elements , (2) *Root IS Transposition* with IS length of three elements, and (3) *Gene Transposition*. The *Recombination* procedure plays the same role as the the crossover operator in MA by gene crossover on generating offsprings. Three recombination methods were involved in our application, which are 1-point recombination, 2-point recombination, and Gene recombination. The

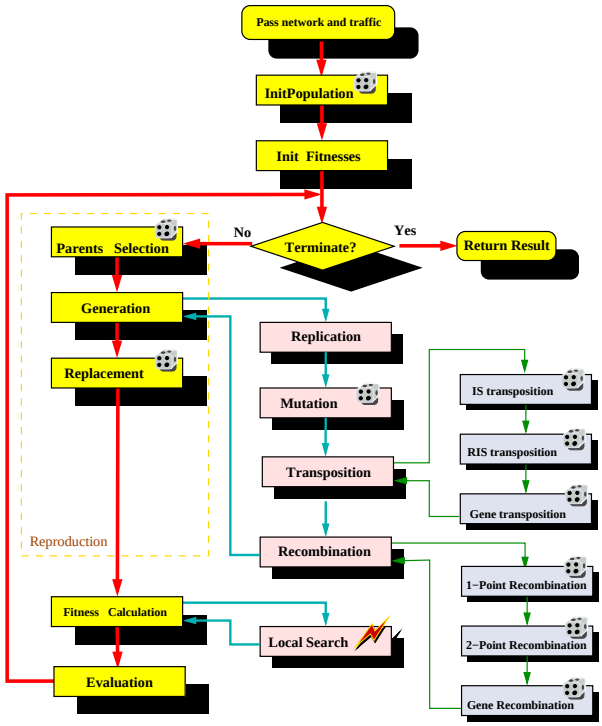


Figure 3.2: "GEP algorithm"

Replacement procedure selects the population from the offsprings and original populations based on its fitness to give birth to a new generation. Consequently the rest candidates are dead after replacement.

The *mutation* procedure in GEP algorithm uses a *ChromosomeMutation* procedure to perform 1-point mutation on a chromosome by assigning a randomly selected terminal to a randomly selected position of the chromosome. There are different other procedures associated with the *Fitness calculation* in GEP to allow the mapping of a population into a set of link weights pattern and implement translate a link weight into a chromosome. These include (1) a *TranslateChromosomes* procedure that translates a population set into a set of link weights (2) a *CtoW* procedure which translates a chromosome into a link weights pattern and a *ReformChromosome* procedure which reforms the link weights modified by *LocalSearch* into reformed into a chromosome.

3.4 Related work

We presented in [20] an application of the GEP algorithm to communication networks to assess the relevance of using GEP as routing optimizer in IGP engineering (destination based routing). We extended this work in [21] to MPLS engineering (flow-based routing) to assess the relevance of using the IGP metric as static cost metric in MPLS engineering. The results obtained in these papers revealed that GEP outperformed GA on several performance parameters and confirmed that IGP weights could be used as replacement for dynamic weights

with the same performance. Heuristic solutions using analogies with natural and social systems have been proposed to optimise IGP routing [31, 29, 39]. Different types of penalty function methods were discussed in [19] varying in the computational cost to attain the optimal solutions and handling the constraints in addition to minimizing the objective functions. [31] uses a fitness function similar to ours and applying this fitness function in an offline setting. Our work [20, 21] adopts the same genetic optimization model as [18, 19] but applies to telecommunication while the former applies to engineering.

Chapter 4

IP+MPLS routing

The dualism of IGP and MPLS routing has raised a debate separating the IP community into divergent groups with different views concerning how the future Internet will be engineered [8]. On one hand, there have been proponents of the destination-based TE model pointing to the ability of the Internet to support substantial increases of traffic load without the need for TE mechanisms such as proposed by MPLS and the capability of traditional IGP routing to optimize routing by using appropriate adjustments to the link weights [38]. On the other hand, the advocates of the MPLS standard were arguing for a flow-based TE model using source- or flow-based routing that provides more flexibility and the potential to optimize the network globally. Though MPLS has won the battle of the core of the Internet, it has become a common ISP practice to implement a hybrid network configuration combining IGP and MPLS engineering.

This chapter presents a TE model referred to as Hybrid which uses an IGP+MPLS routing approach to achieve efficient routing of flows in IP networks. These flows can be high bandwidth demanding flows (HBD) such as real-time streaming protocol flows or low bandwidth demanding flows (LBD) such as best-effort FTP flows. We assume in this chapter an on-line network design process where an optimal network configuration is built around optimality, survivability, scalability and compatibility. We adopt a path selection model where traffic flows are classified into LBD and HBD classes at the ingress of the network and handled differently in the core using a network control strategy where the LBD flows are routed using traditional IGP routing while HBD flows are carried over MPLS tunnels. This approach improves the network optimality and survivability through MPLS engineering and scalability by using IGP engineering.

4.1 The hybrid routing problem

Consider a network represented by a directed graph $(\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ is a set of N nodes and $\mathcal{L}$ is a set of L links. Assume that the network carries IP flows that belong to a set of service classes $S = \{LBD, HBD\}$. LBD and HBD define the class of low bandwidth demanding (LBD) and high bandwidth demanding (HBD) flows respectively. Let C_ℓ denote the capacity of link ℓ and let $\mathcal{P}_{i,e}$ denote the set of paths connecting the ingress-egress pair (i, e). Assume a flow-

differentiated services where a request $r_{i,e} = (i, e, d_{i,e}, s)$ to route a service class $s \in S$ flow of $d_{i,e}$ bandwidth units between an ingress-egress pair (i, e) is received in an on-line setting where future demands concerning IP flow routing requests are not known.

For each path p let $L_p = \sum_{\ell \in p} L_\ell(n_\ell, f_\ell)$ denote the path cost and $L_\ell(n_\ell, f_\ell)$ the cost of link ℓ when carrying n_ℓ flows. f_ℓ is the total bandwidth reserved by the IP flows traversing link ℓ and $(\alpha(s), \beta(s))$ is a pair of network calibration parameters depending on the flow service class s.

Problem 4.1. The flow routing problem consists of finding the least cost path $p_s \in \mathcal{P}_{i,e}$ (e.g $L_{p_s} = \min_{p \in \mathcal{P}_{i,e}} L_p$) such that

$$d_{i,e} < \min_{\ell \in p_s}(C_\ell - f_\ell) \tag{4.1}$$

Equation (4.1) expresses the feasibility of the flows. We consider a flow routing algorithm using a route optimization model based on a mixed cost model and a path selection model based on flow differentiation. This route optimization model uses a cost metric which combines optimality and survivability to route the IP flows so that fewer flows are rejected under heavy load conditions (optimality) and fewer flows are re-routed under link failure (survivability).

4.1.1 Using the link loss to express survivability.

The main survivability objective of our routing approach is to minimize the damage to the network transport layer under failure. This damage is expressed by the number of re-routed flows under failure. Let F be a set of possible failure patterns, w_f the probability of the failure pattern $f \in F$ and n_f the number of re-routed flows under failure f. The expected number of re-routed flows under the set of failure patterns F is defined by

$$W = \sum_{f \in F} W_f = \sum_{f \in F} w_f n_f \tag{4.2}$$

where $W_f = w_f n_f$ expresses the damage to the network transport layer under failure event f. Assuming that a fiber cut is the most likely failure event in optical networks, we consider the set of failure events $F = \mathcal{L}$ and define a measure of survivability expressing the link loss by

$$W_\ell = w_\ell \sum_{r \in R} \delta_{\ell,r} \tag{4.3}$$

where w_ℓ is the probability for the IP flows to traverse link $\ell \in L$ referred to as the *link loss probability*, $R = \cup_{i,e} R_{i,e}$ is the set of flows carried by the network, $R_{i,e}$ is the subset of flows from node i to node e, $n_\ell = \sum_{r \in R} \delta_{\ell,r}$ is the total number of flows carried by link ℓ referred to as its *interference* and

$$\delta_{\ell,r} = \begin{cases} 1 & \text{flow } r \text{ traverses link } \ell \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

4.1.2 Using congestion distance to express optimality.

Most routing algorithms which maximize bandwidth (optimality) assume a fair bandwidth sharing process where the different flows receive the same service on a link ℓ. This service is expressed by the link residual bandwidth $C_\ell - f_\ell$. We consider a measure of optimality referred to as the *link congestion distance* defined by

$$D_\ell(r_\ell, \beta(s)) = C_\ell - \beta(s)(f_\ell + d_{i,e}) \tag{4.5}$$

where $\beta(s)$ is a calibration parameter expressing the subscription for bandwidth to the network link. In contrast to the residual bandwidth $C_\ell - f_\ell$ which is independent of the demand, the *link congestion distance* includes in its definition the bandwidth demand $d_{i,e}$ which can be low for LBD flows and high for HBD flows. It is expected that by introducing unfairness among flows, our measure of optimality will lead to a link sharing model which improves the overall network performance by allowing the different flows to meet their service needs.

The routing optimization adopted in this chapter is based on the assumption that a link metric minimizing the link loss (or equivalently the number of flows on a link) and maximizing the link congestion distance (or equivalently its inverse) can balance the number and magnitude of flows over the network to reduce rejection under heavy load conditions and re-routing under link failure. This objective is achieved by multiplying the link loss probability by power values of the link interference and the link congestion distance to form the mixed metric expressed by

$$L_\ell(n_\ell, f_\ell, \alpha(s), \beta(s)) = w_\ell {n_\ell}^{\alpha(s)} / (C_\ell - \beta(s)(f_\ell - d_{i,e}))^{1-\alpha(s)} \tag{4.6}$$

where $0 \leq \alpha(s) \leq 1$ is a calibration parameter expressing the trade-off between survivability and optimality.

4.2 Cost-based service differentiation.

The basic idea behind our path selection model is to differentiate flows into classes based on their bandwidth requirements and route these flows using different cost metrics according to their service needs. The IP flows are classified into LBD and HBD traffic classes depending on their bandwidth requirements ($d_{i,e}$). The two traffic classes are defined by

$$S_{LBD} = \{\text{flows requesting } d_{i,e} \text{ bandwidth units} \mid d_{i,e} < \tau\} \tag{4.7}$$

$$S_{HBD} = \{\text{flows requesting } d_{i,e} \text{ bandwidth units} \mid d_{i,e} \geq \tau\} \tag{4.8}$$

where each flow bandwidth demand $d_{i,e}$ is uniformly distributed in the range $[1, M]$ and $1 \leq \tau \leq M$ is a cut-off parameter defining the limit between LBD and HBD flows.

The IP flows are routed using different routing metrics expressing their service needs: the paths followed by LBD flows are found using the IGP-based OSPF model while HBD flows are routed over MPLS Label Switched Paths (LSPs). This is achieved using the link cost (4.6) which can lead to different routing metrics depending on the values of the link loss probability w_ℓ and the

set of parameters $(\alpha(s), \beta(s))$. The routing metrics leading to IGP-based OSPF and MPLS routing are obtained by setting the link loss probability to either the value of the link flow carrying probability such as described in [13] or a constant value expressing an equal probability assumption where each link ℓ has the same probability $w_\ell = 1/L$ to carry any traffic flow. The set of parameters $(\alpha(s), \beta(s))$ is set to $(\alpha, 1)$ for MPLS routing and $(0, 0)$ for IGP-based OSPF routing.

4.3 The Hybrid routing algorithm

Consider a request to route a class s flow of $d_{i,e}$ bandwidth units between two nodes i and e. The algorithm proposed executes the following steps to route this flow

1. **Network calibration**.

   Set $(\alpha(s), \beta(s)) = (0, 0)$ if $s \in LBD$, or

   Set $(\alpha(s), \beta(s)) = (\alpha, 1)$ if $s \in HBD$.

2. **Path selection**.

   (a) **Traffic aggregation.** if $(\alpha(s), \beta(s)) = (\alpha, 1)$.

   - Find an existing LSP which can accommodate the new HBD request,
   - If found then (a) set $p_s := p$ where p is the path carrying the existing LSP and (b) goto step 3.
   - **Prune the network.** Set $L_\ell(n_\ell, f_\ell) = \infty$ for each link ℓ whose link slack $C_\ell - f_\ell \leq d_{i,e}$.

   (b) **Find a new least cost path.** Apply Dijkstra's algorithm to find a new least cost path $p_s \in P_{i,e}$.

3. **Route the request**.

   - Assign the traffic demand $d_{i,e}$ to path p_s.
   - Update the link occupancy and interference. For each link $\ell \in p_s$ set $f_\ell := f_\ell + d_{i,e}$ and $n_\ell := n_\ell + 1$.

Note that the path selection algorithm has the same complexity as Dijkstra's algorithm: $O(|L| \log |N|)$.

4.4 Related work

We proposed three different approaches for routing tunnels in hybrid IP/MPLS settings in [24, 25, 26]. The basic model proposed in [24] uses a cost metric which is based on a power based composition rule while [25] considers an additive cost metric which can be applied in delay-sensitive routing environments. In [26], IGP+MPLS routing is revisited with the objective of evaluating the impact of bandwidth aggregation and bandwidth request inflation on bandwidth growth. A trace-based analysis of the complexity of a hybrid IGP+MPLS network

was presented in [27]. This analysis uses a trigger-based mechanism to (1) differentiate flows based on their measured bandwidth during the last minute and (2) route these flows differently according to their bandwidth characteristics. Though using an on-line setting similar to ours, the focus of this approach is on traffic measurement and protocols evaluation while our approaches are based on network modelling and performance evaluation. Loosely related works were proposed in [28, 29, 30] to address the problem of routing IP flows using offline IGP+MPLS approaches where the network topology and traffic matrix are known a priory.

Part II

MPLS/MPλS aware routing

Chapter 5

Contention aware resilience

The emerging generation data/optical networks are based on multi-layer routing architecture requiring cooperation between its different single layers to achieve efficient routing of the offered traffic under different traffic profiles, resilience and reliability upon failure and improved operation Administration and Management (OA&M). Designing efficient rerouting schemes to be deployed in different layers and their interplay is an important issue upon which the efficiency of a multilayer resilience scheme depends.

This chapter revisits the concept on path multiplexing/separation and its impact of the recovery performance when rerouting bandwidth tunnels in converged data/optical networks where an MPLS network is layered above an MPλS network. We formulate the re-routing of failed tunnels as a path set finding problem subject to Quality of Service (QoS) and network control constraints. We build upon the stochastic principle that different links of a network may be provided different traffic flow carrying probabilities to present a heuristic solution where converged MPLS/MPλS networks are engineered to achieve path multiplexing/separation when re-routing LSP and λSP tunnels. We consider the interplay between layers using a mixed scheme where path restoration is implemented in the MPLS layer to complement path switching in the MPλS layer.

5.1 The contention aware resilience problem

Consider a network represented by a directed graph $G(\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ is a set of N nodes and $\mathcal{L}$ is a set of L links. Let C_ℓ denote the capacity of link ℓ and let $\mathcal{A}_{i,e} \in \mathcal{A}$ denote the set of active paths connecting the ingress-egress pair (i, e) while $\mathcal{B}_{i,e} \in \mathcal{B}$ denote the set of backup paths connecting (i, e). Assume an on-line recovery model where a set of failed tunnels $\mathcal{F}_{\tilde{\ell}} \in \mathcal{A}$ are rerouted or switched to a subset of backup paths $\mathcal{R}_{\tilde{\ell}} \in \mathcal{B}$ upon failure of link $\tilde{\ell}$. Assume that each re-routing tunnel k requiring $d_{i,e}(k)$ bandwidth units between an ingress-egress pair (i, e) is received and that demand concerning switching/re-routing requests are flooded throughout the network through Fault Indication Signal (FIS) messages.

Let $L_p = \sum_{\ell \in p} L_\ell(n_\ell, f_\ell)$ denote the cost of path p where $L_\ell(n_\ell, f_\ell)$ is the cost of link ℓ when carrying n_ℓ flows, f_ℓ is the total bandwidth reserved by the

IP flows traversing link ℓ.

Problem 5.1. The tunnel re-routing problem consists of finding the set of re-routing paths $\mathcal{R}_{\bar{\ell}} \in \mathcal{B}$ where for each path $p \in \mathcal{R}_{\bar{\ell}}$ belongs to the set of least cost paths (e.g. $L_p = \min_{k \in \mathcal{R}_{\bar{\ell}}} L_k$) such that

$$d_{i,e}(p) < \min_{\ell \in p}(C_\ell - f_\ell) \tag{5.1}$$

$$I_\ell(p) < \min_{k \in \mathcal{P}_{i,e}} I_\ell(k) \tag{5.2}$$

$$n_\ell \approx \sum_s w_s n_\ell^s \tag{5.3}$$

where $s \in \{a, b\}$ belongs to the set of tunnels classes, n_ℓ is the number of tunnels carried by the link ℓ referred to as the *link interference* and n_ℓ^* is a threshold assigned to the maximum number of tunnels carried by link ℓ and $I_\ell(p) = Prob(n_\ell \rightarrow n_\ell^*)$ is the probability for the interference on a link to approach its maximum value. n_ℓ^s is the number of type s tunnels carried by link ℓ while w_s is the flow carrying probability (weight) assigned to class s paths on the links to express their contribution of the link congestion. This contribution can be high or low depending on the values of w_s. Note that equations (5.1) expresses the QoS routing constraints in terms of bandwidth usage maximization while (5.2) is related to the interference among competing flows and (5.3) expresses the network controls in terms of path multiplexing and/or separation.

Since the solution to the optimisation problem above is $NP-$hard, we can use a polynomial transformation of *Problem* 5.1 [32] into a single constrained optimization problem expressed by

Problem 5.2. The tunnel re-routing problem consists of finding the set of re-routing paths $\mathcal{R}_{\bar{\ell}} \in \mathcal{B}$ where each path $p \in \mathcal{R}_{\bar{\ell}}$ belongs to the set of least cost paths (e.g. $L_p = \min_{k \in \mathcal{R}_{\bar{\ell}}} L_k$) such that

$$d_{i,e}(p) < \min_{\ell \in p}(C_\ell - f_\ell) \tag{5.4}$$

$$n_\ell \approx \sum_s w_s n_\ell^s \tag{5.5}$$

where the path cost $L_p = \sum_\ell L_{\ell \in p}$ and

$$L_\ell = \frac{n_\ell^\alpha}{(C_\ell - f_\ell)^{1-\alpha}}. \tag{5.6}$$

α is a calibration parameter balancing the impact of the two constraints in the link cost. It can be set to a high value ($\alpha \rightarrow 1$) to achieve minimize the interference or low ($\alpha \rightarrow 0$) to optimize bandwidth usage.

5.1.1 Differentiating MPLS and MPλS routing

Equation (5.6) can be stated differently in an MPLS over MPλS network where the two networks may be implementing different bandwidth sharing policies on the links. Let $n_\ell = \sum_s n_{s,\ell}$ denote the link interference and let $f_\ell = \sum_s b_s n_{s,\ell}$ denote the link flow where b_s denotes the per tunnel bandwidth unit for a class

s application on link ℓ and $n_{s,\ell}$ is the number of class s tunnels carried by that link.

Let $f(d_{i,e}) = r$ denote a map where $d_{i,e}$ is the size of a bandwidth guaranteed tunnel and r is the index of an application class. A differentiated bandwidth sharing policy may be applied in MPLS routing where different application classes may have different bandwidth demands so that for each class $r \neq s$ the per tunnel bandwidth unit $b_r \neq b_s$. Likewise an equal bandwidth sharing policy may be considered in MPλS routing to express the same granularity of the per tunnel bandwidth unit so that for $r \neq s$ the per tunnel bandwidth unit $b_r = b_s = b$. These assumptions leave equation (5.6) unchanged when routing in the MPLS network and rewriting the cost metric as

$$L_\ell(n_\ell) = \frac{n_\ell^\alpha}{[b(\hat{n}_\ell - n_\ell)]^{1-\alpha}} \tag{5.7}$$

under MPλS routing where $\hat{n}_\ell = C_\ell/b$ denotes the maximum number of tunnels that the link ℓ can carry. Note that while (5.6) represents the MPLS routing cost as a function of both the link load f_ℓ and the link interference n_ℓ, the MPλS routing cost in (5.7) is a function of the link interference n_ℓ only.

As described in [32], the value of the calibration parameter α may be estimated by using a functional analysis of (5.6) to its critical values. These values belong to a function defining a critical curve representing a set of calibration parameters where the network is operating in optimal mode. This set of parameters referred to as *network operation parameters* are for MPLS routing

$$\alpha = \begin{cases} 0 & \text{under light load} \\ \\ 1 & \text{under heavy load} \end{cases} \tag{5.8}$$

and for MPλS routing

$$\alpha = \begin{cases} 0 & \text{under light load} \\ 1 & \text{under heavy load} \\ 0.5 & \text{under congestion} \end{cases} \tag{5.9}$$

5.2 Achieving path separation/multiplexing.

Converged data/optical networks are expected to carry a mixture of traffic flows into bandwidth tunnels routed along active or backup paths which are commonly link or node disjoint. We build upon this assumption to consider a cost model where the link interference n_ℓ is approximated by $n_\ell \approx \beta n_\ell^a + (1-\beta)n_\ell^b$ to express the relative importance given to the tunnels carrying active traffic (n_ℓ^a) and the tunnels carrying the backup traffic (n_ℓ^b). Under this approximation the link cost (5.6) is given by

$$L_\ell = \frac{(\beta n_\ell^a + (1-\beta)n_\ell^b)^\alpha}{(C_\ell - f_\ell)^{1-\alpha}} \tag{5.10}$$

where $s \in \{a, b\}$ to express active tunnels $s = a$ and backup tunnels ($s = b$), n_ℓ^b is the number of backup tunnels carried by link ℓ referred to as the *backup link interference*, n_ℓ^a is the number of active tunnels carried by link ℓ referred to as

the *active link interference*, β and $1-\beta$ are flow carrying probabilities (weights) assigned to backup and active paths on the links to express their contribution of the link congestion, α expresses the trade-off between interference minimization and congestion distance maximization. Note that the cost function (5.10) may lead to three different QoS routing schemes depending on the value of β. These include (1) path separation/multiplexing for active/backup paths when $\beta = 1$ by moving the active paths away from the set of highly active path interfering links where n_ℓ^a is high and setting no constraints on the backup path interference n_ℓ^b, (2) path multiplexing/separation for active paths when $\beta = 0$ by moving the backup paths away from the set of highly backup path interfering links where n_ℓ^b is high and imposing no constraint on the active path interference n_ℓ^a or (3) mixed separation/multiplexing for both tunnel types when $0 < \beta < 1$.

5.3 The contention aware routing algorithm.

Building upon the common knowledge that in multi-layer networks, restoration is commonly achieved at the upper layer (data network) while protection is implemented by the lower layer (optical network), this paper considers a mixed recovery scheme achieving path protection using the *"1:1"* model in the optical layer and path restoration using the *"0:1"* model in the upper layer.

"1:1" protection provides partial protection to the traffic by creating a active and a backup path but where the normal traffic is carried over only one path (active or recovery) at a time. In this path protection and restoration scheme, extra traffic can be transported using the backup resources. The *"1:1"* model implemented in this chapter is based on a static scheme where the active paths are found by setting the link cost inversely proportional to the link capacity as proposed in CISCO's implementation of the OSPF [23] protocol and each backup path is link disjoint from the active path. In this way the active paths can share the same links (multiplexing).

"0:1" protection does not provide any protection to the traffic prior to a failure event. It is a path restoration scheme where an alternate path is computed on-the-fly upon failure to carry the traffic of the working path. It does not apply to path protection. We consider in the rest of this paper a *"0:1"* where different values of the β parameter are used to assess the impact of the path separation/multiplexing on the multilayer resilience process. A *Bottom-UP escalation strategy* is used to achieve inter-working between layers by having path restoration triggered in the data layer when either the optical layer can not recover a failed link or can only provide sufficient resources to route part of the failed tunnels. The details of the implementation of this strategy are beyond the scope of this paper. Unless specified, the acronym *tunnel* will be used in the rest of this paper to express an LSP or a λSP.

5.3.1 Achieving tunnel rerouting

We extend the *Least Interference Optimization Algorithm (LIOA)* to present a tunnel rerouting algorithm where upon failure (1) the set of failed tunnels is computed to evaluate the demand for bandwidth on different I-E pairs (2) this set is ordered in increasing $FIS(x)$ order to find the rerouting order of the different tunnels and (3) each bandwidth demand is carried over an existing

backup tunnel or rerouted on a new computed path. A high level description of the tunnel rerouting algorithm for each failed tunnel $k \in F_{\tilde{\ell}}$ that reserved $d_{i,e}(k)$ bandwidth units on link $\tilde{\ell}$ is as follows

1. **FIS propagation.** Send a FIS message to the origin of the k.
2. **Build a demand set $D_{i,e}$.** Find the demand $d_{i,e}(k)$ from each source i to each destination e and build the demand set $D_{i,e}$.
3. **Update the network topology.** Update the network topology $\mathrm{G}(N, L)$ to remove $\tilde{\ell}$. For each path $p \in F_{\tilde{l}}$ and each link $\ell \neq \tilde{\ell}$ such that $\ell \in p$ set
   - $n_\ell = n_\ell - 1$
   - $f_\ell = f_\ell - d_{i,e}(p)$
4. **Reroute the tunnels.** For each demand $d_{i,e}(k)$
   (a) **1:1 rerouting.** find an existing backup path $k \in B_{i,e}$ from i to e such that $d_{i,e}(k) < \min_\ell C_\ell - f_\ell$ and goto step (c) if found.
   (b) **0:1 rerouting.** prune the network to discard all links ℓ such that $d_{i,e}(k) > C_\ell - f_\ell$ and compute the shortest path p on the pruned network.
   (c) **Update flow and interference.** for each link $\ell \in p$ set $n_\ell = n_\ell + 1$ and $f_\ell = f_\ell + d_{i,e}(p)$.

The grooming of LSPs into λSPs may be implemented using different techniques including "packing" and "balancing" as proposed by [54] or other traffic grooming techniques proposed in the literature. These techniques are beyond the scope of this paper. The rest of this paper assumes a first-fit grooming model where in the presence of multiple wavelengths along an I-E pair, an LSP is groomed using the first fit wavelengths available in the MPλS network.

5.4 Related work

The work presented in this chapter has been published in [32] where it is revealed that the network efficiency may be improved by separating the backup tunnels on the links to leave room for the active tunnels. This work also reveals that tunnel multiplexing/separation may improve traffic growth when compared to widely deployed algorithms such as Widest Shortest Path [55] and CSPF [11].

Chapter 6

Inter-layer visibility

The design of efficient network control strategies to be deployed in the different layers of a data/optical network is as important as the rerouting mechanisms used by the different layers to recover from failure. The problem of the interplay between these layers to make the errors of a layer visible to another is another important issue upon which the efficiency of a multilayer network depends.

Signaling issues were addressed in [50] using different approaches referred to as "escalation strategies". These include (1) "uncoordinated approach" where different recovery schemes are deployed in the multiple layers without any coordination (2) "sequential approach" using either a Bottom-UP escalation strategy starting recovery actions from the lowest detecting layer where the failure is detected and escalating upward or a Top-Down escalation strategy where the escalation process is initiated in the highest possible layer and goes downward in the layered network and (3) an *"integrated approach"* that ensures coordination between the recovery mechanisms in different layers by combining these mechanisms in one integrated multilayer recovery scheme. The advantages and drawbacks of these approaches are well documented in [51]. There are three main questions related to fast signaling:

1. how are the errors in the different layers of a multi-layer network detected ?

2. how are the errors appearing in a layer of a multi-layer network made visible to another layer ?

3. how are appropriate recovery actions taken to achieve efficient rerouting of the failed tunnels ?

While a lot of work have been done in error detection techniques and rerouting mechanisms, the issue of achieving inter-layer visibility has been less researched or was addressed using techniques which might lead to poor signaling performance.

This chapter addresses the problem of multi-layer resilience by presenting an integrated approach using the interplay between the routing of IP tunnels, the rerouting of these tunnels and error signaling to improve multi-layer resilience. This integrated model uses virtual tunnel pre-emption to achieve optimality, tunnel protection classification to achieve fast signaling and a "hold-off timer"

escalation procedure to achieve inter-layer visibility and compatibility with currently deployed recovery mechanisms.

6.1 The inter-layer visibility model

This section presents the inter-layer visibility model where tunnels are classified and hold-off timer marked at tunnel routing, they are priority ranked and rerouted based on priority upon failure and protected in the MPλS or restored into the MPLS layer based on their classes.

6.1.1 The tunnel routing problem

Consider a network represented by a directed graph $G(\mathcal{N}, \mathcal{L})$ where $mathcal N$ is a set of N nodes and $\mathcal{L}$ is a set of L links. Let $\mathcal{A}_{i,e} \in \mathcal{A}$ and $\mathcal{B}_{i,e} \in \mathcal{B}$ denote respectively the set of active and backup tunnels connecting the ingress-egress pair (i, e). Assume that a request $r = (i, e, k, d_{i,e}(k))$ to route a tunnel $p \in \mathcal{A}_{i,e}$ requiring $d_{i,e}(p)$ bandwidth units between an ingress node i and the egress node e is received and that the tunnel is preferably protected by a backup tunnel $\tilde{p} \in \mathcal{B}_{i,e}$. Consider an on-line setting where future demands concerning routing tunnel requests are not known in advance.

Problem 6.1. The tunnel routing problem consists of finding the least cost active path $p \in \mathcal{A}_{i,e}$ and the least cost backup path $\tilde{p} \in \mathcal{B}_{i,e}$ such that

$$d_{i,e}(p) < \min_{\ell \in p}(C_\ell - f_\ell) \tag{6.1}$$

$$\gamma d_{i,e}(\tilde{p}) < \min_{\ell \in \tilde{p}}(C_\ell - f_\ell) \tag{6.2}$$

$$L_p = \min_{k \in \mathcal{A}_{i,e}} L_k \tag{6.3}$$

$$L_{\tilde{p}} = \min_{k \in \mathcal{B}_{i,e}} L_k \tag{6.4}$$

where γ is a protection parameter expressing whether the active paths are protected using the "1:1" mode for $\gamma = 0$ or the "1+1" mode for $\gamma = 1$ and the path cost $L_p = \sum_{\ell \in p} L_\ell$ while the link cost L_ℓ is expressed by

$$L_\ell = \frac{n_\ell^\alpha}{(C_\ell - f_\ell)^{1-\alpha}} \tag{6.5}$$

6.1.2 The tunnel rerouting problem

Consider an on-line recovery model where a set of failed tunnels $\mathcal{F}_{\tilde{\ell}} \in \mathcal{A}$ are rerouted or switched to a subset of backup tunnels $\mathcal{R}_{\tilde{\ell}} \in \mathcal{B}$ upon failure of link $\tilde{\ell}$. Assume that each request $r = (i, e, k, d_{i,e}(k))$ to reroute a tunnel k requiring $d_{i,e}(k)$ bandwidth units between an ingress node i and the egress node e is received and that demand concerning switching/re-routing requests are flooded throughout the network through Fault Indication Signal (FIS) messages. Let $L_p = \sum_{\ell \in p} L_\ell(n_\ell, f_\ell)$ denote the cost of path p where $L_\ell(n_\ell, f_\ell)$ is the cost of link ℓ when carrying n_ℓ tunnels and f_ℓ denote the total bandwidth reserved by the tunnels traversing link ℓ whose capacity is C_ℓ. n_ℓ is referred to as the link

interference.

Problem 6.2. The tunnel rerouting problem consists of finding the least cost recovery path $\tilde{p}$ such that

$$d_{i,e}(\tilde{p}) < \min_{\ell \in \tilde{p}}(C_\ell - f_\ell) \tag{6.6}$$

$$Sig(x,y) \rightarrow \overline{Rout}(x,y) \tag{6.7}$$

where

$$\overline{Rout}(x,y) = \begin{cases} Rout(x,y) & prio(x) > prio(y) \\ Rout(y,x) & \text{otherwise} \end{cases} \tag{6.8}$$

where $prio(x)$ is a function defining the priority allocated to the request x. $Sig(x,y)$ and $Rout(x,y)$ are two precedence operators which define the relation between the FIS message signaling to the ingress of the failed tunnels and the rerouting of these tunnels. They are used in (6.7) to define either a First Signaled First Rerouted (FSFR) policy or a Highest Priority First Rerouted (HPFR) policy. In FSFR, the first signaled tunnels are rerouted first: *"if the rerouting request x is signaled to the ingress of the network before request y ($Sig(x,y)$ holds), then the request x will be rerouted before request y ($Rout(x,y)$ holds)"*. In HPFR, the highest priority tunnels are rerouted first independently of the signaling order: *"if the rerouting request y has higher priority compared to x ($prio(x) < prio(y)$ holds), then the request y will be rerouted before request x ($Rout(y,x)$ holds)"*. Note that the *FSFR* policy defines the normal rerouting model while *HPFR* is used to effect routing pre-emption.

6.1.3 The fast signaling problem

The survivability requirements of a multi-layer network are closely related to the signaling model used to notify the failure to the different layers by using both FIS messages propagation and inter-layer signaling. FIS messages propagation has been addressed by different authors in the context of "sub-second IP convergence". Inter-layer signaling was addressed using the escalation strategies described above. The signaling problem consists of finding a signaling model or building upon existent signaling models to define network control strategies that can reduce the error dissemination time. It can be formally expressed as follows:

Consider a set of signaling strategies $S = \{s\}$ where the s denotes a signaling model such as FSFR or HPFR in intra-layer signaling or an escalation method in inter-layer signaling.

Problem 6.3. The signaling problem consists of finding a strategy $s \in S$ such that

$$\min_{s \in S} \sum_{k \in \mathcal{R}_{\tilde{\ell}}} Fis(s,k)$$

subject to

$$Fis(s,k) = Fis(s \leftarrow k) + Fis(s \uparrow k) \tag{6.9}$$

where $Fis(s \leftarrow k)$ denotes the time taken to signal a failure of tunnel k when strategy s is applied to propagate the FIS messages in intra-layer signaling and $Fis(s \uparrow k)$ is the time taken by strategy s to signal the failure of tunnel k between layers.

6.2 Network control strategies

This section describes the different network control strategies used to achieve intra- and inter-layer and the coordination between these strategies to improve multi-layer resilience. We assume a data/optical network where an MPLS network is layered above an MPλS network. We consider that upon failure, inter-layer signaling is achieved using a bottom up escalation model based on "hold-off timer".

It is commonly known that achieving resilience in the MPλS layer has several advantages. These include (1) easier and faster fault management by switching fewer larger blocks of traffic on backup paths and (2) faster activation of the resilience mechanism as compared to higher layers where the resilience mechanisms have to wait for the propagation of the alarms from the lower layers to the upper layers. On the other hand, MPLS resilience has also several advantages. These include (1) the capability to protect specific equipment, particular customers and services and (2) a finer granularity of the recovery process that maximizes spare transmission capacity. Building upon this common knowledge, we adopt the "0:1" restoration model in the MPLS layer while "1+1" protection switching is used in the MPλS layer.

6.2.1 Achieving inter-layer visibility

Given the recovery mechanisms described above, the requirements to achieve $1+1$ protection in the MPλS layer are met when at setup time both an active and a recovery λSPs are found. They are not met when either both the active and backup λSP can not be found or only the active λSP can be found. In the former case, the origin of the active λSP p is marked with a higher hold-off timer time ($HoT(p) = high$) to indicate the rerouting over an existent λSP in the MPλS layer upon failure. The case where only the active λSP could be found leads to marking the origin of the λSP with a lower hold-off timer ($HoT(p) = low$) indicating the need to restore this tunnel in the MPLS layer based on the $0:1$ model. By having the rerouting of tunnels in the MPLS layer follow the $0:1$ model while the rerouting of λSP follows either $1+1$ protection in the MPλS layer or $0:1$ in the upper layer, this process will make the protection mode of the MPλS layer visible to the MPLS layer to reduce inter-layer signaling operations time. This is different from normal inter-layer signaling operations based on hold-off timer operations which can last some minutes. The inter-layer signaling operation time resulting from a strategy s using the newly proposed "inter-layer visibility" process is thus $Fis(s \uparrow k) \approx 0$ for each failed tunnel k. This differs from normal bottom up escalation procedures using hold-off timer model which can last some minutes.

6.2.2 Deploying virtual pre-emption

The rerouting of failed tunnels may lead to a competition for bandwidth on the rerouting links upon failure resulting in a situation of starvation where the least important tunnels are rerouted first and starve the most important as illustrated by Figure 6.1. This figure depicts a situation where two IP tunnels are setup between two source destination pairs: the most important tunnel $(1, 2, 6, 7, 9)$ from S_1 to I and the least important tunnel $(1, 3, 7, 9)$ from S_2 to I, both

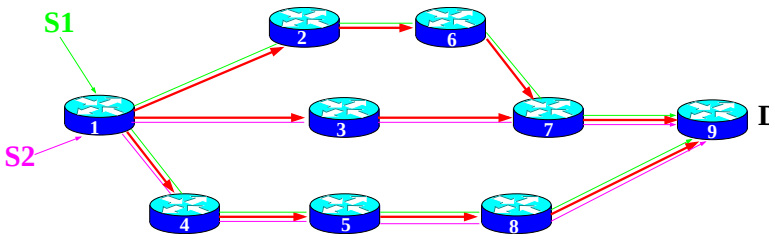


Figure 6.1: "Bandwidth competition"

sharing a common link $(7, 9)$. A racing situation may happen upon failure of the common link $(7, 9)$ if the least important tunnel (S_2, I) is signalled first to the common source node 1 and rerouted first. In case that only limited resources (e.g bandwidth) are available on the common rerouting path $(1, 4, 5, 8, 9)$, this leads to the least important tunnel (S_2, I) tie-ding up resources (bandwidth) that the most important tunnel (S_1, I) could use and a subsequent denial to reroute this tunnel. To solve this problem, we adopt a tunnel pre-emption strategy where instead of implementing the FSFR model described earlier, HPFR is implemented by having the tunnels rerouted based on their importance: "the most important tunnels are rerouted first".

6.2.3 Fast signaling

Delaying the rerouting of the least important tunnels can lead to routing instabilities which can outweigh the benefit of the virtual tunnel preemption strategy. This problem is solved by using fast signalling strategies [70] where normal processing operations are pre-empted to leave room for fault signaling operations to reduce the time taken by the fault indication messages.

The total delay induced by the preemption process when using strategy s to reroute a tunnel x may be expressed by $PT(s, x)$ and the time to recover this tunnel $TTR(s, x)$ expressed by

$$TTR(s, x) = Fis(s \leftarrow x) + PT(s, x) + Fis(s \uparrow x) \tag{6.10}$$

where $Fis(s \uparrow x) \approx 0$ as described above.

Building upon the assumption that the time spent in preemption $PT(s, x) << Fis(s \leftarrow x)$ and using "sub-second convergence" to reduce the FIS propagation time from $Fis(s \leftarrow x)$ to $Fis^{'}(s \leftarrow x) << Fis(s \leftarrow x)$ leads to rewriting equation (6.10) as

$$TTR^{'}(s, x) \approx Fis^{'}(s \leftarrow x) + PT(s, x) \tag{6.11}$$

where $Fis^{'}(s \leftarrow x)$ is the FIS message propagation time when a network is implementing "sub-second convergence". Note that since $Fis^{'}(s \leftarrow x)$ is very small compared to the normal fault signalling indication time $Fis(s \leftarrow x)$, the following relation holds

$$TTR(s, x) > TTR^{'}(s, x) \tag{6.12}$$

6.3 Routing, rerouting and signaling algorithms

A high level description of the routing, rerouting and signaling algorithms associated with these controls are

1. Routing algorithm

   Find a least cost path p

   If found {

   - Prune the network to discard the links of path p in backup path selection
   - Find a backup path $\tilde{p}$
   - If found set $HoT(p) = high$ to allow the λ layer take over when rerouting this tunnel upon failure else Set $HoT(p) = low$ to allow the MPLS layer take over when rerouting this tunnel.

   }.

2. Rerouting algorithm

   Build a demand set $D_{i,e}$ of failed tunnels

   For each tunnel $k \in D_{i,e}$ do {

   - If $(HoT(k) == high)$ reroute k on its $(1 + 1)$ backup tunnel
   - If $(HoT(k) == low)$ signal the demand of k to its origin where k will be rerouted based on a the virtual tunnel pre-emption model by
     - ranking of the tunnels based on their importance, and
     - rerouting these tunnels based on the order set.

   }

3. The signaling algorithm is based on

   (a) Selection of the rerouting layer based on HoT.

   (b) Propagation of the FIS messages to the edge of the network: the origin of the failed tunnels.

6.4 Related work

The design of efficient re-routing schemes to be deployed in the different layers of a hybrid data/optical architecture is an issue which has been widely addressed in [48, 49, 50, 51, 52, 53] and other papers such as [58] and [60] and is still debated by the research community. Fast signalling strategies achieving "subsecond convergence" have been proposed in [70]. We proposed in [66] and [67] an "inter-layer communication" model which uses static classification of the errors that may happen at different layers of a multi-layer network and identifies the layers where they might be handled. This classification process is used in a token based escalation procedure to notify the appropriate layer to handle the error and thus reduce the waiting time that could be induced by time out actions.

Part III

MPλS/Fiber aware routing

Chapter 7

Availability aware TE

GMPLS is based on a hybrid architecture where different logical networks are layered above a unique physical (Fiber) network to form a *multi-layer routing architecture* and a set of paths located in a higher layer of the multi-layer network are multiplexed into paths located into a lower network layer to form a multi-layer grooming architecture. By making the status of the lower layers visible to the upper layers through its sets of routing and signaling protocols (OSPF-TE and RSVP-TE), GMPLS has opened the way to a multi-layer control and management approach allowing the cooperation between different single layers to achieve specific network functions. These include the efficient routing of the offered traffic and grooming of data under different traffic profiles, recovery from failure, priority handling and preemption and improved operation Administration and Management (OA&M). However, the design and implementation of QoS routing mechanisms which consider the interplay between layers has been only scarcely addressed by the IP community.

The main idea behind the availability aware routing paradigm is to improve the availability of connections by having the traffic carried over the most available tunnels and provide availability guarantees by having the IP tunnel requests to be routed over the links that guarantee the availability level requested by these tunnels. Figures 7.1 and 7.2 depict two network configurations where each link ℓ is assigned a label indicating its availability. We consider a routing scheme where three IP tunnels are setup on three source-destination pairs (S_1, D_1), (S_2, D_2), and (S_3, D_3). The two network configurations depicted by Figures 7.1 and 7.2 reveal that availability aware routing may lead to two routing configurations: (1) a path separated routing configuration depicted by Figure 7.1 where the three tunnels are routed over three different most available least cost paths and (2) a path multiplexed network configuration depicted by Figure 7.2 where the three tunnels are routed over three different least cost paths which share the same link $(7, 8)$. In contrast to the path separated network, the upper and lower paths $(1, 2, 3, 5, 5)$ and $(10, 11, 12, 13, 14)$ of the path multiplexed network present links which are least available. This leads to selecting the most available links. The path separated network configuration may be preferred to the path multiplexed network configuration since it leads to higher load balancing of the traffic over the network infrastructure by having each tunnel routed over a different path and fewer rerouting upon failure by having only the tunnel (S_2, D_3) to

be rerouted upon a failure of the link $(7, 8)$. The path multiplexing routing configuration would lead to multiplexing the three tunnels on the link $(7, 8)$ and the rerouting of all the three tunnels upon failure of link $(7, 8)$. However it would be preferred in situations where the availability of the tunnels is important and in proactive resilient networks where a failure does not degrade highly the network quality of service.

Figure 7.1: Path separated network

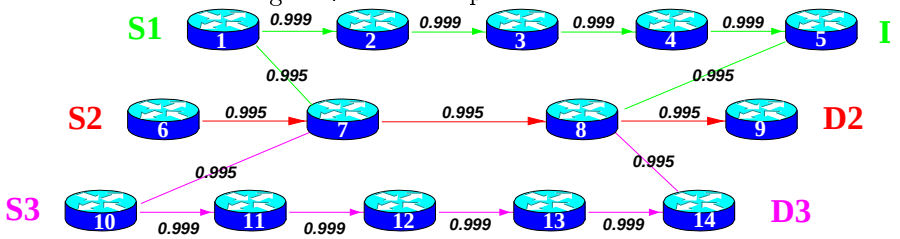
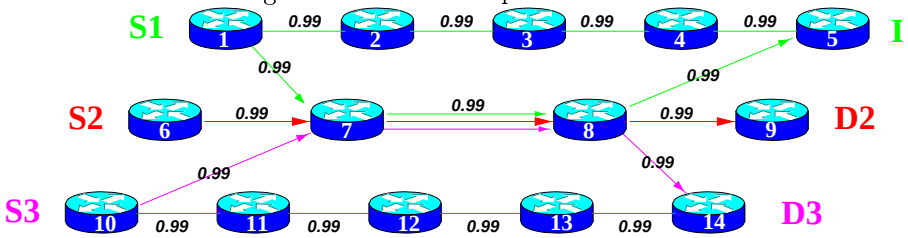


Figure 7.2: Path multiplexed network



This chapter presents QoS routing mechanisms to achieve TE in the MPLS and MPλS layers by using Photonic characteristics of the underlying physical (Fiber) network such as the link availability to conduct the routing in the upper layers: MPLS and MPλS layers. We consider a link capacity subscription (LCS) model where the links of a network are under-subscribed based on availability to reduce their routing cost in order to carry more tunnels on the most available links. A link and tunnel coloring (LTC) model is also considered where the links and the requests to set up tunnels are classified into color classes to guaranty that the tunnels will be routed over the links that provide higher availability.

7.1 A link capacity subscription (LCS) model

Consider a network represented by a directed graph $(\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ is a set of N nodes, $\mathcal{L}$ is a set of L links. Let $\overline{C}_\ell$ denote the capacity of link ℓ and let $\mathcal{P}_{i,e}$ denote the set of paths connecting the ingress-egress pair (i, e). Assume that each link $\ell \in L$ of the network has availability a_ℓ and a request $r_{i,e} = (i, e, d_{i,e})$ to route a tunnel of $d_{i,e}$ bandwidth units between an ingress-egress pair (i, e) is received. Consider an on-line setting where future IP tunnel routing requests are not known. Assume that $L_p = \sum_{\ell \in p} L_\ell(n_\ell, f_\ell)$ denote the path cost where $L_\ell(n_\ell, f_\ell)$ is the LIOA cost metric defined previously as a function of the interference on links n_ℓ and the total flow f_ℓ.

7.1.1 The LCS routing problem

Problem 7.1. The tunnel routing problem consists of finding the least cost path $p \in \mathcal{P}_{i,e}$ (e.g $L_p = \min_{p \in P_{i,e}} L_p$) such that

$$d_{i,e} < \min_{\ell \in p}(\overline{C}_\ell - f_\ell) \tag{7.1}$$

$$P_r(u_k \leq u_\ell | a_k \leq a_\ell) \geq P_r(u_k \leq u_\ell | a_\ell \leq a_k) \tag{7.2}$$

where u_ℓ is the utilization of link ℓ and $P_r(u_k \leq u_\ell | a_k \leq a_\ell)$ is the conditional probability that the utilization of link ℓ is superior to the utilization of link k ($u_k \leq u_\ell$) if link ℓ is more available than link k ($a_k \leq a_\ell$). Note that while equation (7.1) expresses a bandwidth usage maximization constraint equation (7.2) represents the availability aware routing constraint consisting of maximizing the probability of routing the traffic over the most available links.

7.1.2 Scaling the link capacity to improve availability

Constraint (7.1) is often associated with the problem of bandwidth growth through minimization of the link utilization. It is a common network operator practice [63] to solve this latter problem by either under-subscribing the links to control the utilization of a specific link or by inflating the bandwidth requirement of the tunnel to control the utilization of all the links. It was shown in [26] that the deployment of bandwidth inflation in an IGP+MPLS setting improves the routing efficiency. Link under-subscription is used to achieve bandwidth growth by applying to each link of a network ℓ a bandwidth under-subscription parameter $0 < \gamma_\ell < 1$ to reduce its maximum reservable bandwidth $\overline{C}_\ell$ so that its subscribed capacity is $C_\ell = \gamma_\ell \overline{C}_\ell$.

We propose an availability aware routing model where the capacity scaling factor γ_ℓ is set proportionally to the value of the link availability with the objective of reducing the cost L_ℓ of the most available links to route more traffic on the these links and reduce the load of the least available links. This may improve the availability of the connections. This leads to an expression of the LIOA link cost previously defined but where the link capacity has been scaled based on availability as follows

$$L_\ell = \frac{n_\ell^\alpha}{(\gamma_\ell \overline{C}_\ell - f_\ell)^{1-\alpha}} \tag{7.3}$$

where $\gamma_\ell = a_\ell / \max_{k \in L} a_k$. Note that the LIOA routing algorithm may be used to achieve availability aware routing using the link cost above.

7.1.3 The LCS routing algorithm

A sample on-line constrained routing algorithm using the LCS model is as follows

1. Compute the availability of each link of the network.
2. Compute for each link ℓ the scaling factor γ_ℓ.
3. Route a tunnel routing request as follows.
   (a) Prune all the unfeasible links based on constraint (7.1).

(b) Run Dijkstra's algorithm on the pruned network and using the link cost (7.3) to find the least cost path that meets constraint (7.2).

(c) Adjust the used bandwidth on links to reflect the current bandwidth availability.

The algorithm above may be extended to achieve offline constrained routing. The offline algorithm is as follows

1. Sort the tunnels in decreasing order of their availability to allow the most available tunnels to take the best possible path.

2. For each tunnel in the ordered set run the on-line algorithm above.

7.1.4 Computing the link availability

Building upon the asymptotic (steady state) availability for a continuously operating, repairable item with no protection, the reliability block diagram (RBD) of an optical link can be expressed by a simple series structure consisting of n spans and $n-1$ nodes. It can be used to express the availability of a physical link a_ℓ as in [71, 74]

7.2 The Scaled Link Cost (SLC) model

Consider a network represented by a directed graph $(\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ is a set of N nodes, $\mathcal{L}$ is a set of L links. Let $\overline{C}_\ell$ denote the capacity of link ℓ and let $\mathcal{P}_{i,e}$ denote the set of paths connecting the ingress-egress pair (i, e). Assume that each link $\ell \in L$ of the network has availability a_ℓ and a request $r_{i,e} = (i, e, d_{i,e})$ to route a tunnel of $d_{i,e}$ bandwidth units between an ingress-egress pair (i, e) is received. Consider an on-line setting where future IP tunnel routing requests are not known. Assume that $L_p = \sum_{\ell \in p} L_\ell(n_\ell, f_\ell)$ denote the path cost where $L_\ell(n_\ell, f_\ell)$ is the LIOA cost metric defined previously as a function of the interference on links n_ℓ and the total flow f_ℓ.

7.2.1 The SLC routing problem

Problem 7.2. The tunnel routing problem consists of finding the least cost path $p \in \mathcal{P}_{i,e}$ (e.g $L_p = \min_{p \in P_{i,e}} L_p$) such that

$$d_{i,e} < \min_{\ell \in p}(\overline{C}_\ell - f_\ell) \tag{7.4}$$

$$I_\ell(p) = \min_{k \in \mathcal{P}_{i,e}} I_\ell(k) \tag{7.5}$$

$$\tilde{n}_\ell = \phi_\ell n_\ell \tag{7.6}$$

where the link interference n_ℓ denotes the number of tunnels carried by the link ℓ and n_ℓ^* is a pre-assigned maximum number of tunnels that can be carried by link ℓ. $I_\ell(x) = Prob(n_\ell \to n_\ell^*)$ denotes the probability that the link interference approaches its maximum value. Note that while equation (7.4) expresses a bandwidth usage maximization constraint equation (7.6) represents the availability aware routing constraint while equation (7.5) is a constraint on

the interference among competing tunnels as previously defined. *Problem* 7.2 may be reduced into a single constrained optimization problem by integrating the two constraints (7.5) and (7.6) into a cost metric using a polynomial transform similar to chapter 1. We propose a polynomial transform using a link cost scaling model where the link interference is scaled by a measure of the link unavailability to reduce the cost of the most available links and increase the cost of the least available links to achieve availability aware routing. This will lead to routing more traffic on the most available links and reduce the load of the least available links. The resulting link cost model is expressed by

$$L_\ell = \frac{\tilde{n}_\ell^\alpha}{(\overline{C}_\ell - f_\ell)^{1-\alpha}} \tag{7.7}$$

where $\tilde{n}_\ell = \phi_\ell n_\ell$, $\phi_\ell = \overline{a}_\ell / \max_{k \in L} \overline{a}_k$ and $\overline{a}_\ell = 1 - a_\ell$ is the link unavailability.

7.2.2 The SLC routing algorithm

A sample on-line constrained routing algorithm using the LCS model is as follows

1. Compute the availability of each link of the network.
2. Compute for each link ℓ the scaling factor ϕ_ℓ.
3. Route a tunnel routing request as follows.
   (a) Prune all the unfeasible links based on constraint (7.4).
   (b) Run Dijkstra's algorithm on the pruned network and using the link cost (7.7) to find the least cost path that meets constraint (7.6).
   (c) Adjust the used bandwidth on links to reflect the current bandwidth availability.

7.3 A link and tunnel coloring (LTC) model

Given the network described above, we assume a tunnel-differentiated services model where a request $r_{i,e} = (i, e, d_{i,e}, s, a_{i,e}(s))$ to route a service class $s \in S$ tunnel of $d_{i,e}$ bandwidth units between an ingress-egress pair (i, e) is received. Consider that the application carried by this request requires $a_{i,e}(s)$ availability and each link ℓ has availability a_ℓ. Assume an on-line setting where future demands concerning IP tunnel routing requests are not known.

For each path p let $L_p = \sum_{\ell \in p} L_\ell(n_\ell, f_\ell)$ denote the path cost where $L_\ell(n_\ell, f_\ell)$ is the cost of link ℓ when carrying n_ℓ flows and $f_\ell = \sum_s f_{\ell,s}$ is the total bandwidth reserved by the IP flows traversing link ℓ while $f_{\ell,s}$ is the bandwidth reserved by a class s tunnel on link ℓ.

7.3.1 The LTC routing problem

Problem 7.3. The tunnel routing problem consists of finding the least cost path $p_s \in P_{i,e}$ (e.g $L_{p_s} = \min_{p \in P_{i,e}} L_p$) such that

$$d_{i,e} < \min_{\ell \in p_s}(C_\ell - f_\ell) \tag{7.8}$$

$$f_{\ell,s} = \delta_{\ell,s} f_\ell \tag{7.9}$$

$$\delta_{\ell,s} = \begin{cases} 1 & a_{i,e}(s) < a_\ell \\ 0 & \text{otherwise} \end{cases} \tag{7.10}$$

Note that $\delta_{\ell,s}$ expresses the fact that the links are carrying only tunnels whose availability requirements are inferior to the availability of the link: e.g. a applications requiring for example 0.9 availability will be carried only on links ℓ which have higher or the same availability ($0.9 <= a_\ell$).

7.3.2 The LTC routing algorithm

Link coloring is a simple mechanism that allows the grouping of links in differentiated path selection settings where flows are classified into color classes and the links of a network are grouped into similar classes to either exclude or include them during path selection. Link coloring may be useful when routing for example delay-sensitive flows in a network by assigning to long delay links such as satellite links a link coloring different from the color assigned to fiber links. In a link color setting where the demands are differentiated into color classes, the demand is redefined as $d_{i,e}(c)$ to include the color c allocated to the request. Each tunnel routing request will be allocated a color reflecting its availability and each link is colored based on its availability.

A sample on-line constrained routing algorithm using the LTC model is as follows

1. Compute the availability of each link of the network.
2. Assign to each link ℓ a color corresponding to its availability.
3. Route the tunnel request as follows.
   (a) Prune all the unfeasible links based on the set of constraints (7.8) and (7.9).
   (b) Run Dijkstra's algorithm on the pruned network to find the least cost path.
   (c) Adjust the used bandwidth on links to reflect the current bandwidth availability.

7.4 Related work

The work presented in this chapter addresses the issue of availability by using a cost-based optimization approach in an on-line setting. It builds upon several other works done in the context of availability [71, 72, 73, 74] to evaluate the availability of a link.

Chapter 8

Availability aware NE

GMPLS has opened the way for self-adaptive networks that relieve the optical network manager from the complex and time-consuming manual network planning and configuration required by traditional communication networks. Self-adaptation encompasses automated management functions such as connection creation, connection provision, connection modification and connection deletion by allowing cost-effective and short-term deployment of Bandwidth on Demand (BoD) [79]. These functions allow the routing of bandwidth-guaranteed tunnels (LSPs/λSPs) under various traffic profiles and applications requirements, the re-routing of these tunnels upon failure or congestion and the re-sizing of these tunnels upon redefinition of Service Level Agreements (SLAs) between users and the network manager. The routing of tunnels in single and multi-service settings are TE mechanisms consisting of moving the traffic to where bandwidth is available in the network. A situation can happen in the emerging BoD settings where the established tunnels have to be re-sized to adapt to traffic fluctuations or to meet new Service Level Agreements (SLAs). This NE mechanism is achieved by moving bandwidth where the traffic is offered to the network.

We propose in this chapter an artificial economy where the resizing of the pre-established MPLS tunnels is done by pricing bandwidth based on congestion and availability and trading this bandwidth between artificial market agents to control the allocation of bandwidth on the tunnels. The tunnel resizing can be performed at random instances of time as proposed in [79] or periodically to re-optimize the network under traffic fluctuations or based on trigger to adjust these tunnels to new SLAs. We propose a route capacity subscription (RCS) model where the capacity of a route is scaled based on the route availability to reduce the price on the most available route to allow these routes to sell more bandwidth. We also consider a differentiated availability pricing (DAP) model where the routes of a network are classified into availability classes which are priced differently and handled differently according to their prices to trade the bandwidth on the most available routes in a "highest available market" while the bandwidth on the least available routes is traded in a "least available market".

8.1 The route capacity subscription (RCS) model

In our artificial economy, a broker referred to as a "bandwidth manager" is assigned to each tunnel. Periodically, based on trigger or at random time instants the manager calculates the expected revenue, over a short period of time (known as the *planning horizon*) that the route would gain, if the route were to acquire U additional units of bandwidth (the "buying price") and also the expected value, over the same short period of time, that the route would lose should it give up U units of bandwidth (the "selling price").

We will use the LIOA cost function presented in previous chapters to determine the cost of a unit of bandwidth. Let B_r denote the capacity of route r and, with an abuse of notation, let n_r and f_r denote the number of LSPs in service on route r and the bandwidth assigned to route r respectively.

When acquiring (buying) U units of capacity the route capacity will increase from B_r to $B_r + U$. Likewise when releasing (selling) U units of capacity the route capacity will decrease from B_r to $B_r - U$.

A manager that decides to buy bandwidth on route r will thus lead the link cost to

$$L_r = \frac{n_r^\alpha * \theta_r \tau_r}{(B_r + U - f_r)^{1-\alpha}}. \tag{8.1}$$

while a manager that sells bandwidth on route r will lead the route cost to

$$L_r = \frac{n_r^\alpha * \theta_r \tau_r}{(B_r - U - f_r)^{1-\alpha}} \tag{8.2}$$

where $\theta_r = |r|\theta$, $|r|$ is the length (hop-count) of route r, τ_r is the planning horizon on route r and θ is the revenue earned per unit time on a single hop connection.

8.1.1 Pricing bandwidth

We build upon the assumptions above to define the "selling price" on route r as

$$V_r = \left(\frac{n_r^\alpha}{(B_r^- - U - f_r)^{1-\alpha}} - \frac{n_r^\alpha}{(B_r^- - f_r)^{1-\alpha}} \right) \theta_r \tau_r$$

and the "buying price" on route r as

$$K_r = \left(\frac{n_r^\alpha}{(B_r^- - f_r)^{1-\alpha}} - \frac{n_r^\alpha}{(B_r^- + U - f_r)^{1-\alpha}} \right) \theta_r \tau_r$$

where B_r^- is the subscribed bandwidth on route r, $B_r^- = A_r^+ * B_r$, $A_r^+ = A_r / \max_{k \in R} A_r$, A_r is the availability of route r. A route manager will buy bandwidth on another manager's route if the selling price is inferior to its buying price.

8.1.2 The route availability

In contrast to the link availability which is computed based on a simple series structure reliability block diagram (RBD), the route availability is computed using a parallel structure reliability block diagram (RBD) which considers the

protection available on each route. It can be computed using the end-to-end availability model proposed in [74].

8.2 A differentiated availability pricing (DAP) model

We consider a multi-service model where the tunnels are classified into less available tunnels (lat) and highly available tunnels (hat) and handled differently by using a different pricing function for each class. The tunnel classification is expressed by

$$S_{hat}(\tau) = \{\text{tunnels } r \text{ such that } A_r > \tau\} \quad (8.3)$$

$$S_{lat}(\tau) = \{\text{tunnels } r \text{ such that } A_r <= \tau\} \quad (8.4)$$

where A_r is the availability of route r and τ is an availability threshold. We consider a pricing model where the "selling price" is expressed by

$$V_r = \left(\frac{n_r^\alpha}{(B_r^- - U - \delta_s f_r)^{1-\alpha}} - \frac{n_r^\alpha}{(B_r^- - \delta_s f_r)^{1-\alpha}} \right) \theta_r \tau_r$$

and the "buying price" on route r as

$$K_r = \left(\frac{n_r^\alpha}{(B_r^- - \delta_s f_r)^{1-\alpha}} - \frac{n_r^\alpha}{(B_r^- + U - \delta_s f_r)^{1-\alpha}} \right) \theta_r \tau_r$$

where $\delta_s = 1$ for *hat* tunnels and $\delta_s = 0$ for *lat* tunnels. Note that since f_r represents the congestion level of a route r, the tunnel classification and differentiated handling above will lead to the creation of two different Virtual Private Networks: a congestion aware VPN ($\delta_s = 1$) used by the most available routes and a congestion free VPN ($\delta = 0$) used by the least available routes. Note also that despite the cost-based separation between VPNs resulting from the use of the parameter δ_s, this separation will be effective and lead to two different markets only when the congestion on routes is high $f_r \to B_r$. When the routes are lightly loaded $f_r \approx 0$, the two pricing functions will be the same.

8.3 Cooperative routing using TE+NE

A hybrid TE+NE strategy may be implemented in GMPLS aware networks to allow a network operator to sell "LSPs on demand" to some clients of an MPLS network and "λSPs on demand" to other clients in a DWDM (MPλS) network. The MPLS network will be built by having a set of LSPs set up on demand and thereafter re-sized either periodically or at random periods of time to achieve network re-optimization. These LSPs may also be re-sized based on trigger to adapt to new SLAs. They are groomed (multiplexed) into some λSPs in the DWDM network. The MPλS network will be built by having a set of λSPs setup in the MPλS layer based on the DWDM client's demand. When a request to re-size some of the LSPs which are groomed in λSPs arrives, these LSPs are signaled to the MPLS layer to be re-sized as a normal LSP resizing operation using bandwidth trading and re-groomed either in the same λSPs or relocated into other λSPs.

8.3.1 λSP re-sizing using LSP relocation

Note that though the re-sizing of the LSPs is done through trading in the MPLS layer, the finer trading granularity of the MPLS network cannot be applied to the MPλS layer where bandwidth exchanges are done at coarser granularity. The λSPs will be re-sized when they are experiencing high blocking upon attempting to groom the LSPs. The re-sizing of the λSPs will be done by relocating some of their groomed LSPs with the objective of leading the blocking in these λSPs (the grooming λSPs) at a given threshold pre-designed by the network operator. It follows the simple algorithm described by

while ($blocking > threshold$)

- relocate an LSP,
- adjust the number of groomed LSPs in the λSP.

8.4 Related work

Part of the work presented in this chapter has been published in [79] where the basic model describing the artificial economy market is presented. The evaluation of the route availability used as scaling factor in the RCS model may be built upon different studies done in the context of end-to-end availability [71, 72, 73, 74]. The work the most related to ours are [64] and [65]. We showed in [79] that our model achieves the same performance as [64] but at a lower computational price. The pricing model adopted by [65] uses a bandwidth sharing model similar to ours but a pricing model based on auctions.

Chapter 9

Risk Group Aware Routing

Current generation Constraint Based Routing (CBR) [11] algorithms are based on a myopic path selection model where the least cost paths are computed to route the traffic on a set of resources (links and nodes) which can probably fail together. This may lead to network configurations which can hide the weakness of the underlying routing approaches while highlighting their strengths.

This chapter reconsiders the problem of traffic engineering to propose a new constraint based routing approach which includes the probability of the links/nodes of a network to fail together as an additional constraint in CSPF routing. Building upon the stochastic principle that different links/nodes of a physical network may follow similar failure patterns, we introduce the Failure Risk Group Avoidance (FRGA) paradigm and propose a new algorithm referred to as Constrained Failure Risk Group Avoidance (CFRGA) algorithm that maximizes bandwidth usage as in normal CSPF routing while routing the traffic over failure risk free paths.

9.1 Failure Risk Group Avoidance (FRGA)

Failure Risk Groups are states of a network where a set of its resources may fail together with a given probability. Figures 9.1 and 9.2 depicts two network configurations where each link ℓ is assigned a label indicating the risk group G_ℓ to which that link belongs. We consider a routing scheme where three IP tunnels are setup on three source-destination pairs (S_1, D_1), (S_2, D_2), and (S_3, D_3) subject to a " failure risk group avoidance (FRGA)" constraint. This constraint imposes that each of the three tunnels be routed on a least cost path whose links belong to different risk groups: e.g. for each link $\ell \in p$ and $\ell^{'} \in p$ such $\ell \neq \ell^{'}$ the inequality $G_\ell \neq G_{\ell^{'}}$ holds. The two network configurations depicted by Figures 9.1 and 9.2 reveal that FRGA routing may lead to two routing configurations: (1) a path separated routing configuration depicted by Figure 9.1 where the three tunnels are routed over three different risk group free least cost paths and (2) a path multiplexed network configuration depicted by Figure 9.2 where the three tunnels are routed over three different least cost paths which share the same link $(7, 8)$. In contrast to the path separated network, the upper and lower paths $(1, 2, 3, 5, 5)$ and $(10, 11, 12, 13, 14)$ of the path multiplexed network present links which share the same risk group: links

$(1,2)$ and $(4,5)$ belong to risk group G_1 while the links $(10,11)$ and $(13,14)$ belong to risk group G_4. The path separated network configuration may be preferred to the path multiplexed network configuration since it leads to higher load balancing of the traffic over the network infrastructure by having each tunnel routed over a different path and fewer rerouting upon failure by having only the tunnel (S_2, D_3) to be rerouted upon a failure of the link $(7,8)$. The path multiplexing routing configuration would lead to multiplexing the three tunnels on the link $(7,8)$ and the rerouting of all the three tunnels upon failure of link $(7,8)$.

Figure 9.1: Path separated network

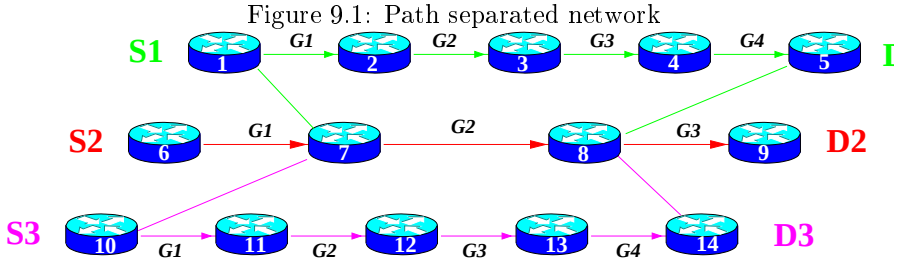
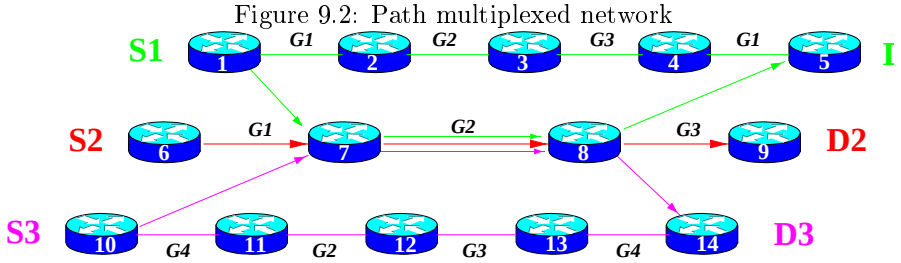


Figure 9.2: Path multiplexed network



The illustration above shows that the selection of efficient routing configurations is an important issue upon which the efficiency of the emerging and next generation IP networks depends. It also reveals that these routing configurations depend on the probability for the links/nodes of the network to achieve similar failure patterns. As the emerging networks are based on DWDM technology allowing the transport of terabytes of traffic over a single optical link, these networks require high survivability which can be achieved by deploying proactive recovery algorithms to reduce the risk of failure and reactive recovery mechanisms providing efficient rerouting capabilities upon failure. The Shared Risk Link Group (SRLG) [80, 81, 82] has been extensively researched by the optical community as a single failure mode that results in the failure of multiple links. However, investigation studies on whether multiple failures can occur in a network and the recovery actions to be taken upon these failures have only been scarcely addressed [83, 84, 85]. The constrained failure risk group avoidance (CFRGA) problem proposed in this paper and its algorithmic solution are an extension to the SRLG which include the probability of multiple failures of a set of links/nodes in path computation.

9.2 Failure risk group finding problem

Consider a network $G = (\mathcal{N}, \mathcal{L})$ consisting of a set of N nodes $\mathcal{N}$ interconnected by a set $\mathcal{L}$ of L links. Assume that the nodes of the network are perfect while the links $i \in [1 \dots L]$ are operating in binary mode: failed or working. Assume that a network state s represents a particular condition where a certain number of links are failed while the other are in working condition. Consider that the network moves from one state to another if the current state of any of the links changes. Assume that each link ℓ can fail independently of others with probability $q_\ell = 1 - p_\ell$, and the state space denoted $\mathcal{J}$ contains all possible states. Assume that the links are numbered from the least reliable to the most reliable as follows

$$1/2 \leq p_1 \leq p_2 \leq \dots \leq p_L \leq 1$$

while the ratios $R_i = q_i/p_i$ are numbered in non increasing order from the highest value to the least value as follows

$$1 \geq R_1 \geq R_2 \geq \dots \geq R_L \geq 0$$

9.2.1 The failure risk group finding problem

Problem 9.1. The state space enumeration consists of finding the smallest subset of states $\mathcal{V} \subseteq \mathcal{J}$ of the network such that

$$\sum_{s \in \mathcal{V}} p(s) \geq \pi \tag{9.1}$$

where $p(s)$ is the probability of state $s \in \mathcal{V}$ given by

$$p(s) = \prod_{\ell \in \overline{S}} p_\ell \prod_{\bar{\ell} \in S} q_{\bar{\ell}} = \prod_{\ell=1}^{L} p_\ell \prod_{\bar{\ell} \in S} R_{\bar{\ell}} = (\prod_{\ell=1}^{L} p_\ell) R(S).$$

$R(S)$ is the R-value of state S, π is the state space coverage, and $S \subseteq \mathcal{L}$ is the set of failed components occurring in the state s and $q_{\bar{\ell}} = 1 - p_{\bar{\ell}}$.

9.2.2 The failure risk group finding algorithm

We consider an algorithm that generates the most probable states of a network in non-increasing order as proposed in [86, 87, 88]. It is based on an offline model that finds the most probable states of a network and their probabilities using a given state coverage and the link failure probabilities preassigned for a given network topology. As defined earlier, the most probable states found are expressed in terms of sets of links that fail simultaneously.

The algorithm is based on a two-step process illustrated by Figure 9.3: an initialization step and an iterative step. It uses a Hasse diagram and two sets (a candidate and an active set) to generate the most probable states of a network in non-increasing order as proposed by [86]. The Hasse diagram illustrated by Figure 9.4 for a four link network is a representation of a partially ordered set in

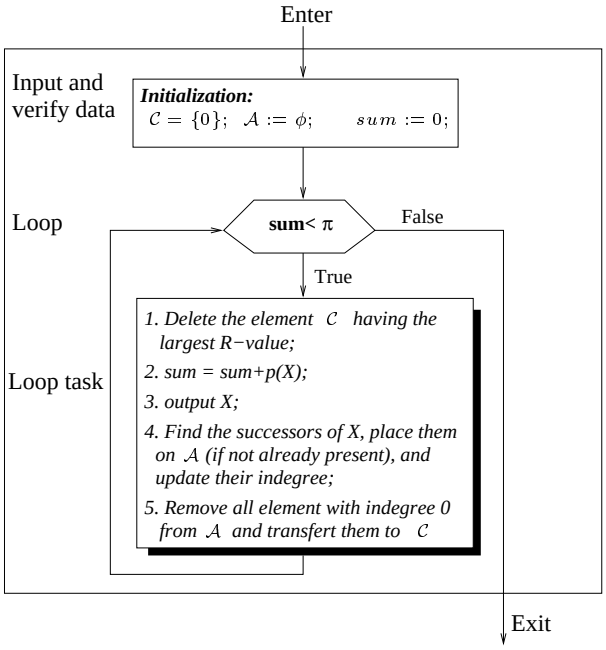


Figure 9.3: Algorithm GENERATE

which distinct elements are represented by distinct points. In the Hasse diagram, two comparable elements X and Y which are related by the relation $X \geq Y$ are joined by a line segment descending from X to Y, but the relationship implied by transitivity are not explicitly drawn. Note that as illustrated by the shaded lines in the Figure, the four-link Hasse diagram includes a copy of a three-link Hasse diagram. This property can be generalized to an arbitrary number of components n: an $n+1$ Hasse diagram contains a copy of an n Hasse diagram. The successor Y of a node $X = k_1 k_2 ... k_j$ in the Hasse diagram is find as follows:

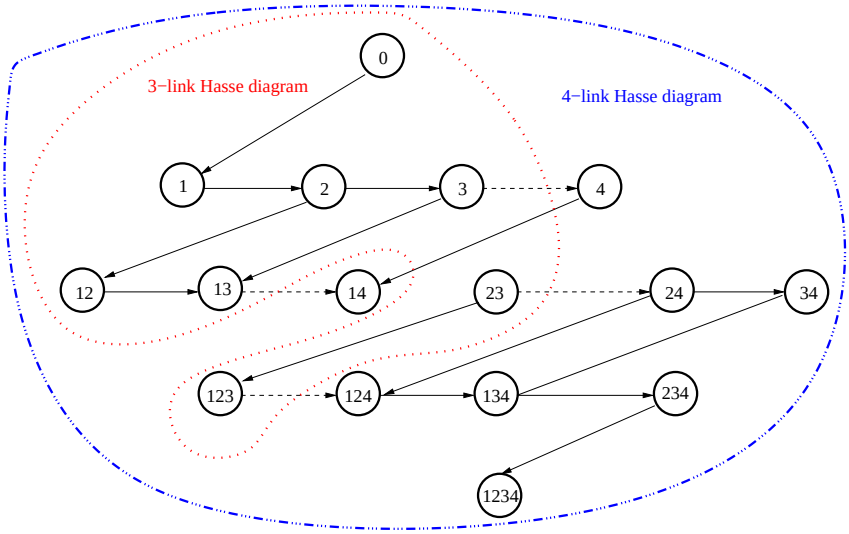
- If X is on the same level as Y then $Y = k_1^{'} k_2 ... k_j$ where $k_1^{'} = k_1 + 1$.
- If X and Y are on different levels then $Y = k_1^{'} k_2 ... k_j$ where $k_1^{'} = 1 k_1$.

The Risk Group Finding algorithm uses a two-step process : an initialization step and an iterative step. The candidate set $\mathcal{C}$ contains all the states with in-degree zero while the active set $\mathcal{A}$ contains all states whose predecessors have just been deleted from the Hasse diagram. A variable *sum* is used to record the sum of the probabilities $p(X)$ of all states that have been deleted from the Hasse diagram.

During the **initialization step** the candidate set $\mathcal{C}$, the active set $\mathcal{A}$, and the *sum* of state probabilities are initialized. The candidate set $\mathcal{C}$ is initialized with the only state 0 of the Hasse diagram since it has in-degree zero while the active set $\mathcal{A}$ will be empty since no element has been yet deleted from the Hasse diagram. The variable *sum* is also initialized to 0 since initially no state have yet been deleted from the Hasse diagram.

The **iterative step** is executed into a loop while the sum of all state

Figure 9.4: Hasse diagram



probabilities sum is less than the probability coverage π. It follows a sequence of (1) deleting from the candidate set $\mathcal{C}$ the element that has the highest $p(X)$-value (2) adding the probability $p(X)$ of the deleted element to sum (3) placing the deleted element X into the output list (4) finding the successor(s) of the deleted element X and place them into $\mathcal{A}$, if they are not already in $\mathcal{A}$ and decrement their in-degree by 1 and (4) finally placing all elements with in-degree 0 from the active set $\mathcal{A}$ into the candidate set $\mathcal{C}$.

9.3 Constrained failure risk group avoidance

Consider a network represented by a directed graph $(\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ is a set of N nodes and $\mathcal{L}$ is a set of L links. Let C_ℓ denote the maximum reservable bandwidth of link ℓ and let $P_{i,e}$ denote the set of feasible paths connecting the ingress-egress pair (i, e). Assume that a request $r_{i,e} = (i, e, d_{i,e}, \pi^*)$ to setup bandwidth-guaranteed tunnels (LSPs or λSPs) of $d_{i,e}$ bandwidth units between an ingress-egress pair (i, e) is received and that future demands concerning tunnel setup requests are not known. Assume that these tunnels may be either unprotected or diversely computed tunnels which are constrained by a failure risk group avoidance probability π^*.

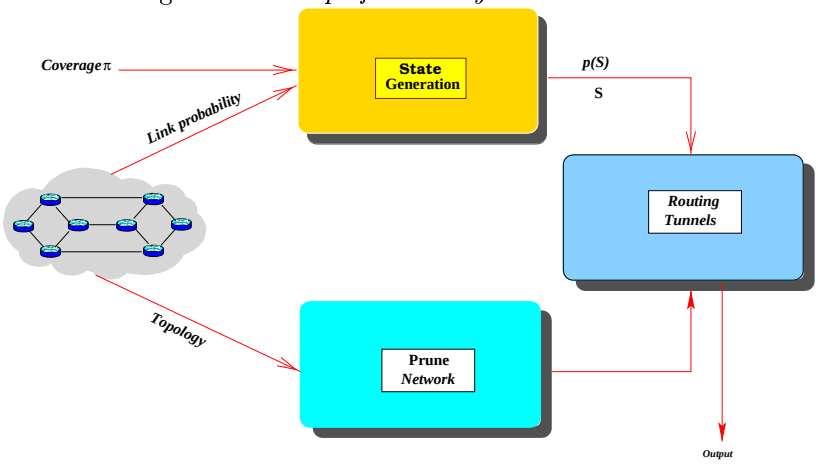
Let $L_p = \sum_{\ell \in p} L_\ell(n_\ell, f_\ell)$ denote the cost of path p where $L_\ell(n_\ell, f_\ell)$ is the cost of link ℓ when carrying n_ℓ tunnels and f_ℓ is the total bandwidth reserved by the tunnels traversing link ℓ.

Problem 9.2. The tunnel routing problem consists of finding the least cost path p (eg. $L_p = \min_{k \in P_{i,e}} L_k$) satisfying the routing constraints

$$d_{i,e}(p) < \min_{\ell \in p}(C_\ell - f_\ell) \tag{9.2}$$

$$P^{\pi}_{\ell,\ell'}(p) < \pi^* \tag{9.3}$$

Figure 9.5: *The performability evaluation model*



where $P^{\pi}_{\ell,\ell'}(p)$ is the probability that two links $\ell \in p$ and $\ell^{'} \in p$ belong to a same risk group computed at coverage π while π^* is a threshold defining the level of risk allowed. Note that while constraint (9.2) expresses a bandwidth maximization, constraint (9.3) is a risk group avoidance constraint expressing how risk group free the links of the computed tunnel should be.

Problem 9.3. The diversity tunnel routing problem consists of finding the least cost path p (e.g. $L_p = \min_{k \in P_{i,e}} L_k$) satisfying the routing constraints and its least cost backup $\tilde{p}$ (e.g. $L_{\tilde{p}} = \min_{k \in P_{i,e}} L_k$) satisfying the routing constraints

$$d_{i,e}(p) < \min_{\ell \in p}(C_\ell - f_\ell) \quad (9.4)$$

$$\gamma d_{i,e}(\tilde{p}) < \min_{\ell \in \tilde{p}}(C_\ell - f_\ell) \quad (9.5)$$

$$P^{\pi}_{\ell,\ell'}(p,\tilde{p}) < \pi^* \quad (9.6)$$

where $P^{\pi}_{\ell,\ell'}(p,\tilde{p})$ is the probability that two links $\ell \in p$ and $\ell^{'} \in \tilde{p}$ belong to a same risk group computed at coverage π while π^* is a threshold defining the level of risk allowed. Note that the parameter γ in equation (9.5) expresses the type of protection required for the active tunnel: $1:1$ for $\gamma = 0$ and $1+1$ for $\gamma = 1$. Constraint (9.3) is a risk group avoidance constraint expressing how risk group free the links of the two computed tunnels should be.

9.3.1 Constrained risk group avoidance algorithm

We consider a constrained risk group avoidance algorithm which is based on the following key features

- Offline computation of the the failure risk groups.
- Routing based on bandwidth and risk group constraints.

- Path selection based on selecting the best risk group free paths: the best path may be for example the widest, the least utilized or the least delay path.

Figure 9.5 depicts a routing architecture which can be used to compute constrained risk group avoidance paths. It includes three main blocks corresponding to the three main routing processes involved in the path computation. In *State generation*, the FRGs are computed based on a coverage π likely provided by the network administrator and the link failure probabilities $p(\ell)$. The *State generation* process produces a set of states $\{s\}$ and their probabilities $p(s)$. These states and probabilities are used in the *Routing Tunnels* process to compute risk group constrained paths. The *Prune network* process prunes the network based on the network topology and available bandwidth on links to meet the bandwidth maximization constraint. In the *Routing tunnels* the bandwidth pruned network provided by *Prune network* is pruned again based on the FRGs and their probabilities to meet the risk group constraints. *Routing tunnels* performs the least cost calculation on a bandwidth and risk group pruned network to find the least cost paths.

Consider a demand for $d_{i,e}$ bandwidth units between two nodes i and e at risk group probability π^*. Assume that $QoP(p)$ denotes the quality of path p expressed in terms of path delay, path width, maximum link utilization of the path or other performance parameter and QoP is a variable expressing the quality of the best path found. CRGA executes the following steps in routing this demand

1. **State enumerate.** Enumerate the states of the network based on a predefined coverage π and link failure probabilities.
2. **Prune the network.** Eliminate all links with residual capacities less than $d_{i,e}$ to form a reduced network whose links have sufficient spare capacity to carry the demand $d_{i,e}$.
3. **Find the new least cost path.**
   - $BestPath = \emptyset$ and QoP=0.
   - For each state s that meets the risk group constraint $p(s) < \pi^*$
     - Prune the network again to meet the risk group constraints
     - Use Dijkstra's algorithm to find the new least cost path p from i to e in the reduced network.
     - If $(QoP(p) \geq QoP)$ set $BestPath = p$ and $QoP = QoP(p)$.
4. **Route the traffic demand.** Assign the traffic demand $d_{i,e}$ to the best path *Bestpath*.
5. **Update the link flows and interference.** For each link $\ell \in Bestpath$: $f_\ell := f_\ell + d_{i,e}$ and $n_\ell := n_\ell + 1$.

9.4 Related work

While using constraint based routing in an on-line setting, the work presented builds upon the offline state enumeration models proposed in [86, 87, 88] to find failure risk groups.

Chapter 10

Conclusions and future work

Building upon different frameworks and research fields, this thesis revisits the problem of traffic engineering and network engineering to present QoS routing mechanisms and network control strategies to be deployed at the main network layer interfaces of the emerging and next generation IP stack. We present IGP, MPLS and IGP+MPLS routing approaches to be deployed at the IP/MPLS network layer interface by combining the strengths of native IP (scalability) and MPLS engineering (optimality and survivability) to improve the QoS received by the applications. We also consider multi-layer routing approaches to be deployed at the interface between MPLS and MPλS networks. These include contention-aware routing approaches based on LSP/λSP separation/multiplexing mechanisms and LSP/λSP rerouting mechanisms which use fast signaling and inter-layer visibility to improve multi-layer resilience. Finally, we propose Photonic aware routing approaches where the knowledge of the underlying fiber characteristics is used to conduct the routing and rerouting of LSPs/λSPs in MPLS/MPλS networks. These include failure risk group aware and availability aware routing mechanisms. There is room for further research to extend and complement the work presented in this thesis.

10.0.1 Deploying (G)MPLS in metro-, access and private networks

MPLS has won the battle of the core of the Internet by using its label-switched paths as a natural fit for virtual private networks and constraint based routing and fast-rerouting capabilities to collapse ATM and SONET/SDH in an emerging IP stack which is futured with an IP/(G)MPLS over DWDM architecture and possibly a thin SONET layer between the two layers. Its move into metro-, access- and even some private networks is driven by the promises to increase service provider revenue by providing the capability of multiplexing IP, frame relay/ATM and Ethernet traffic into shared tunnels. This emerging IP stack requires the redesign of the QoS routing mechanisms and network control strategies defined for the core to achieve different QoS requirements into metro,

access and private networks. This is a direction for future research work.

10.0.2 A step forward into "IP over Photons" deployment

The next generation "IP over Photons" networks are expected to layer native IP directly on top of cross connect. This will be possible by redefining the IP layers to move some of their functionalities in the IP layer and other in the DWDM layer, by designing new mechanisms and redesigning the existent IP/Optical routing mechanisms such as Optical Burst Switching (OBS) and Optical Packet Switching (OPS) to improve IP delivery on top of cross connect. While OBS has been adopted as an intermediate solution between fast circuit switching and packet switching, OPS is widely recognized as a promising technology that will allow IP packets to be routed over fiber in their native connectionless nature. It is expected that the integration of efficient wavelength conversion mechanisms such as waveband routing and OBS/OPS would improve greatly the optical networking performance. This integration is still in its infancy. It requires further research studies.

10.0.3 Deploying QoS beyond the Internet domain boundaries

BGPv4, the de-facto standard protocol for inter-domain routing was designed to achieve policy routing and topology discovery, discounting the end-to-end QoS in a business environment where there was no strong financial justifications for inter-domain TE. The emerging multi-service Internet requires that both real-time and best-effort traffic be carried within and beyond a Service Provider's domain boundaries with acceptable QoS guarantees. This challenging problem has raised debates within the IP community concerning how inter-domain traffic engineering will be implemented. On one hand there is a school of thought pointing to the ability of the current generation BGP4 protocol to achieve inter-domain TE with minimum changes by using clever manipulations of the BGP decision process parameters or through minor extensions to BGP4. On the other hand, centralized strategies based on a client-server model, once abandoned by the IP community for scalability reasons , have been reconsidered for IP packets delivery beyond AS boundaries to support QoS guarantees. The wide deployment of this model in the Internet requires new mechanisms for classifying traffic flows into classes and handling these flows according to their QoS requirements. The localization of these mechanisms in the emerging client-server inter-domain architecture is another issue to be solved before QoS routing becomes effectively deployed in the inter-domain picture. This deployment requires further research studies.

Part IV

PART IV: Summary of the original work

This chapter presents a short description of the publications appended to this thesis by highlighting my contribution and presenting other author's contribution in case of co-authored papers. While Papers $1-7$ describe and evaluate the performance of QoS routing mechanisms and network control strategies to be deployed at the IP/MPLS network layer interface, Papers $8-11$ are related to the MPLS/MPλS network layer interface. While chapter 9 contain unpublished material on the failure risk group avoidance paradigm, Paper 12 is strongly related to the unpublished work on availability aware routing presented in chapters 7 and 8. It can be easily extended to include the different availability aware routing approaches.

10.1 IGP/MPLS

10.1.1 Paper 1: ICC04

A.B. Bagula, M. Botha, A.E. Krzesinski, "Online traffic engineering: the least interference optimization algorithm", Proceedings of the ICC2004 Conference, June 2004.

This paper presents a model of routing bandwidth-guaranteed tunnels in MPLS networks to reduce the interference among competing tunnels (optimality) and the rerouting upon failure (survivability) with minimal changes to the existing routing algorithms (compatibility). A novel cost based optimization scheme is proposed to improve the acceptance of routing requests while rerouting fewer LSPs upon failure at reduced computation time and using the classical constraint shortest path first (CSPF) routing algorithm. Simulation revealed the relative efficiency of the routing scheme compared to several other flow-based routing algorithms. Besides proposing the main ideas, I wrote the paper while the co-authors contributed to the discussions and simulation code.

10.1.2 Paper 2: SACJ05

A.B. Bagula, "Hybrid Traffic Engineering: From Constraint Shortest Path First to Least Path Interference", South African Computer Journal, Volume 34, Pages 2-10, June 2005.

Building upon the work done in Paper 1, this paper revisits the problem of bandwidth-guaranteed tunnels routing in MPLS networks with the objective of achieving optimality, survivability and simplicity. I proposed in this paper a hybrid cost optimization scheme combining offline approximation of the survivability objective and and on-line computation of the optimality objective. Simulation revealed that this scheme can achieve network performance similar to the model presented in Paper 1.

10.1.3 Paper 3: CIS05

A.B. Bagula and H.F. Wang, "On the Relevance of Using Gene Expression Programming in Destination-based Traffic Engineering", Lecture Notes in Computer Sciences, Volume 3801, Pages 224-229, December 2005.

This paper assesses the relevance of using Gene Expression Programming in destination-based routing. We applied GEP for the first time to achieve IGP routing in communication networks with the objective of finding a set of optimal IGP weights which minimize the maximum link utilization of a network. In this paper GEP is presented as a Memetic algorithm which performs better than classical memetic algorithms based on genetic algorithms (GA) in terms of the quality of the routing paths and connectionless based routing performance. Besides proposing the main ideas of the paper, I wrote the paper, part of the simulation code while the second author contributed to part of the simulation code.

10.1.4 Paper 4: NOMS06

A.B. Bagula, "Traffic Engineering Next Generation IP Networks Using Gene Expression Programming", Proceedings of the 2006 IEEE/IFIP Network Operations & Management Symposium, Pages 230-239, April 2006.

This paper revisits the relevance of using the genetic optimization framework in fine-tuning IGP and MPLS routing. Unlike classical Constraint Shortest Path First (CSPF) routing schemes which are based on dynamic routing metrics, static routing metrics are used in this paper to route bandwidth-guaranteed tunnels in MPLS networks. The underlying routing paradigm consisting of using a static IGP metric as a TE metric was inspired by best current practices (BCP) in provider's networks [17] but tested for the first time in this paper. Simulation revealed that using the best current practice leads to performance improvements similar to those found in Paper 1.

10.1.5 Paper 5: QOFIS04

A.B. Bagula, "Online traffic engineering: a hybrid IGP/MPLS routing approach", Lecture Notes in Computer Science, Volume 3266, Pages 134-143, September 2004.

A hybrid IGP+MPLS routing approach is presented where the requests to route bandwidth-guaranteed tunnels are classified into low bandwidth demanding (LBD) and high bandwidth (HBD) demanding requests. These requests are routed differently in an on-line setting where the LBD requests are routed over IGP computed paths while the HBD requests are carried over MPLS tunnels. While related works [28, 29, 30, 31] use an offline setting, this paper was the first to address the dualism IGP/MPLS in an on-line intra-domain setting and show that the approach could perform better than both IGP and MPLS routing. The work in [27] uses a tunnel classification and handling models similar to ours but in a trace-based inter-domain setting where cost-based optimization is discounted.

10.1.6 Paper 6: QOSIP05

Antoine B. Bagula. "Hybrid IGP+MPLS Routing in Next Generation IP Networks: An Online Traffic Engineering Model", Lecture Notes in Computer

Science, Volume 3375, Pages 325-338, February 2005.

This paper extends the work done in Paper 5 to present a cost-based optimization scheme that builds upon the stochastic property that the links of a network may have different traffic carrying probabilities to achieve different network configurations in a hybrid IGP+MPLS setting. Besides highlighting the importance of the composition rule of the cost metric, this paper shows that an additive cost metric may be obtained from a logarithmic transform of a power-based cost metric and reveals through simulation that an additive cost metric can be used to achieve the same performance as a power-based cost metric.

10.1.7 Paper 7: COMCOM06

A.B. Bagula, "Hybrid Routing in Next Generation IP Networks", Elsevier Computer Communications Volume 29, Number 7, Pages 879-892, April 2006.

Building upon network operation practices, this paper revisits the relevance of using a hybrid IGP+MPLS approach to route bandwidth guaranteed tunnels in IP networks by inflating the routing requests to achieve bandwidth protection of the links of a network. In contrast to Papers 5 and 6 , this paper uses a routing algorithm which is based on the aggregation of the HBD requests to reduce the signaling overheads resulting from setting up and tearing down the LSPs in an MPLS network.

10.2 MPLS/MPλS

10.2.1 Paper 8: JON06

A.B. Bagula and M. Botha, "On achieving LSP/LambdaSP multiplexing/separation in converged data/optical networks", OSA Journal of Optical Networking, Volume 5, Number 4, Pages 280-292, April 2006.

In this paper, the rerouting of bandwidth guaranteed tunnels (LSPs/λSPs) upon failure is presented as an NP hard multi-constrained path set finding problem. This problem is solved using a heuristic solution which is based on classical CSPF implementation. A novel cost metric using tunnel type differentiation is proposed to achieve path separation/multiplexing on the links for the active and/or backup tunnels. Our simulation experiments revealed the relevance of packing (multiplexing) active tunnels and separating the backup tunnels on links when rerouting failed tunnels. Besides proposing the main ideas, I wrote the paper and part of the simulation code while the second author contributed to part of the simulation code.

10.2.2 Paper 9: JSAC07

A.B. Bagula, "On achieving Bandwidth-aware LSP/LambdaSP multiplexing/separation in Multi-layer networks", to appear in the IEEE Journal on Selected Areas in Communications (JSAC): special issue on Traffic Engineering for Multi-Layer Networks, second quarter 2007.

This paper reconsider the rerouting of bandwidth guaranteed tunnels (LSPs/λSPs) upon failure to present an NP hard multi-constrained path set finding problem which is solved by a heuristic solution which uses classical CSPF implementation. A novel cost model novel cost metric using bandwidth aware tunnel type differentiation is proposed to achieve path separation/multiplexing on the links for high bandwidth demanding (HBD) and low bandwidth demanding (LBD) tunnels. Our simulation experiments revealed that the performance of a network is improved by packing (multiplexing) LBD tunnels and separating the HBD tunnels on links when rerouting failed tunnels.

10.2.3 Paper 10: SPIE06

A. Muchanga, A.B Bagula and L. Wosinska, "Inter-layer Communication for Faster Restoration in a 10 Gigabit Ethernet-based Network", In Proceedings of SPIE, Volume 6193, May 2006.

This paper discusses methods to improve restoration time in optical networks and proposes inter-layer communication mechanisms to be implemented in 10 Gigabit Ethernet-based networks in order to reduce the restoration time. My contribution in this paper consisted more in the discussions of the ideas and correcting the camera ready version of the paper.

10.2.4 Paper 11: OFC/NFOEC07

A. Muchanga, A.B Bagula and L. Wosinska, "On Using Fast Signalling to improve Restoration in Multi-layer Networks", Submitted to the OFC/NFOEC'07 conference.

This paper reveals the relevance of using fast signaling and inter-layer communication to reduce restoration operation times in multi-layer networks. Besides the simulation coding, I wrote the performance evaluation part of the work and contributed to the discussions and writing of other parts of the paper.

10.3 MPλS/Fiber

10.3.1 Paper 12: BoD06

A.B. Bagula and A.E. Krezinsiski, "Traffic and Network Engineering in Emerging Generation IP Networks: A Bandwidth on Demand Model", to appear in the proceedings of the First IEEE Workshop on Bandwidth on Demand, San-Francisco/USA, November 2006.

This paper presents a new hybrid TE+NE strategy where network engineering based on bandwidth trading mechanisms is used to complement traffic engineering under QoS mismatches between the available resources and the offered traffic. Building upon network's operation practices, the paper proposes the use of link under-subscription to protect some of the links of a network from being overloaded. This TE model presented in this paper can be extended to achieve availability aware routing as described in chapter 7 of this thesis. The NE model presented in this paper can also be extended to achieve availability

aware NE as described in chapter 8 of this thesis. Besides writing the paper and coding part of the simulation program, i proposed the cost and pricing models. The co-author contributed to the bandwidth trading simulation code, writing part of the paper and discussions of the ideas.

Bibliography

[1] L. Berger, "Generalized Multiprotocol Label Switching (GMPLS) Signalling Functional Description", Request for Comments, http://www.ietf.org/rfc/rfc3471.txt, 2003.

[2] S.J. Ben Yoo, "Optical-label Switching, MPLS, MPLambdaS, and GMPLS" Optical Networks Magazine, Pages 17-31, May/June 2003.

[3] K. Seppanen, "ALL IP - ALL Optical: Are Networks Converging or Diverging and Will There Ever Be All Optical Networks?", Proceedings of the IPSI-2003 conference, Pages 1-10, 2003.

[4] A. Banerjee et al, "Generalized Multiprotocol Label Switching: An overview of Routing and Management Enhancements", IEEE Communications Magazine, Volume 39, Number 1, Pages 144-150, January 2001.

[5] A. Banerjee et al, "Generalized Multiprotocol Label Switching: An overview of Signaling Enhancements and Recovery Techniques", IEEE Communications Magazine, Volume 39, Number 7 , Pages 144-151, July 2001.

[6] E. Kubilinskas, "Designing Resilient and Fair Multi-layer Telecommunication Networks", Licentiate Thesis, Department of Communication Systems, Lund Institute of Technology, Lund/Sweden, 2005.

[7] E.C. Rosen, A. Viswanathan and R. Callon, "Multiprotocol Label Switching Architecture", Request for comments, http://www.ietf.org/rfc/rfc3031.txt, 2001.

[8] D. Awduche and Y. Rekhter, "Multiprotocol Lambda Switching: Combining MPLS Traffic Engineering Control with Optical Crossconnects", IEEE Communications Magazine, Volume 39, Number 3, Pages 111-116, March 2001.

[9] C. Xin et al, "On an IP-Centric Optical Control Plane", IEEE Communications Magazine, Volume 14, Number 7, Pages 88-93, September 2001.

[10] Z. Wang, W. Crowcroft, "Quality-of-service routing for supporting multimedia applications", IEEE Journal on Selected Areas in Communications, Volume 14, Number 7, Pages 1228-1234, September 1996.

[11] E. Crawley et al, "A Framework for QoS-based Routing in the Internet", Request for comments, http:// www.ietf.org/rfc/rfc2386.txt, August 1998.

[12] A. Bagula, M. Botha and A.E. Krzesinski, "Online Traffic Engineering: The Least Interference Optimization Algorithm", Proceedings of IEEE ICC2004, June 2004.

[13] A.B. Bagula, "Hybrid Traffic Engineering: From Constraint Shortest Path First to Least Path Interference", South African Computer Journal, Volume 34, Pages 2-10, June 2005.

[14] K. Kar, M. Kodialam and T.V. Lakshman, "Minimum Interference Routing with Application to MPLS Traffic Engineering", Proceedings of IEEE INFOCOM2000, Volume 2, Pages 884-893, March 2000.

[15] M. Kodialam and T.V. Lakshman. "MPLS Traffic Engineering Using Enhanced Minimum Interference Routing: An Approach Based On Lexicographic Max-Flow", Proc. Eighth International Workshop on Quality of Service (IWQoS), Pages 105-115, June 2000.

[16] S. Suri, M. Waldvogel and P. Warkhede, "Profile-based Routing: a new framework for MPLS Traffic Engineering", Lecture Notes in Computer Science, Volume 2156, September 2001.

[17] F. Faucheur et al., "Use of interior gateway protocol (IGP) metric as a second traffic engineering(TE) metric", Request for comments, http:// www.ietf.org/rfc/rfc3785.txt, May 2004.

[18] A.B. Hadj Alouane and J.C.Bean, "A genetic Algorithm for the Multiple Choice Integer Program", Operations Research, Volume 45, Number 1, Pages 92-101, 1997.

[19] C.A.Coello, "Theoretical and Numerical Constraint-Handling Techniques used with Evolutionary Algorithms: A Survey of the State of the Art", *http : //www.citeceer.com*, 2002.

[20] A.B. Bagula and Hong F. Wang, "On the Relevance of Using Gene Expression Programming in Destination-based Traffic Engineering", Lecture Notes in Computer Sciences, Volume 3801, Pages 224-229, December 2005.

[21] A.B. Bagula, "Traffic Engineering Next Generation IP Networks Using Gene Expression Programming", Proceedings of the 2006 IEEE/IFIP Network Operations & Management Symposium, Pages 230-239, April 2006.

[22] F. Faucheur et al., "Use of interior gateway protocol (IGP) metric as a second traffic engineering(TE) metric", Request for comments, http:// www.ietf.org/rfc/rfc3785.txt, May 2004.

[23] J. Moy, "OSPF Version 2", Request for comments, http://www.ietf.org/rfc/rfc1583.txt, March 1994.

[24] A.B. Bagula, "Online traffic engineering: a hybrid IGP/MPLS routing approach", Lecture Notes in Computer Science, Volume 3266, September 2004.

[25] A.B. Bagula, "Hybrid IGP+MPLS Routing in Next Generation IP Networks: An Online Traffic Engineering Model", Lecture Notes in Computer Science, Volume 3375, Pages 325-338, February 2005.

[26] A.B. Bagula, "Hybrid Routing in Next Generation IP Networks", Elsevier Computer Communications, Volume 29, Number 7, Pages 879-892, April 2006.

[27] S. Uhlig and O. Bonaventure. "On the cost of using MPLS for interdomain traffic", Lecture Notes in Computer Science, Volume 1922, Pages 141-152, September 2000.

[28] W. Ben-Ameur et al.,"Routing Strategies for IP-Networks", Telektronikk Magazine, Volume 2, Number 3, March 2001.

[29] E. Mulyana and U. Killat, "An Offline Hybrid IGP/MPLS Traffic Engineering Approach under LSP constraints", Proceedings of the 1st International Network Optimization Conference INOC2003, October 2003.

[30] S. Koehler, A. Binzenhoefer, "MPLS Traffic Engineering in OSPF Networks- A Combined Approach", 18th ITC Specialist Seminar on Internet Traffic Engineering and Traffic Management, August-September 2003.

[31] A. Riedl, "Optimized routing adaptation in IP networks utilizing OSPF and MPLS", Proceedings of IEEE ICC2003, May 2003.

[32] A.B. Bagula & Marlene Botha, "On achieving LSP/LambdaSP multiplexing/separation in converged data/optical networks", OSA Journal of Optical Networking, Volume 5, Number 4, Pages 280-292, April 2006.

[33] M. Vigoureux et al. "Multilayer Traffic Engineering for GMPLS-Enabled Networks", IEEE Communications Magazine, Volume 43, Number 7, Pages 44-50, July 2005.

[34] Cisco Systems. http://www.cisco.com.

[35] I.W. Widjaja et al, "Online Traffic Engineering with Design-Based Routing", 15th ITC Specialist Seminar on Internet Traffic Engineering and Traffic Management, July 2002.

[36] G. Apostolopoulos et al. "QoS routing mechanisms and OSPF extensions", Request for comments, http://www.ietf.org/rfc/rfc2676.txt, August 1999.

[37] D. Katz, K. Kompella and D. Yeung, "Traffic Engineering Extensions to OSPF Version 2", Request for comments, http://www.ietf.org/rfc/rfc3630.txt, October 2002.

[38] B. Fortz and M. Thorup, "Internet Traffic Engineering by Optimizing OSPF Weights", Proceedings of IEEE INFOCOM2000, Vol 2, Pages 519-528, March 2000.

[39] L.S. Buriol et al, "A memetic algorithms for OSPF routing", Proceedings of the 6th INFORMS Telecom, Pages 187-188, 2002.

[40] C. Ferreira, "Gene Expression Programming: A New Adaptive Algorithm for solving problems", Complex Systems, Volume 13, Number 2, Pages 87-129, 2001.

[41] Network Simulator, http://www.isi.edu/nsnam/ns/

[42] R. Bush, "Complexity - The Internet and the Telco Philosophies, A Somewhat Heretical View", *NANOG/EUGENE http://psg.com/ randy/021028.nanog-complex.pdf*, October 2002.

[43] F. Faucheur et al., "Enhanced Interior Gateway Routing Protocol", http://www.Cisco.com.

[44] J.E Burns et al, "Path Selection and bandwidth allocation in MPLS networks", Performance evaluation, Volume 52, Numbers 2-3, Pages 133-152, 2003.

[45] A. Baruani, A.B. Bagula and A. Muchanga, "On Routing IP Traffic Using Single- and Multi-objective Genetic Optimization", In Proceedings Southern African Telecommunication Networks and Applications Conference, South Africa, September 2006.

[46] T.M. Fathelrahman and A.B. Bagula, "On Routing IP Traffic Using Multi-constrained Genetic Optimization with Penalty Functions", In Proceedings Southern African Telecommunication Networks and Applications Conference, South Africa, September 2006.

[47] W. Lai and D. McDysan, "Network Hierarchy and Multilayer Survivability", Request for Comments, http://www.ietf.org/rfc/rfc3386.txt, November 2002.

[48] Piet Demeester et al, "Resilience in Multilayer Networks", IEEE Communications Magazine, Volume 37, Number 8, Pages 70-76, August 1999.

[49] Didier Colle et al, "Data-centric Optical Networks and Their Survivability", IEEE Journal on Selected Areas in Communications, Volume 20, Number 1, Pages 6-20, January 2002.

[50] S. De Maesschalck, "Intelligent Optical Networking for Multilayer Survivability", IEEE Communications Magazine, Volume 40, Number 1, Pages 42-49, January 2002.

[51] B. Puype et al, "Benefits of GMPLS for Multilayer Recovery", IEEE Communications Magazine, Volume 43, Number 7, Pages 51-59, July 2005.

[52] J. Vasseur, M. Picavet and P. Demeester, "Network Recovery: Protection and Restoration of Optical, SONET-SDH, IP, and MPLS", Morgan Kaufman publishers, ISBN:0-12-715051-X, 2004.

[53] M. Pickavet et al, "Recovery in Multilayer optical Networks" Journal of Lightwave Technology, Volume 24, Number 1, Pages 122-134, January 2006.

[54] R. Sabella et al, "Strategy for Dynamic Routing and Grooming of Data Flows into Lightpaths in New Generation Network Based on the GMPLS Paradigm", Photonic Network Communications , Volume 7, Number 2, Pages 131-144, June 2004.

[55] R. Guerin, D. Williams and A. Orda, "QoS Routing Mechanisms and OSPF Extensions" , Proceedings of Globecom1997, Volume 3, Pages 1903-1908, November 1997.

[56] L. Sahasrabuddhe, S. Ramamurthy and B. Mukherjee, "Fault Tolerance in IP-Over-WDM Networking: WDM protection vs. IP restoration", IEEE Journal on Selected Areas in Communications, Volume 20, Number 1, Pages 21-33, January 2002.

[57] S. Ramamurthy and B. Mukherjee, "Survivable WDM mesh networks. Part-I Protection", Proceedings of IEEE INFOCOM1999, Volume 6, Pages 1015-1020, 1999.

[58] J. Wang, L. Sahasrabuddhe and B. Mukherjee, "Path vs. subpath vs. link restoration for fault management in IP-Over-WDM networks: performance comparisons using GMPLS control signaling", IEEE Communications Magazine, Volume 40, Number 11, Pages 80-87, November 2002.

[59] J. Zhang and B. Mukherjee, "Review of Fault Management in WDM Mesh Networks: Basic Concepts and research Challenges", IEEE Networks, Volume 18, Number 2, Pages 41-48, March 2004.

[60] A. Fumagali and L. Valcarenghi, "IP Restoration vs. WDM Protection: Is There and Optimal Choice?", IEEE Networks, Volume 14, Number 6, Pages 34-41, November/December 2000.

[61] A. Iselt, A. Kirstadter and R. Chahine, "Bandwidth Trading – A Business Case for ASON?", 11th International Telecommunications Network Strategy and Planning Symposium (Networks 2004), Vienna, Austria, Pages 63-68, June 2004.

[62] Å. Arvidsson et al, "A Distributed Scheme for Value-Based Bandwidth Re-Configuration", can be downloaded at http://www.cs.sun.ac.za/∼aek1/COE/dowloads/ four_authors.pdf, submitted 2006

[63] X. Xiao et al, "Traffic Engineering with MPLS in the Internet", IEEE Networks, Volume 16, Number 2, Pages 28-33, March/April 2000.

[64] B.A. Chiera and P.G. Taylor, "What is a Unit of Capacity Worth?", Probability in the Engineering and Informational Sciences, Volume 16, Number 4, Pages 513-522, 2002.

[65] C. Courcoubetis et al, "An Auction Mechanism for Bandwidth Allocation over Paths", 17th International Teletraffic Congress (ITC), Dec 2001.

[66] A. Muchanga, A.B Bagula and L. Wosinska, "Inter-layer communication for faster restoration in a 10Gigabit Ethernet based network", Proceedings of The SPIE Photonics Europe, Strasbourg France, April 2006.

[67] A. Muchanga, A.B Bagula and L. Wosinska, "On Using Fast Signalling to improve Restoration in Multi-layer Networks", Submitted to the OFC/NFOEC'07 conference.

[68] M. Pioro and D. Medhi, "Routing, Flow, and Capacity Design in Communication and Computer Networks", Morgan Kaufmann, ISBN 0-12-557189-5, 2004.

[69] R. Bhandari, "Survivable Networks- Algorithms for Diverse Routing", Springer, First edition, ISBN: 0792383818, 1999.

[70] P. Francois et al, "Achieving Sub-second IGP Convergence in large IP networks", SIGCOMM Comput. Commun. Rev., Volume 35, Number 3, Pages 35-44, 2005.

[71] M. Held et all, "Consideration of connection availability optimization in optical networks", Proceedings of DRCN2003, Pages 173-180, October 2003.

[72] B. Mikac, "Modelling and Availability Evaluation of Optical WDM Networks". COST 270 Meeting, GRAZ, April 8-9, 2002.

[73] M. Tornatore, G. Maier and A. Pattavina, "Availability Design of Optical Transport Networks", IEEE Journal on Selected Areas in Communications, Volume 23, Number 8, Pages 1520-1532, August 2005.

[74] M. To and P. Neusy, "Unavailability Analysis of Long-Haul Networks", IEEE Journal on Selected Areas in Communications, Volume 12, Number 1, Pages 100-109, January 1994.

[75] Y.G Huang et al, "A Generalized Protection Framework Using a New Link-State Availability Model for Reliable Optical Networks", IEEE Journal of Lightwave Technology, Volume 22, Number 11, Pages 2536-2547, November 2004.

[76] X. Yang, "Availability-differentiated service provisioning in free-space optical access networks", OSA Journal of Optical Networking, Volume 4, Number 7, Pages 391-399, July 2005.

[77] C. Davis, I. Smolyaninov and S. Milner, "Flexible Optical Wireless Links and Networks", IEEE Communications Magazine, Volume 41, Number 3, Pages 51-57, March 2003.

[78] S. Bloom et al, "Understanding the performance of free-space optics", OSA Journal of Optical Networking, Volume 2, Number 6, Pages 178-200, June 2003.

[79] A.B. Bagula and A.E. Krzesinski, "Traffic and Network Engineering: A Bandwidth on Demand Model", to appear in the proceedings of the first IEEE worshop on Bandwidth on Demand, sand Francisco/USA, November 2006.

[80] P. Sebos et al, "Effectiveness of shared risk link group auto-discovery in optical networks", Proceedings of OFC'02, 2002.

[81] E. Bouillet et al, "Stochastic approaches to compute shared mesh restored lightpaths in optical network architectures", Proceedings of INFOCOM'02, Pages 801-807, 2002.

[82] Y. Liu, D. Tipper and P. Siripongwutikorn, "Approximating optimal spare capacity allocation by successive survivable routing", Proceedings of INFOCOM'01, Pages 699-708, 2001.

[83] D. Andersen et al, "Resilient overlay networks", Proceedings of SOSP 2001, October 2001.

[84] O. Crochat, J.Y. Le Boudec and O. Gerstel, "Protection Interoperability for WDM Optical Networks", IEEE/ACM Transactions on Networking, Volume 2, Number 3, Pages 384-395, June 2000.

[85] W. Cui, I. Stoica and R.H. Katz, "Backup Path Allocation Based on A Correlated Link Failure Probability Model IN Overlay Networks", Proceedings of IEEE ICNP 2002, 2002.

[86] Charles J. Colbourn, "The Combinatorics of Network Reliability", Oxford University Press ISBN 0-19-5404920-9, 1987.

[87] V.O.K Li, J.A Silvester, "Performance Analysis of Networks with Unreliable Components", IEEE Transactions on Communications, Volume 32, Number 10, Pages 1105-1110, October 1984.

[88] C.L. Yang and P. Kubat, "Efficient Computation of Most Probable States for Communication Networks with Multimode Components", IEEE Transactions on Communications, Volume 37, Number 5, Pages 535-538, 1989.