UNIVERSITAT POLITÈCNICA DE CATALUNYA (UPC)

Advanced Routing Mechanisms in ASON/GMPLS Networks

by

Fernando Agraz Bujan

A thesis submitted in fulfillment for the
degree of Doctor of Philosophy

in the

Optical Communications Group (GCO)
Signal Theory and Communications Department (TSC)

Advisors: Dr. Salvatore Spadaro and Dr. Jordi Perelló Muntan

July 2012

Oficina de Doctorat
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Acta de qualificació de tesi doctoral

Curs acadèmic:

Nom i cognoms

DNI / NIE / Passaport

Programa de doctorat

Unitat estructural responsable del programa

Resolució del Tribunal

Reunit el Tribunal designat a l'efecte, el doctorand / la doctoranda exposa el tema de la seva tesi doctoral titulada

_____

_____.

Acabada la lectura i després de donar resposta a les qüestions formulades pels membres titulars del tribunal, aquest atorga la qualificació:

☐ APTA/E ☐ NO APTA/E

(Nom, cognoms i signatura) President/a		(Nom, cognoms i signatura) Secretari/ària
(Nom, cognoms i signatura) Vocal	(Nom, cognoms i signatura) Vocal	(Nom, cognoms i signatura) Vocal

_____, _____ d'/de _____ de _____

El resultat de l'escrutini dels vots emesos pels membres titulars del tribunal, efectuat per l'Escola de Doctorat, a instància de la Comissió de Doctorat de la UPC, atorga la MENCIÓ CUM LAUDE:

☐ SI ☐ NO

(Nom, cognoms i signatura) Presidenta de la Comissió de Doctorat	(Nom, cognoms i signatura) Secretària de la Comissió de Doctorat

Barcelona, _____ d'/de _____ de _____

Contents

List of Figures

List of Tables

List of Algorithms

Abbreviations and Symbols

Abbreviations

AS	Autonomous System
ASON	Automatically Switched Optical Network
ATM	Asynchronous Transfer mode
BE	Best Effort
BER	Bit Error Rate
BGP	Border Gateway Protocol
BP	Blocking Probability
BRPC	Backward Recursive Path Computation
CAPEX	Capital Expenditures
CC	Connection Controller
CDF	Cummulative Distribution Function
CCDF	Complementary Cummulative Distribution Function
CCI	Connection Controller Interface

CP	Control Plane
CR-LDP	Constrained Routing Label Distribution Protocol
CSPF	Constrained Shortest Path First
CWDM	Coarse Wavelength Division Multiplexing
DCN	Data Communications Network
DWDM	Dense Wavelength Division Multiplexing
EGP	Exterior Gateway Protocol
E-NNI	External Network-to-Network Interface
EON	European Optical Network
FA	Forwarding Adjacency
FPGA	Field Programmable Gate Array
FSC	Fiber Switching Capable
GbE	Giga-bit Ethernet
GMPLS	Generalized Multi-protocol Label Switching
IA-RWA	Impairment-Aware Routing and Wavelength Assignment
ICT	Information and Communication Technologies
IETF	Internet Engineering Task Force
IGP	Interior Gateway Protocol
I-NNI	Internal Network-to-Network Interface
ION	Intelligent Optical Network
IP	Internet Protocol

IS-IS-TE	Intermediate System to Intermediate System with TE Extensions
ITU(-T)	International Telecommunication Union(-Telecom Standardization)
LMP	Link Management Protocol
LRM	Link Resource Manager
LoL	Loss of Light
LSC	Lambda Switching Capable
LSP	Label Switched Path
MIB	Management Information Base
MPLS	Multi-protocol Label Switching
ND	Node Degree
NE	Network Element
NMI-A	Network Management Interface for the ASON control plane
NMI-T	Network Management Interface for the ASON Transport plane
NMS	Network Management System
NNI	Network-to-Network Interface
NRZ	Non-Return to Zero
OBS	Optical Burst Switching
OCC	Optical Connection Controller
OCS	Optical Circuit Switching
OIF	Optical Internetworking Forum
OLSA	Opaque Link State Advertisement

OPEX	Operational Expenditures
OPS	Optical Packet Switching
OSNR	Optical Signal to Noise Ratio
OSPF-TE	Open Shortest Path First with TE Extensions
OTN	Optical Transport Network
OXC	Optical cross Connect
PCE	Path Computation Element
PCEP	Path Computation Element Communication Protocol
PCF	Path Computation Flooding
PPD	Physical Parameters Database
PPP	Point-to-Point Protocol
PSC	Packet Switching Capable
QoS	Quality of Service
QoT	Quality of Transmission
RA	Routing Area
RC	Routing Controller
RSVP-TE	Resource reSerVation Protocol with TE Extensions
RWA	Routing and Wavelength Assignment
ROADM	Reconfigurable Optical Add/Drop Multiplexer
SCN	Signaling Communications Network
SDH	Synchronous Digital Hierarchy

SONET	Synchronous Optical NETwork
TCP	Transmission Control Protocol
TDMC	Time Division Multiplexing Capable
TE	Traffic Engineering
TED	Traffic Engineering Database
TLV	Type Length Value
TP	Transport Plane
UDP	User Datagram Protocol
UNI	User to Network Interface
VoIP	Voice over IP
WC	Wavelength Conversion
WCC	Wavelength Continuity Constraint
WDM	Wavelength Division Multiplexing
WSON	Wavelength Switched Optical Networks
WSS	Wavelength Selective Switch

Symbols

C_{FA}	Forwarding Adjacency creation cost
d	Destination node in path computation
D	Number of single domains in a multi-domain network

E	Set of WDM links in the network graph G
E^i	Set of WDM links in the domain network graph G^i
G	Multi-domain network graph
G^i	Network graph for domain i
h	Tunable parameter to foster/penalize the use of O/E ports
H	Number of hops spanned by a Forwarding Adjacency
L	Network load
p_H	Probability that an incoming demand has H hops
s	Source node in path computation
v_d^j	Destination node belonging to domain j
v_m^i	m-st node in domain i
v_s^i	Source node belonging to domain i
V	Set of nodes in the network graph G
V^i	Set of nodes in the domain network graph G^i
δ_{mn}^{ij}	Inter-domain link connecting nodes v_m^i and v_n^j
ΔL	Network load increment
λ	Mean connection arrival rate per unit of time
μ	Mean connection holding time
γ^i	Number of nodes in G^i

Resum

Les xarxes de comunicacions actuals segueixen un model distribuït de quatre capes que permet la transmissió de tràfic de dades heterogeni de forma eficaç sobre la mateixa infraestructura de transport. En aquesta arquitectura, la capa superior correspon a la capa IP, seguida per dues capes intermèdies que són la capa ATM (Asynchronous Transfer Mode) i la capa SDH (Synchronous Digital Hierarchy) i, finalment, la capa WDM (Wavelength Division Multiplexing). Encara que les capes ATM i SDH proporcionen beneficis com qualitat de servei i protecció i restauració, també comporten una sèrie de problemes com interaccions complexes entre elles i un excés d'informació de control. Això fa necessari evolucionar cap a una arquitectura més lleugera on els fluxos IP es puguin enviar directament a través de la capa WDM. En aquest nou model IP/WDM les funcionalitats proporcionades inicialment per les capes ATM i SDH es traspassen al domini òptic.

Des d'un punt de vista operacional, la naturalesa estàtica del model en quatre capes fa que els temps de provisió de serveis siguin molt llargs (d'hores o dies) i, per tant, incompatibles amb els patrons de tràfic IP que són altament dinàmics. Per superar aquestes limitacions, la ITU-T va proposar l'arquitectura ASON (Automatically Switched Optical Network) que utilitza un pla de control per proporcionar connexions en la xarxa òptica de forma ràpida i fiable. A més, la tecnologia GMPLS (Generalized Multi-Protocol Label Switching), definida per la IETF, apareix com el model més prometedor per implementar les funcionalitats del pla de control d'ASON.

Tot i que el paradigma ASON/GMPLS suposa un avenç significatiu per aconseguir un model flexible i fàcil de mantenir, encara presenta una sèrie de limitacions que es tracten en aquesta tesi.

Primer, degut a la granularitat de longitud d'ona oferida per la tecnologia de transport WDM, els circuits òptics establerts queden infrautilitzats quan transmeten fluxos de dades de client de granularitat inferior. Per aquest motiu una solució multi-capa, en la que aquests fluxos de client puguin ser agregats a la capa IP i després transportats pel mateix circuit en la capa òptica, sembla la més apropiada. A més, la tecnologia GMPLS resulta molt adient per implementar el pla de control en aquesta arquitectura multi-capa, ja que permet la gestió de diferents tecnologies de commutació de forma integrada. En aquest context, aquesta tesi descriu el desenvolupament i validació experimental d'una xarxa de transport òptic multi-capa controlada per un pla de control GMPLS.

En segon lloc, encara que la transmissió purament òptica basada en WDM proporciona un gran ample de banda amb un cost reduït, també és molt sensible als efectes negatius que afecten al senyal òptic transmès i, per tant, a la informació transportada. En aquest context esdevé necessari un pla de control que sigui capaç de manegar la informació associada a la capa òptica. Aquesta tesi presenta i avalua de forma experimental diferents solucions per implementar un pla de control GMPLS amb gestió dels impediments físics associats a la capa òptica.

Finalment, degut al creixement de les infraestructures de xarxa, aquestes s'acostumen a particionar ja sigui per raons tecnològiques, administratives o d'escalabilitat. A més, la interconnexió entre dominis de diferents operadors, necessària per proporcionar connectivitat a llarga distància, implica problemes de confidencialitat entre operadors. Tot això dóna lloc a les xarxes multi-domini. En xarxes òptiques multi-domini controlades amb GMPLS, aquest particionat afecta a la provisió de serveis extrem a extrem, que es veu afectada pel limitat intercanvi d'informació causat per la necessitat de satisfer els requeriments esmentats. El tercer objectiu d'aquesta tesi consisteix en proporcionar mecanismes de provisió de connexions extrem a extrem, escalables i efectius, en aquest escenari multi-domini.

Summary

Current network infrastructures are supported on a layered model whereby heterogeneous data traffic can be seamlessly transported. In this architecture, where client/server relationships are established between adjacent layers, there exists an IP layer on top, two intermediate Asynchronous Transfer Mode (ATM) and Synchronous Digital Hierarchy (SDH) layers, and a Wavelength Division Multiplexing (WDM) layer at the bottom. Despite the benefits provided by the intermediate ATM and SDH layers (i.e., QoS and resilience), the complex interaction between them and the introduced overhead motivate an evolution towards a lighter model, where IP flows are directly sent through the WDM layer. In this new IP/WDM model the functionalities formerly provided by the ATM and SDH layers are moved to the optical domain.

From an operational perspective, the static nature of current transport networks, which leads to long service provisioning times (i.e., hours or days), becomes incompatible with the dynamic patterns associated to the prevalent IP traffic. To overcome these limitations, the ITU-T proposed the Automatically Switched Optical Network (ASON) architecture, which utilizes a control plane to provide fast and reliable lightpaths within the optical transport network. In addition, the Generalized Multi-Protocol Label Switching (GMPLS) model defined by the IETF appears as the most promising technology to implement the functionalities of the ASON control plane.

Although being a significant advance towards flexible and easy-to-maintain transport network architectures, the ASON/GMPLS paradigm still presents three major open issues that are addressed in this thesis.

First, due to the coarse granularity offered by the WDM transport technology, established connections remain underused when sub-wavelength client flows must be transmitted. In light of this, a multi-layer approach, where different higher level data flows are aggregated in the IP layer and transmitted over the same optical path, appears as the most valid solution. Besides, the GMPLS technology is more than indicated to implement the control plane in this multi-layer architecture, since it allows the management of different switching technologies in an integrated way. In this regard, this thesis reports the experimental design and validation of a GMPLS-controlled multi-layer optical transport network.

Second, even though pure optical transmission enabled by WDM provides high bandwidth in a cost-effective way, it is also very sensitive to the physical layer impairments that affect to the transmitted optical signal and, thus, to the transported data. Hence, a control plane capable of managing physical layer information to provide feasible lightpaths becomes a must. This thesis introduces and experimentally evaluates different architectural solutions to implement an impairment-aware GMPLS-based control plane.

Finally, as transport network infrastructures grow, they are typically segmented into domains due to administrative, technological, or scalability reasons. Furthermore, the interconnection of network infrastructures managed by different operators is mandated by the need to provide long distance connectivity. Here, confidentiality and reliability concerns become of paramount importance. All this justifies the need for multi-domain networks. In GMPLS-controlled multi-domain optical transport networks, this partitioning affects to end-to-end service provisioning, which is hindered by the reduced information exchange between domains resulting from the fulfillment of segmentation criteria. As the third goal of this thesis, different mechanisms for scalable and effective end-to-end service provisioning are proposed in this multi-domain scenario.

Dedicated to my parents, my brother and Maria.

Special thanks to my supportive advisors
Jordi Perelló and Salvatore Spadaro.

Chapter 1

Introduction

Current network infrastructures are supported on a layered model whereby heterogeneous data traffic can be seamlessly transported. In this architecture, where client/server relationships are established between adjacent layers, there exist an IP layer on top, two intermediate Asynchronous Transfer Mode (ATM) and Synchronous Digital Hierarchy (SDH) layers, and a Wavelength Division Multiplexing (WDM) layer at the bottom. Despite the benefits provided by the intermediate ATM and SDH layers (i.e., QoS and resilience), the complex interaction between them and the introduced overhead motivate an evolution towards a lighter layered architecture, where IP flows are directly sent through the WDM layer. In this new IP/WDM model the functionalities formerly provided by the ATM and SDH layers are moved to the optical domain.

From an operational perspective, the static nature of current transport networks, which leads to long service provisioning times (i.e., hours or days), becomes incompatible with the dynamic patterns associated to the prevalent IP traffic. To overcome these limitations, the International Telecommunication Union (ITU-T) proposed the Automatically Switched Optical Network (ASON) architecture, which utilizes a control plane to provide fast and reliable lightpaths within the optical transport network. In this context, the Generalized Multi-Protocol Label Switching (GMPLS) model defined by the Internet Engineering Task Force (IETF) appears as the most promising technology to implement the functionalities of the ASON control plane.

Although being a significant advance towards flexible and easy-to-maintain transport network architectures, the ASON/GMPLS paradigm still presents three major open issues. First, due to the coarse granularity offered by the WDM transport technology, established connections remain underused when sub-wavelength client flows must be transmitted. In light of this, a multi-layer approach, where different higher level data flows are aggregated in the IP layer and transmitted over the same optical path, appears as the most valid solution. Moreover, the GMPLS technology arises as the best positioned candidate to implement the control plane in this multi-layer architecture, since it allows the management of different switching layers in an integrated way. Second, even though pure

optical transmission enabled by WDM provides high bandwidth in a cost-effective way, it is also very sensitive to the physical layer impairments that affect to the transmitted optical signal and, thus, to the transported data. Hence, a control plane capable of managing physical layer information to provide feasible lightpaths becomes a must. Third, as transport network infrastructures grow, they are typically segmented into domains due to administrative, technological, confidentiality, or scalability reasons, giving birth to multi-domain networks. In GMPLS-controlled multi-domain optical transport networks, this division affects to end-to-end service provisioning, which is hindered by the reduced information exchange between domains resulting from the fulfillment of segmentation criteria. Therefore, mechanisms for effective end-to-end service provisioning are needed in multi-domain scenarios.

These open issues are addressed along this thesis, which devotes an individual chapter to each one of them. In this process, each chapter introduces the architectural solution to face the stated problem, together with a review of the related standards and previous work. Next, the contributions (i.e., proposed extensions and mechanisms) to solve the problem are provided along with the experimental validation. The main achievements are highlighted in the last section of each chapter.

1.1 Overview of the thesis

This thesis is structured into two introductory chapters (2 and 3) and three chapters dedicated to the research that has been conducted (chapters 4, 5 and 6). Chapter 2 provides some background on next-generation optical networks. Chapter 3 is devoted to the description of the ASON/GMPLS CARISMA test bed, which is the scenario where the deployments and experiments that compose this thesis have been carried out. Chapters 4, 5 and 6 address different open issues on ASON/GMPLS networks.

Chapter 2, *Overview of optical transport networks*, surveys enabling technologies to next-generation optical transport network deployment. It reviews point-to-point WDM transmission systems, ring and meshed all-optical networks, which are based on Optical-Add-Drop Multiplexers (OADMs) and Optical Cross Connects (OXCs) respectively, the ASON standard proposed by the ITU-T for dynamic wavelength-routed optical networks and, finally, future OPS and OBS switching models. Being ASON/GMPLS networks the main target of this thesis, a section has been dedicated to explain GMPLS fundamentals. The open issues presented by such ASON/GMPLS paradigm, addressed in forthcoming chapters, are finally introduced.

Chapter 3, *Thesis scenario: The CARISMA test bed*, introduces the ASON/GMPLS CARISMA network, which is the framework where the experimental evaluation of the contributions of this thesis has been conducted. The test bed is also used in this chapter as a driver for an extensive explanation of the GMPLS protocols and functionalities.

Chapter 4, *Design and validation of a GMPLS-controlled multi-layer optical transport network*, concentrates on the efficient utilization of the network resources in ASON/GMPLS networks. Being OCS networks essentially, the wavelength granularity provided by ASON/GMPLS networks leads to a tremendous waste of bandwidth when sub-wavelength client flows have to be transported. In this context, multi-layer networks with traffic grooming capabilities are identified as a solution to offer an improved usage of network resources. As the main contributions of this chapter, GMPLS protocol and functional extensions are proposed at the control plane level to enable an integrated management of a GMPLS-controlled multi-layer optical transport network.

Chapter 5, *Impairment-aware routing over a GMPLS-controlled optical transport network*, focuses on the impairment awareness in ASON/GMPLS networks. All-optical networks are very sensitive to the physical effects associated to the WDM transmission technology. In this context, different architectural solutions are proposed and evaluated in this chapter to provide an ASON-based transparent optical transport network with an impairment-aware GMPLS control plane.

Chapter 6, *PCE-based routing in GMPLS-controlled multi-domain optical transport networks*, studies the GMPLS-based end-to-end path computation in a multi-domain optical network. Scalability and confidentiality requirements in multi-domain networks lead to reduced routing information exchange between domains. This lack of topological information hinders the operation of the path computation entities and prevents an optimal end-to-end service provisioning, which results into an inefficient network performance. In view of this, end-to-end path computation mechanisms are proposed and evaluated in the framework of a GMPLS-controlled multi-domain optical network.

Finally, chapter 7 summarizes this thesis and opens future lines of research.

Chapter 2

Overview of optical transport networks

This chapter gives an overview of technologies enabling new generation optical transport network deployment. To begin, a quick look at optical network evolution is firstly taken. Therein, several evolutionary stages are identified, ranging from former static optical networks to the expected optical packet switching, passing through the current wavelength switching models. The focus is then put on dynamic wavelength-routed optical transport networks, specifically on those deployed according to the ITU-T ASON architecture and provided with a GMPLS control plane as defined by the IETF. Providing high bandwidth capacity by means of a mature optical transmission technology, this network architecture still presents a set of challenges which are outlined in the third section of the chapter and faced throughout this thesis.

2.1 Evolution of optical transport networks

Since late nineties, data traffic has experienced an exponential growth as a consequence of the Internet explosion. The evolution of end-users' requirements moving from voice-centric low bandwidth-consuming services to IP-centric high bandwidth-demanding applications (such as real time services, HDTV, etc.) has forced a transformation of the transport network schemes towards models capable to efficiently cope with such amount of traffic. This section gives a concise detail of such progression from the traditional architectures to a newer wavelength switched paradigm, which embraces the scope of this thesis, keeping in mind that the ultimate goal in optical transport network evolution, namely the Optical Packet Switching (OPS), is still far in time. Figure 2.1 illustrates the expected time-line in the optical transport evolution and is used as a road map throughout this section.

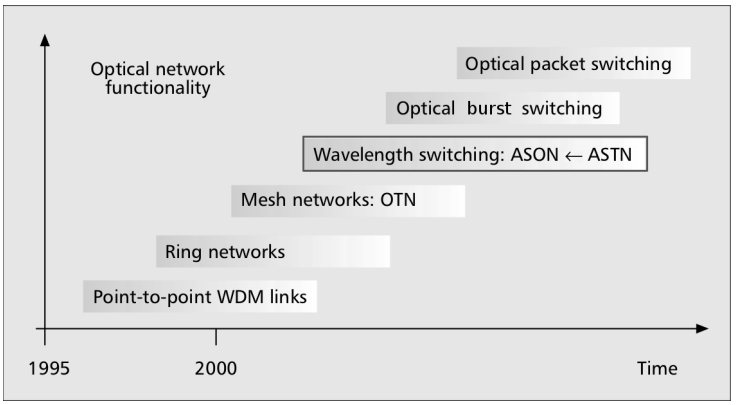


FIGURE 2.1: Optical network evolution road map.

2.1.1 First-generation optical transport networks

Following the evolution depicted in Figure 2.1, the first approach to optical transport networks was done by means of static point-to-point Wavelength Division Multiplexing (WDM) optical links, which interconnected the different nodes in the network. This optical transport occupied the lower level of an evolved four-layer network architecture capable to manage the increasing IP data traffic demands while keeping traditional voice-centric traffic support. In such architecture, IP flows are encapsulated in Asynchronous Transfer Mode (ATM, [RFC2225]) cells that, in turn, are added into Synchronous Optical NETwork/Synchronous Digital Hierarchy (SONET/SDH, [G.707]) frames and finally transported over an optical transport network. Functionally, IP packets transport application data, ATM provides traffic engineering (TE) and quality of service (QoS), SONET/SDH is responsible for network resilience, and WDM is used for pure optical transmission.

Regarding the transport technology, a single digital signal was initially transmitted within one fiber. Nonetheless, the SONET/SDH architecture took advantage of the emergent WDM transmission technology, which offers high bandwidth capacity, to interconnect the SONET/SDH Add/Drop Multiplexers (ADMs) through point-to-point links. The WDM technology splits the optical transmission spectrum into a number of non-overlapping wavelength (or frequency) bands, each one supporting a single communication channel, that are transported within the same fiber [Muk00]. Different WDM (e.g., Dense or Coarse) technologies can be found depending on the spacing between the channels to be transmitted.

Hence, increasing the bandwidth capacity thanks to the WDM technology and being capable to manage heterogeneous data traffic, the IP/ATM/SDH/WDM model was widely accepted as the one to be the new generation network infrastructure. However, this model presents multiple drawbacks such as overlapped functionalities, complex inter-working between layers and huge overhead. Besides, although data transmission through the links is optical (e.g., by means of WDM), data has to be electrically processed at each node

so upper layers' tasks (such as monitoring, QoS providing, etc.) can take place. To do this, optical-electrical (O/E) conversions have to be performed at each node for incoming traffic processing, and electrical-optical (E/O) conversions are needed for outgoing traffic. Such an inefficient data processing increases the network deployment costs (i.e., capital expenditures, CAPEX) due to the high number of required transmitter-receiver pairs. Moreover, the advantages of high optical transmission rates can not be fully exploited since the slower electronic data processing acts as a bottleneck of the overall transmission capacity. Last but not least, due to the static nature of this network architecture, operational aspects become a limitation too. Specifically, connection provisioning is made in a fully-manual way so each node has to be configured by an operator leading to long provisioning times (e.g., days) as well as increased operational expenditures (OPEX).

The aforementioned drawbacks can be redefined as the requirements for an evolved and more efficient optical transport network. In summary, reducing the data processing in intermediate nodes as well as moving towards a more efficient, less layered and easy to maintain network scheme arise as the most important action points.

2.1.2 Wavelength-routed optical transport networks

Wavelength-routed optical transport networks (also known as Optical Circuit Switching networks, OCS) are the result of the technological improvements in the transport plane added for overcoming the limitations presented by the previously illustrated first generation optical networks.

The optical transport plane evolution came by means of the Optical Add Drop Multiplexer (OADM), a device capable of inserting/extracting certain wavelengths to/from the network and let the remainder pass through it optically. In this way, end-to-end optical circuits (or lightpaths) are established keeping the data transmission in the optical domain from source to destination, thus avoiding O/E/O conversions of the traffic not terminating in the node. Therefore, the OADM enabled the transition from the legacy SONET/SDH-based ring networks to the all-optical OADM-rings (second stage in Figure 2.1). Besides, in order to introduce a certain degree of dynamism into OADM-rings, Reconfigurable OADM devices (ROADMs), which allow remote configuration of the wavelengths to be inserted/extracted/bypassed, were developed. In the physical plane, the transition from SONET/SDH to OADM rings also means a conceptual evolution from opaque networks (i.e., networks having hop-by-hop O/E/O conversion) to transparent networks where the signal is optically transmitted from source to destination.

From an architectural perspective, as the increasing IP-centric traffic model started to prevail over the voice-centric one, efforts were focused on obtaining a simplified model to substitute the overlayed IP/ATM/SDH/WDM scheme. In this context, a two-layer model where IP flows could be directly conveyed over a WDM transport plane appeared to be the most efficient solution. Nonetheless, such an evolution should be done without

losing the benefits offered by the ATM and SONET/SDH layers. In this regard, the Optical Transport Network (OTN) architecture introduced by the ITU-T in [G.872] provides flexible handling of client signals as well as switching, multiplexing and resilience capabilities directly to the optical domain.

In addition, a further step was done in wavelength-routed optical networks thanks to the Optical Cross-Connect device (OXC, Figure 2.2), a device capable of switching a wavelength from any input port to any output one bearing, in this way, the multi-degree optical switch. These optical devices, equipped with the novel OTN technology, paved the way to mesh optical networks providing more flexibility and better resource utilization than the preceding OADM-based ring topologies. It is worth noting at this point that if O/E ports are removed from the network, and in the absence of wavelength converters, the so-called *wavelength continuity constraint* (WCC) has to be fulfilled when establishing an end-to-end lightpath. This means that the same wavelength must be used in all the segments composing the route.

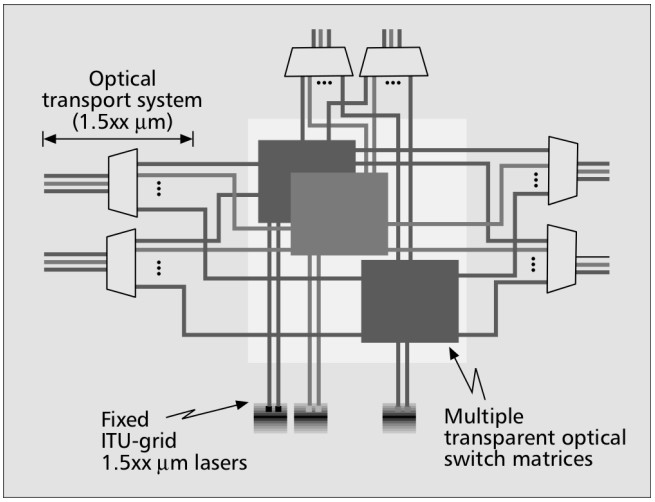


FIGURE 2.2: Optical Cross-Connect (OXC) architecture [ELW+04].

Although providing improved transport, wavelength-routed optical transport networks still lack a completely automated control and management. Hence, connection provisioning and releasing as well as monitoring is done manually, usually from a centralized Network Management System (NMS). Such a static model does not fit the network requirements claimed by the bursty traffic patterns, which demand flexibility and agility in connection provisioning, as well as QoS, which requires effective protection and restoration mechanisms. In light of this, a technology-independent generic functional architecture for transport networks is defined in [G.805]. Therein, a dynamic control plane situated over the transport network is proposed. Besides, [G.807] describes the requirements for the control plane of Automatically Switched Transport Networks (ASTN), devoted to set up and release connections across the transport network, but leaving to other recommendations the technical details required to implement these networks for particular transport technologies. Concretely in optical transport networks, the Automatically Switched Optical Network (ASON) architecture ([G.8080]) arises as the standard proposal to overcome

the drawbacks presented by static wavelength-routed optical transport networks. As a matter of fact, [G.807] was withdrawn after being merged with [G.8080]. For these reasons, ASON takes up the fourth stage in the optical transport network evolution depicted in Figure 2.1.

2.1.3 Dynamic wavelength-routed optical networks: The ASON architecture

The Automatically Switched Optical Network architecture is defined by the ITU-T in its recommendation [G.8080]. There, the utilization of a control plane is proposed to provide fast and reliable services within the optical transport network. Taking as a goal the development of an architecture that is independent from the data and control planes technologies, the ASON definition is expressed in terms of functional components, thus maintaining a high level of abstraction. As will be further detailed, this leverages the introduction of GMPLS as a valid implementation for the ASON control plane.

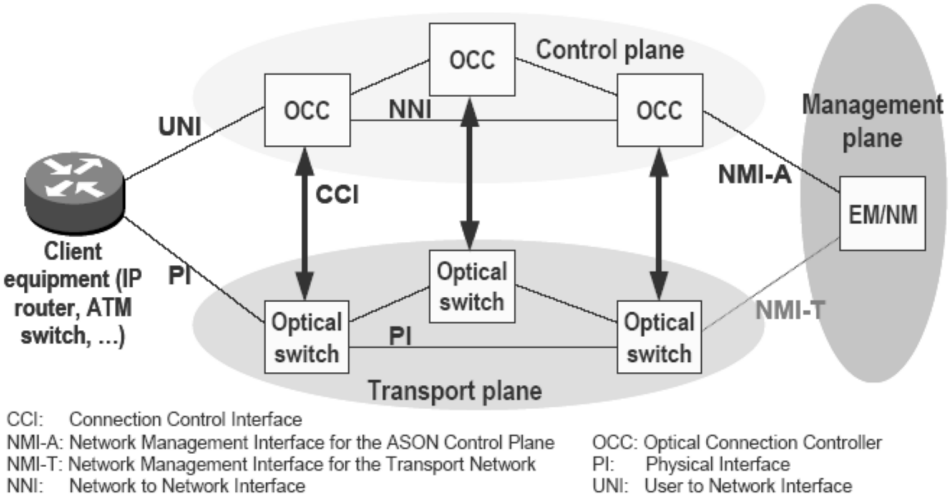


FIGURE 2.3: The ASON architecture: management, control and transport planes.

From an architectural perspective, ASON defines three planes, namely transport, control and management, and the interactions between them (Figure 2.3). The transport plane includes all the hardware (such as optical switches and fibers) that composes the optical transport network whereby data traffic is transmitted. The control plane is the responsible for providing intelligence to the network. To this end, Optical Connection Controllers (OCCs) are deployed to manage the elements in the optical transport network (i.e. transport plane). Each OCC is composed of a set of modules which provide routing, signaling and resource management capabilities, thus allowing automated connection provisioning, maintenance and release. The communication between OCCs, needed for routing and signaling, is supported over the Network to Network Interface (NNI). Furthermore, a communication interface is required between each OCC and its controlled

network elements (NEs). The Connection Controller Interface (CCI) implements this communication, which enables the configuration and release of optical connections from the control plane. Finally, the management plane aims at the management of the other two planes. Moreover, the manual setup, tear-down and supervision of optical channels is enabled from the management plane bypassing the control plane. To this end, ASON defines two Network Management Interfaces (NMI). While, the NMI-A interface interconnects the management and the control planes, the NMI-T interconnects the management plane and the transport one.

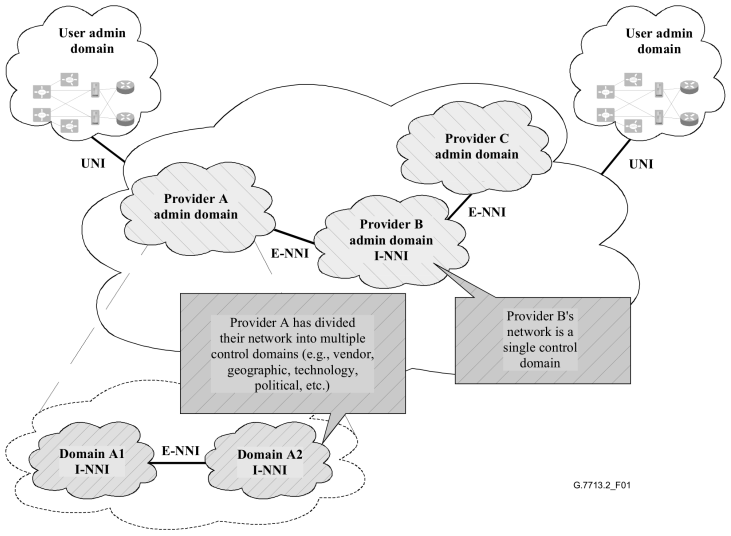


FIGURE 2.4: Overall view of control plane partition [G.7713.2].

From a functional point of view, ASON contemplates splitting the optical network due to administrative, geographical or technological reasons. This division can be horizontal or vertical. The former, called *partitioning*, implies creating different domains, where a domain represents a collection of network entities that are grouped for a particular purpose (Figure 2.4). To enable the interaction between partitions, the *reference point* concept is introduced, which defines an abstract functional interface between components of the network. There are three reference points defined in [G.8080]: the User to Network Interface (UNI), the Internal NNI (I-NNI) and the External NNI (E-NNI). The UNI is used to request a service from the network. It is an untrusted interface, so only information required by the service request is exchanged. The I-NNI is a trusted reference point used within a domain to exchange both routing and signaling information. The E-NNI interconnects different ASON domains. Being a semi-trusted reference point, only a limited degree of information is exchanged through the E-NNI, which is intended to provide scalability as well as confidentiality between operators. The vertical division or *layering* responds to the network service provisioning model defined in ASON. In this *overlay* model (Figure 2.5), network layers follow a client/server relationship where the client network requests a service to the server one through the UNI interface [KLXF08]. It is worth mentioning that partitioning and layering can co-exist within the same scenario.

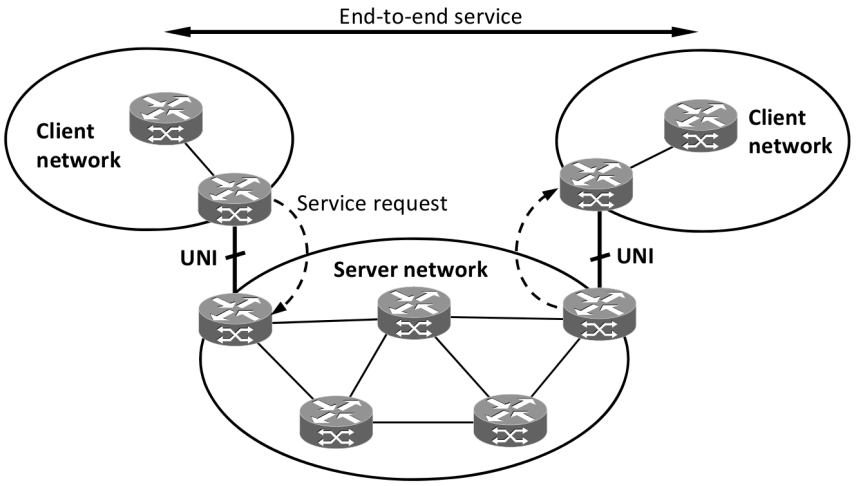


FIGURE 2.5: ASON overlay model.

Connection provisioning in ASON can be performed in three different ways. First, the traditional manually configured permanent connections are supported. These connections are established using the management plane by configuring each optical network device through the NMI-T interface. In addition, two novel transport services are provided in ASON: soft-permanent and switched connections. In the former, the connection request is processed by the management plane which contacts the control plane through the NMI-A. The control plane, in turn, automatically calculates the route for the requested connection and starts the signaling process, in which OCCs configure the optical devices composing that route by means of the CCI. In switched connections, the request is sent from the client directly to the control plane by means of the UNI. Then, the same mechanism as in soft-permanent connections is performed.

2.1.4 Third-generation optical transport networks

Although ASON holds the focus of this thesis as a short to mid-term solution to cope with current traffic requirements, Optical Packet Switching (OPS) is expected to be the ultimate paradigm in optical transport networks. In brief, OPS [Yoo06] aims at bringing the packet-switching functionalities offered by electronic routers to the optical domain. In this way, IP packets are mapped into optical packets each one having its own header and payload. Then, these packets are processed entirely in the optical domain by OPS routers in a hop-by-hop basis from source to destination. Note that most of OPS proposals assume full wavelength conversion capability in the nodes. Hence, optical packets can be forwarded to the next hop in whatever wavelength available, thus achieving a better usage of the available bandwidth than OCS networks. Therefore, OPS adapts better than OCS to the bursty IP traffic dynamics.

However, OPS presents two important technological limitations that prevent its implementation. First, the absence of optical buffers restricts the store-and-forward operation

as performed by electronic routers. Second, all-optical packet processing is currently unfeasible and, as previously illustrated, O/E conversions do not keep the pace with optical transmission speed.

While technological lacks make from OPS a long-term solution, a relaxed model known as Optical Burst Switching (OBS, [QY99]) is presented as an intermediate architecture between OPS and OCS, which could be deployed in the near future. In OBS, packets are aggregated into bursts at the edge routers before being sent through the optical network. To enable optical switching, a Burst Control Packet (BCP), which is electrically processed in the downstream node, is sent an offset time before the burst. The BCP acts in this case as a header, which contains the information needed by the OBS router to forward the burst to the next hop. In this way, being the offset time between the BCP and the burst precisely calculated, when the burst arrives, downstream nodes have already processed the BCP so the burst can be directly forwarded. Nonetheless, the burst contention problem appears when several bursts need to be processed at the same time in the same node. In the absence of optical buffers, this contention may lead to burst losses, that affect negatively to the QoS provided by OBS networks. Different contention resolution mechanisms can be applied to minimize these losses [YMYD03].

2.2 A control plane for ASON: The GMPLS protocol suite

Defined by the IETF in [RFC3945], the Generalized Multi-Protocol Label Switching is the result of extending the existing MPLS set of protocols aiming to implement a common control plane, able to manage several switching regions in an integrated way. Unlike in traditional MPLS, where only Packet Switched Capable (PSC) interfaces can be managed, a Label Switched Path (LSP) hierarchy is defined in GMPLS (Figure 2.6).

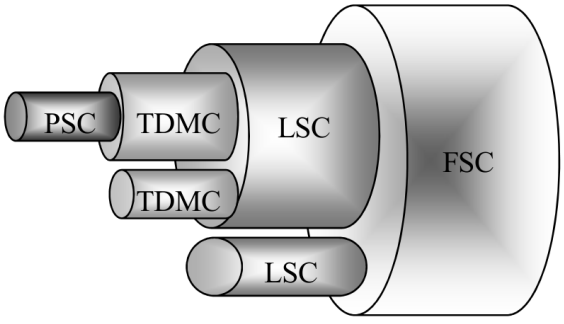


FIGURE 2.6: The GMPLS LSP Hierarchy.

In this way, Time Division Multiplexing Capable (TDMC), Lambda Switched Capable (LSC), or even Fiber Switched Capable (FSC) interfaces can be managed in addition to the PSC ones. Besides, the concept of Traffic Engineered (TE) links is introduced

to optical networks enabling the establishment of LSPs (i.e., Lambda Switched Paths or lightpaths) that take into account network constraints and service preferences. This makes GMPLS the most accepted solution for implementing the control plane functionalities in ASON. As a matter of fact, ITU-T and IETF have been joining efforts towards ASON networks with a GMPLS-capable control plane (e.g., see [Ala04, RFC4139, RFC4394, RFC4397]), which will be hereafter referred to as ASON/GMPLS networks.

There are three main functionalities a GMPLS-based control plane has to deal with: routing, signaling and resource management. To deal with the first one, the Open Shortest Path First (OSPF) and the Intermediate System to Intermediate System (IS-IS) protocols have been provided with Traffic Engineering (TE) extensions to enable the dissemination of the topological information related to the new transport network scenario ([RFC3630, RFC3784]). Due to its extended use in IP networks, OSPF-TE is the most accepted choice to be implemented in the control plane. For signaling purposes, the extended versions of the Resource reSerVation Protocol with Traffic Engineering (RSVP-TE, [RFC3473]) and the Constraint Routed Label Distribution Protocol (CR-LDP, [RFC3472]) are proposed, being the former the most accepted solution. Finally, the Link Management Protocol (LMP, [RFC4204]) has been introduced in the GMPLS protocol stack to take charge of the resource management tasks. The IP-based network infrastructure to support these control functionalities is known as the Data Communication Network (DCN, [G.7712]).

2.2.1 Routing extensions for GMPLS

In the context of ASON/GMPLS networks the routing involves two main tasks: topological information dissemination and path computation. The former is performed by the routing protocol (e.g., OSPF-TE) and enables the latter, which is realized by the route computation engine of the OCC. Being routing a *leitmotiv* in this thesis, special attention is paid here to the requirements in support of GMPLS routing protocols in order to facilitate the computation of traffic engineered paths. Therefore, a few requirements introduced by the IETF are summarized since they are important to understand the mechanisms and extensions explained later on.

As said, one of the main tasks assigned to the routing module in the control plane has to do with physical topology dissemination, which especially involves link information. Like in MPLS, GMPLS links are identified by the interface they are connected to. In MPLS the interface identifiers are typically IP addresses, since IP flows between two nodes are normally transported through a single link (wire). However, in an optical network a huge number of fibers each carrying tens (or even hundreds) of wavelengths can be deployed between two network elements (e.g., optical switches). It is then clear that using IP addresses to identify such a huge number of elements is unfeasible. For this reason, the concept of *Unnumbered TE link* is introduced in [RFC4202]. Basically, an unnumbered TE link is a TE link whose interface identifier is a non-zero 32 bits number which is unique in the scope of the node. Therefore, being the node identified by a unique *Router*

Id, a TE link can be unambiguously identified in the network by the tuple [Routed Id, TE link interface identifier].

It has been also introduced the ability of GMPLS to cover different switching technologies. Therefore, the *Interface Switching Capability Descriptor* (ISCD, [RFC4202]) is proposed to disseminate the type of traffic supported by the TE link (e.g., PSC, LSC, ...). It is noteworthy that, since one TE link, or more precisely the interfaces it is connected to, may support several switching capabilities, it may have more than one associated ISCDs.

Furthermore, two routing extensions are defined in [RFC4202] to support protection and restoration in GMPLS-controlled networks. First, the *Link Protection Type* is used to disseminate the protection capability of a TE link. Second, the *Shared Risk Link Group (SRLG)* is used to identify a set of TE links sharing a resource whose failure would affect all the TE links in such group.

Finally, although still a work in progress, [WSON-Enc] sets up the basis to disseminate information related to the data channels (i.e., wavelengths or, logically speaking, *data links*) contained in a TE link. This enables the combined routing and wavelength assignment (RWA, [ZJM00, RFC6163]) mechanism in the routing module of the OCC. In this method, path computation does not only comprise the calculation of the spatial route of the path (seen as a sequence of nodes and/or interfaces), but also designates the wavelength to be used.

2.2.2 A new routing paradigm: The Path Computation Element

Path computation is a fundamental building block for traffic engineering in GMPLS-controlled networks. As network scenarios grow and evolve into multi-layer or/and multi-domain deployments, route computation becomes more complex. Hence, new information flows and cooperation between routing entities is needed. In addition, new mechanisms such as adaptive routing, load balancing or route re-optimization are proposed aiming to satisfy the requirement of a better use of network resources.

In light of this, the former distributed model where each OCC was provided with a routing entity assuming the responsibility for computing routes departing from that OCC, moves towards an architecture where a specialized element monopolizes the route computation task. In the framework of the IETF, the Path Computation Element (PCE, [RFC4655]) arises as the new routing entity within the control plane. The PCE collects topological information disseminated by OCCs through the implemented routing protocol (e.g., OSPF-TE) and calculates end-to-end routes requested by the OCCs in the network. Moreover, the communication to the PCE is standardized in [RFC5440] bearing the Path Computation Element Communication Protocol (PCEP).

In summary, the PCE is the result of an attempt to standardize route calculation mechanisms within the GMPLS-enabled control plane as well as to improve path computation

power. It is worth pointing out that several collaborative PCEs can be deployed in the same GMPLS control plane instance, thus adding distributeness to path computation, so that a scalable model is obtained. All this makes from the PCE-based scheme a more powerful, flexible and easy to interconnect routing paradigm. Further detail is given on the PCE in forthcoming chapters.

2.3 Challenges on new generation optical transport networks

This section raises the open issues that affect to dynamic wavelength-routed optical transport networks. In this context, there are three key words, namely efficiency, transparency and multi-domain, that have to be kept in mind. Efficiency is related to the utilization of the coarse bandwidth granularity provided by these OCS networks which usually exceeds single client's needs. Transparency is associated to the WDM-based all-optical transmission technology which is affected by a number of physical effects, the so-called physical layer impairments (PLIs). The latter leads with the problems that arise when transport networks are segmented. In multi-domain networks, scalability and confidentiality requirements impact on end-to-end path computation and, thus, on service provisioning.

2.3.1 Improving efficiency: Multi-layer optical networks

Dynamic wavelength-routed optical networks have been presented in section 1 as a significant advance towards higher capacity, more flexible and easy-to-maintain transport networks. In particular, the ASON architecture provided with a GMPLS control plane, which is capable to dynamically establish and release end-to-end lightpaths, becomes the most accepted solution. Nonetheless, the wavelength granularity (e.g., tens of Gbps) provided by these lightpaths results in low bandwidth utilization when supporting sub-wavelength data flows (e.g, hundreds of Mbps). Multi-layer architectures try to bridge this bandwidth gap by nesting incoming low-bandwidth client flows into high-bandwidth wavelength channels. In this way, architectures like MPLS/WDM or SDH/WDM enable better lightpath utilization than pure dynamic OCS networks.

In this scenario, the ASON specification defines a reference layered transport network architecture where each network layer holds a twofold role: a server role to the client layer above it, and a client role to the network layer below it. This aforementioned *overlay* model leads to the cooperation between different control planes (one for each defined layer). However, since this collaboration is done by means of the UNI interface, a low trust relationship is assumed between layers so that a poor information exchange is performed. In consequence, a highly opaque, inflexible and inefficient architecture is obtained. Two alternative models, namely *peer* and *augmented* [RFC3717], relax these

data exchange limitations to increase the efficiency in front of the *overlay* one. The *peer* model, also known as *integrated* [RFC3945] or *unified* [DVP04], relies on a full information exchange between layers. Eventually, this comes out into a common control plane which manages all layers in an integrated way. The *augmented* model, in turn, arises as an intermediate solution since it allows a certain information flow between layers which are interconnected by means of a semi-trusted interface (i.e. E-NNI).

In a context where layering responds to switching capabilities criteria, a peer model seems to be the most appropriate solution since it proposes a single control plane to manage all the different transport layers seamlessly. In this way, the use of resources belonging to different layers can be optimized to provide a more efficient network operation. In addition, the avoidance of inter-layer communication simplifies the architecture also making it much more flexible. This is endorsed by the previous work that has undertaken the study of integrated TE mechanisms in multi-layer networks. As a matter of fact, from an architectural point of view, most of the research assumes a peer model control plane (e.g., see [SYT+02, CMP+03, ZHLL04]). It is also worth to mention that GMPLS is the most assumed implementation for the control plane.

In light of the above, this thesis addresses the integrated management of a GMPLS-controlled multi-layer optical network. In particular, the focus is placed on the experimental development and assessment of several extensions, which are applied to the standard GMPLS protocol stack, targeting at an efficient exploitation of the network resources. Moreover, different TE mechanisms are proposed at the control plane level to reduce both network deployment and operation costs.

Looking at the literature, some efforts have been focused on theoretical aspects of the problem. In this regard, mathematical formulations of the traffic aggregation problem (e.g., [DR02, ZM02]) or novel integrated routing algorithms for the new multi-layer scenario (e.g., [YADA01, ZHLL04, YR05b, GR08]) have been proposed. In contrast, the work carried out in this thesis concentrates on the experimental evaluation of an integrated GMPLS-based control plane for multi-layer service provisioning. In this line, more practical research has been also performed on the transport and control planes in previous works. For instance, [SYT+02] proposes an architecture for a grooming-capable node, that is a node capable of aggregating different electrical client flows into an optical circuit. Furthermore, experimental validations of the traffic aggregation architecture are detailed in [OOI+05, MKH+05]. However, [SYT+02, MKH+05] underline the work realized on the transport plane while avoiding the impact of multi-layer on the control plane, which is the focus of this thesis. Similarly, [OOI+05] provides an experimental validation of a multi-layer scenario from the PCE perspective but no detail is given about the other elements of the control plane.

2.3.2 The transparency issue: Impairment-aware optical networks

Wavelength routed all-optical networks are expected to run away from the opaqueness of first generation optical transport networks, where signals transported over point-to-point WDM channels are electrically terminated and regenerated at every intermediate node of the end-to-end path. By moving data transmission and switching operations entirely in the optical domain, operators are relieved from deploying expensive and not scalable intermediate electronic stages, resulting in high speed, highly-scalable and cost-effective backbone networks [BSBS08, SS11]. Furthermore, thanks to the introduction of a control plane, transparent optical networks are enhanced with automated connection (i.e., lightpath) provisioning and restoration functionalities, easing the adoption of reliable on-demand transport services at low cost [Jai05]. Costs savings come here from the avoidance of manual resource provisioning operations plus the high network capacity efficiency provided by the restoration schemes [Gro04].

However, the transition from opaque to transparent optical networks poses some challenges that should be properly addressed. First of all, transparency implies the transmission of signals over very long physical distances without electrical regeneration. This may lead to unfeasible lightpaths due to the physical layer degradations that accumulate on the signal along the path. To address this issue, regenerators (i.e., optical-electronic-optical, O/E/O) can be strategically placed at a selected number of sites. This increases the total distance that can be spanned by the lightpaths, thus making them feasible. Networks following this configuration are typically referred to as *translucent* networks (e.g., see [YR05a]).

At the control plane level, either in fully-transparent or translucent optical networks, PLI information can be taken into account in the RWA algorithms, turning them to Impairment-Aware RWA (IA-RWA) algorithms. In this way, the physical feasibility of the lightpaths under establishment can be predicted and regeneration, if available, can be utilized. Numerous IA-RWA strategies can be found in the literature differing on the accounted physical effects, the approach to the RWA problem or the implementation (e.g., see [APMS07, PPK08, HHM05, HBPPS07, MBL+08, RSM09]). In order to enable IA-RWA in the control plane , the OCCs have to be provided with the mechanisms necessary to collect and disseminate the information related to physical layer characteristics. In light of this, one of the goals of this thesis is the definition, development and final validation of a GMPLS-based impairment-aware control plane. To this end, several extensions are incorporated to the GMPLS protocol stack.

Another challenge inherent in transparent networks is that, without electrical termination at network nodes, Loss of Light (LoL) alarms stemming from a single failure propagate downstream from the failure point. This issue complicates failure localization and isolation procedures, which are of paramount importance when targeting at fast lightpath

restoration. A straightforward solution to address failure localization and isolation problems in all-optical networks is the deployment of dedicated out-of-band Optical Supervisory Channels (OSCs) per fiber link. This solution, however, requires a WDM channel and a transmitter-receiver pair per link only for monitoring purposes. In order to improve the performance of this basic scheme, more sophisticated out-of-band techniques have been also proposed, aiming to minimize network monitoring costs while detecting failures without ambiguity (e.g., see [ZH04, WHY09, TWH09]). In addition, in-band monitoring schemes have also been presented [MTT05, SSC+07], which allow finer soft-failure detection and localization per WDM channel. To address this open issue, a dynamic impairment-aware restoration scheme is proposed in this work. This solution is based on monitoring trails (m-trails), which minimize the number of out-of-band OSCs that need to be deployed in the network in order to achieve unambiguous failure localization. Thus, once the failure is localized and isolated, restoration actions can be triggered for all those affected lightpaths.

2.3.3 Service provisioning in multi-domain optical networks

Envisaging the scalability problems associated to the growth of network infrastructures, and the need for interconnecting facilities belonging to different operators, the ITU-T standardization for ASON considers the possibility to divide the network into domains. These domains exchange routing and signaling information by means of the previously introduced reference points. Hence, network partitioning [G.805] yields to scalable scenarios where, thanks to interconnections based on semi-trusted interfaces (i.e., E-NNI), different administrative domains can cooperate to provide end-to-end network services also respecting confidentiality between them. However, the resulting lack of full topological and TE information complicates path computation. For this reason, models providing a good trade-off between information exchange limitations and the quality of the computed routes are needed.

A domain is defined as a collection of network elements within a common area of path computation responsibility. Specifically, the ITU-T categorizes a domain as a *Routing Area* (RA) whose responsible routing entity is the *Routing Controller* (RC), and its architecture and requirements are detailed in [G.7715]. From an architectural viewpoint, horizontal partitioning in RAs may result into a set of collaborative RCs, or a multi-level hierarchy where lower-level (child) RAs are contained in higher-level (parent) ones. This hierarchical deployment allows routing information abstraction, that is, child RCs send summarized topological information to their parents so they can compute end-to-end multi-area routes.

As a consequence, two approaches can be considered to face topological information dissemination and later route computation in multi-domain networks (Figure 2.7). On the one hand, the hierarchical deployment of RAs proposed in [G.7715], which is normally

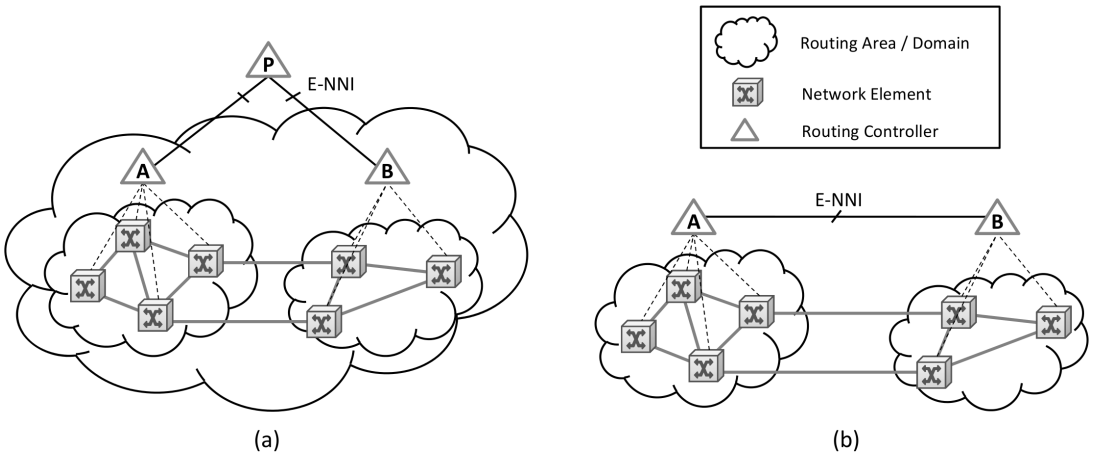


FIGURE 2.7: Routing models: (a) Hierarchical model, (b) Flat model.

supported on topology abstraction mechanisms, can be investigated. The Optical Internetworking Forum (OIF) proposes in [OIF11] an implementation of the E-NNI following this model. This implementation is based on a hierarchical version of the OSPF-TE protocol. In the same line, the IETF defines in [RFC5787] a set of experimental extensions to OSPF-TE in support of the RA hierarchy. In this context, different topology aggregation policies may lead to a wide range of results in terms of blocking probability, resource utilization, network overhead, etc. Studies based on this approach can be found in [SLMBMT+05, LKG+06, LKGG06]. On the other hand, a flat model in which RAs are placed at the same level can be considered. In this case, OSPF-TE can be used to implement this approach by mapping RAs into OSPF areas ([RFC2328]). Reachability information is then exchanged between such areas. A similar solution can be implemented by means of an Exterior Gateway Protocol (EGP) such as the Border Gateway Protocol (BGP, [RFC4271]) if RAs are mapped into Autonomous Systems (ASes). However, extensions are needed by these routing protocols in order to disseminate the information necessary for TE end-to-end path computation (e.g., [OOO08, MRBD09, BGP-TE]).

Regarding the control plane implementation, the main efforts of the IETF in multi-domain are not only focused on protocol extensions but also on the architectural side. In particular, the IETF proposes its own version of the ITU-T Routing Controller: the PCE. Introduced in section 2.2.2, the PCE is gaining momentum as the *de facto* standard routing entity in the GMPLS control plane. The PCE can be deployed following the two aforementioned hierarchical and flat models [PCE-Hierarc], and its applicability in an ASON environment is depicted in [G.7715.2] and [RFC4655].

Hence, in a context where different PCE-based GMPLS-controlled optical networks are interconnected, multi-domain end-to-end path computation, and thus service provisioning, is hindered by a set of control plane functional requirements such as scalability and confidentiality. Standard approaches to the multi-domain end-to-end service provisioning problem try to be compatible with the existing protocol developments (i.e., OSPF-TE, BGP, PCEP, etc.) [RFC5152, RFC5441]. However, this leads to sub-optimal solutions

that affect to the performance of the network. The sixth chapter of this thesis faces the path computation problem in a PCE-based multi-domain network. Specifically, a scenario where different PCEs, each one controlling a single domain, cooperate to obtain end-to-end multi-domain routes is considered. Therein, different path computation mechanisms are proposed, and their scalability, performance and impact on the standardization are evaluated.

2.4 Chapter summary

This chapter has provided a review on the evolution of optical transport networks. Such a review has paid special attention to dynamic automatically managed wavelength-routed optical network architectures, since they are envisaged as a short to mid term solution to overcome the limitations of the overlayed traditional model. Following the ITU-T ASON definition, this paradigm promises flexibility and rapidness in service provisioning supported on the operation of a control plane, as well as high bandwidth capacity and reliability in data transmission thanks to the WDM transport technology. Furthermore, the GMPLS protocol stack has been introduced in section 2 as the most accepted solution to implement the functionalities of the ASON control plane, namely routing, signaling and resource management.

The third section has underlined the limitations still presented by the ASON/GMPLS paradigm, which are addressed along this thesis. First, the performance of ASON/GMPLS networks has been identified as inefficient since the wavelength granularity provided by the established lightpaths leads to poor bandwidth utilization when sub-wavelength client flows must be transported. To address this, a GMPLS-controlled multi-layer optical network architecture, where different client flows are aggregated and transmitted within the same optical path, has been introduced. Next, the technological limitations associated to WDM-based all-optical transmission have been highlighted. Although offering huge bandwidth capacity at low cost, this technology is very sensitive to a set of physical effects that impose penalties to the transmitted optical signal and, in consequence, affect to transported data. In view of this, a control plane capable of managing the physical layer information has been presented in order to provide lightpaths with guaranteed optical signal quality. Finally, the end-to-end path computation problem in multi-domain optical networks has been described. There, scalability and confidentiality requirements lead to reduced topological information exchange between domains, which complicates the operation of the routing entities and prevents an optimal end-to-end service provisioning. In this context, a PCE-based architectural solution has been introduced aimed to provide effective and scalable path computation mechanisms.

Chapter 3

Thesis scenario: The CARISMA test bed

The main characteristics of the ASON/GMPLS CARISMA test bed are presented in this chapter. Located at the Universitat Politècnica de Catalunya (UPC) premises, this scenario has been the basis for implementing and further experimentally evaluating most of the proposals in this thesis. The chapter starts with a brief explanation of the CARISMA test bed history and its main features. Next, the CARISMA GMPLS-based control plane implementation is detailed going through the modules, interfaces and protocols that take part of the OCC. In addition, the PCE is after introduced as a new control plane entity aimed to provide enhanced path computation methods. Moreover, the PCE standardizes the route request messaging within the GMPLS-enabled control plane by means of the PCEP protocol. Running over the control plane, the NMS operation in the CARISMA network is finally outlined.

3.1 Overview of the CARISMA test bed

The development of the CARISMA test bed started in 2002 within the framework of the homonymous project "Conexión y Acceso a RedIri2 Mediante Anillo óptico multi-canal" (CARISMA, TIC2000-0304-P4-04). The key goal of this project was to design and implement a three-node ring-based intelligent and dynamically switched Dense WDM (DWDM) all-optical network, able to provide Bandwidth on Demand (BoD) between different IP clients (Figure 3.1). To achieve this, ASON was the candidate architecture chosen to be deployed thanks to the support received by the standardization bodies. Likewise, the emerging GMPLS protocol stack was the best positioned control plane implementation. Furthermore, a Network Management System was planned to be developed as well as the manufacture of three ROADMs for the transport plane.

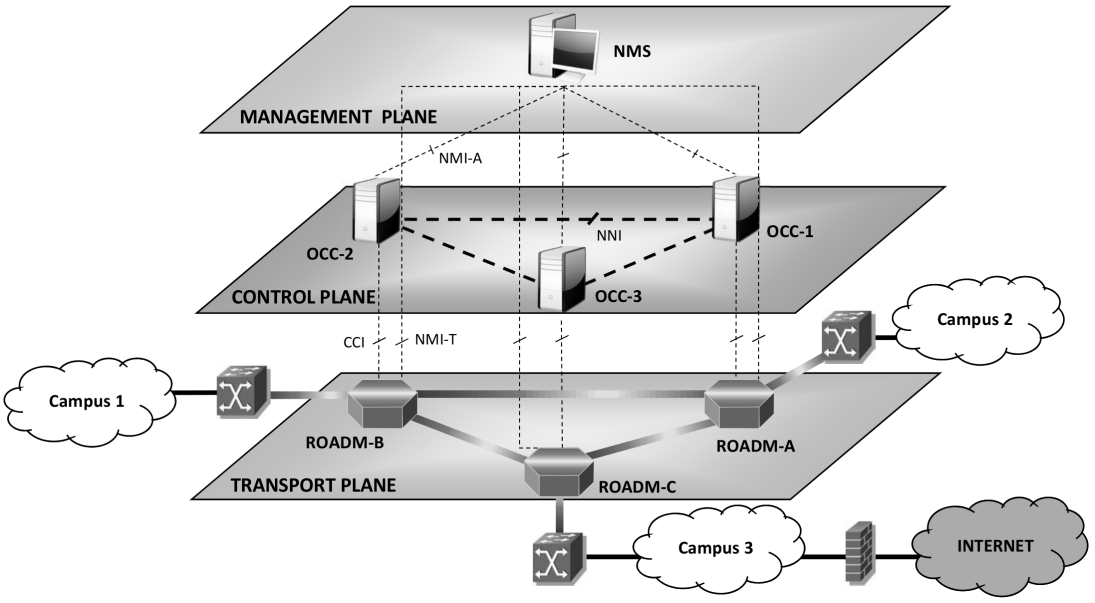


FIGURE 3.1: The CARISMA network.

Once the CARISMA project finalized, research efforts remained with aims to validate the ASON/GMPLS paradigm. Having a real test bed, the work was specially fated to experimental implementations and assessments of the open issues identified in such a model. In this way, the CARISMA test bed kept evolving to its current status. A glance at the test bed characteristics is taken below following the ASON plane division.

- The transport plane is composed of three 2-degree ROADMs based on Wavelength Selective Switching (WSS) technology. In addition, an OXC emulator software is used to test larger meshed optical networks.
- The control plane comprises fourteen OCCs each one running over a Linux-based Pentium 4 machine with 512 MB of RAM memory. The topological deployment of the DCN is completely configurable, so that different network scenarios can be tested. Thanks to virtualization mechanisms the number of OCCs can be increased to configure larger networks.
- The management plane is implemented by means of the NMS, which enables permanent and soft-permanent connection setup and tear-down, as well as network monitoring. Formerly developed as a web-based application, it was re-implemented as a standalone client/server application using Java technology.

3.2 Optical Connection Controller architecture

The Optical Connection Controller (OCC, Figure 3.2) is the key element in the control plane since it brings the intelligence to the optical transport network devices (i.e., the

OXCs). Designed following the architectural specification of the ASON recommendation G.8080, the CARISMA OCC is composed of three well-defined modules: the Link Resource Manager (LRM), the Routing Controller (RC) and the Connection Controller (CC), each one implementing the IETF proposed LMP, OSPF-TE and RSVP-TE protocols, respectively. The interaction of these software elements fulfills the aforementioned requirements of automated and rapid lightpath provisioning, as well as effective failure recovery.

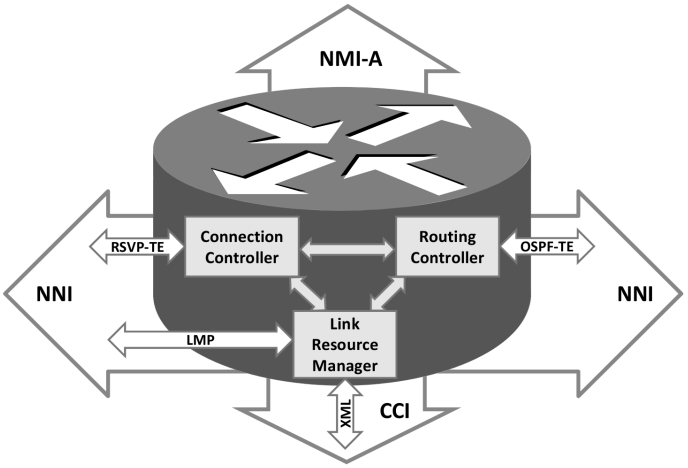


FIGURE 3.2: Optical Connection Controller (OCC) architecture.

3.2.1 Link Resource Manager

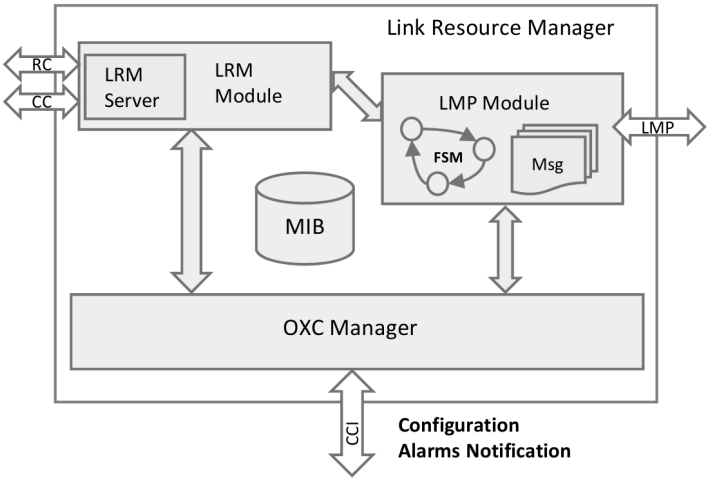


FIGURE 3.3: Link Resource Manager architecture.

The Link Resource Manager (LRM, Figure 3.3) maintains the local information of the data plane elements it is responsible for. More precisely, the LRM keeps the information related to the components of its managed OXC (such as fibers, ports and wavelengths) by storing them in its Management Information Base (MIB). The key information elements here are the Traffic Engineering links (TE links) and the data links, since they represent

the physical connection of the local OXC to its neighbors. It is worth mentioning that, along this work, a TE link represents a fiber link between two OXCs. Besides, each wavelength channel that can be transported through it is logically represented by a data link. The maintenance of such resources' status is crucial for connection establishment and failure detection and localization. Therefore, the LRM implements the Link Management Protocol (LMP, [RFC4204]) which is responsible for monitoring the status of the local TE and data links by means of *control channels*. Moreover, aiming at complete network automation, the protocol also implements a neighbor discovery mechanism. The IETF establishes four standard functionalities for LMP:

- The *Control Channel Management* mechanism establishes the IP communication interface between two adjacent OCCs (i.e., the control channel). This control channel is thereafter used to exchange the LMP messages between these two nodes. The health of the control channel is monitored by means of a so-called Hello protocol.
- The *Link Property Correlation* is responsible for assuring that the properties of the TE and data links are the same at both the local and remote ends of the link.
- The *Link Connectivity Verification* ensures the physical connectivity of the data links between two neighboring nodes. This mechanism is also used for resource discovery.
- The *Fault Management* embraces fault detection and localization procedures.

In addition, concerning the OCC internal operation, the LRM has a twofold role. Firstly, it updates the Routing Controller with the local data plane information whenever a change happens. This information is afterwards disseminated by the RC. Moreover, the LRM interfaces the requests coming from the Connection Controller for allocating/releasing the local resources during lightpath setup/tear-down and configures the OXC through the Connection Controller Interface (CCI).

3.2.2 Routing Controller

The Routing Controller (RC, Figure 3.4) is responsible for providing routing adjacency to the OCC within the DCN. Being an IP network, the routing in the DCN is done by means of the standard OSPF version 2 protocol ([RFC2328]). In this way, the activity (i.e. route calculation, signaling and link management) in the whole control plane is enabled.

Furthermore, the RC is the main actor in the path computation tasks. In this regard, when the OCC (concretely, the CC) receives a connection request, a path computation request is sent to the RC. Since this calculation comprises the routing and the wavelength assignment, each RC manages a complete graph of the optical transport network to obtain

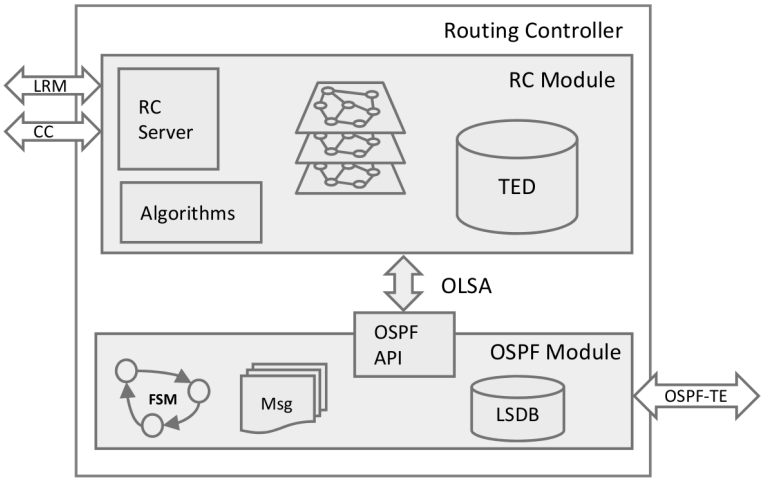


FIGURE 3.4: Routing Controller architecture.

end-to-end traffic engineered routes. To build the global map of the network, the RCs in the different OCCs cooperate sharing the local data plane information provided by their respective LRMs. This is achieved thanks to the OSPF protocol engine and the TE extensions it is provided with (OSPF-TE). Thus, the RC disseminates the TE and data links information by means of OSPF-TE and processes the one coming from its neighbors. This information is then stored in the Traffic Engineering Database (TED) and used to build the network graph over which route computations are performed.

As said, the OSPFv2 protocol for IP networks is defined in [RFC2328]. Nonetheless, a set extensions is applied to the protocol so it is able to fulfill the traffic engineering requirements of new generation optical transport networks. In this context, [RFC3630] defines a number of Traffic Engineering attributes aimed to cope with the MPLS technology routing requirements and, by extension, the GMPLS ones. Note that extensions in OSPF are usually introduced as Type/Length/Value (TLV) objects, which are inserted in the *Link State Advertisement* (LSA) messages to convey specific topological information. Particularly, to disseminate TE information, the *Opaque LSA* (OLSA, [RFC2370]) message is extended with two top-level TLVs, namely *Router Address* and *Link* which, in turn, are provided with a set of sub-TLVs. It is clear that whilst OLSAs conveying a Router Address TLV disseminate information related to the node, the ones provided with a Link TLV disseminate data related to a TE link. For sake of conciseness, the Router Address TLV is left out of the scope of this document. Table 3.1 shows the parameters related to a TE link, which are mainly focused on link identification (through a set of identifiers) and bandwidth occupancy. More precisely, sub-TLV type 1 characterizes the link type as point-to-point or multi-access. Sub-TLVs from 2 to 4 identify the TE link through a unique link identifier, and the interfaces of the routers it is connected to. Sub-TLV 5 allows setting a user-defined cost to the link. Sub-TLVs ranging from 6 to 8 carry the link bandwidth occupancy. Finally, sub-TLV 9 is set by the network administrator and identifies the administrative group assigned to the TE link.

Type	**Name**
1	Link Type
2	Link Id
3	Local Interface IP Address
4	Remote Interface IP Address
5	Traffic Engineering Metric
6	Maximum Bandwidth
7	Maximum Reservable Bandwidth
8	Unreserved Bandwidth
9	Administrative Group

TABLE 3.1: TE attributes (sub-TLVs) assigned to the Link TLV [RFC3630].

In light of the stated in [RFC4202], additional extensions for OSPF-TE in support of GMPLS are proposed in [RFC4203]. Like in the previous case, they consist of new sub-TLVs which are added to the above-mentioned Link top-level TLV. These extensions are depicted in Table 3.2 and the list below highlights the most relevant ones for the research done in this work:

- The *Link Local/Remote Identifier* is proposed to support unnumbered links as reported in section 2.2.1. In contrast to the previously depicted Local and Remote Interface IP Address sub-TLVs (type 3 and 4 in Table 3.1), which were defined to carry the numbered interfaces (i.e., IP addresses) of the two routers connected by the link, this new sub-TLV carries user-defined local and remote identifiers assigned to the TE link.
- The *Link Protection Type* contains the protection capabilities of such TE link.
- The *Interface Switching Capability Descriptor* gives information about the switching technologies (i.e., packet, TDM, wavelength, fiber) associated to the interfaces the TE link is physically connected to.
- The *Shared Risk Link Group* is used for protection and restoration. The *SRLG* identifies TE links having the same exposure to failures. This attribute is not used in the scope of this thesis.

Section 2.2.1 has introduced the work that is being currently done regarding the definition of routing protocols extensions for wavelength information dissemination, which

Type	**Name**
11	Link Local/Remote Identifier
14	Link Protection Type
15	Interface Switching Capability Descriptor
16	Shared Risk Link Group

TABLE 3.2: TE attributes (sub-TLVs) assigned to the Link TLV in support of GMPLS [RFC4203].

facilitates the combined RWA within the RC. Based on those proposals, a new non-standard *Wavelength availability* sub-TLV is added to the CARISMA OSPF-TE protocol implementation, in particular to the Link top-level TLV. In this way, the information related to the wavelengths (i.e. data links) of each TE link is flooded at network boot-up and under changes occurred in the TE link status. Specifically, the new sub-TLV conveys a bitmap representing the status (free or occupied) of the wavelengths supported by the TE Link, so the routing engine is aware of the availability of such resources. Hence, each wavelength is represented by a bit in the bitmap being set to 1 for free channels and to 0 for the occupied ones. The proposed encoding of this sub-TLV is based on the defined in [WSON-Enc], and its building fields, headed by the type and length, are shown in Figure 3.5. Therein, the *Action* field gives information about the type of data contained in the sub-TLV. In this case it is always set to 4 indicating that the sub-TLV conveys a bitmap (the interested reader may refer to [WSON-Enc] for further detail). Next, the number of wavelengths contained in the TE Link is conveyed. *Length* indicates the number of bytes remaining in the sub-TLV and is used to process the sub-TLV properly. The next four bytes indicate the lowest channel of the TE Link encoded following the established in [RFC6205]. The lowest channel is mapped in the bit position zero while each succeeding bit position represents the next frequency a channel spacing (CS) above the previous. The bit map is made up of 32-bit words to assure that the sub-TLV is multiple of four bytes so, in some cases, there will be unused bits in the bitmap which will be set to zero and ignored. Note that, with this sub-TLV format, there is no differentiation between occupied and unused wavelengths, since both types are ignored during route computation. Further, extensions could be applied to the sub-TLV for more precise wavelength information.

More specific details will be given on these extensions when necessary in forthcoming chapters.

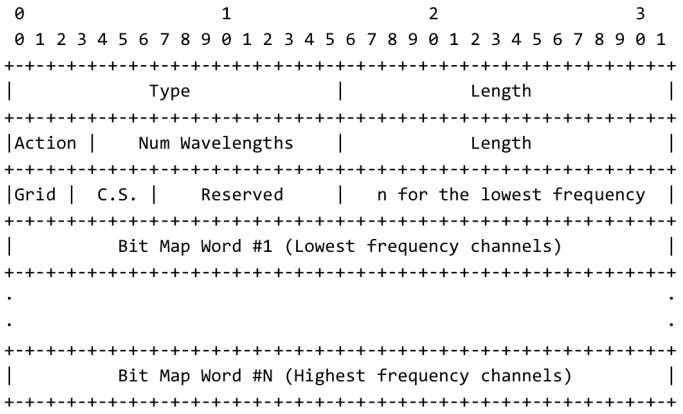


FIGURE 3.5: Wavelength availability sub-TLV

3.2.3 Connection Controller

The Connection Controller (CC, Figure 3.6) is the element that centralizes the activity in the OCC. It receives the LSP requests coming from either the NMS or a client network, coordinates the establishment of the lightpath, and keeps the lightpath status as long as it is active. The LSP information is stored in the CC by means of the Path State Block (PSB) structure. In particular, the CC stores each LSP whose signaling has processed.

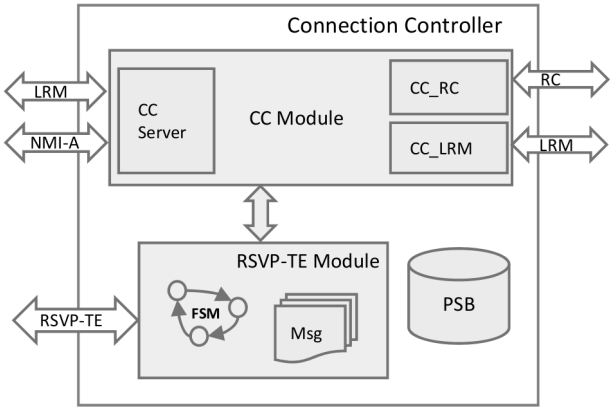


FIGURE 3.6: Connection Controller architecture.

Upon the arrival of an LSP request, the CC asks the RC for a route computation. Once the RC sends the route back to the CC, it starts the signaling of the LSP using the RSVP-TE protocol [RFC3473]. The signaling is performed in two phases. In the first one, a *PATH* message is sent from the source OCC to the destination one in a hop-by-hop basis. To do this, the *PATH* is provided with an *Explicit Route Object* (*ERO*, [RFC3209]) which contains the route calculated by the RC and the interfaces of the TE links to be used. In addition, the *LABEL_SET*, *SUGGESTED_LABEL* and *UPSTREAM_LABEL* objects convey the information regarding the wavelength to be used. The *PATH* message is processed by each OCC contained in the route, which checks the availability of the local resources to be used (by contacting the LRM) and forwards the message to the next hop. Moreover, the CC stores the information of the LSP in a PSB structure. If an error occurs

during the *PATH*, a *PATH_ERROR* message is sent back hop-by-hop to the source OCC. In that case, the PSB entry is removed from each CC and the already reserved resources are freed. If no error occurs, when the *PATH* reaches the destination OCC and it is properly processed, the second phase of the signaling, aimed to the actual allocation of the resources and the activation of the lightpath, is initiated. To this end, the destination OCC sends a *RESV* message that travels hop-by-hop towards the source OCC. Upon the reception of the message, each CC requests to the LRM the allocation of the requested local resources and the appropriated configuration of the OXC. Finally, once the source OCC has processed the *RESV* message, and the LSP is successfully configured, the source CC notifies the requester, that is, the NMS by means of the NMI-A or the client network by means of the UNI. If an error occurs during signaling the requester is also notified accordingly.

Similarly, if an LSP tear-down request arrives to the CC, a *PATH_TEAR* message is sent along the route of such LSP. Then, each CC requests its respective LRM to unallocate the used resources and to deconfigure the OXC, and removes the corresponding PSB. In addition, the source CC notifies about the tear-down to the requester. Besides this, the RC of each affected OCC disseminates the information of the newly available resources through OSPF-TE.

3.3 Path Computation Element

The increase of computational complexity associated to arising multi-layer and multi-domain networks is the main driver for the adoption of a Path Computation Element (PCE)-based architecture. In brief, this model relies on a centralized entity which takes up the path computation responsibility in the whole network.

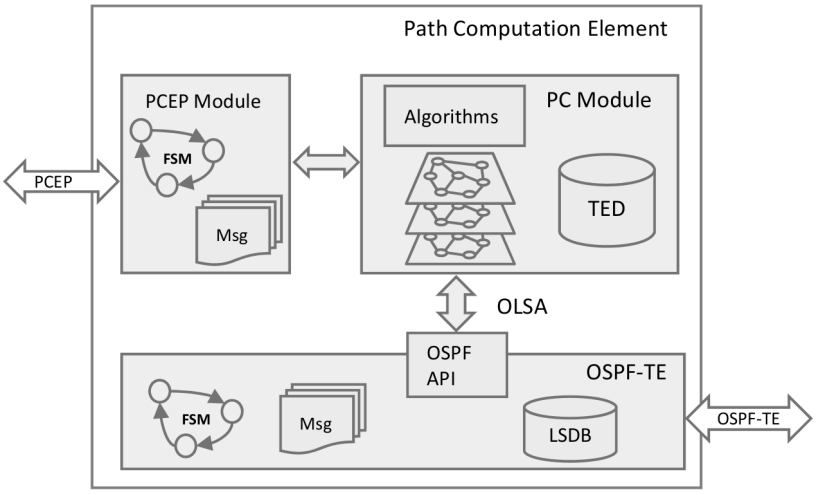


FIGURE 3.7: Path Computation Element architecture.

As stated in [RFC4655], the PCE (Figure 3.7) implements an OSPF-TE instance which is used to collect the topological and traffic engineering information disseminated by each OCC. This information is used by the routing engine for path computation. The communication between the PCE and its Path Computation Clients (PCCs) is done by means of the Path Computation Element Communication Protocol (PCEP, [RFC5440]), which operates over TCP for reliable messaging. Hence, stateful connections are maintained between the PCE and its PCCs.

As could be expected, the modules of the OCC need to be adapted to the new PCE-based model. First, a PCC is needed in the OCC to enable the communication with the PCE. This PCC is added to the CC (Figure 3.8 left) since, as seen in previous section, it is the coordinator of lightpath establishment. Therefore, in the new architecture the CC asks for route computations to the PCE instead of its RC. Consequently, the RC (Figure 3.8 right) is relieved from route calculation tasks, but it is still responsible for topological and traffic engineering information dissemination.

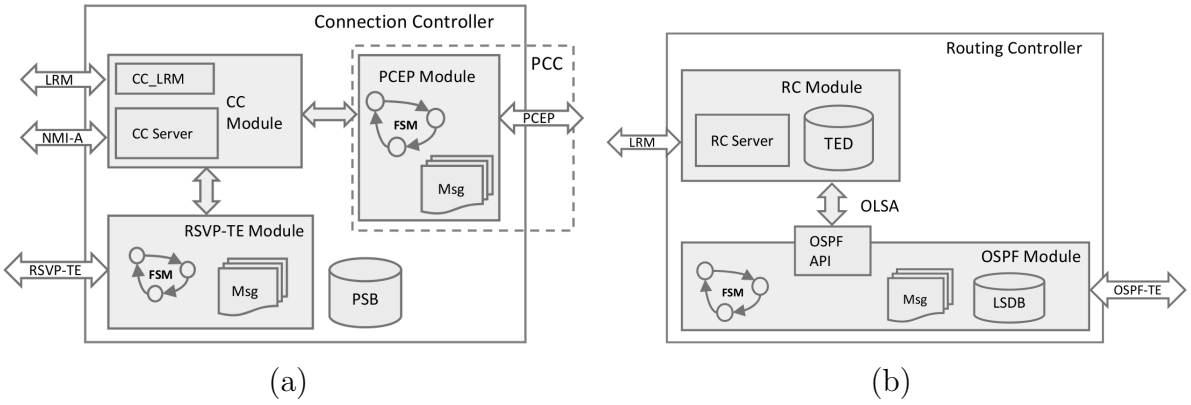


FIGURE 3.8: (a) CC module provided with a PCC; (b) Simplified RC module for PCE-based control plane architecture.

The PCEP protocol results from the efforts done by the IETF targeting the standardization of the interface between the CC and the path computation entity. As said, [RFC4655] defines a stateful client/server communication between the PCE and its PCCs. To this end, a set of messages is defined and exchanged following a finite state machine ([RFC5440]). Therein, the communication is initiated by means of an *Open* message sent by either end, and it is replied by the other side (following the stated request/response model). In order to maintain the statefulness of the connection, *Keepalive* messages are exchanged periodically along the connection activity. For route computation, a pair of *Path Computation Request/Reply* (hereafter, *PCReq* and *PCRep*) messages is exchanged between the PCC and the PCE. To finalize the connection a *Close* message is sent by the closure initiator and replied by the other end.

3.4 The Network Management System

The Network Management System (Figure 3.9) is a software platform that allows remote automated configuration and monitorization of the optical transport network. To this goal, the NMS implements the NMI-A and NMI-T interfaces, which enable the communication with the control and data planes, respectively. A Graphical User Interface (GUI) facilitates the network management providing a human-friendly environment.

Aiming to provide remote configuration capabilities, the NMS architecture was designed as a client/server application where the client provides the graphical interface and it is connected to the server, which contains the functionalities of the system. Besides, the server side is connected to the DCN and, thus to the OCCs. Similarly, the NMS server is connected to the OXC management network, which allows the configuration and monitorization of such devices.

For implementation, Java technology was used. In particular, Enterprise Java Beans (EJBs) were used in the server for the logical representation of the objects of the application. Thanks to the EJB technology, these objects can be remotely invoked from the client side and depicted in the graphical view. The latter was implemented by means of the Standard Widget Toolkit (SWT) provided by the Eclipse community. Finally, MySQL was used as database management system to provide object and data persistence.

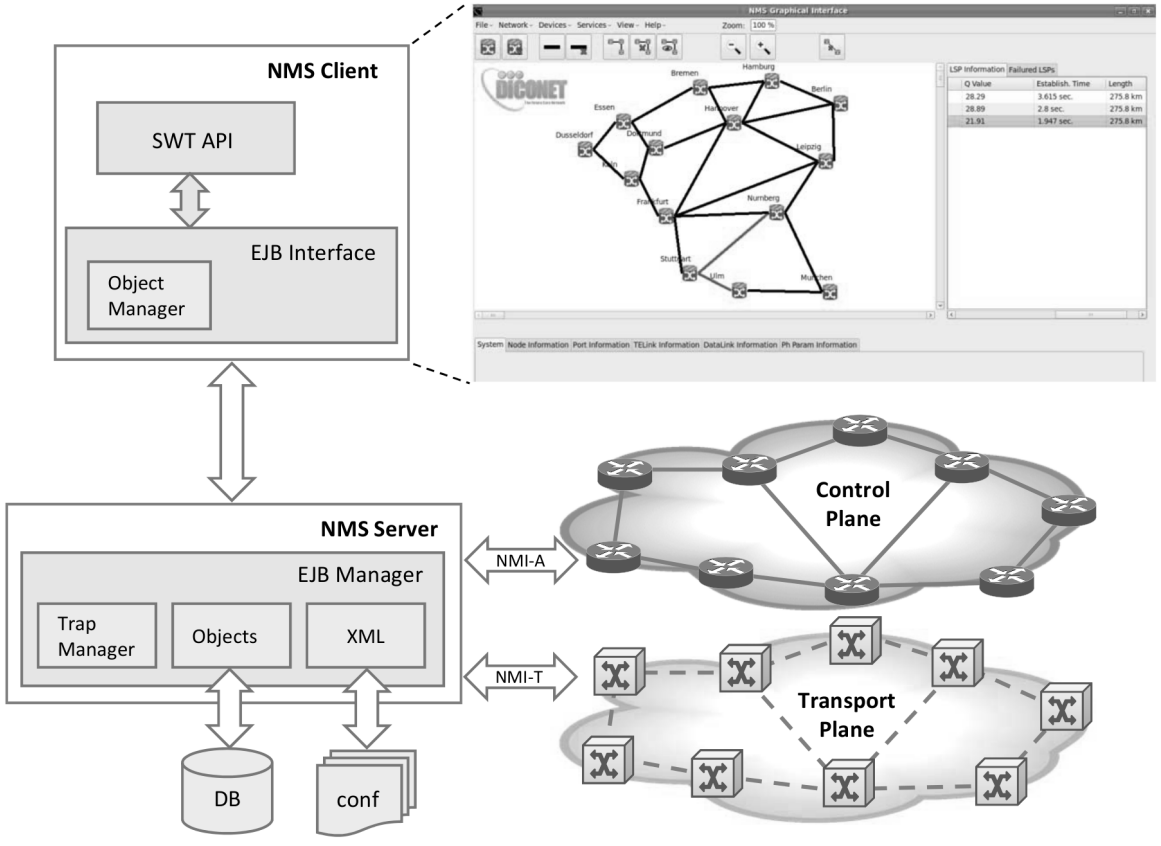


FIGURE 3.9: Network Management System architecture.

To establish a soft-permanent connection from the NMS, the user selects the setup option in the graphical menu also choosing the parameters of the connection (e.g., source and destination nodes, bandwidth, etc.). Once the lightpath characteristics have been configured, the *OK* button triggers the establishment process. By this means, the client side sends the lightpath request to the server which, in turn, forwards it to the source OCC of the connection. Then, the CC starts the process detailed in section 3.2.3. When the server side of the NMS receives the confirmation of the lightpath setup, it stores the information in its database and notifies the client side, that depicts the new lightpath in the GUI. If an error occurs, a message is displayed on the screen.

3.5 Chapter summary

The main features of the ASON/GMPLS CARISMA test bed have been illustrated in this chapter. The test bed presented here deploys the functionalities of a GMPLS-controlled all-optical transport network fulfilling the basic requirements imposed by the ITU-T in the ASON recommendations, and the IETF with respect to the GMPLS-enabled control plane. In this regard, the initial steps of the CARISMA test bed development are put in context first. Next, the modules that compose the OCC implementing the GMPLS-capable control plane are thoroughly described in section 2. Section 3 presents the IETF proposed Path Computation Element as the new routing entity to be incorporated to the control plane. It is worth mentioning here that the CARISMA GMPLS control plane is the framework wherein the developments and experiments that compose this thesis have been realized. The test bed overview is completed with a brief description of the Network Management System.

Chapter 4

Design and validation of a GMPLS-controlled multi-layer optical transport network

This chapter describes the experimental assessment of a GMPLS-controlled multi-layer optical network that has been conducted over the ASON/GMPLS CARISMA network. To this end, the architectural view of the scenario under study is given in the first section, and the main character in GMPLS-controlled traffic grooming is introduced: the *Forwarding Adjacency*. Furthermore, the work carried out by the standardization bodies in support of multi-layer is presented. Subsequent sections detail the contributions of this thesis to the subject. In particular, two aspects have been evaluated. First, an experimental validation of the standard multi-layer mechanisms is realized. To achieve this, several novel mechanisms have been introduced in the control plane modules, which allow the implementation of the standard protocol proposals. As a result, an experimental standard compliant control plane is obtained. Second, different traffic engineering mechanisms are experimentally evaluated and compared in terms of network performance. The obtained achievements are summarized in the last section of the chapter.

4.1 A GMPLS-controlled multi-layer network architecture

In chapter 2 wavelength-routed optical networks were presented as a solution to support the increasing bandwidth demand. The ASON architecture was also introduced as a valid model for the implementation of such high-capacity transport networks. However, given that the provided wavelength (i.e., lambda) granularity is very coarse, typically supporting 10, 40 Gbps or even more, these networks lack the flexibility to support sub-wavelength

traffic demands, thus leading to a poor bandwidth usage when low data-rate client demands must be allocated in the network (i.e., an entire wavelength might be needed to provision a client demand of merely few tens of Mbps). In this regard, as detailed in section 2.3.1, the ASON architecture contemplates the co-existence with networks having different switching technologies following the so-called overlay model. Moreover, two alternative models (peer and augmented) providing a more flexible management were introduced as well.

From the GMPLS perspective, its enhanced traffic engineering protocols pave the way to a peer multi-layer network architecture controlled by means of a common control plane [RFC5212]. As a matter of fact, since GMPLS provides a framework for controlling different switching capabilities, a single instance of the control plane may be needed. Therefore, being an enabler for such a vertical integration ([DVP04]), GMPLS allows a rapid service provisioning as well as efficient traffic engineering across all switching capabilities (i.e. layers).

In this context, the term traffic grooming identifies the process of packing several low-speed traffic streams into higher-speed streams trying to maximize optical channels bandwidth usage [ZM02].

The key entity for traffic grooming here is the Forwarding Adjacency (FA, [RFC4206, CMP+03]). In GMPLS, those already established lower-layer LSPs (e.g., λ-LSPs) are advertised as FA-LSPs, which can be used to transport new higher-layer client LSPs. In this way, lower-layer resources can be more effectively utilized.

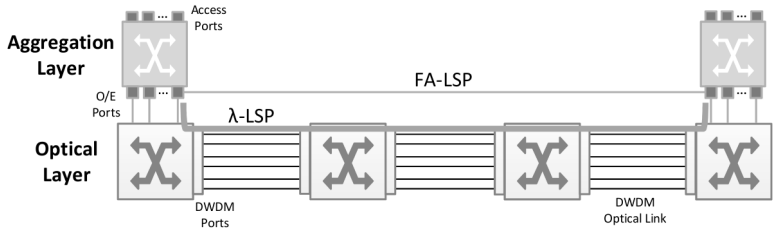


FIGURE 4.1: Example of a two-layered optical network where a λ-LSP has been set up. A FA-LSP with the residual bandwidth is advertised.

Without loss of generality, a two-layered network peer architecture is assumed in this thesis (Figure 4.1). In particular, an optical server layer and a client aggregation layer on top (e.g., SONET/SDH, MPLS, GbE, etc.) are considered. This architecture allows the mapping of the client traffic to be transported over the DWDM physical layer. At the bottom, optical nodes provide network ports (DWDM) as well as client access ports, used to inject an aggregated client flow to the network. In this context, the incoming signal is adapted, switched to a network port, multiplexed into a DWDM bundle and finally transmitted to the downstream optical node. On top, the client aggregation layer includes generic nodes providing electrical switching, flow aggregation and other features. Client nodes are connected to optical nodes through the client access ports.

4.1.1 The GMPLS grooming entity: The Forwarding Adjacency

As said, the traffic grooming strategy is based on the utilization of the FA concept introduced in GMPLS. IP electrical flows, also referred to as packet-LSPs [RFC3471], use optical channels (λ-LSPs) to reach their destination. In classical IP routing, packet-LSPs (from now on client LSPs, for sake of generalization) are forwarded to the downstream node in a hop-by-hop basis until they reach their destination. The introduction of MPLS mechanisms has made this process easier as intermediate nodes only have to read a label to forward the data to the next node. Conversely, assuming a fully-transparent optical network (i.e., GMPLS-controlled WDM transport network), lightpaths are established from the source to the destination node and flows are tunneled through them [BDL+01]. However, problems may arise in this case when a lightpath has to be created even if it has to transport small traffic flows. Hence, a new scenario, where all the constraints (traffic pattern, network topology, resources status and performance objectives) would be considered for lightpath establishment, seems to be a right choice. The main idea consists of, upon a client LSP request, considering the possibility of re-using already active λ-LSPs rather than establishing a new one. Although this could lead to longer end-to-end paths (i.e., the shortest path might not be the chosen option) if an existent direct optical connection between source and destination was not available, a more efficient use of the resources would be achieved.

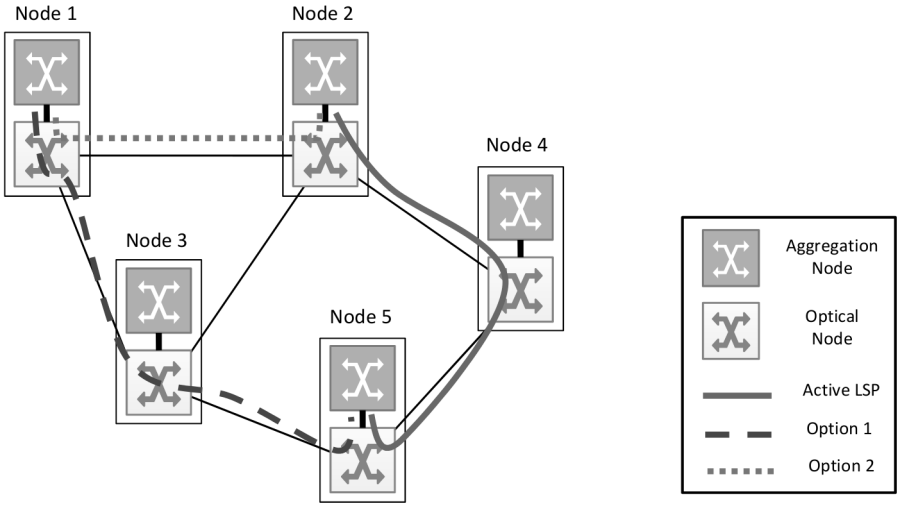


FIGURE 4.2: Example of IP/WDM grooming using the FA concept [CMP+03].

An example of the traffic grooming strategy is depicted in Figure 4.2. There, node 1 has to create a new client LSP to node 5. If a typical shortest path algorithm is used at the optical layer, a new λ-LSP through node 3 seems to be the most efficient choice (discontinuous trace, option 1, in the figure). However, there may be some situations (e.g., node 3 resources are highly busy), where it may be better to create a new λ-LSP from 1 to 2 (dotted line, option 2) and then group this flow with the already active λ-LSP from node 2 to node 5 (continuous trace), which is assumed to have enough available bandwidth. Moreover, a better use of the resources would be achieved in this second case.

Then, from the transport perspective, the client flow tunneled following option 2 would be extracted at the tail-end of the dotted λ-LSP (i.e., node 2), and inserted again at the same node to be tunneled through the already active λ-LSP (depicted by the solid trace).

The key point here consists of leveraging the FA concept at the optical layer. To this end, λ-LSPs are advertised to the network as FA-LSPs. By definition, an FA is a TE link between two GMPLS nodes whose path transits one or more other GMPLS nodes in the same instance of the GMPLS control plane [RFC4206]. Then, as all network nodes maintain a complete network state information database, they are able to calculate the most appropriate routes, assuming that the databases are properly updated and the routing algorithm manages the information adequately. Therefore, the grooming functionality here dealt with is achieved. Particularly in this thesis, λ-LSPs are supposed to be FA-LSPs so that client LSPs can be nested into them even if their source is different than the λ-LSP's. At this point, it is worth noting that only the source and the destination nodes of an FA-LSP can access to the lower order LSPs (client LSPs) carried within that FA-LSP (the intermediate nodes just cross-connect the optical channel). The rationale behind this is that the head and tail ends of FA-LSPs are the only points where data is inserted or extracted.

4.1.2 GMPLS protocol extensions for traffic aggregation

[RFC3477] proposes a new RSVP-TE object named *LSP_TUNNEL_INTERFACE_ID* to be added to the *PATH* and *RESV* messages when signaling a new FA-LSP. This object contains two fields: the router identifier and an interface identifier for the FA-LSP.

In the event of a new λ-LSP to be set up, the head-end OCC must allocate an identifier for the interface associated to the yet to be created FA-LSP. Next, it originates a *PATH* message containing the *LSP_TUNNEL_INTERFACE_ID* object filled with the selected local interface identifier, along with the local optical node identifier. When the *PATH* message arrives to destination, the tail-end OCC must allocate an identifier for that FA-LSP end. This one is called the remote FA-LSP interface identifier, and it is reported back to the head-end within the *RESV* message. As soon as the λ-LSP has been created, the head-end OCC advertises it as a forwarding adjacency by means of the OSPF-TE protocol. Being the FA-LSP bidirectional, it is also advertised by the tail-end OCC. All OCCs receiving the advertisement update their link state database adding a new FA link between the involved nodes.

Besides, [RFC4206] defines the mechanism to signal an LSP to be tunneled over an existing FA-LSP. Here, the *PATH* message must contain an *IF_ID RSVP_HOP* object [RFC3473] carrying the interface identifier of the FA-LSP to be used. Once the LSP has been created, the updated information regarding to the FA resource is disseminated through OSPF-TE by its head-end OCC.

From the routing perspective, no specific extensions are defined since an FA constructed out of an LSP is disseminated as a regular unnumbered link [RFC3477]. However, some guidelines are given in [RFC4206] about how to use the traffic engineering parameters related to an FA. The list below highlights the most relevant TE routing parameters in the context of this thesis. Therein, the parameters related to the bandwidth use are checked by the OCC during operation to assure that the FA resource is still able to tunnel client LSPs. In addition, the TE metric parameter fosters the use of FAs in front of regular unallocated data links, in order to achieve a better resource utilization.

- Link Local/Remote Identifier: This attribute takes the values of the local and remote interface identifiers exchanged during the signaling process (by means of the *LSP_TUNNEL_INTERFACE_ID* object), which identify the FA as an unnumbered link.
- Traffic Engineering Metric: By default the TE metric on the FA is set to *max(1, (the TE metric of the FA-LSP path) - 1)* so that it attracts traffic in preference to setting up a new LSP.
- Maximum Bandwidth: By default, the Maximum Reservable Bandwidth and the initial Maximum LSP Bandwidth for all priorities of the FA is set to the bandwidth of the FA-LSP.
- Unreserved Bandwidth: The initial unreserved bandwidth for all priority levels of the FA is set to the bandwidth of the FA-LSP. This field is updated as new client LSPs are tunneled across the FA by subtracting the client LSP requested bandwidth from the previous value.

4.2 Evaluation of an experimental GMPLS-enabled multi-layer control plane

In section 4.1, the extensions defined to the GMPLS protocol stack by the standardization bodies have been explained. This section reports all the modifications and developments that have been applied to the CARISMA control plane to provide it with multi-layer support including not only protocol extensions, but also the procedures added to the control plane modules.

The traffic aggregation scheme involves the three main blocks of the control plane, namely the LRM since an FA is a new kind of resource that has to be stored and managed, the RC which advertises this new resource and uses it in route calculation, and the CC as the responsible module for signaling both regular and FA LSPs.

Besides, modifications to signaling and routing protocols have been developed. In contrast, the LMP protocol is not affected by the inclusion of such traffic aggregation capabilities.

For sake of clarification, the extensions applied to the control plane functional blocks have been separated from the description of the extensions added to the routing and signaling protocol implementations.

4.2.1 Forwarding adjacencies mechanism implementation

In current GMPLS standardization there is an intrinsic association between the signaling of a new client LSP and the creation of the required λ-LSP to support it [RFC4206]. Hence, a route from source to destination is computed upon client LSP request, which may be constituted of both unallocated data links and already existent FA-LSPs. In the case that no FA-LSP is comprised along the route, a new λ-LSP is typically set up from source to destination to support the incoming request. Otherwise, λ-LSPs are set up to provide connectivity on those route segments where no FA-LSP is yet established. In light of this, three possible scenarios can be found when establishing a client LSP in a multi-layer environment:

- A client LSP which only uses data links: In this case an end-to-end FA-LSP is created and logically associated to the client LSP. In the transport plane, the client data flow is tunneled through the new λ-LSP.
- A client LSP which uses both data links and forwarding adjacencies: New FA-LSPs are created in the sections composed of consecutive data links (satisfying the WCC). Each used FA-LSP is associated to the new client LSP and the client data flow is transported through the concatenation of λ-LSPs.
- A client LSP which only uses forwarding adjacencies: No FA is created. Like in the previous case, the used FA-LSPs are associated to the new client path and the occupied bandwidth of each one is updated. From the transport plane perspective, the client LSP is tunneled through the concatenation of λ-LSPs.

Figure 4.3 (a) depicts the first situation in the list. There, a new client LSP request from E to D arrives to the CC of the source node E. The CC requests a route to its RC which calculates the shortest path using the information contained in the traffic engineering database (i.e. data links and existing FAs). Particularly in this first case, the route is only composed by data links so, an end-to-end λ-LSP from E to D will be created to tunnel the client LSP. Furthermore, the newly established λ-LSP will be associated to a new FA-LSP resource ($FA_{E\text{-}D}$) which, in turn, will be available to tunnel as many client LSPs as its bandwidth capacity can afford.

The RC builds the *ERO* from the computed route and passes it back to the CC which starts the signaling process by means of the RSVP-TE protocol. As the source node of the yet to be created λ-LSP, the CC in A pops the first *ERO* subobject containing the node and the interface identifiers of the data link that has to be reserved by the LRM. Additionally, the CC asks to the LRM for the local identifier to be associated to the new FA resource (E1 in the figure). This value is used to create the *LSP_TUNNEL_INTERFACE_ID* object which is added to the *PATH* message. Finally, the *PATH* message is forwarded to the next hop of the *ERO*. Each intermediate node (i.e., F and G) of the yet to be created λ-LSP proceeds as in the usual single-layer signaling mechanism to reserve the requested local resource (data link). Upon the arrival of the *PATH* message, the destination node D is responsible for assigning a remote identifier (D1) to $FA_{E\text{-}D}$ in the same manner as done in node A, that is, by asking to the LRM for a value for the remote *LSP_TUNNEL_INTERFACE_ID* object. This remote identifier is sent to the source node through the *RESV* message initiated by the destination node. Like in the single layer scenario, the *RESV* message is also used to allocate the data links assigned to the new λ-LSP. Once the *RESV* message reaches the source node, the new client and λ LSPs are established and the new FA-LSP resource is created in the scope of the LRM. Eventually, the LRM notifies about the created resource to the local RC which, in turn, disseminates it as a new Opaque LSA to the remainder OCCs in the network by means of the extended OSPF-TE protocol. Note that the newly created FA resource presents an allocated bandwidth equal to the one requested by the client LSP.

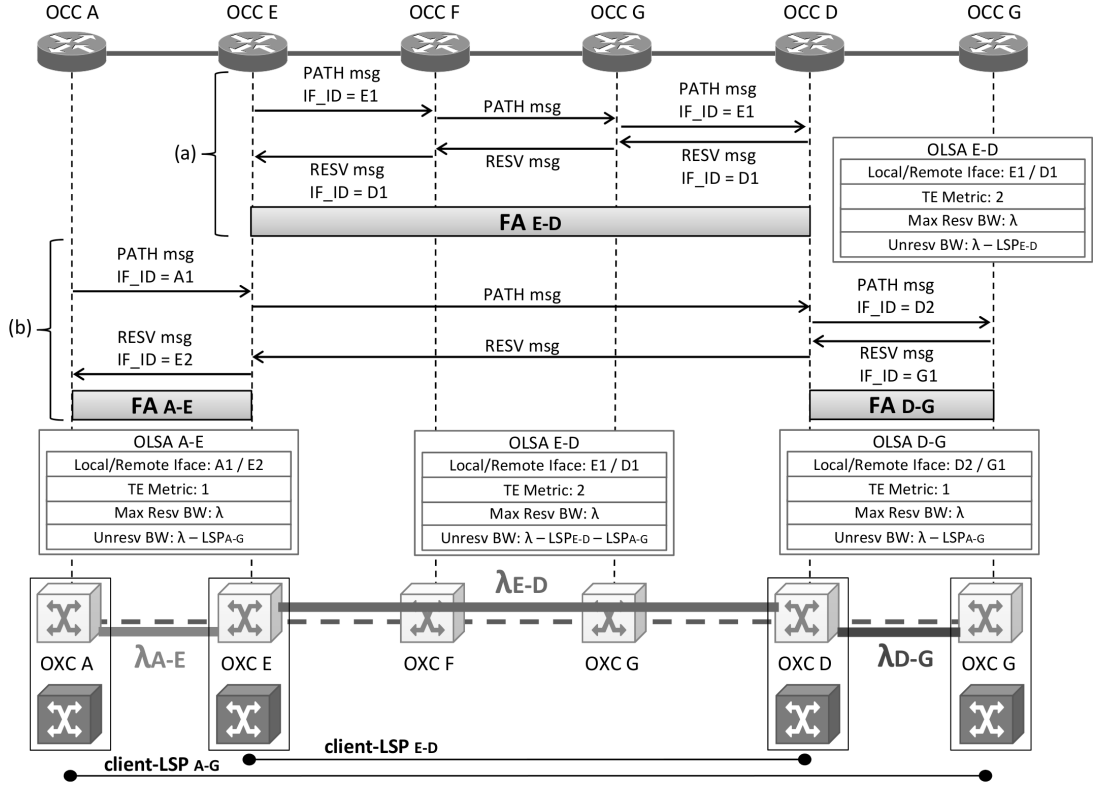


FIGURE 4.3: Example of end-to-end FA-LSP establishment.

In the second case, shown in Figure 4.3 (b), a client LSP between nodes A and G is requested. There, the best calculated route contains data links (one from A to E and another one from D to G) and the previously created FA ($FA_{E\text{-}D}$). In this situation, new FA-LSPs (each one associated to a λ-LSP) are created from the segments of the route composed by consecutive data links, and the bandwidth requested by the new client LSP has to be allocated from the already active $FA_{E\text{-}D}$. As shown in the figure, the mechanism for creating FAs in this context is exactly the same as the one described for the first scenario. Specifically, two FAs (from A to E and from D to G) are created in this example. Once the *ERO* has been created in node A, the signaling process is started by assigning the identifier A1 to $FA_{A\text{-}E}$. Upon the reception of the *PATH* message, the CC at node E notices that the incoming resource is a data link (so a new FA will be created) and the outgoing resource corresponds to the already existing $FA_{E\text{-}D}$. Therefore, being the destination node of $FA_{A\text{-}E}$, node E requests an identifier E2 that will be sent back to node A by means of the *RESV* message. Furthermore, node E allocates the client LSP requested bandwidth from $FA_{E\text{-}D}$ and forwards the *PATH* message to the downstream node which, in that case, results to be node D since it is the tail-end of the used $FA_{E\text{-}D}$. When the *PATH* reaches node D, the local CC checks that the incoming resource is an FA but the outgoing one is a data link so, the reverse process from node E has to be initiated. Herein, node D requests an identifier D2 for the new FA (from D to G) and forwards the *PATH* message. Finally, node G assigns the remote identifier G1 to $FA_{D\text{-}G}$. As explained in the previous scenario, the *RESV* message is used to disseminate remote FA identifiers as well as for allocating the network resources. The new FA resources ($FA_{A\text{-}E}$ and $FA_{D\text{-}G}$) are disseminated along with the new available bandwidth of the already existing $FA_{E\text{-}D}$.

As can be extracted from the previous case, in a scenario using only FAs the head-end node of each FA allocates the requested bandwidth during the signaling but no FA creation process is triggered. The used FA resources with their updated bandwidth information are flooded by means of the extended OSPF-TE.

4.2.2 CARISMA control plane module extensions

Targeting at the proper management of the FA entity, the functional blocks of the control plane need to be extended. In this regard, new structures are necessary to keep the new FA resource type, as well as new communication flows need to be established between control plane modules in order to exchange FAs related information. These new functionalities are thoroughly detailed in next paragraphs.

In a multi-layer context lightpaths become FA-LSPs that may contain several client paths (client LSPs) each one using a portion of the whole FA's bandwidth. As the responsible module for maintaining lightpaths' related information, the CC needs to be extended to manage an LSP hierarchy [RFC4206] (a two level LSP hierarchy in the case of study). To do this, the structures of the CC and, in particular, the PSB need to be modified to not only keep FA-LSP information, but also the information of the client LSPs they are

carrying. In this way, during the signaling of a new client LSP, each traversed CC creates a new client PSB instance containing the data related to such LSP. If a new FA-LSP is going to be created, a new FA PSB instance will be allocated as well and associated to the client PSB. Otherwise, if an existing FA is going to be used, the already existing FA PSB instance will be associated to the newly created client PSB.

The FA creation process described in section 4.2.1 details the interaction between the CC and the LRM to assign an interface identifier to the newly created FA, as well as to allocate the requested bandwidth of an existing one. This interaction is an extension implemented to the already existing communication between the CC and the LRM (whose former version was detailed in 3.2.1 and 3.2.3).

As previously mentioned, once created, FA-LSPs become new routing resources that have to be stored and managed by the control plane. In section 3.2.1, the LRM is presented as the responsible module for keeping local resources' information (i.e. local TE links and data links). In the new scenario, where the FA resource type is introduced, it is clear that the LRM has to be provided with the capacity for both storing the local node FA resources and managing them properly. Thus, while in a single layer network (where a whole wavelength is allocated for each client LSP) the resources have essentially two states, namely free or occupied, in a multiple layer scenario (where a client LSP can be established using just a portion of the total bandwidth of the transport resource) a more sophisticated resource management is needed. Concretely, an available bandwidth control has to be added to the former free/occupied status check. Furthermore, an association between the FA and the data links it uses is needed, and the integrity of the information associated to both resources (e.g. the available and used bandwidth) has to be maintained.

Additionally, the communication interface between the RC and the LRM (sections 3.2.1 and 3.2.2) needs to be extended so the RC can be notified about the creation, deletion or modification of the local FAs. To this end, FA related operations are added to such interface's communication protocol.

From the RC viewpoint, the most relevant modifications necessary to enable grooming capabilities include, first of all, the above-mentioned extensions to the RC-LRM communication interface at the RC side. Additional structures and mechanisms to manage FA resources are needed as well. For instance, the TED has to be extended to store the new resource type. Moreover, in order to disseminate FAs' information by means of the OSPF-TE protocol (as illustrated in the previous subsection) new procedures have to be developed to transform each FA resource, as it comes from the LRM, to an OLSA.

Last but not least, RWA algorithms have to be extended to use FAs. In this regard, the management of the so-called wavelength continuity constraint changes according to the new routing scenario, since wavelength continuity can be broken between two segments of a lightpath linked by a forwarding adjacency. This is due to the fact that the optical signal is extracted at the tail-end of an FA and electrically processed, so that transported flows ending at this point can be forwarded to the upper layer client. Nonetheless,

the remainder client flows, which end at different locations, have to be inserted again into the optical layer. Then, since an implicit O/E/O conversion is performed, different wavelength channels can be used by each FA-LSP.

4.2.3 CARISMA control plane protocol extensions

Like the control plane modules, the GMPLS protocol stack also needs to be modified in order to cope with the requirements posed by traffic aggregation schemes. As detailed in section 4.1.2, the IETF defines a set of mechanisms to provide both routing and signaling protocols with such traffic aggregation capabilities. These proposals have been taken as the basis for the development of the grooming techniques in the CARISMA GMPLS protocol suite implementation.

With respect to the RSVP-TE protocol, the previously presented extensions, namely the *LSP_TUNNEL_INTERFACE_ID* and the *IF_ID RSVP_HOP* objects, plus their associated procedures have been implemented. As an example, Figure 4.4 shows the use of the *LSP_TUNNEL_INTERFACE_ID* during the signaling process of a new FA-LSP. The picture contains the trace of the *PATH* message created by the head-end of the FA, and underlines the conveyed local interface identifier (i.e., "0x8000001A", or "-2147483622" in the signed integer format).

```
▷ Frame 1: 232 bytes on wire (1856 bits), 232 bytes captured (1856 bits)
▷ Linux cooked capture
▷ Internet Protocol Version 4, Src: 192.168.101.10 (192.168.101.10), Dst: 192.168.101.11 (192.168.101.11)
▽ Resource ReserVation Protocol (RSVP): PATH Message. SESSION: IPv4-LSP, Destination 192.168.101.12, Tunnel ID 1, Ext ID 0. SEND
  ▷ RSVP Header. PATH Message.
  ▷ SESSION: IPv4-LSP, Destination 192.168.101.12, Tunnel ID 1, Ext ID 0.
  ▷ HOP: IPv4 IF-ID. Control IPv4: 192.168.101.10. Data If-Index: 192.168.101.10, 34144522.
  ▷ TIME VALUES: 30000 ms
  ▽ EXPLICIT ROUTE: Unnum 192.168.101.11/34013444, Unnum 192.168.101.12/16908546
      Length: 28
      Object class: EXPLICIT ROUTE object (20)
      C-type: 1
    ▷ Unnumbered Interface-ID - 192.168.101.11, 34013444, Strict
    ▷ Unnumbered Interface-ID - 192.168.101.12, 16908546, Strict
  ▷ LABEL REQUEST: Generalized: LSP Encoding=SDH ITU-T G.707 / SONET ANSI T1.105, Switching Type=Time-Division-Multiplex Capable
  ▷ LABEL SET: Inclusive list, Generalized Label: 637534212
  ▷ SENDER TEMPLATE: IPv4-LSP, Tunnel Source: 192.168.101.10, LSP ID: 12289.
  ▷ SENDER TSPEC: IntServ, Token Bucket, 0 bytes/sec.
  ▽ RECORD ROUTE: Unnum 192.168.101.10/34144522
      Length: 16
      Object class: RECORD ROUTE object (21)
      C-type: 1
    ▽ Unnumbered Interface-ID - 192.168.101.10, 34144522,
        Type: 4 (Unnumbered Interface-ID)
        Length: 12
      ▷ Flags: 0x00
        Router-ID: 192.168.101.10
        Interface-ID: 34144522
  ▷ SUGGESTED LABEL: Generalized: 0x26000004
  ▷ UPSTREAM LABEL: Generalized: 0x26000004
  ▽ LSP INTERFACE-ID: Unnumbered, Router-ID 192.168.101.10, Interface-ID -2147483622
      Length: 12
      Object class: LSP-TUNNEL INTERFACE-ID object (193)
      C-type: 1 - Unnumbered interface
      Router ID: 192.168.101.10
      Interface ID:-2147483622
```

FIGURE 4.4: PATH message for the establishment of a new FA-LSP.

In addition, Figure 4.5 shows the signaling trace of a second client LSP that uses the previously created FA. Therein, the *IF_ID RSVP_HOP* object, which contains the router

and interface identifiers of the used FA is highlighted. Looking at the *Record Route Object* (*RRO*, [RFC3209]), the use of the FA resource can be observed as well.

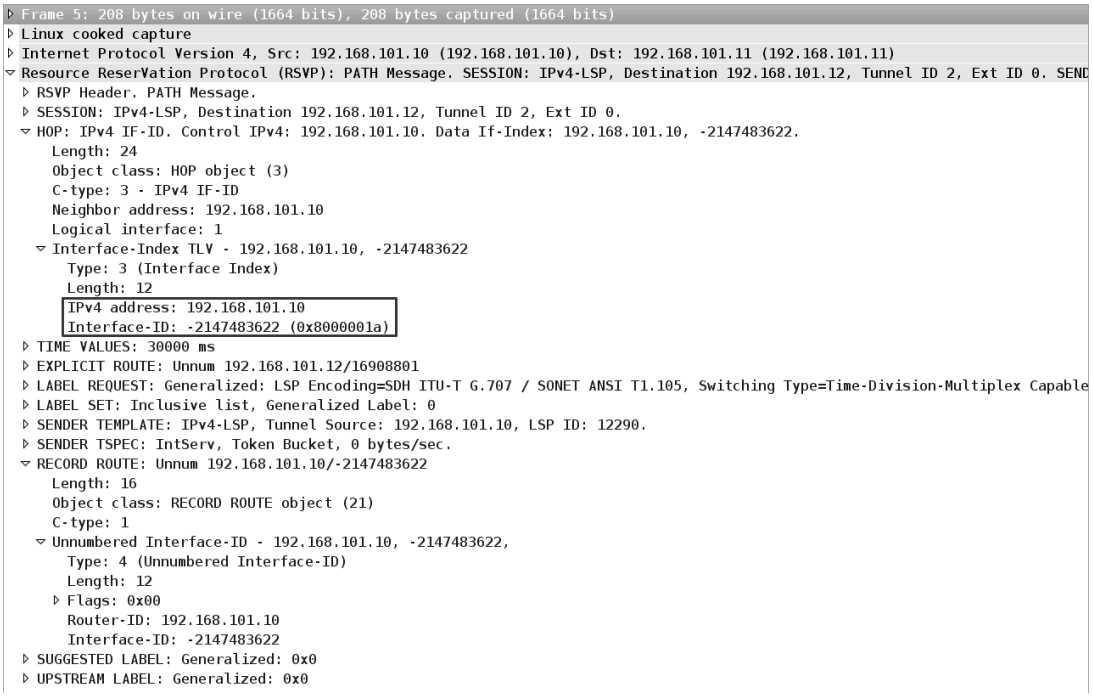


FIGURE 4.5: PATH message of an LSP using an already active FA.

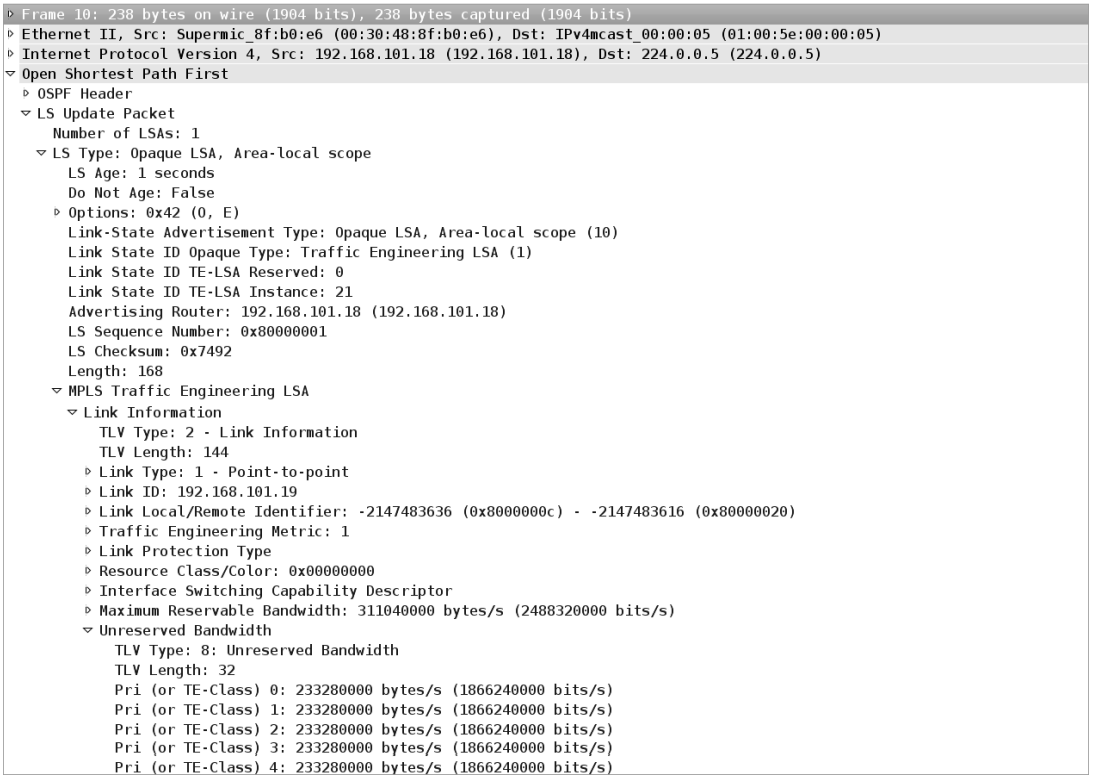


FIGURE 4.6: Opaque LSA associated to a just created forwarding adjacency.

Similarly, the CARISMA OSPF-TE implementation has been updated according to the proposed in [RFC4206]. Figure 4.6 depicts the OLSA resulting from a newly created

FA-LSP. There, the unreserved bandwidth field has been updated subtracting to its previous value the bandwidth requested by the new client LSP. As stated by the IETF, the maximum reservable bandwidth keeps the maximum value corresponding to the whole bandwidth assigned to the FA-LSP. Besides this, the traffic engineering metric field is conveyed in the OLSA and it is assigned the value resulting from *max(1, (the TE metric of the FA-LSP path) - 1)*. It is also noteworthy that the link local and remote identifiers of the OLSA take the values of the FA identifiers, which have been previously exchanged during the signaling phase.

4.2.4 Experimental results

The experimental assessment of the extensions described along this section, has been carried out over the ASON/GMPLS CARISMA test bed (introduced in chapter 3). To perform this evaluation, a 9-Node meshed optical transport network with 8 wavelengths per link (following the topology depicted in Appendix A.1) has been considered. The same topology has been assumed in the control plane, that is, the configured control plane has been deployed associated with the nine-node grooming-capable optical transport network, as shown in Figure 4.7.

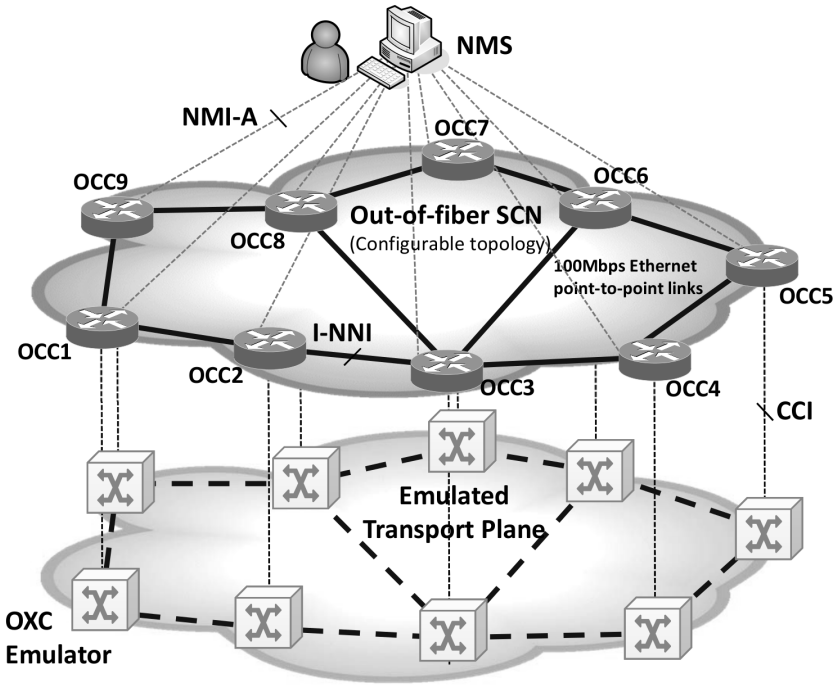


FIGURE 4.7: The ASON/GMPLS CARISMA test bed describing the 9-Node network scenario.

Departing from this network scenario, the FA-LSP mechanism implementing the GMPLS-controlled traffic grooming has been analyzed and compared to the all-optical solution, where an end-to-end lightpath allocating the whole wavelength capacity is established per client LSP request.

For the traffic characteristics, client LSP requests follow a Poisson process with mean Inter-Arrival Time (IAT) equal to $1/\lambda$. For each request source and destination nodes are uniformly selected. For connection duration, an exponential distribution with mean Holding Time (HT) set to $1/\mu$ has been assumed. Besides, the requested bandwidth of all incoming client LSP requests has been considered to be 1/4 of the total wavelength capacity.

Figure 4.8 plots the obtained client LSP blocking probability, for each architectural solution, as a function of the load offered to the network ($\lambda/\mu = HT/IAT$), normalized to a value of 200. As expected, significantly better resource usage is achieved when implementing FA-LSP capabilities in the network. For instance, if 0.5% client LSP blocking probability would have to be assured, a maximum load L=0.04 could be offered to a pure all-optical wavelength-routed optical network. Conversely, it could be drastically increased to L=0.5 if FA-LSPs were implemented. This ΔL=0.46 experimentally assesses the FA-LSP capabilities to automatically manage grooming actions in future transport networks, given the limitations of pure wavelength-routed optical networks to allocate incoming sub-lambda client LSP requests. In fact, as a whole wavelength is allocated for each client LSP request demanding 1/4 of the wavelength total capacity, 3/4 of the total network capacity is directly wasted. Qualitatively speaking, this approximately results in four times less traffic carried by the network.

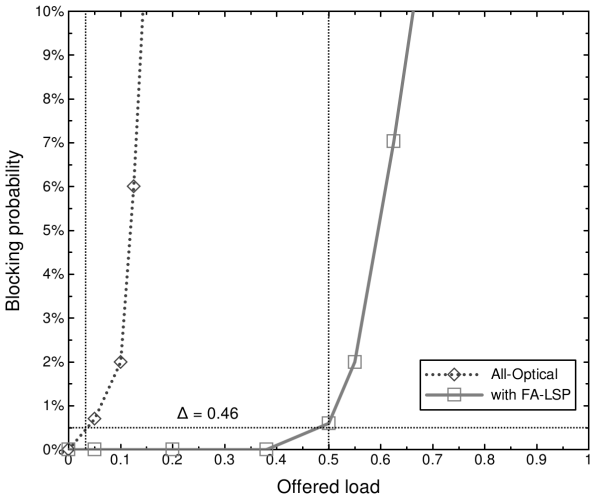


FIGURE 4.8: Blocking probability vs. offered load.

4.3 Traffic Engineering in GMPLS-controlled multi-layer optical networks

4.3.1 Forwarding adjacency creation cost function

In the current GMPLS standardization [RFC4206] there is an intrinsic association between the signaling of new client LSPs and the creation of the required λ-LSP to support them. In consequence, an association between the client LSP establishment and the FA resource resulting from the λ-LSP establishment can be inferred, since the establishment of a new λ-LSP results into the creation of a new FA-LSP. As detailed in section 4.2.1, a route from source to destination is computed upon client LSP request, which may be constituted of both unallocated data links and already existent FA-LSPs. In the case that no FA-LSP is comprised along the route, a new λ-LSP is typically set up from source to destination to support the incoming request. Otherwise, λ-LSPs are set up to provide connectivity on those route segments where no FA-LSP is established yet. This operation, however, may lead to resource waste in the network. Notice that long FA-LSPs connecting far-off nodes may be limited to be only reused by incoming LSP requests between remote endpoints. Hence, it appears more appropriate to separate the signaling functionality from the λ-LSP creation, so that λ-LSP placement can be decided based on network characteristics.

In the present work, the routing metric of the already created FA resources is set to be *max (1, FA-LSP hops-1)*, as described in [RFC4206]. Besides, the routing metric assigned to the unallocated data links spanning one single physical hop is set to 1. Nonetheless, aiming at better resource utilization, λ-LSP establishment is dissociated from network signaling functionality in the following way. Once the route from source to destination is calculated, the heuristic cost function depicted in Equation 4.1 is applied to the route segments where connectivity is not yet existent, standing H for the number of hops of the yet-to-be-created λ-LSP and p_H for the probability that any incoming demand in the network has a certain number of hops H.

$$C_{FA}(H) = H[(1 - p_H) + h/H] \tag{4.1}$$

The cost function provides the most appropriate λ-LSP configuration to optically connect the yet uncovered route segment. As will be later depicted by example, the term $(1-p_H)$ encourages those λ-LSP lengths close to the average network distance, thus being more likely to be reused. The term $1/H$ identifies the use of O/E port pairs per hop, so that the longer the λ-LSP, the lower the use of O/E ports to connect its endpoints. The tunable h parameter fosters/penalizes the use of O/E ports in the network. Finally, the total cost is multiplied by H as long as λ-LSPs need a higher number of unallocated data links. As an example in this context, the establishment of a new λ-LSP, which

will afterwards act as an FA-LSP, is assumed. This LSP has to be established between a node-pair distancing 4 hops and no intermediate FA resources are comprised in the calculated route. Supposing that a 2 hops client LSP length is the most likely in the network, the combination $C_{FA}(2) + C_{FA}(2)$, that is, two λ-LSPs spanning each one two hops, could have lower cost than $C_{FA}(4)$, meaning one single end-to-end λ-LSP. In the subsection below (4.3.1.1), $C_{FA}(H)$ is particularized for the scenario under study. As will be highlighted, very short FA-LSP establishment (e.g., 1 hop) is also penalized $C_{FA}(H)$, due to the large amount of required expensive O/E ports, as well as the large amount of bypass traffic to be electrically processed.

Figure 4.9 exemplifies the proposed scenario where the λ-LSP establishment is dissociated from the client LSP set up procedure, allowing in this way the establishment of several underlying λ-LSPs while signaling only one client LSP request. Specifically in the example, a client LSP (from A to D) is tunneled through two new FA-LSPs, each one spanning two hops, instead of creating a single four-hops FA-LSP.

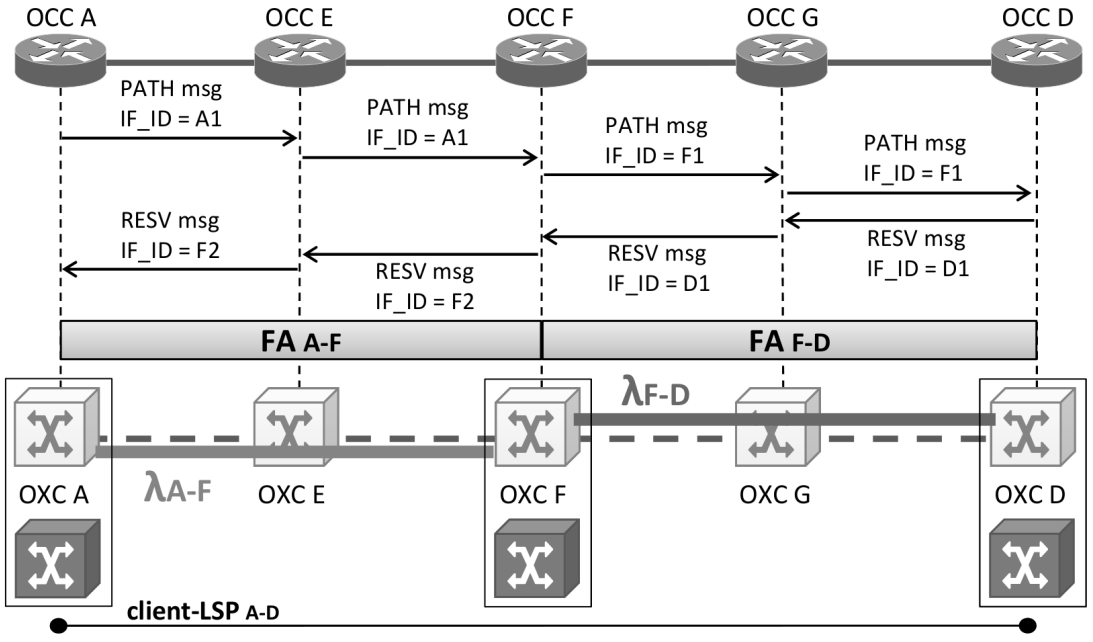


FIGURE 4.9: Establishment of two FA-LSPs to tunnel one single client LSP.

From the implementation perspective, to permit this separation between client LSP and λ-LSP setup, the head-end OCC evaluates the accumulated optical length, and uses loose hops in the *ERO* object [RFC3477] when it decides to divide the whole route segment into two or more underlying λ-LSPs. The same mechanism is used also when an intermediate λ-LSP must be created. Every intermediate OCC receiving a *PATH* message with a next hop set as loose must compute the next route segment possibly signaling a new λ-LSP.

It is worth pointing out that λ-LSPs created during client LSP setup signaling procedures have no specific RSVP-TE refresh messages. In fact, the λ-LSP must be torn down when releasing the last supported client LSP. In order to maintain the association between client LSPs and λ-LSP everywhere, the signaling of client LSPs using an already-existing

FA-LSP follows, in this implementation, the same path in the control plane as the signal at the optical layer. This means that the same OCC can process a single RSVP-TE *PATH* or *RESV* message more than once. For example, in Figure 4.10 nodes E and D are directly connected through $FA_{E\text{-}D}$. Supposing that the route of a new client LSP between nodes A and G includes $FA_{E\text{-}D}$ together with $DL_{A\text{-}E}$ and $DL_{D\text{-}G}$ links, two new FA-LSPs ($FA_{A\text{-}E}$ and $FA_{D\text{-}G}$) are created to support the new client LSP from A to G. Then, node G processes the RSVP-TE messages of the client LSP A-G two times, that is, the first time as an intermediate OCC and the second time as the tail-end OCC.

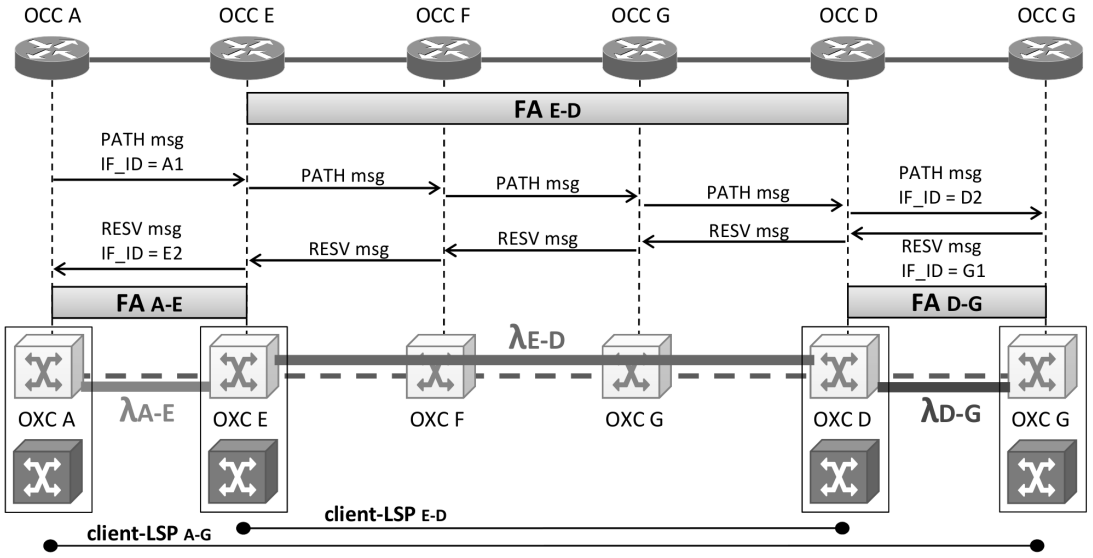


FIGURE 4.10: FA-LSP signaling example.

4.3.1.1 Experimental results

The GMPLS integrated routing performance has been evaluated in the ASON/GMPLS CARISMA test bed, which implements all FA-LSP related functionalities. For the evaluation, the previously used 9-Node scenario illustrated in Figure 4.7 with 8 bidirectional wavelengths per link has been assumed. There, uniformly distributed bidirectional client LSP requests demanding 1/4 of the total wavelength capacity arrive to the network following a Poisson process. In particular, λ-LSPs are routed in the network so that they accomplish the wavelength continuity constraint. Three different network architectural solutions have been analyzed: generic FA, FA with the creation cost function, hereafter referred as FA (CCF), and opaque. The first solution stands for a GMPLS-controlled optical transport network where end-to-end FAs are created per each client LSP request (as detailed in section 4.2). The second solution, FA (CCF), identifies the GMPLS-controlled traffic grooming exactly as explained in the previous subsection. Finally, the third solution contemplates an opaque transport network, where only single-hop λ-LSPs are allowed. Once established, these LSPs are advertised as FA-LSPs, thus permitting their re-use.

For the FA(CCF) solution, Figure 4.11 depicts the obtained $C_{FA}(H)$ function for the scenario under study. The bar graph plots the probability that an incoming client LSP request has a certain number of hops, assuming availability of resources through the shortest path. As seen, there is a 40% probability that an incoming request traverses 2 hops from source to destination. In contrast, only 5% of the incoming requests would traverse 4 hops. This validates the previous assumption, which stated that by splitting very long FA-LSPs into shorter ones resources are much more likely to be reused. Values greater than the network diameter (i.e., 4 hops in our scenario) have $p_H = 0.0$ and are not depicted in the figure. To finally obtain $C_{FA}(H)$, h is fixed to 0.5, as it provided the best network performance, while fulfilling the design criteria: $C_{FA}(2) + C_{FA}(2) < C_{FA}(4)$ and $C_{FA}(1) + C_{FA}(2) < C_{FA}(3)$.

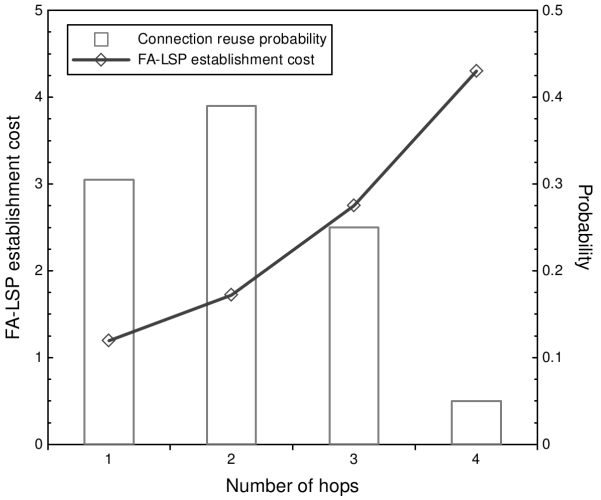


FIGURE 4.11: FA-LSP creation cost function for the 9-Node topology under study. Bar graph shows the probability that a client LSP request has 1, 2, 3 or 4 hops.

Figure 4.12 plots the obtained client LSP blocking probability as a function of the load offered to the network ($\lambda/\mu = HT/IAT$) normalized to a value of 200. The three above-mentioned scenarios have been represented: The *Opaque* curve depicts the results corresponding to an opaque transport network architecture where O/E/O conversions are performed at each node of the path, i.e. only single-hop FAs are allowed. The *Generic FA* curve represents the scenario where the standard-defined grooming solution, as presented in section 4.2, is applied. Finally, the curve *FA(CCF)* shows the figures obtained for a grooming capable scenario when the defined FA creation cost function policy is applied to the aggregation mechanism.

As could be expected, the best results in terms of blocking probability are achieved by the *Opaque* solution as it gives the highest flexibility regarding the wavelength usage. Conversely, the *Generic FA* solution offers the highest blocking as the creation and use of the FA is closely connected to the random nature of the offered connections. The *FA(CCF)* mechanism arises as an intermediate solution to the traffic aggregation problem since it offers a better FA utilization than *Generic FAs* resulting in a lower blocking

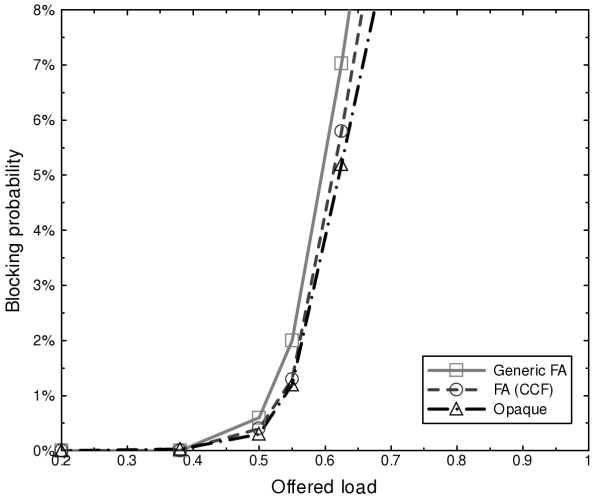


FIGURE 4.12: Blocking probability vs. offered load in the 9-Node network topology.

probability. As a matter of fact, looking at the load values ranging from 0.4 to 0.65 it can be noticed that the *FA(CCF)* solution performance is very close to the *Opaque* one. For example, the blocking probability is reduced from 2% (in the *Generic FA*) to around 1% in *FA(CCF)* when the load is 0.55, almost overlapping the figures obtained in the *Opaque* solution. Furthermore, as will be demonstrated in next subsection, these good blocking results are obtained using a reduced number of O/E ports, which drastically reduces the network deployment costs.

4.3.2 Resource utilization in a GMPLS-controlled grooming-capable optical transport network

Heretofore, the improved performance of traffic grooming solutions has been demonstrated in terms of the supported traffic load, which is increased thanks to a more efficient bandwidth usage. In this context, a step beyond has been done in the previous subsection (4.3.1), by searching an optimal length (in hops) for created FA-LSPs, which leads to an even better resource utilization.

In the same line, a further study is conducted in this section to analyze the resource utilization of the proposed traffic aggregation solutions in front of the existing architectures, namely all-optical and opaque. The figure of merit here is the number of O/E ports needed by each approach, which is used as an indicator of the network capital expenditures. Note that, the better use given to the ports, the lower number of ports would be necessary in the network, thus lowering the cost of its deployment (i.e., the CAPEX).

4.3.2.1 Experimental results

Two different transport network topologies have been deployed in the CARISMA test bed for evaluation purposes. On the one hand, the already used 9-Node topology (Appendix A.1, Figure 4.7) has been considered in order to obtain a complete view of the advantages and drawbacks of the models under study. On the other hand, a scenario composed of 16 nodes and 23 links (Appendix A.2) has been utilized to evaluate the performance in larger and more meshed transport networks. The test bed deployment for the latter case is depicted in Figure 4.13.

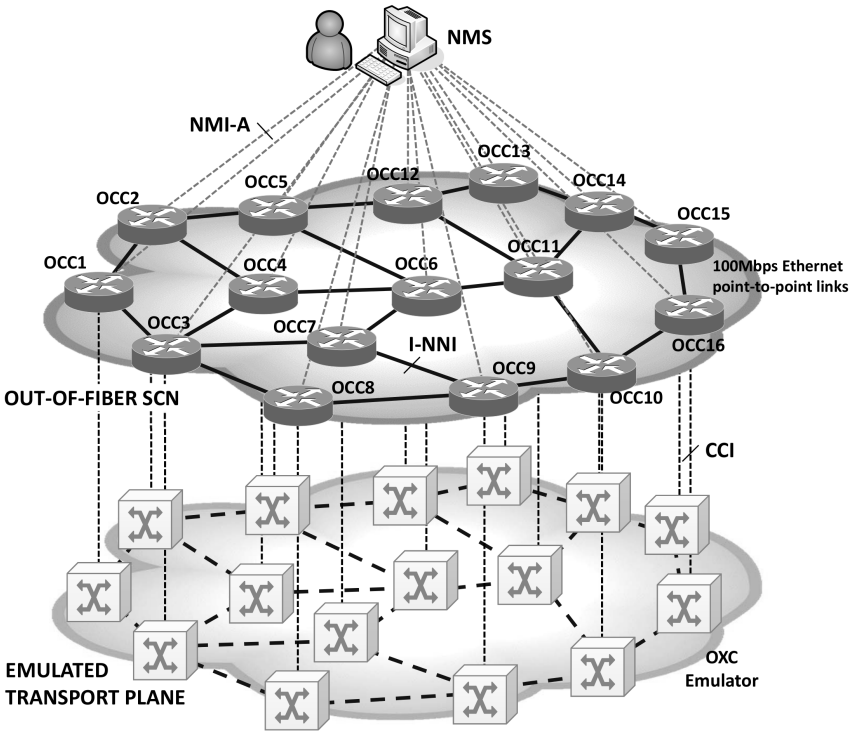


FIGURE 4.13: The ASON/GMPLS CARISMA test bed describing the 16-Node network scenario.

The assumed traffic characteristics are analogous to the ones used in the previous experiments, that is, uniformly distributed client LSP request arrivals following a Poisson process with a mean IAT of $1/\lambda$, and connections' duration following an exponential distribution with a mean HT set to $1/\mu$. As also assumed in previous tests, the requested bandwidth of all incoming client LSPs is considered to be 1/4 of the total wavelength capacity.

Figure 4.14 shows the network O/E port usage results for each solution in the 9-Node network topology, as an indicator of the related network CAPEX. To this end, the Y axis has been normalized to the maximum number of O/E ports that could be equipped in the network (one per output wavelength at each node). Here, the scenario implementing an all-optical network architecture has been added to give a lower boundary on such port usage. As seen, besides pleading for huge electronic routers, able to process all the

incoming information at the nodes, the opaque solution rapidly requires a large number of O/E ports, since an optical bypass is not possible in the network. On the other side, although the all-optical solution leads to the lowest port usage, this is mandated by its poor bandwidth efficiency, which forces it to consume all available wavelength channels. Interestingly, the FA-based solution lies in between. As seen before, it results into a blocking probability similar to the opaque architecture. Nonetheless, thanks to the optical bypass capability, it leads to a lower O/E port usage, thus reducing the total CAPEX. Note that for an offered load of 0.5, leading to a similar BP around 0.5% (Figure 4.12), the O/E port usage is decreased by 20%.

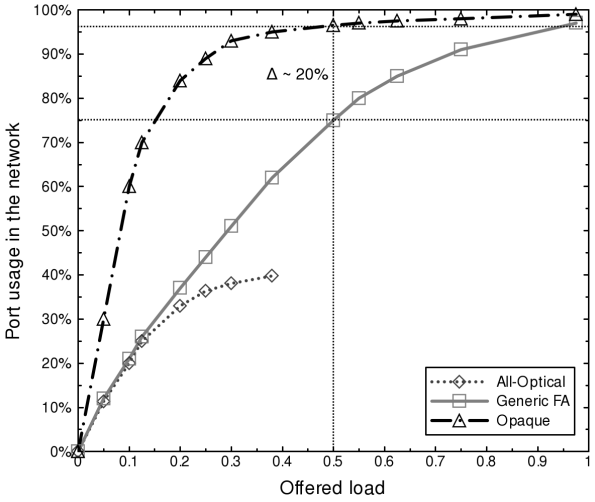


FIGURE 4.14: O/E port usage vs. offered load in the 9-Node network topology.

Figure 4.15 (a) shows the blocking probability results for each architectural solution, now in the 16-Node network topology. Again, the blocking probability figures are drastically reduced when implementing the FA-LSP functionality in the network. Here, fixing a blocking probability value around 0.5%, the offered load to the network in the generic FA and opaque solutions can be increased by 0.6 compared to the all-optical one.

Furthermore, in order to envisage the related network CAPEX in all-optical, generic FA and opaque solutions, Figure 4.15 (b) quantifies the O/E port usage in the 16-Node network topology. As done for the 9-Node network, the Y axis has been normalized to the maximum number of O/E ports that could be equipped in the network. Once again, the opaque solution results in the highest port usage, due to the O/E conversion forced at each node. Besides, as seen in the first network scenario, the FA-based solution leads to the best trade-off between bandwidth usage efficiency (in terms of blocking probability) and port usage. For instance, fixing the offered load to 0.7, which leads to a BP≈0.5% for both opaque and FA solutions, the O/E port usage is decreased by 15%.

For sake of completion, Figure 4.16 plots the O/E port utilization of the FA solution using the cost creation function for the 9-Node network scenario. As done in the sections above, this proposed mechanism is here compared to the all-optical, generic FA and opaque ones. For an offered load of 0.55, the figure shows that the FA(CCF) solution leads to the same

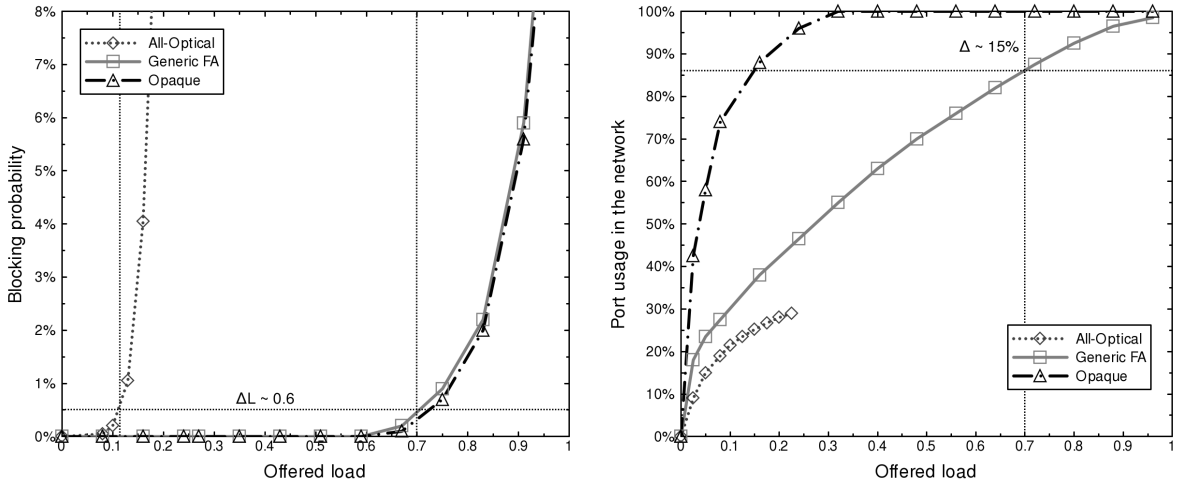


FIGURE 4.15: Experimental results for the 16-Node network topology: a) blocking probability vs. offered load; b) O/E port usage vs. offered load.

BP as the opaque one (around 1%, Figure 4.12), but drastically outperforming its O/E port usage, which is decreased by 15%. As could be expected, due to the FA-LSP length limitation performed by the creation cost function, the number of O/E ports used by this proposal is slightly increased compared to the generic FA mechanism, where the creation of longer FAs (spanning 3 or 4 hops) is not penalized, however resulting in a worse bandwidth utilization.

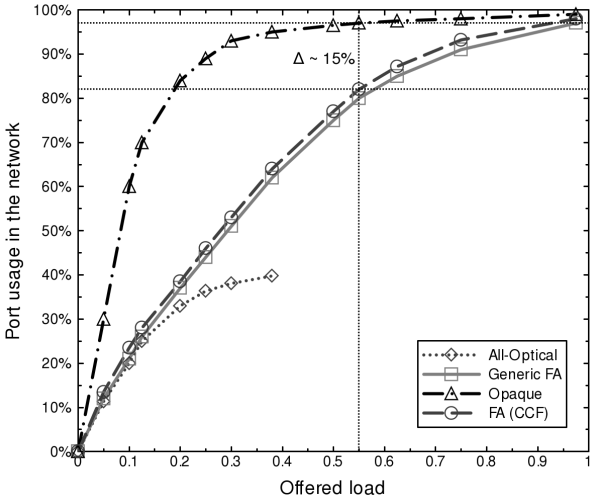


FIGURE 4.16: O/E port usage vs. offered load in the 9-Node network topology.

4.4 Chapter summary

This chapter has introduced traffic aggregation as the main driver to improve the efficiency of next-generation wavelength-routed optical transport networks. To start with, the traffic grooming concept and its architectural requirements have been detailed in the first section. Therein, the forwarding adjacency has been presented as the entity enabling the logical management of the transport plane grooming capabilities from the control plane. Moreover, standardization efforts already done in this context have been described.

Later on, the extensions applied to the CARISMA control plane to implement traffic aggregation support have been explained. Additionally, former experimental assessment has been presented in the second section.

To conclude the chapter, a mechanism to optimize the utilization of network resources has been proposed. While the standard FA support mechanism creates an FA which is assigned to and end-to-end lightpath, the proposed mechanism allows the establishment of several shorter FAs which are then concatenated to provide the service to the end-to-end client lightpath. Furthermore, aiming at better bandwidth utilization, a forwarding adjacency creation cost function has been defined. This cost function calculates the best FA splitting so the created FAs are more likely to be re-used.

Both the standard and proposed schemes have been compared in terms of blocking probability and O/E port usage, which are taken here as figures of merit to measure network efficiency. For evaluation, the all-optical and opaque network schemes have been taken as benchmarks for port usage and blocking, respectively. The experiments have been conducted over two network scenarios to evaluate the behavior of the mechanisms in different topological conditions. The obtained results show a dramatic outperformance of the traffic aggregation schemes in front of the all-optical scenario in terms of blocking probability. Moreover, traffic aggregation leads to similar blocking as the opaque scenario but drastically reducing the port usage. Besides, the results demonstrate that the proposed FA management scheme gives a better trade-off between port usage and blocking probability than the standard FA mechanism.

Chapter 5

Impairment-aware routing over a GMPLS-controlled optical transport network

In recent years the development of high-speed, transparent and reconfigurable optical networks, involving the employment of WDM line systems and optical switching technology, has attracted substantial attention. This approach, introduced in section 2.1 and featured by the removal of O/E/O conversion in the intermediate nodes of the optical network, allows efficient transportation of any type of data traffic, support of traffic growth, and brings tremendous cost savings [Jai05]. Although transparency, where signals propagate from source to destination through intervening nodes without opto-electronic conversions, is an important evolutionary step in fiber optic communications it nevertheless leads to some undesirable effects that are addressed along this chapter.

First, transparent networks are very sensitive to signal degradations since optical pulses are transmitted through the optical layer without any electrical regeneration. To face this problem, different architectural solutions are presented here to implement a control plane capable to provide lightpaths assuring an appropriate quality of transmission (QoT). Specifically, a GMPLS-based control plane implementation is extended and experimentally validated for each scheme.

Besides, in all-optical networks, failures propagate downstream through the lightpaths and therefore cannot be easily localized and isolated. The huge amount of information transported in optical networks, makes rapid fault localization and isolation a crucial requirement for providing guaranteed QoS and bounded unavailability times. The identification and location of failures in transparent optical networks is complex due to three factors: fault propagation, lack of digital information and large processing effort. In this regard, a fast lightpath restoration mechanism is presented and experimentally evaluated in this chapter.

5.1 Physical layer impairment constraint in optical transport networks

In transparent optical networks, the signal experiences the impact of a variety of quality degrading phenomena that are introduced by different types of signal distortions. These physical layer impairments (PLIs) accumulate along the path and limit the system reach and the overall network performance. There are distortions of almost *deterministic* type only related to the pulse stream of a single channel, while others have a statistical nature and are classified as *random*. Besides, some of these effects depend on the network topology and are categorized as *static*, while others are *dynamic* since they depend on the characteristics of the lightpaths already active in the network. Finally, PLIs may be classified into *linear* if they are independent of the power level of the signal, or into *non-linear* otherwise.

The list below gives a brief description of the impairments that are taken into account to determine lightpaths' QoT. In addition, Table 5.1 summarizes such impairments according to the above-mentioned characteristics. Further information about these effects can be found in [Agr02, AKM+09, RFC4054].

- *Chromatic Dispersion (CD).* Also known as *Group Velocity Dispersion* (GVD), the CD describes the broadening of the optical pulse as a consequence of the speed variation in the spectral components of the emitted optical power. Due to this phenomenon different optical pulses may overlap leading to Inter-Symbol Interference (ISI), thus increasing the bit error rate (BER). Although CD can be compensated using dispersion compensation fibers (DCF), there is an inherent interdependence between CD and non-linear effects (e.g., XPM) which also contributes to ISI.

- *Filter Concatenation (FC).* In optical transport networks, the signal traverses numerous optical components (such as WDM multiplexers and demultiplexers) each containing optical filters. Due to device imperfections, temperature variations and aging, the concatenation of optical filters causes the narrowing of the spectral width of the signal. As a static effect, the penalty induced by FC depends on the selected route since it will determine the number of optical filters to be traversed.

- *Self Phase Modulation (SPM).* The SPM is given by the fact that the phase of the signal is modulated by its own pulse intensity due to the *Kerr effect* ([Agr02]). More precisely, the non-linear variations of the refractive index of the fiber affecting to different intensity portions of the pulse result into different phase shifts, which give raise to the chirping (i.e., variation of the frequency spectrum) of the pulse.

- *Amplifier Spontaneous Emission Noise (ASE).* The ASE noise is produced by the optical amplifiers placed at the end of each fiber span to boost the power of the signal and, in this way, compensate fiber attenuation. The ASE noise, which acts

cumulatively through the chain of amplifiers, is propagated along with the signal consequently degrading the system performance.

- *Polarization Mode Dispersion (PMD).* Fiber irregularities (such as asymmetry) may cause different polarizations of the optical signal to travel at different speed and, in consequence, reach the receiver at different times. This effect can spread the pulse to make it overlap with other pulses or even change its own shape making it undetectable to the receiver. The difference between signal polarization modes' arrival times is called difference group delay (DGD), and it is used as a measure of the PMD.

- *Cross Phase Modulation (XPM).* The effect of the non-linear phase modulation of the signal caused by its own pulse intensity induced by the SPM can be increased because of the intensities of other signals. This phenomenon is referred to as XPM and may result into spectrum broadening and distortion of the pulse shape.

- *Four Wave Mixing (FWM).* When three optical signals co-propagate through the fiber, a fourth one is generated producing the FWM effect. If the frequency of the generated wave falls into the transmission window of the original ones, the FWM power interferes with the transmitted signals, imposing penalties.

Physical Layer Impairment	**Type**
Chromatic Dispersion (CD)	Deterministic / Static / Linear
Filter Concatenation (FC)	Deterministic / Static / Linear
Self Phase Modulation (SPM)	Deterministic / Static / Non-linear
Amplifier Spontaneous Emission (ASE)	Random / Static / Linear
Polarization Mode Dispersion (PMD)	Random / Static / Linear
Cross Phase Modulation (XPM)	Random / Dynamic / Non-linear
Four Wave Mixing (FWM)	Random / Dynamic / Non-linear

TABLE 5.1: PLI classification summary.

5.2 Impairment-aware lightpath provisioning strategies in optical networks

Two major strategies can be found in the literature to face the impairment-awareness in optical transport networks. The first one (static or offline) is based on traditional network design strategies, which have been widely used by network operators, consisting

of a looped process where the required network resources are carefully quantified for the forecast traffic demands and, thereafter, the network is upgraded during operation with the increase in traffic requirements. Much work can be found in the literature devoted to the static network design taking into account the PLIs that may affect to lightpaths (e.g. see, [YS99, AT02, SIL+98, BCM07]). In most of these contributions, RWA algorithms become either mathematical or heuristic solutions aimed to provide the optimal deployment of optical components to fulfill the requirements of a given traffic matrix within a given network topology, however, differing on the considered PLIs and feasibility constraints (BER, PMD, Q-factor, OSNR, etc.).

Nonetheless, the static network design presents some limitations. Even though network upgrades can be considered to adapt the network to new traffic requirements, these upgrades may affect the optical layer performance, which had been designed for a very specific scenario. Moreover, re-design tasks stemming from such upgrade are expensive and time-consuming. To lessen these drawbacks, network nodes capable of learning about the optical properties of the already deployed elements (as proposed in [CFH+06]) would be appropriated. Besides, in dynamic scenarios, where the traffic matrix is unknown or unpredictable, static solutions are difficult to apply.

The second strategy relies on a dynamic process where online IA-RWA computations are conducted upon the arrival of lightpath establishment requests. Different procedures can be followed here. Some studies (such as [RDF+99, HHM05, CCM05]) propose a two-step path computation, where the RWA is performed first and the optical feasibility of the computed lightpath is checked afterwards. Alternatively, a three-step mechanism, where spatial routes are calculated first, the wavelength assignment is then performed and the optical feasibility is checked in the last stage, is also evaluated in [HBP06a, HBP06b, HBPP+07]. A third strategy consists of considering PLI estimations during the RWA process (e.g., see [KTMT05, MBL+08, RSM09]).

In a context where flexibility in lightpath provisioning is mandatory due to the bursty characteristics of the traffic to be served, static network design becomes insufficient to provide an effective operation. Therefore, dynamic impairment-aware routing strategies take the token once the network is put into service. Moreover, in an ASON/GMPLS scenario, the dynamic model can be applied to the GMPLS-enabled control plane [MPC+06], thus leading to an impairment-aware GMPLS-based control plane.

Looking at the literature, two main strategies are proposed to introduce impairment-awareness in automatically managed transparent optical transport networks: centralized and distributed [MPC+06, CCV+07]. Centralized approaches assume the existence of a centralized server that is reachable by all the network elements and aware of the complete network topology, resource availability and physical layer parameters, which are used during the routing and wavelength assignment process. Hence, centralized approaches are able to guarantee and satisfy a set of lightpath specific requirements such as bandwidth, path diversity, QoT, latency, etc. Distributed approaches assume that all NEs in the

network use a common distributed control plane which manages the required procedures for establishing lightpaths (i.e. routing and signaling).

Assuming a GMPLS-controlled optical transport network, either centralized or distributed impairment-aware control plane architectures may require some protocol extensions in order to be implemented. These extensions can be applied to the signaling protocol (signaling-based approach) or to the routing protocol (routing-based approach). In the former, the signaling protocol (e.g., RSVP-TE) is extended to include PLI information which is collected along the path (during the signaling phase) until the destination node that evaluates the feasibility of the lightpath. In contrast, the routing protocol (e.g., OSPF-TE) is kept in its standard version. Contrariwise, in the routing-based scheme, the routing protocol is extended to disseminate PLI information (e.g., [RFC4054]). Thus, global PLI information is kept by the routing entity and used for route calculation. Herein, RSVP-TE protocol is not extended since the lightpath feasibility is known at route calculation time. Different intermediate approaches applying extensions to both routing and signaling can be also considered [SS09]. Hence, different proposals can be found in the literature depending on the selected architectural solution (centralized or distributed) and the protocol to be extended. For example, [DS05] evaluates by simulation two centralized adaptive routing algorithms assuming OSPF-TE to disseminate Q-factor information. Conversely, [LHDF07], [PH06], and [SYS+09] propose OSPF-TE extensions for impairment-aware route computation in a distributed control plane. In addition, [SYS+09] compares this proposal to another distributed one, however, based on RSVP-TE extensions. This architectural solution is also studied in [CSA+08]. A fourth option is taken into consideration in [CPVC07], where a PCE-based centralized scheme is provided with RSVP-TE extensions to evaluate lightpaths feasibility. Finally, the authors in [PSLR+11] introduce the *probe-based IA-RWA* concept. In probe-based IA-RWA, simplified QoT estimation models are used during the RWA process. Then a lightpath *pre-setup* is done, and the QoT performance of the reserved lightpath is verified prior to starting the transmission of client data. This verification is done through sending probe traffic over the lightpath. Only if the QoT measured at the egress node is acceptable is the lightpath considered suitable and activated for client traffic. This probe-based IA-RWA is applied to and evaluated in both routing and signaling-based approaches.

In general, although providing more accurate lightpaths, routing-based schemes present convergence limitations in front of the scalability of the signaling-based, due to the flooding process necessary for PLI information dissemination. Therefore, different proposals try to achieve the best trade-off between network performance and scalability.

In this chapter, a centralized architectural proposal implementing a GMPLS-enabled control plane is presented and studied to deal with the physical layer impairment constraint in dynamic optical transport networks. Such scheme is validated against a distributed one which is also described. The standardization implications are detailed as well.

5.3 Development scenario: The DICONET project

The research presented along this chapter has been realized in the framework of the "Dynamic Impairment Constraint Networking for Transparent Mesh Optical Networks" (DICONET, FP7-216338) project. In this section a description of the project, its main objectives and actors is done.

5.3.1 Project description: Motivation and objectives

The evolution of core networks towards all-optical models has been motivated in previous sections. As said, this evolution aims at improved cost economics, since all-optical networks reduce the number of O/E/O devices to be used (i.e. the CAPEX), also leading to reduced operation efforts and increased scalability due to the simplification of the network architecture. In this context, the DICONET project was funded by the ICT Framework Program 7 of the European Commission to contribute to the strategic objective "The Network of the Future" by supporting innovative solutions and technologies for intelligent and transparent optical networks. The goal of the DICONET project was to deal with the main limitations inherent to these networks which are the physical layer impairments that accumulate along the optical path, and failure localization since failures propagate in such a transparent environment making them difficult to localize and isolate. In light of this, the vision of the project was that the intelligence in core optical networks, which embraces the functions that are positioned in the management and control plane of the network, should be fed with information related to the data plane on the optical layer using a cross-layer approach. Such an approach leans on the monitoring and management of the physical layer impairments plus using the collected information during the network operation, that is, in RWA and failure localization procedures. These tasks were thought to be carried out by the network planning and operation tool (NPOT, described in next subsection) which would reside in the nodes of the core network and would be integrated into a GMPLS-enabled control plane. Hence, new impairment-aware routing and wavelength assignment algorithms to manage PLI information, efficient failure localization to lead with transparency, and the extension of the GMPLS protocols for a standardized performance arose as the key challenges of the project.

5.3.2 The Network Planning and Operation Tool (NPOT)

The main objective of the DICONET project was to design and develop a Network Planning and Operation Tool (NPOT, [Azo10]). The NPOT would incorporate real-time assessments of optical layer performance into impairment-aware RWA, component placement, and failure handling algorithms and would be integrated into a unified GMPLS-based control plane. While network planning is more focused on the details of how to

accommodate the traffic that will be carried by the network, which typically occurs before a network is deployed [APA+09, MCK+09]; in network operation phase, which is the focus of this thesis, the demands are processed upon their arrival and one at a time. It is assumed that the traffic must be accommodated using whatever equipment already deployed in the network [KCMV09, APA+10b].

Figure 5.1 depicts the anatomy of the DICONET NPOT which includes network description repositories, a QoT estimator (Q-Tool), IA-RWA engines, a component placement module and a failure handling module as its key building blocks. The planning mode specific modules are accessible through a command line interface (CLI). The communication of the NPOT with the other entities in the DICONET control plane (operation mode) is realized by means of a messaging protocol layer designed and implemented on top of the standard TCP/IP socket interface. The modular design of the NPOT paves the way to enhance or upgrade each of its building blocks without affecting the overall functionality of the tool. In fact, the DICONET NPOT evolves as new modules or algorithms are replaced with the existing ones. In the sequel, each of these building blocks is presented.

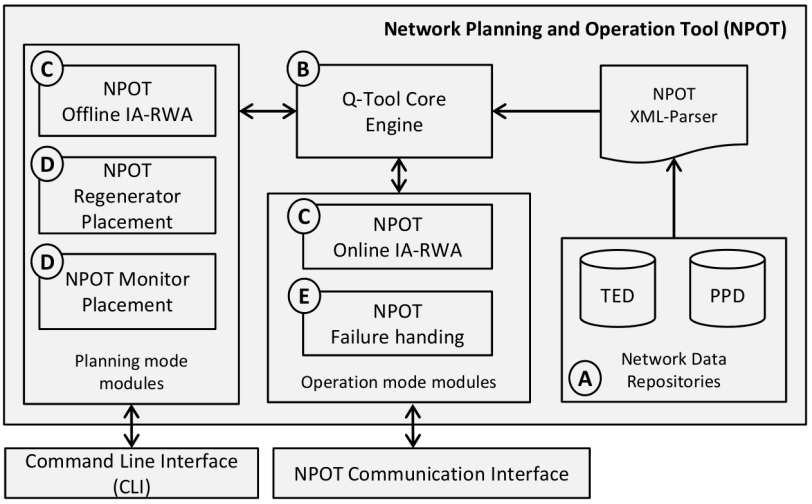


FIGURE 5.1: Anatomy of the DICONET NPOT: A) Network description repositories, B) QoT estimator, C) IA-RWA engines, D) Component placement modules, and E) Failure handling module.

5.3.2.1 Network description repositories

The network description (at both physical layer and topology level) is included in two main repositories that are kept as external databases to the NPOT. The Physical Parameters Database (PPD) is the master repository, which includes the physical characteristics of the links, nodes and components in the transport network. More specifically, physical characteristics of nodes, amplifiers, attenuators, fibers (both transmission fibers and Dispersion Compensation Fiber, DCF, modules), transmitters, receivers and eventually the definition of physical links are kept in the PPD (expressed in XML format). The

Traffic Engineering Database (TED) includes the nodes and detailed network topology in XML format. The connectivity of the nodes in the network and the definitions of network nodes (e.g. node ID, node IP address, node names) are kept in this repository. The NPOT XML parser is responsible for parsing the XML repositories and transforming the network description (physical specification and network topology) into the internal data structures, which are stored inside the NPOT memory. The NPOT PPD and TED manager is in charge of the management of the PPD and TED data structures, while they are residing inside NPOT memory.

5.3.2.2 Quality of Transmission (QoT) estimator: The Q-Tool

To assess the QoT of a lightpath in the NPOT, a Q-Tool was developed [Ang12]. Given the topology, physical characteristics and state (i.e., active lightpaths) of the network, the Q-Tool evaluates the dominant impairments that affect to a candidate lightpath integrating them into a single figure of merit: the *Q-factor*. Thus, the NPOT Q-Tool offers fast assessment of the QoT of a lightpath given a specific set of lightpaths, and physical topology based on analytical and numerical simulations. The building blocks of the NPOT Q-Tool are presented in Figure 5.2.

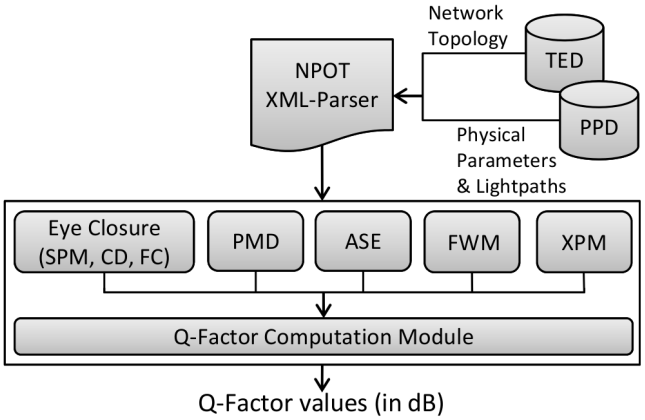


FIGURE 5.2: Building blocks of the NPOT Q-Tool module

The NPOT Q-Tool provides the required access to the physical layer performance evaluator. In this way, the tool receives a set of lightpaths (at least one) and then computes the Q-factors of those lightpaths and returns them back to the calling module. The Q-Tool is developed in MATLAB and interfaced to the NPOT through a shared library.

The Q-factor for a lightpath is a QoT indicator related to the signal's BER; assuming an On-Off Keying (OOK) modulated signal, it can be expressed as:

$$BER = \frac{1}{2} erfc\left(\frac{Q}{\sqrt{2}}\right) \tag{5.1}$$

where the Q factor is defined as [Agr02]:

$$Q = \frac{P_1 - P_0}{\sigma_1 + \sigma_0} \tag{5.2}$$

In (5.2), P_1 and P_0 are the mean of the distributions (assumed to be Gaussian) of the received samples corresponding to the sent 1 and 0 bits, and σ_1 and σ_0 are the respective standard deviations.

The overall QoT of the signal is reflected in the Q-factor, which considers the impact of the mentioned impairments as a single figure of merit (Q_{est}):

$$Q_{est} = 20log\left(\frac{I_{1,min} - I_{0,max}}{\sigma_{0,ASE} + \sigma_1}\right) - Q_{penPMD} \tag{5.3}$$

$$\sigma_1 = \sqrt{\sigma_{1,ASE}^2 + \sigma_{XPM}^2 + \sigma_{FWM}^2} \tag{5.4}$$

The numerator of (5.3) refers to the difference of the minimum detected current at the level of 1 and the maximum at the 0 level which defines the distortion induced by SPM, CD and FC on the signal (Eye diagram closure effect). The addend in the denominator of (5.3) includes the variances of all the noise impairments that add up to the total signal power variance. $\sigma_{0,ASE}$ and $\sigma_{1,ASE}^2$ in (5.3) and (5.4) refer to the variance of the detected spaces (0's) and marks (1's) due to ASE noise. XPM and FWM are assumed only to add noise at the level of 1 and therefore the non-linear induced degradation is expressed by σ_{XPM}^2 and σ_{FWM}^2. Finally, since the Q-factor is in fact a figure of merit, the PMD-induced penalty has to be subtracted from the total estimated Q-factor.

In order to quantify the impact of the static impairments (SPM, CD and FC) on the signal quality, the NPOT Q-Tool utilizes a detailed numerical Split-Step Fourier (SSF) method. During this process every optical signal is treated as if there was only single-channel propagation, thus accounting only for SPM, CD and FC. This approach achieves an accurate computation of the state of the optical signal at the receiver-end without considering the presence of neighboring channels, that would impose a prohibitive time penalty in the process. Thus, the proposed Q-Tool introduces a balance between speed and accuracy by numerically simulating the single-channel signal propagation.

The analytical model utilized to estimate the power of the ASE noise of a cascade of inline (EDFA) amplifiers is similar to the models in [RDF+99, OZW05] to assess the accumulation of this effect and its impact on the lightpath QoT. The Gaussian distribution of the ASE-signal beating noise facilitates the incorporation of its contribution to the Q-factor by considering the noise variances that are superimposed on the levels of marks and spaces.

The NPOT Q-Tool considers the impact of XPM on the performance of a single link according to the Cartaxo analytical model [Car99], which is properly modified to match

the specific link architecture. The analytical expression of σ^2_{XPM} is derived using the approach reported in [PRSV06].

The NPOT Q-Tool treats XPM and FWM as random noise that affects the QoT identically as the ASE noise, imposing fluctuations that typically occur at the mark level. As shown in [WW04] this can be considered as a good approximation particularly at the regime of relatively high Q and in the same order of magnitude as ASE noise. In [WW04] the XPM-induced distortion is estimated using the frequency response of XPM-induced intensity modulation from a modulated pump channel to a continuous wave (CW) probe channel ([Car99], "Cartaxo model"), as done in the Q-Tool. It was shown in [WW04] that FWM is well approximated with a Gaussian distribution. Assuming a random behavior is particularly applicable in FWM due to the fact that many independent channels contribute to the total FWM power. The work in [WW04] utilizes the model reported in [INOT94] and is extended for a multi-span system, which is implemented in the NPOT Q-Tool.

To calculate the PMD-induced penalty two different methods have been generally followed [ZLS+06]. One of them relies on a sophisticated statistical modeling method that approximates the all-order PMD, while the other considers only the first-order PMD using an analytical or numerical approach. Despite the restriction of the analytical model to the first-order PMD, a comparison for a 10Gb/s non return to zero (NRZ) signal in [ZLS+06] demonstrated that its penalty was in good agreement with the penalty obtained with the statistical modeling. In fact, NRZ signals are primarily dominated by the first-order PMD rendering the choice for the analytical approach valid for the framework of the DICONET NPOT. However, in a higher bit-rate system, in which robust modulation formats to PMD are utilized (e.g. RZ), the mentioned approach could not be applied.

5.3.2.3 Impairment-aware RWA engines

During the network planning phase, the demand set (traffic matrix) is already known at least partially, enabling the network operator to perform the resource allocation task offline. Since in all-optical networks bandwidth is allocated under the form of lightpaths (i.e., the combination of a route between two nodes, and a wavelength), the problem of pre-planned resource allocation in such networks is called static or offline RWA problem [AKM+09]. The offline IA-RWA algorithm utilized in the NPOT is named offline *Rahyab*, and it is adopted from [APA+09]. There are two IA-RWA engines for the operation (online) mode of the NPOT that correspond to both NPOT integration approaches: distributed and centralized.

In the distributed scheme, which will be defined in detail in section 5.4, as the NPOT IA-RWA module receives a demand it computes k shortest routes from source to destination. In this scheme the route calculation is done without considering the impact of the PLIs, since the QoT is checked during the signaling process. In case of a demand with 1+1

protection requirement, this module computes k diverse pairs of primary and backup paths. The caller module (in this particular case the OCC of the control plane) tries to establish the lightpath from source to destination using an extended version of the signaling protocol. If a lightpath (or a pair of lightpaths) is (are) established, the NPOT updates its global data structure (via the TED and PPD manager). If none of the k candidate routes are feasible, the caller sends the corresponding error code back to the NPOT.

In the centralized integration scheme (detailed in section 5.5) the NPOT IA-RWA engine [KCMV09] computes the lightpath from source to destination, assigns a wavelength to it and returns it back to the caller (i.e., OCC). In this case, the NPOT assures that the computed route is feasible, that is, the Q-factor of the yet to be established lightpath and the already established ones is above the given threshold. Then, the OCC tries to establish the lightpath using the standard signaling protocol and returns the result of the lightpath establishment process back to the NPOT. If the establishment is successful, the NPOT updates the list of active lightpaths in the network and also updates the topology data structure. If the lightpath establishment fails, the source OCC notifies the NPOT accordingly.

The IA-RWA engines in general can be used for transparent or translucent optical networks. In the case of translucent networks, the information regarding the regeneration sites (decided by the regenerator placement module of the NPOT) is available through the PPD repository. Particularly in the distributed scheme, the utilized IA-RWA engine is not constrained by the regeneration sites, whereas the IA-RWA algorithm of the centralized case is able to exploit them. However, the experimental setup represents a transparent optical network and both centralized and distributed schemes are evaluated under similar conditions.

5.3.2.4 Component placement modules

Regenerator and monitor placement modules are two component placement modules of the NPOT, which are mainly designed for planning mode. The main assumption in an all-optical network is that the network is truly transparent (all-optical), thus all intermediate O/E/O conversion is eliminated. However, given that the longest connections for instance in a North American backbone network are on the order of 8000 km, clearly some regeneration is still required. The NPOT regenerator placement module is responsible for optimizing the number of regeneration sites and modules that are going to be deployed in the network. This module is developed according to the COR2P (Cross Optimization for RWA and Regenerator Placement) algorithm [YAZG10a, YAZG10b]. This module receives a demand set (traffic matrix) along with the network description and minimizes the number of required regeneration sites and regeneration ports in the network. Besides, the COR2P algorithm utilizes the NPOT Q-Tool in order to evaluate the performance of the physical layer.

In addition to the regenerator placement, the NPOT exploits a special purpose monitor placement algorithm, which deploys optical impairment/performance monitors (OIM/OPM) on the network links. The OIM/OPM equipment can be utilized for enhanced QoT estimation, compensation of QoT estimation inaccuracy [APA+10b] or failure localization.

5.3.2.5 Communication interfaces: NPOT-OCC protocol

The integration of the NPOT in the ASON/GMPLS test bed is realized by means of a proprietary protocol which allows the communication between the NPOT and the other elements of the data communication network. In particular, the NPOT interacts with the OCCs through the NPOT-OCC interface, which is used by the OCC to request route computations, update the NPOT repositories (i.e., PPD and TED), request Q-factor calculations and to notify failures detected in the transport plane. This interface implements a TCP-based communication called *NPOT-OCC* protocol. Although the general behavior of the protocol is the same for the different integration schemes (centralized and distributed), slight changes are applied to adapt it to the specific requirements of each approach. Therefore, the functionalities of the protocol associated to each integration scheme are specifically described in each architecture's corresponding section.

In addition, an NPOT-NMS interface is defined for the NPOT to communicate directly with the management system, specially for failure notification purposes. As a matter of fact, the protocol that implements such an interface contains a single message aimed to notify the NMS about failed links located by the NPOT after receiving a set of LoL alarms from the control plane. Upon the reception of this message, the NMS shows the failure on the graphical view.

5.3.3 Control plane extensions and integration

So far in this third section, the NPOT architecture has been detailed. Besides, an overview of different impairment-aware control plane schemes was shown in section 5.2. Nonetheless, this thesis focuses on the enhancements that have to be applied to the GMPLS-based control plane to cope with the requirements of impairment-awareness. These enhancements refer to the extensions needed by the control plane protocols to manage physical layer information, and to integrate the NPOT. While the former includes the extensions applied to the existing GMPLS protocol stack, the NPOT integration is achieved by means of a new communication protocol between the OCCs and the NPOT. In this regard, sections 5.4 and 5.5 describe the defined integration schemes. An exhaustive description of the GMPLS protocol extensions and the new NPOT-OCC communication is included in these sections.

5.4 Distributed GMPLS-based impairment-aware control plane scheme for dynamic optical networks

In the distributed architecture proposed within the DICONET project (Figure 5.3), each OCC is equipped with an NPOT, which interacts with the running modules of the control plane through the NPOT-OCC communication protocol. By doing so, the RC is relieved from the route computation task (now assigned to the NPOT), but keeps the topological information dissemination responsibility. Regarding the GMPLS protocols, both RSVP-TE and OSPF-TE are extended to provide a good trade-off between network performance, control overhead and operational complexity.

The OSPF-TE protocol needs to be extended to enable the dissemination of the wavelengths available for each TE link. Nonetheless, as shown in section 3.2.2, the CARISMA OSPF-TE already implements the Wavelength availability sub-TLV to facilitate the RWA process. This capability, along with the standard features of the protocol, provides a complete topological view of the network to every OCC which, in turn, populates the global TED (gTED) of its controlled NPOT. On the contrary, PLI information is kept locally in each node's PPD and not disseminated. The OCC feeds the local PPD of the NPOT with the physical layer information provided to the LRM module by the OXC through the CCI. Besides this, the RSVP-TE signaling protocol is extended to collect real-time information of the PLIs during the connection signaling process. In particular, the *PATH* message is responsible for collecting such information which is later used to calculate the lightpath's QoT and, thus, its feasibility.

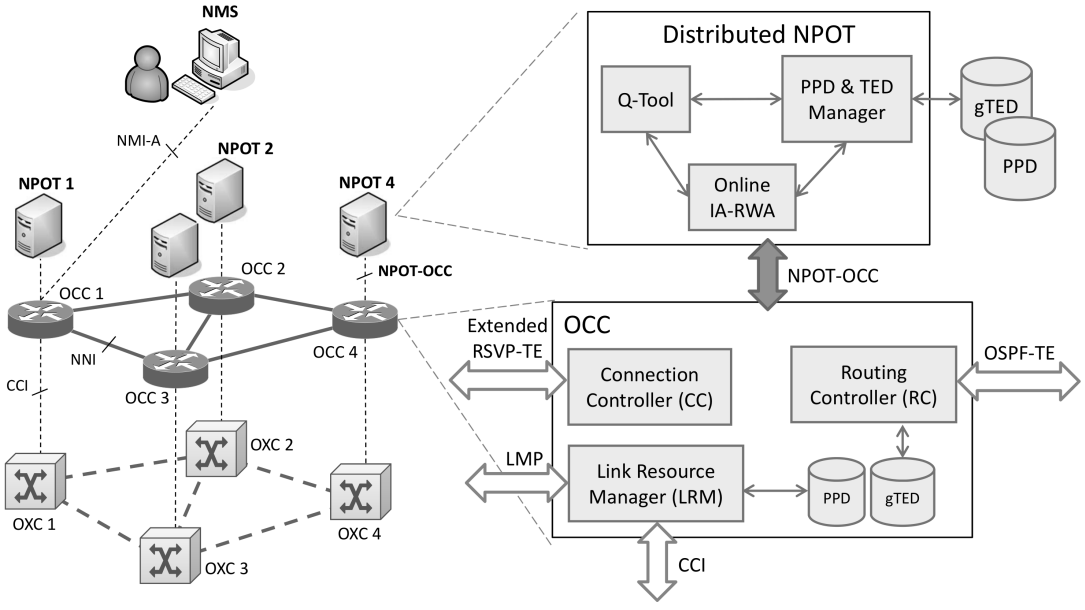


FIGURE 5.3: Distributed control plane integration scheme.

5.4.1 Lightpath establishment

Upon receiving a new connection request (Figure 5.4, step 1), the source OCC requests the online IA-RWA module of the NPOT to compute k link disjoint routes from source to destination (step 2). The wavelength availability information stored in the gTED is used for route computation, which is done following a shortest path policy. However, no information about PLIs is used, since this information is locally kept in each node. Once these k shortest routes are computed, the source OCC triggers the extended RSVP-TE protocol to initiate the lightpath establishment using the first candidate route. The *PATH* message (step 3a) collects the PLIs from source to destination along the candidate path (step 3b) and, at destination, the node requests for a QoT estimation from its corresponding NPOT (step 4). If the QoT of the candidate lightpath is acceptable (i.e., the Q-factor is above a given threshold), then the destination nodes of the potentially disrupted lightpaths (i.e., those lightpaths sharing at least one optical section with the candidate one) are notified to request a QoT estimation from their respective NPOTs (step 5). This verification step (called the Q-check mechanism, detailed in subsection 5.4.4) checks that the Q-factor of the already established lightpaths remains above the required threshold in spite of the establishment of the new lightpath. Once the Q-factor values are re-computed, the destination nodes of these active lightpaths respond back to the destination node of the candidate one. An ACK is sent if the new Q-factor is acceptable. If not, the node sends a NACK. If the QoT of any of the affected lightpaths is not violated, each node updates its local databases with the new lightpath information, the signaling process continues by means of the *RESV* message, and the actual optical cross-connections are properly configured through the CCI interface (step 6). Otherwise, a *PATH_ERROR* is sent back and the source node tries the next route. If none of the k candidate routes meet the required QoT the request is definitively blocked. If the establishment is successful, the OSPF-TE flooding mechanism disseminates the updated topological information and wavelength availability (step 7a) and, in turn, each OCC updates the PPD and gTED of its controlled NPOT (step 7b). Finally, the source OCC informs to the NMS about the lightpath establishment result (step 8).

5.4.2 RSVP-TE protocol extensions

As seen, the RSVP-TE protocol is responsible for carrying the physical layer related information necessary to compute the feasibility of a new lightpath. To this end, the protocol has been extended with a new object (and its corresponding subobjects) which is added to the *PATH* message and finally used at the destination node to compute the QoT of the lightpath.

The information required for impairment evaluation and lightpath feasibility check is the one related to the sections composing the route of the candidate lightpath. In particular, the optical parameters of the components of each section (transmitters, receivers,

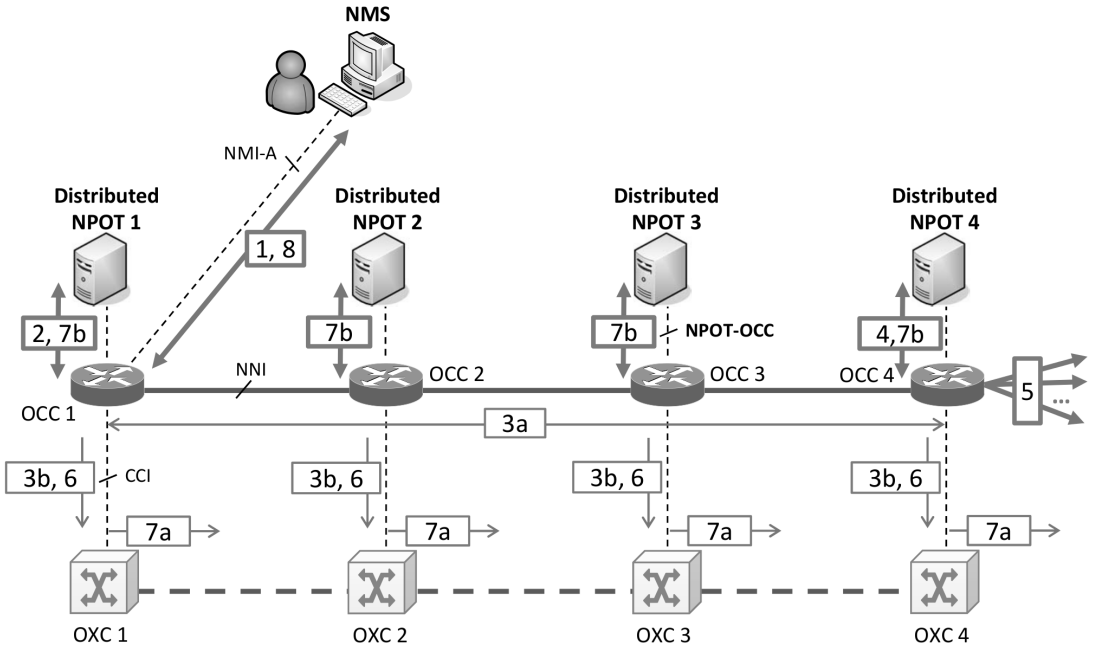


FIGURE 5.4: LSP establishment example for the distributed scheme

amplifiers, etc.), as well as information about the active lightpaths passing through each section are needed. All this information is contained in *Optical Section Description* (*OSD*) objects (Figure 5.5).

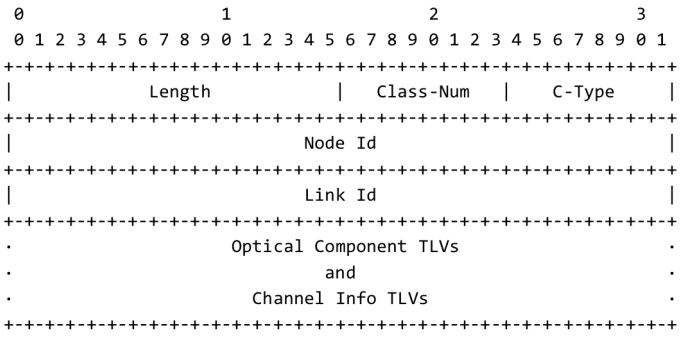


FIGURE 5.5: Optical Section Description object

The new *OSD* object (Class-Num 255, C-Type 1) conveys data corresponding to the optical section of the TE link identified by *Node Id* and *Link Id* (Figure 5.5). Information about each optical component of the section is carried by means of *Optical Component* TLVs (Figure 5.6).

The *Optical Component* TLV type is set to 1 and the component type field is used to differentiate among component categories (i.e. transmitters, receivers, amplifiers, etc.). Each component also has a unique identifier and a set of sub-TLVs (following the diagram depicted in Figure 5.7) that contain the values of the physical parameters of the component. A complete list of the components (with their corresponding parameters) used in this work can be found in Appendix B.

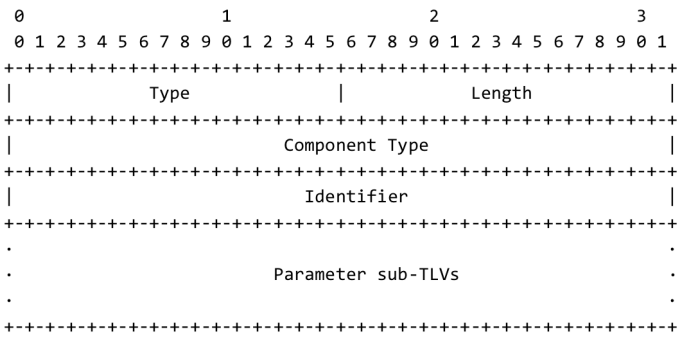


FIGURE 5.6: Optical Component TLV

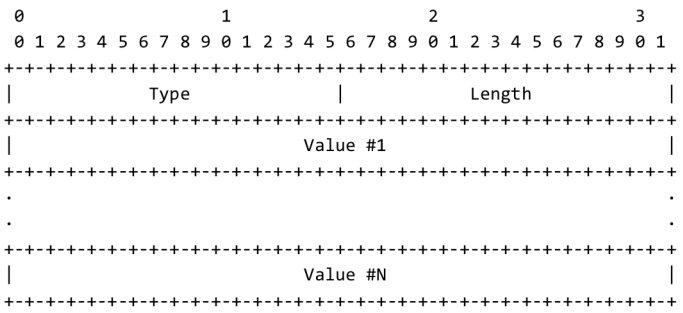


FIGURE 5.7: Parameter sub-TLV

Since information regarding the active lightpaths traversing each optical section is also needed to evaluate the QoT, a *Channel Information* TLV (Figure 5.8) is added to the protocol. In this way, for each channel (encoded as proposed in [RFC6205]) of the optical section that is being used by an active lightpath, the lightpath identifier and the power (if available) are conveyed. Moreover, the Destination Node Id of the lightpath is needed to recompute the Q-factor of the lightpath in order to assure that its QoT is not affected by the new lightpath establishment (this procedure is detailed in next subsection). To facilitate message processing, the P flag is set to 1 if the *Power* value is present in the TLV. Contrariwise, it is set to 0.

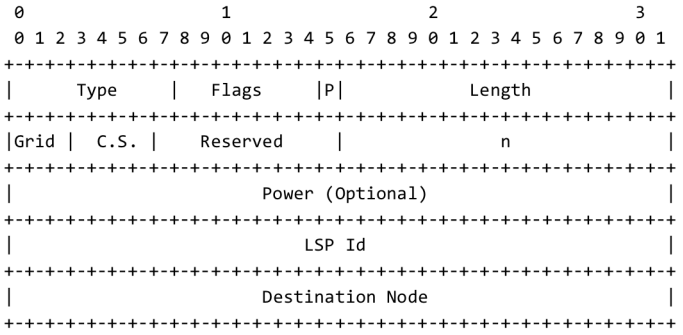


FIGURE 5.8: Channel Information TLV

5.4.3 NPOT-OCC communication protocol

The NPOT-OCC protocol introduced in section 5.3.2.5 is responsible for the interaction between the NPOT and the OCC. The list below details the specific functions defined for the distributed case. Figures 5.9 and 5.10 depict the definition of the messages enabling such functions.

- The route computation message is used by the OCC to request a route computation from the NPOT. Therein, the source and destination nodes are sent to the NPOT which performs the RWA. More precisely, the NPOT calculates a set of k shortest link-disjoint paths with their respective wavelengths. Then, the set of routes is sequentially processed until a feasible lightpath in terms of QoT is obtained. To do this, during the signaling of each candidate route the QoT computation is requested to the NPOT by means of this protocol (see next bullet). If any of the candidate routes is feasible, the connection is finally blocked. Figure 5.9 shows the route request/reply message pair.

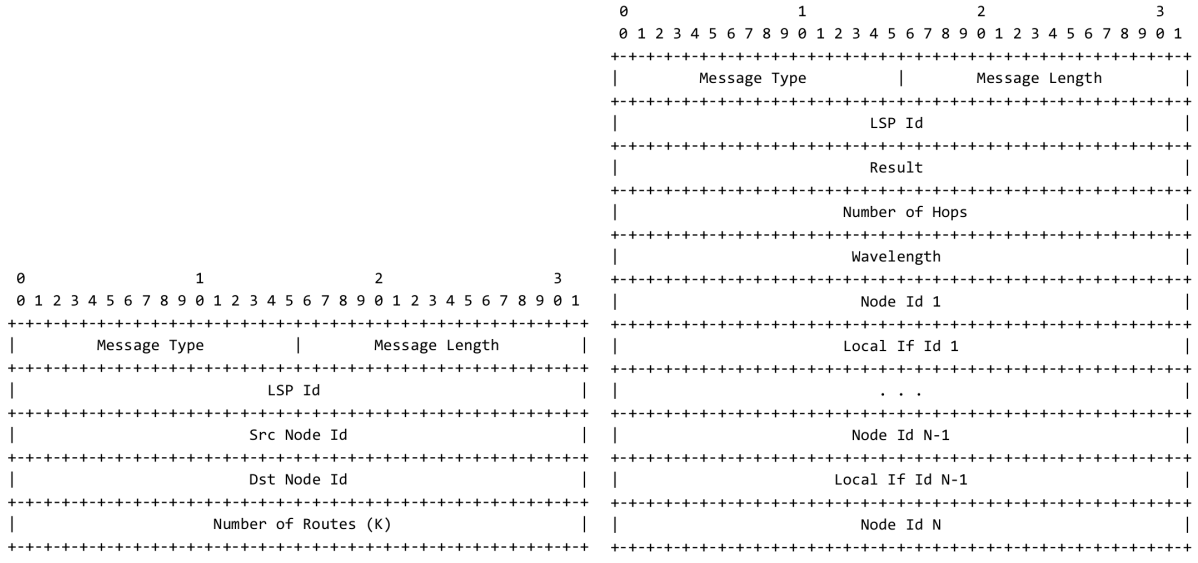


FIGURE 5.9: NPOT-OCC protocol messages: a) route request from the OCC to the NPOT; b) route reply from the NPOT to the OCC.

- Q-factor computation request/reply messages (Figure 5.10) are defined in the protocol for the OCCs to ask for the QoT of one or more lightpaths. This operation is basically used in the distributed approach since, as stated in the previous bullet, the set of routes calculated by the NPOT in such mechanism does not guarantee a sufficient QoT.

- The update PPD and TED operations are used by the OCC to update the PLIs and the topological information, respectively. For sake of efficiency, updates only take

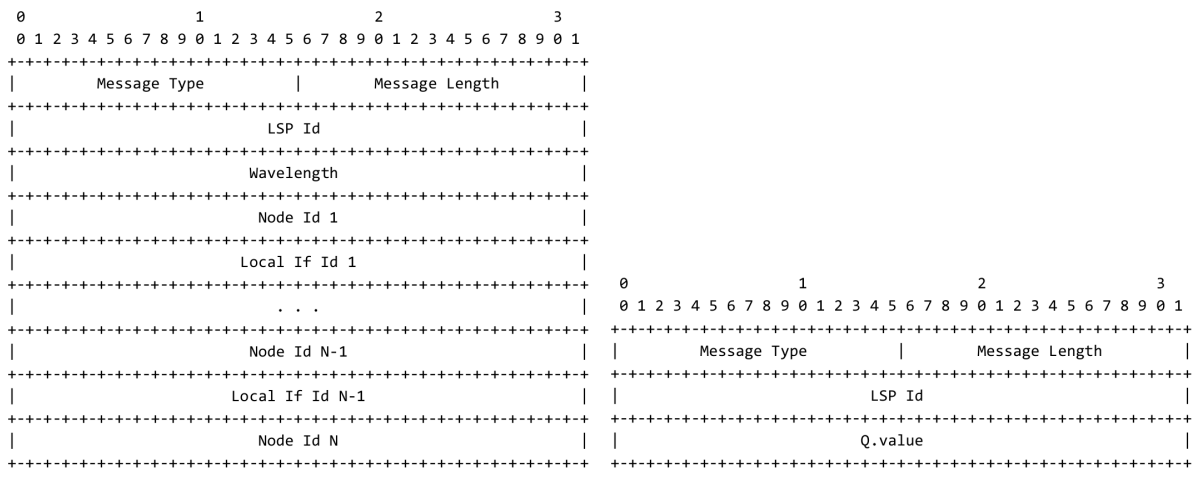


FIGURE 5.10: NPOT-OCC protocol messages: a) Q-factor computation request from the OCC to the NPOT; b) Q-factor computation reply from the NPOT to the OCC.

place if changes occur in any of the stored parameters. Moreover, only the modified parameters are disseminated. For example, if the input power of one receiver has changed, only the power parameter of the affected receiver is disseminated and, in the same way, if one TE link's wavelength has been freed or occupied, only this TE link is flooded and updated in the TED of the NPOT. Due to the space limitations update messages are not shown in this dissertation.

5.4.4 Lightpath feasibility: The Q-check mechanism

The target of the Q-check mechanism ([SSZ+09b, SSZ+09a]) is twofold. Firstly, it is used to estimate the feasibility of the lightpath to be established taking into account the already active lightpaths. Secondly, it checks if any of the working lightpaths that share one or more optical sections with the new one is disrupted by its establishment. In the context of this thesis, the feasibility of a lightpath is measured in terms of QoT, which is translated into a Q-factor value.

The Q-check mechanism is triggered once the *PATH* message reaches the destination OCC of the lightpath to be set up. At this point the message contains the physical layer information about all the sections the new lightpath passes through, in other words, the *PATH* contains the list of the *OSD* objects of the calculated route. This information is then sent to the NPOT controlled by the destination OCC which calculates the Q-factor of the lightpath by means of the Q-Tool module. If the calculated Q-factor is below the minimum threshold, a *PATH_ERROR* message is sent back to the source node and the next route is tried. Otherwise, the second phase of the mechanism is started. This phase consists of checking that the Q-factors of the lightpaths affected by the one to be established are still above the threshold. To this end, the OCC extracts the IP addresses

of the destination OCCs of the affected lightpaths from the *Channel Information* TLVs conveyed in the *OSD* objects. Then, it sends a *Q-factor computation request* containing the *OSD* objects to each destination OCC. On the reception of the *Q-factor computation request*, each OCC extracts from the *OSDs* the information needed to recompute the Q-factor of the lightpaths ending at that OCC and affected by the new one. This information is sent to the NPOT of the OCC which recalculates the Q-factor. If any of the affected lightpaths has a recomputed Q-factor under the threshold, a NACK message is sent to the destination OCC of the new lightpath which sends a *PATH_ERROR*. If there are no affected lightpaths disrupted by the new one, an ACK is sent to the destination OCC which sends the *RESV* message back to the source node, thus finalizing the new lightpath establishment process.

5.4.4.1 Q-check mechanism example

For sake of clarification, Figure 5.11 exemplifies a use case of the Q-factor computation mechanism. Such scenario depicts the situation in which LSP 3 is going to be established whereas LSPs 1 and 2 are already active. During the signaling of LSP 3, the *PATH* message collects the *OSD* objects along its route from OCC_A to OCC_E. In particular *OSDs* belonging to sections AC, CD and DE are gathered. Once the message arrives to destination (node OCC_E), the Q-check mechanism is initiated. First of all, the NPOT controlled by OCC_E calculates the Q-factor of LSP 3 taking into account the already established LSPs 1 and 2 (step 1 in the figure). To this end, *OSDs* from sections AC, CD and DE are used. If the Q-factor is above the defined threshold, the Q-factor recomputation process for LSPs 1 and 2 is executed. In the example, OCC_E requests Q value recomputations to OCC_D since D is the destination of LSP 1, and to OCC_F given that F is the destination of LSP 2. Thus, the NPOT controlled by OCC_D recalculates the Q-factor of LSP 1 (Q_{LSP1}) considering LSPs 2 and 3 as active lightpaths in the network. Similarly, the NPOT of OCC_F recalculates the Q value for LSP 2 (Q_{LSP2}) taking into account the existence of LSPs 1 and 3. If both Q_{LSP1} and Q_{LSP2} are above the minimum threshold that guarantees an acceptable QoT, OCC_D and OCC_F send their respective ACKs to OCC_E, which finishes the lightpath establishment. Contrariwise, a NACK message is sent and OCC_E sends a *PATH_ERROR* message back to OCC_A.

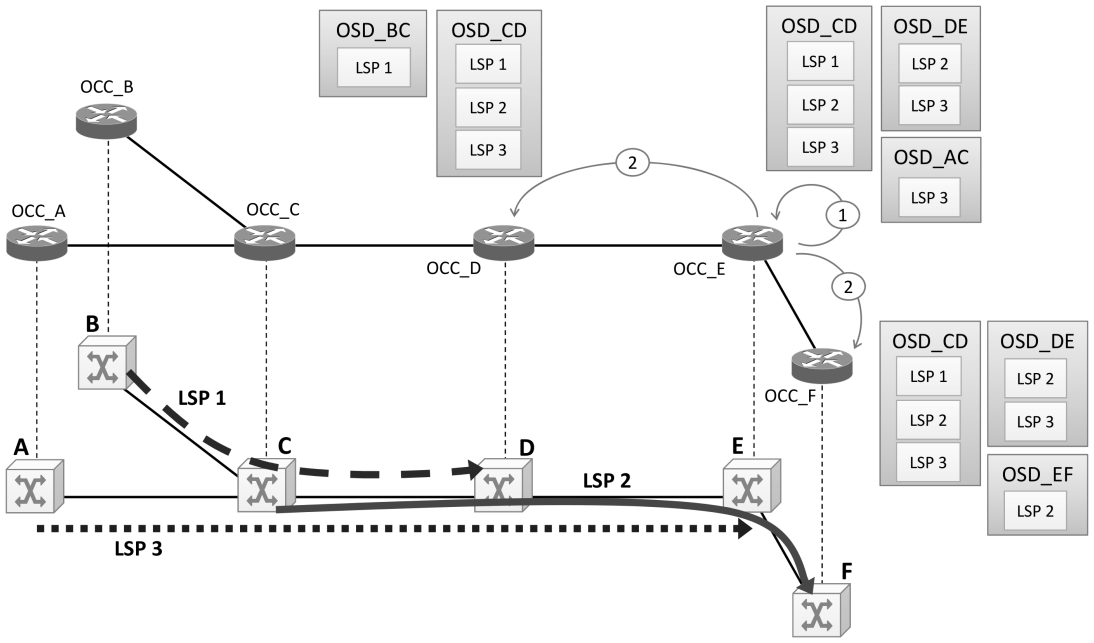


FIGURE 5.11: Q-check mechanism example

5.5 Centralized GMPLS-based impairment-aware control plane scheme for dynamic optical networks

The centralized scheme (Figure 5.12) is featured by the fact that there is one (centralized) NPOT instance running in the control plane. Herein, the NPOT is responsible for the impairment-aware lightpath computation and failure localization, and it serves requests arriving from all the OCCs in the control plane. For route computation, the Online IA-RWA module of the NPOT uses the information stored in both the global TED (gTED) and the global PPD (gPPD), which describe the complete network topology and the physical layer characteristics respectively. Moreover, the Q-Tool is used by the IA-RWA module to check the feasibility of computed routes. Like in the distributed case, a route is considered feasible if the QoT (specifically the Q-factor) of the future lightpath is above a pre-defined threshold and, at the same time, the QoT of the active lightpaths of the network is not disrupted by the new establishment.

The gTED and the gPPD are fed with data coming from the OCCs through an improved version of the OSPF-TE protocol, which has been extended to disseminate links' wavelength availability and the physical layer information. Furthermore, OCCs also keep global topological information in their traffic engineering database but, unlike the NPOT, each OCC only manages information regarding the local physical resources. The latter is collected from the local OXC via the CCI interface, flooded by means of the extended OSPF-TE and finally stored in the local PPD. Besides, OCCs request route computations through the NPOT-OCC interface using a proprietary protocol.

As will be detailed in section 5.7, the centralized scheme takes advantage of the failure localization module of the NPOT to implement a lightpath restoration mechanism. In this mechanism, failures detected in the transport plane are notified (via the CCI) to the OCCs which forward them to the NPOT through the NPOT-OCC interface. This message exchange triggers the failure localization procedure carried out by the NPOT, as well as the lightpath restoration process. Moreover, once the failure has been localized, the NPOT sends a message to the NMS (using the NMS-NPOT interface) indicating the failed link, so the NMS can show it in the graphical view of the optical transport network.

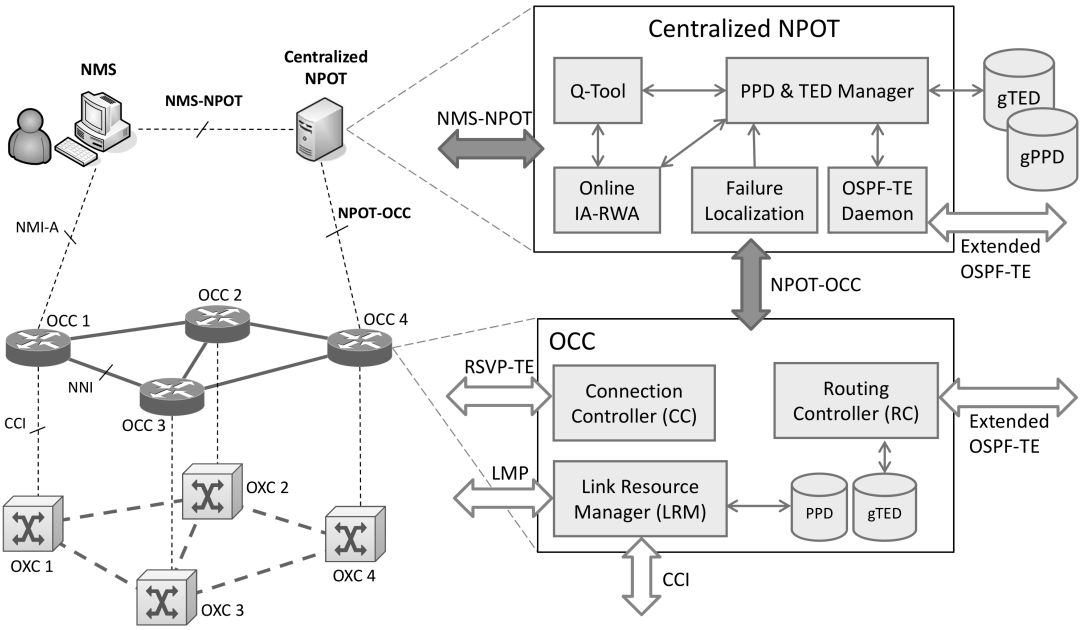


FIGURE 5.12: Centralized control plane integration scheme

5.5.1 Lightpath establishment

Figure 5.13 shows an example of a lightpath establishment in the centralized approach. Therein, upon the arrival of a connection request (step 1), the source OCC contacts the centralized NPOT to request a path computation (step 2). Next, the request is forwarded to the online IA-RWA module, which is responsible for performing the path computation for the given request. Note that, as in the distributed case, the RC is again relieved from path computation responsibility. The online IA-RWA module of the NPOT utilizes the QoT estimator (Q-Tool) and the information of the gPPD and the gTED. Specifically, the Q-Tool is the module within NPOT that quantifies the impact of the PLIs on the lightpaths' QoT. Once the NPOT finds a lightpath with guaranteed QoT (Q-factor value above a pre-defined threshold), the lightpath is returned back to the source OCC and it is established from source to destination using the standard RSVP-TE signaling protocol (step 3a). During the signaling, each OCC configures the OXC resources to be used (3b) by means of the XML-based proprietary protocol implemented for the CCI interface. Upon the successful establishment of a lightpath, the global PPD and TED in

the NPOT, and the local PPD and global TED of every OCC in the network are updated using the extended OSPF-TE protocol (step 4). Finally, the source OCC updates the NMS and the NPOT with the new lightpath establishment (step 5). In case of lack of resources or unacceptable QoT, the demand is blocked and the source OCC informs the NMS accordingly. The NMI-A interface is used to this end.

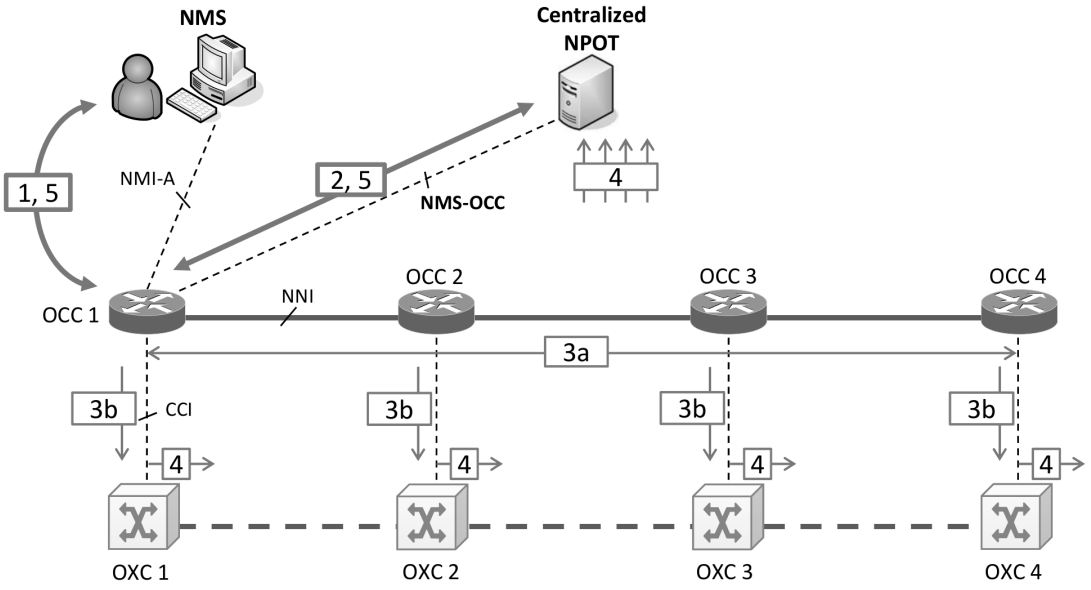


FIGURE 5.13: LSP establishment example for the centralized scheme

5.5.2 OSPF-TE protocol extensions

The function of the OSPF-TE protocol is to provide useful information to enable the RWA process. The current standard OSPF-TE characteristics deal with basic routing constraints such as reachability, cost, protection level, etc. However, extended information is needed by the RWA engines to manage with wavelength assignment and optical impairments constraints. Therefore, this subsection details the extensions that have been defined for the OSPF-TE protocol to meet the extra requirements of the proposed scenario. On the one hand, the previously defined extension to disseminate the wavelength availability in the whole network has been used (Wavelength availability sub-TLV, subsection 3.2.2). In this way, the RWA engine is able to compute routes that fulfill the wavelength continuity constraint. On the other hand, several extensions have been developed to disseminate information regarding the physical parameters of the transport network, so that information about the optical status of the transport plane can be used to compute routes optimizing the QoT of the lightpath. Regarding the storage of these new data, whereas the wavelength availability is kept in the TED since it is considered a topological property of the network, the PLIs are stored in the PPD which is designed to keep the physical layer characteristics of the network.

5.5.2.1 PLI top level TLVs

[RFC3630] defines the Opaque LSA structure to flood basic TE information. Therein, two main classes of OLSA, namely Node and Link are identified, each one conveying specific information regarding transport nodes and TE links respectively. In a similar way, two new OLSA top-level TLVs, *Node PLI* and *Link PLI*, have been proposed to disseminate physical layer information. Moreover, a set of sub-TLVs has been defined to be added to these new TLVs.

The *Node PLI* top level TLV (type 3, Figure 5.14) describes the physical layer parameters related to the node. Each node in the network has its own Node PLI TLV that is flooded into the DCN, and then stored in the NPOT's global PPD. It is worth to mention here that, by design, the OLSA dissemination is done using multicast, so its contained information could be stored by each OCC of the network, thus allowing the OCCs to keep a global PPD. However, given that such global information is useless for the OCC, and in order to simplify the operation at the control plane, the processing of incoming PLI OLSAs is avoided inside the OCCs.

As shown in the figure, the *Node PLI* is composed of a *Node Id* sub-TLV (that identifies the node whose information is being carried) and one *Component* sub-TLV which, in turn, contains a set of sub-sub-TLVs conveying the values of the physical parameters related to the node.

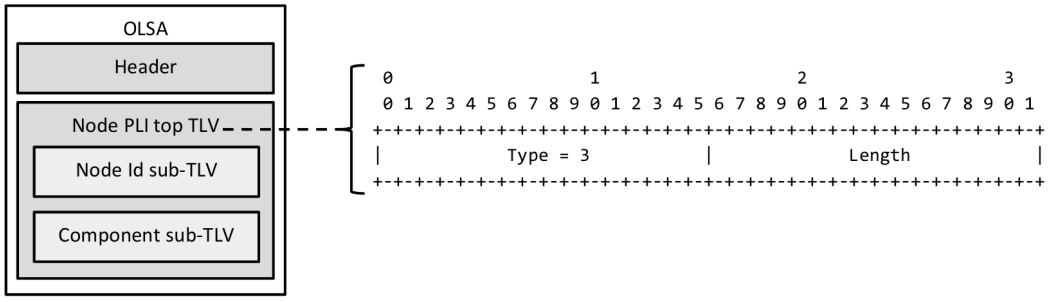


FIGURE 5.14: PLI Node OLSA diagram (left) and Node PLI top level TLV (right)

The *Link PLI* top level TLV (type 4) contains the physical layer parameters related to the link. Each node in the network disseminates a *Link PLI* OLSA for each of its output TE links. These OLSAs are stored in the global PPD of the NPOT but not in the PPDs of the neighboring OCCs (which only manage local PPDs). As depicted in Figure 5.15, the *Link PLI* is composed of a *Link Identifier* sub-TLV, a *Local Interface Identifier* sub-TLV and a *Remote Interface Identifier* sub-TLV which univocally identify the link within the network (section 5.5.2.3). This TLV also contains a set of *Component* sub-TLVs, one for each physical component of the TE link. The *Component* sub-TLV is described in detail in section 5.5.2.4, and it is used by both the *Node PLI* and *Link PLI* OLSAs only differing in the classes of the components.

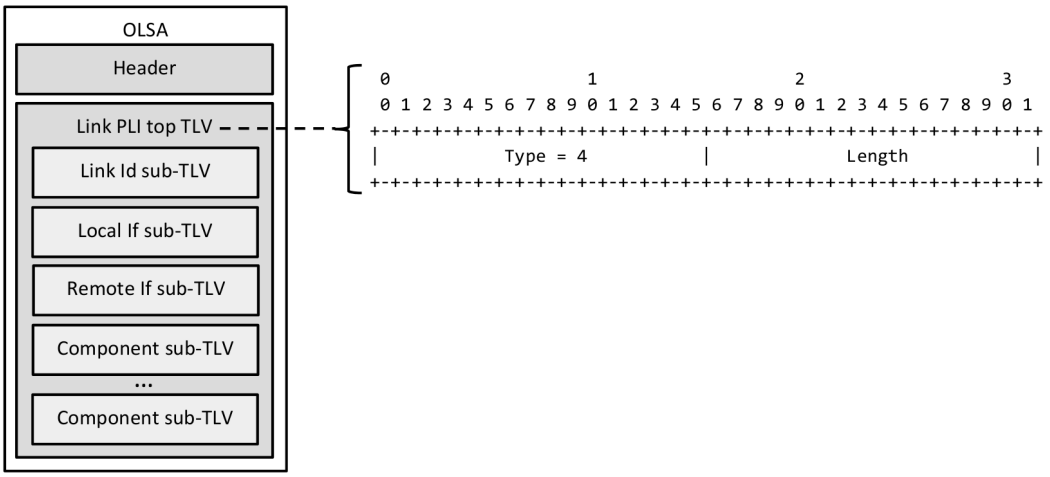


FIGURE 5.15: PLI Link OLSA diagram (left) and Link PLI top level TLV (right)

In both cases the length field of the top level TLV is variable because the number of components and the size of their contents may be different for each node and link.

5.5.2.2 Node PLI sub-TLVs

Two sub-TLVs enclose the *Node PLI* TLV and are used to transport the physical layer information of an optical node (Figure 5.16). The *Node Identifier* sub-TLV (type 1) is used to identify the node whose information is being disseminated. Hence, the *Node Identifier* is a 4 bytes fixed-length field that contains the IP address of the optical node. The *Component* sub-TLV is detailed in section 5.5.2.4.

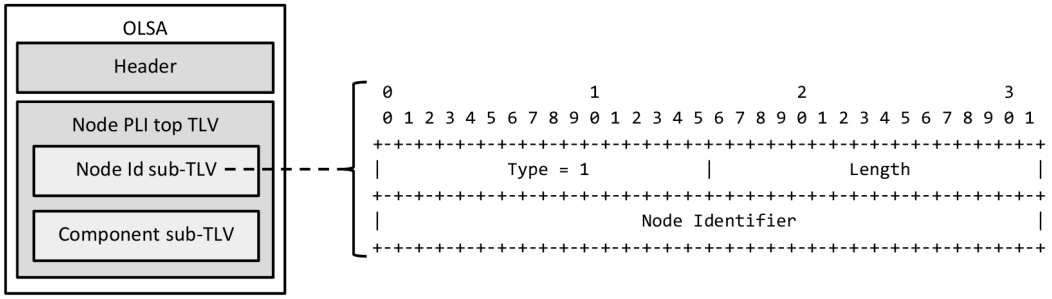


FIGURE 5.16: PLI Node OLSA diagram (left) and Node Identifier sub-TLV (right)

5.5.2.3 Link PLI sub-TLVs

Four sub-TLVs are defined for the *Link PLI* top TLV (Figure 5.17). In order to identify the TE link whose information is being flooded, the *Link Id*, the *Local Interface* and the *Remote Interface* have been taken from the *Link* top level TLV [RFC3630]. Therefore, the *Link Id* contains the neighbor node identifier (i.e. its IP address), and the local and remote interfaces contain the identifiers of the interfaces of the TE link.

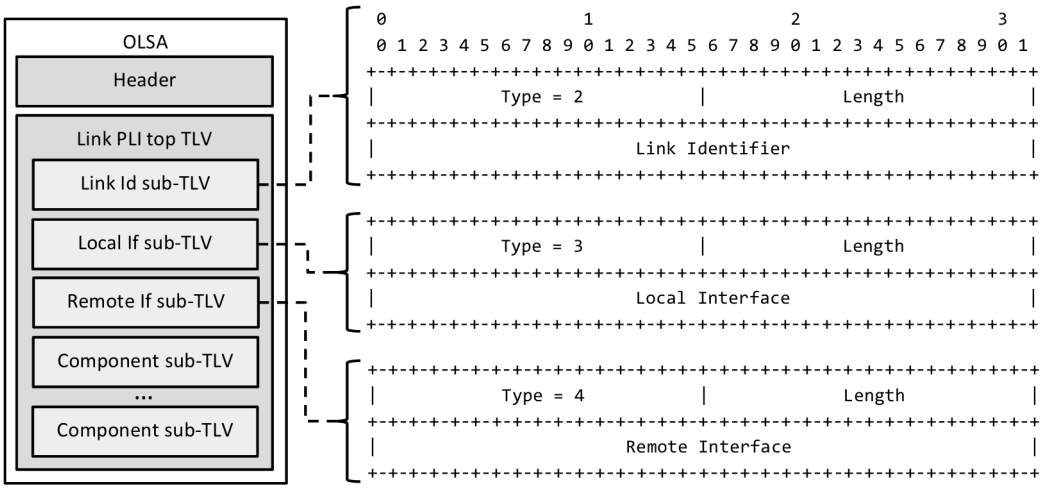


FIGURE 5.17: Link Identifier, Local Interface and Remote Interface sub-TLVs

5.5.2.4 Component sub-TLV

The *Component* sub-TLV (Figure 5.18) contains the values associated to the physical parameters of one component (identified by a *Component Id*) of either a TE link, or a node. Since a TE link can have several components, the *Link PLI* TLV may convey more than one *Component* sub-TLV. Currently, there is only one component (i.e. Physical node) associated to the node but, for the extensibility of the protocol, a variable number of *Component* sub-TLVs is allowed.

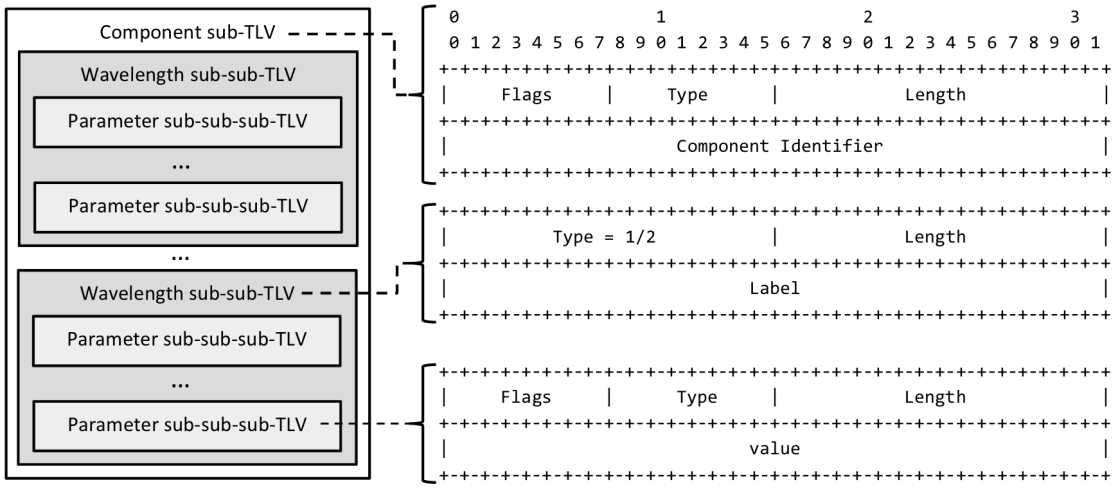


FIGURE 5.18: Component, Wavelength and Physical parameter sub-TLVs

To simplify the processing of the new OLSAs, the physical parameters of each component have been classified into wavelength related and non-related. This means that some of these parameters have specific values for each data channel traversing the component, while others have a single value for the whole component. For example, the Power parameter associated to a Transmitter component would have a value for each transmitted

wavelength, but a Fiber has the same Length value regardless of the traveling wavelengths. Then, as illustrated in the figure, the *Component* sub-TLV is composed of a set of *Wavelength* sub-sub-TLVs which, in turn, contain a set of (wavelength related) *Parameter* sub-sub-sub-TLVs each containing the specific value of such physical parameter for that channel. The *Wavelength* sub-sub-TLV is identified by the *Label* field that encodes the wavelength value as proposed in [RFC6205]. Besides, wavelength non-related parameters lie under a single *Wavelength* sub-sub-TLV having the *Label* set to zero. Note that, all the TLVs have a header containing a type value and the length so they can be unambiguously processed.

A complete list of the components and their associated parameters can be found in Appendix B.

5.5.3 NPOT-OCC communication protocol

As done for the distributed scheme, the list below details the set of messages specifically defined for the centralized case. For the sake of illustration, Figures 5.19, 5.20, and 5.21 depict the definition of the messages.

- The route computation message is used by the OCC to request a route computation to the NPOT. There, the source and destination nodes are sent to the NPOT which calculates a route with an acceptable QoT by means of its implemented IA-RWA algorithm. Figure 5.19 shows the route request/reply message pair definition.

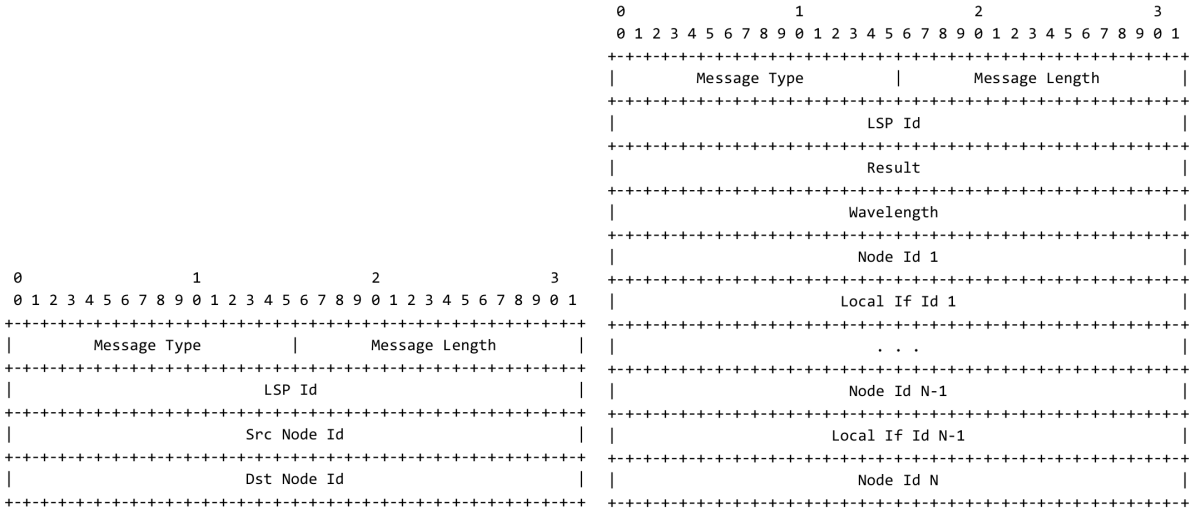


FIGURE 5.19: NPOT-OCC protocol messages: a) route request from the OCC to the NPOT; b) route reply from the NPOT to the OCC.

- The lightpath setup/tear-down function (Figure 5.20) is aimed to communicate to the NPOT that a new lightpath has been established/deleted. This information is needed by the NPOT (and stored in the gPPD) in order to calculate the QoT of subsequent lightpaths.

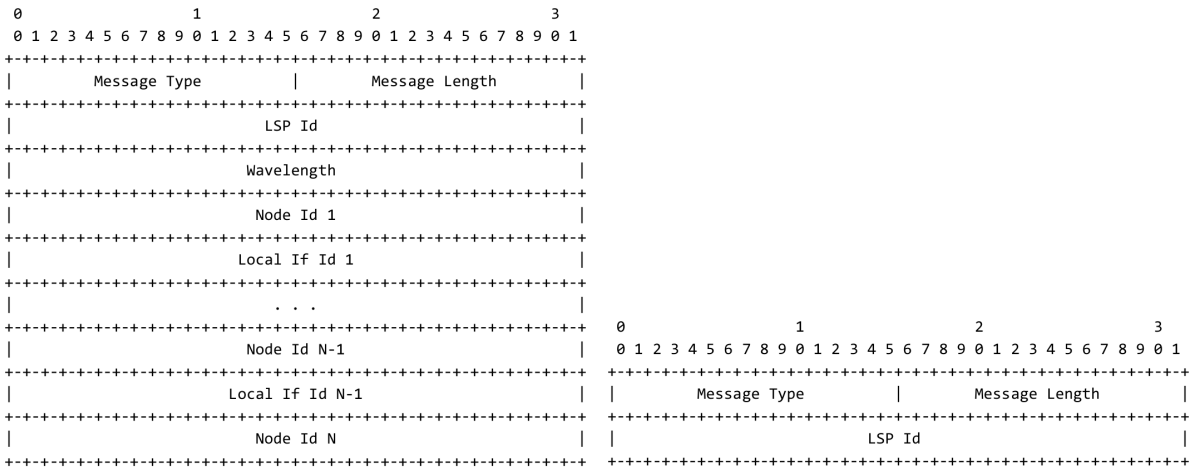


FIGURE 5.20: NPOT-OCC protocol messages: lightpath establishment (a) and tear-down (b) notification from the OCC to the NPOT.

- For failure localization a message is defined (Figure 5.21) so that OCCs forward the LoL alarms received from their controlled OXCs to the NPOT which, in turn, localizes the failure, detects the failed lightpaths, and notifies each protected/restorable lightpath's source OCC about the failure. Then, each OCC triggers the GMPLS standard protection/restoration mechanism. As will be explained further in this chapter, the failure handling mechanism is only enabled for the centralized NPOT, so this is not applied in the distributed case.

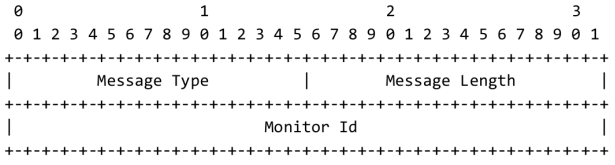


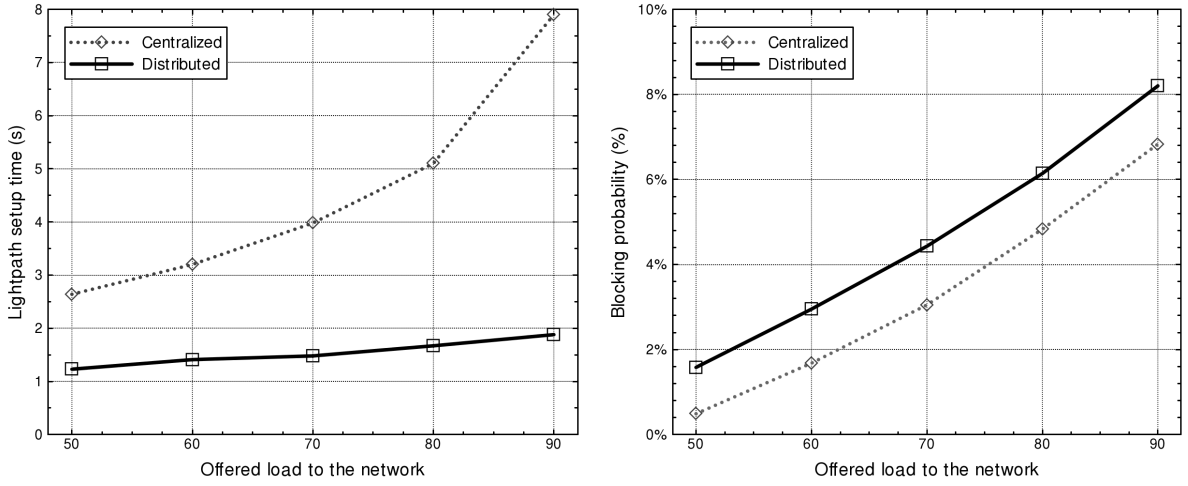
FIGURE 5.21: NPOT-OCC protocol messages: LoL alarm notification.

Note that no update PPD and TED messages are needed in the centralized case, since this task is carried out by the extended OSPF-TE protocol.

5.6 Experimental validation of distributed and centralized impairment-aware control plane schemes

This section presents the experimental assessment of the distributed and centralized schemes detailed in sections 5.4 and 5.5, respectively. Herein, the performance of such architectures is compared in terms of blocking probability and lightpath setup time. The experiments were conducted on the CARISMA test bed.

In this evaluation, a 14-node meshed network configuration describing the same topology as the generic Deutsche Telekom (DT) was configured. The topology of this network and the link lengths are depicted in Appendix A.3. Furthermore, 10 bidirectional wavelengths per link were assumed. Given the physical characteristics of this network, lightpaths can be transparently established between any node pair in the network. However, the heterogeneous characteristics of the fiber links and nodes, and considering the impact of neighboring lightpaths on each other introduce cases in which the QoT of the lightpaths falls below the acceptable threshold. Regarding the traffic characteristics, uniformly distributed lightpaths were assumed following a Poisson process. Moreover, lightpath holding times were exponentially distributed with a mean value of 600 seconds. Different dynamic loads (in Erlang) were thus generated by modifying the connection inter-arrival times accordingly (load = HT/IAT).



(a) (b)

FIGURE 5.22: Experimental results: a) lightpath setup time vs. offered load; b) blocking probability vs. offered load.

Figure 5.22 depicts the setup delay (a) and the blocking probability (b) experienced by the incoming lightpath establishment requests, depending on whether the distributed or centralized approach was deployed in the network. In particular, k = 2 candidate shortest routes were computed in the distributed case. Each result was obtained as the average of 10,000 requests. Looking at the results in Figure 5.22 (a), the distributed approach yields

lower setup times, especially as the offered load to the network increases. To explain this, note that in the centralized scheme only one route computation is allowed at the same time. In such computation, a single NPOT (and its Q-Tool) has to process all the related topological and PLI information, and compute the Q-factors of the candidate and affected lightpaths, which is a computationally intensive task and becomes the bottleneck of the whole process. Furthermore, a sufficient amount of time must be left between two consecutive route computations in order to let the centralized NPOT be fed with the new wavelength availability, PLI and active lightpaths information. Otherwise, subsequent routes might be computed with inaccurate link state information. Hence, the centralized NPOT must delay new incoming requests until the signaling and the respective flooding of the previous connection establishment has been completed (around 2 seconds in the test bed). To this aim, a queue is implemented in the centralized NPOT. Conversely, the distributed approach can benefit from some overlapping between lightpath establishments, since the Q-factor values of the lightpath to be established and the potentially disrupted ones are computed during the signaling phase by collaborative NPOTs each one having a partial view of the network state. This eventually results into very attractive connection setup times, around 1.8 seconds, 1/5th of the setup time reported in [TMKO08]. In contrast, Figure 5.22 (b) shows that the centralized approach leads to lower BP than the distributed solution. In fact, end-to-end routes in the latter are computed only with wavelength availability information. These routes lead in some occasions to an unacceptable Q-factor of candidate or potentially disrupted lightpaths and hence need to be blocked. On the contrary, route computation in the centralized approach relies on complete and updated wavelength availability and PLI information, so the computed routes more likely satisfy the requested Q-factor values.

5.6.1 NPOT enhancements for fast centralized lightpath provisioning

As detailed in previous sections, the online IA-RWA module in the NPOT invokes the Q-Tool in order to guarantee the acceptable QoT of the newly established lightpaths and the potentially disrupted active ones. The Q-Tool takes both linear (ASE, CD, FC, PMD) and non-linear (SPM, XPM and FWM) impairments into consideration. Computing all these impairments by analytical expressions and numerical simulations to come up with a lightpath QoT prediction is a computationally intensive task, which may incur significant running times in a standard computer. This is specially true in the centralized scheme, where a single NPOT is responsible for computing QoT-assured routes for all the network clients using complete and updated topological and PLI information. In order to bridge the setup time gap with the distributed scheme, the Q-Tool module in the centralized NPOT was implemented on a Field Programmable Gate Array (FPGA) hardware and coupled to the NPOT following a client-server model. Specifically, the hardware-accelerated Q-Tool is composed of two different modules: the QoT Estimation Agent (QoTEA) and Estimation Engine (QoTEE). The QoTEE module is deployed in a

Xilinx Virtex IV FPGA and is responsible for the actual QoT estimation. In turn, the QoTEA runs on a 300MHz IBM PowerPC 405 hard core embedded inside the FPGA fabric with 1GB DDR2 memory, and is responsible for receiving the lightpath QoT estimation requests coming from the Q-Tool client in the online IA-RWA module and sending them to the QoTEE. Once the QoTEE has computed the requested Q-factor, it returns the values back to the QoTEA, which finally reports them to the online IA-RWA module. More details about the implementation of the Q-Tool on FPGA hardware can be found in [QCT+10]. Therefore, by reducing the Q-factor computation time, which is the greater component of the whole lightpath setup time, the delay between route computations in the NPOT, is automatically reduced. Hence, lower overall lightpath setup times are achieved.

The new scheme was validated over the CARISMA test bed. To be comparable with the previous experiment the same scenario and conditions were used.

Figure 5.23 depicts the average lightpath setup times obtained for the new improved centralized scheme in front of the results obtained for the former centralized approach, equipped with a software-based Q-Tool, and the distributed one. As seen in the section above, the setup time strongly increases with the offered load when the software based Q-Tool is used (former centralized solution). Interestingly, the FPGA-accelerated Q-Tool allows the centralized scheme to show setup delays similar to the distributed approach, while still keeping the low BP figures characteristic of a centralized impairment-aware route computation with fully updated physical layer and wavelength availability information (as depicted in Figure 5.22, right, from previous section). For instance, for a medium load of 70 Erlang, the FPGA acceleration allows to reduce the setup time from 4 seconds to 2.21 seconds (45% reduction), quite close to the 1.5 seconds of the distributed approach, whereas its BP is 3% compared to the 4.4% in the distributed case. For higher loads, even more pronounced setup time reduction is observed, for example, 60% improvement for 90 Erlang.

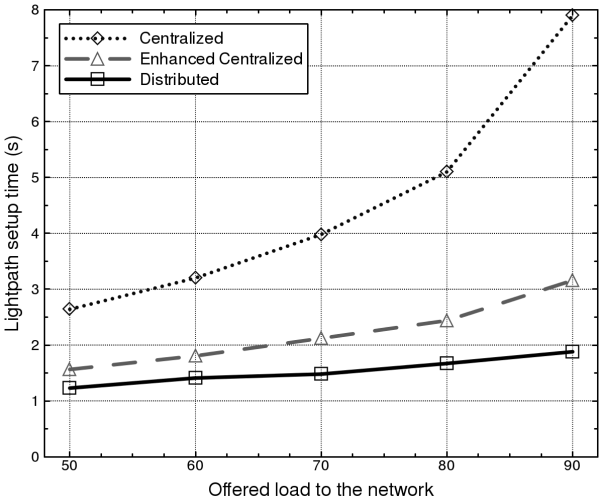
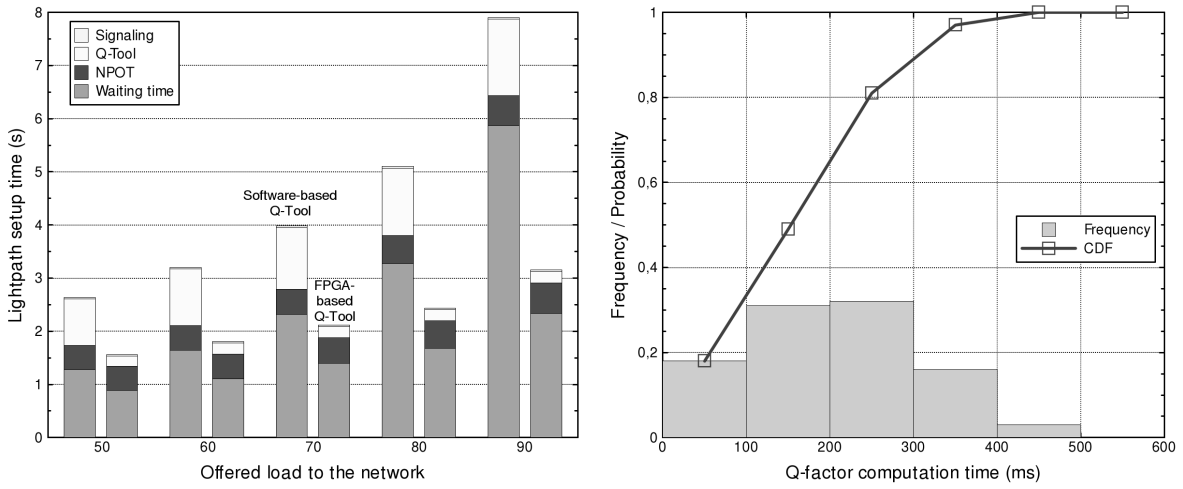


FIGURE 5.23: Average lightpath setup time vs. offered load.

Figure 5.24 (a) breaks down the different contributions to the lightpath setup time. Specifically, the waiting time spent in the queue of the NPOT until the request can be processed, the NPOT processing time, the Q-factor computation time and the control plane signaling time have been considered. The NPOT processing time comprises the total amount of time needed to process the route request, trigger the IA-RWA algorithm and send the route back to the source OCC. Leaving the waiting and signaling times aside (the latter, around 30 ms, is almost imperceptible in the graph), the Q-factor computation can be identified as the route computation bottleneck in case of the software-based Q-Tool. Indeed, it becomes 1.43 s for 90 Erlang, compared to the 570 ms of the NPOT processing time. In contrast, the FPGA-accelerated Q-Tool enables much faster Q-factor estimation, also scaling perfectly with the load offered to the network. As shown, the Q-factor computation time only raises from 190 ms to 218 ms when increasing the offered load from 50 to 90 Erlang, that is, only a 15% increment while almost doubling the load offered to the network. In particular, in a medium-loaded scenario with 70 Erlang, the average Q-factor computation time can be decreased from 1.16 s to 210 ms with the FPGA acceleration (x5.5 faster). On average, the Q-Tool is invoked to compute the Q-factor value of about 10 lightpaths simultaneously. Figure 5.24 (b) plots the Cumulative Distribution Function (CDF) and relative frequency of the FPGA-accelerated Q-factor computation time for the same offered load. Evidently, around 80% of the Q-factor computation invocations experience a response time below 300 ms, while no invocation goes beyond 500 ms. Note that the measured lightpath setup times remain drastically below the 10 seconds of the impairment-aware network management system presented in [TMKO08]. As a matter of fact, having an empty queue in the centralized NPOT, which means a zero delay in the incoming requests, setup times under 1 second would be obtained in every evaluated load scenario.



(a) (b)

FIGURE 5.24: (a) Lightpath setup time breakdown for Software-based (left) and FPGA-based (right) centralized approaches; (b) Q-factor computation time frequency and CDF with FPGA-acceleration (load = 70 Erlang).

5.7 Experimental centralized lightpath restoration for impairment-aware GMPLS-controlled optical networks

Heretofore, novel impairment-aware lightpath establishment mechanisms, as well as involved elements and processes, have been described. Moreover, failure localization and isolation has been presented as one of the main drawbacks associated to end-to-end lightpath transparency, as Loss-of-Light (LoL) alarms stemming from a single failure propagate downstream along the path from the failure point. This section describes the work developed to address this issue, including improved NPOT modules and several strategies to speed up lightpath restoration. The experimental assessment of these studies is presented as well.

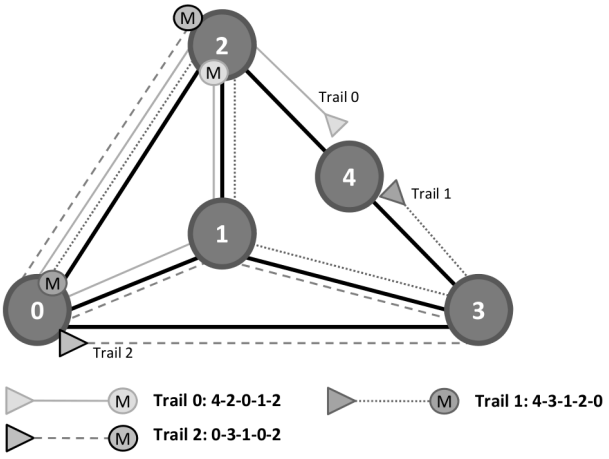
5.7.1 Failure handling in the NPOT

Provided that the centralized scheme equipped with an FPGA-accelerated Q-Tool attains setup times similar to the distributed approach also providing better performance in terms of blocking probability and, thus, better network usage; a centralized mechanism has been proposed to address the failure handling problem.

In this mechanism the NPOT failure localization module relies on the monitor placement one (see Figure 5.1 on page 61), which implements an innovative meta-heuristic, called MeMoTA (Meta-heuristic for Monitoring Trail Assignment), that solves the problem of hard-failure localizations in the case of large networks with high connectivity [HDG10]. The main target of the MeMoTA algorithm is to build a Monitoring Trail solution (m-trail, [WHY09]), able to provide an exact detection of the broken link with reduced monitoring CAPEX. Therefore, the MeMoTA meta-heuristic is executed in the network planning phase to obtain the optimal configuration of m-trails enabling a successful failure detection and localization. Basically, an m-trail is an Optical Supervisory Channel spanning several network links, with a signal transmitter located at the m-trail source node, and a monitor at the m-trail destination node with the receiver. By definition, an m-trail can traverse a certain network node many times, but a link no more than once. When a link failure happens, the monitors placed at the head-end of all those m-trails supported on the failed link detect a LoL and generate a failure alarm with the identifier of the disrupted m-trail. By collecting these alarms, the NPOT failure localization module is able to identify the failed link unambiguously.

Figure 5.25 depicts an example of the m-trail solution for a 5-node network. In this case, three m-trails are enough to detect and localize any failed link in the network without ambiguity. It is worth mentioning here that, if a simple OSC solution (i.e. implementing an OSC per fiber link) would have been deployed, seven monitors instead of three would

have been required, thus increasing the final network CAPEX. The table in the figure reflects the unique code assigned to each network link as a function of the alarms raised upon its failure. Once implemented in the NPOT failure localization module, this table reduces the failure localization process to a simple and fast alarm code mapping where, for example, if monitors from m-trails (0) and (2) send a LoL alarm to the NPOT, it is clear that the failure affects to link (0,1). Note that, since the NPOT requires information about all disrupted m-trails to successfully localize a given failure, a hold-off timer is set once a failure alarm is detected. Once the timer expires, localization is performed using the alarms received so far.



Link	T2	T1	T0
(0,1)	1	0	1
(0,2)	1	1	1
(0,3)	1	0	0
(1,2)	0	1	1
(1,3)	1	1	0
(2,4)	0	0	1
(3,4)	0	1	0

FIGURE 5.25: Example of an m-trail design in a 5-node network consisting of 3 trails. The resulting alarm code table is shown beside

5.7.2 NPOT enhancements for fast lightpath restoration

As experimentally demonstrated in section 5.6, the centralized approach performs significantly better than the distributed one in terms of lightpath blocking probability. Moreover, thanks to the FPGA-accelerated Q-Tool it shows similar setup times in spite of the high computational complexity of the IA-RWA process at the centralized NPOT. However, while setup times in the order of a few seconds may be acceptable for a new lightpath establishment, they would lead to large data losses in case that failure restoration was implemented in the network. In view of this, this thesis targets at enhancing the centralized NPOT so as to successfully fulfill the failure restoration needs for future optical core networks, while still taking benefit from its accurate impairment-aware path computation feature. The improvements introduced here comprise a prioritized NPOT scheduler to efficiently support traffic with differentiated resilience requirements and extensions to the NPOT-OCC protocol (detailed in section 5.5.3) enabling resource pre-reservation to allow consecutive backup path computations. These enhancements, along with the hardware-accelerated Q-Tool, are expected to provide acceptable lightpath restoration times. The enhanced NPOT architecture is depicted in Figure 5.26.

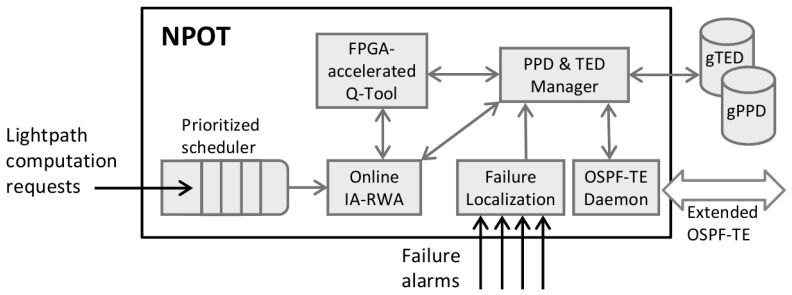


FIGURE 5.26: Network Planning and Operation Tool (NPOT); main blocks involved in impairment-aware dynamic lightpath restoration.

5.7.2.1 Prioritized NPOT scheduler

In the network operation phase, the NPOT scheduler (Figure 5.26) queues the incoming requests until the online IA-RWA module becomes available, since the NPOT is only able to process one path computation at a time. This operation becomes of critical importance during the failure restoration procedures, where a large number of restoration requests must be served almost simultaneously. Besides, as pointed out before, the IA-RWA process is a computationally intensive task. Hence, the online IA-RWA engine requires a significantly large amount of time to process each restoration request (even in the order of a few hundreds of milliseconds with the hardware acceleration features already presented in this thesis). This makes the latest restoration requests to arrive at the centralized NPOT experience high queuing delays, which could even compromise the success of their restoration.

Taking into account the large variety of data applications that are expected to run over next-generation optical networks, requesting each one of them differentiated treatment in case of failure (e.g., a maximum recovery time), the NPOT has been equipped with a prioritized scheduler aiming to serve those restoration requests with higher resilience requirements (i.e., higher priority) first, incurring them lower recovery times. The prioritized scheduler, governed by the NPOT scheduler manager, implements n software-based First-In-First-Out (FIFO) queues, one for each lightpath priority supported on the network. This solution becomes interesting provided that a reasonable number of priorities are defined in the network, since no sorting algorithm is needed in the prioritized scheduler. In particular, two different priorities are considered in this thesis.

The operation of the scheduler is the following. In the moment that the NPOT scheduler manager detects that any queue of the prioritized scheduler has path requests to be served, it checks the availability of the online IA-RWA engine. Being this one free, the manager starts checking the state of the priority queues, from the highest to the lowest priority (e.g., from 1 to n). Finding any restoration request in any of those queues, the manager pops the request that arrived at the queue first, sends the request to the online IA-RWA engine and stops the search. Upon receiving the request, the online IA-RWA engine changes its state to busy, notifying the NPOT scheduler manager accordingly,

and processes the request. Once the request is processed, the online IA-RWA engine changes its state back again to free and notifies the NPOT scheduler manager. Upon this notification, provided that any request is waiting in the scheduler, the same operation described above is performed again and again, thus emptying the queues in the scheduler according to their priority.

5.7.2.2 NPOT-OCC protocol extensions for resource pre-reservation

Multiple requests must be processed one after another during the restoration procedures. Therefore, the state of those resources involved in a backup path computation must be rapidly updated for subsequent requests. In the initial NPOT implementation, a guard time was left between two consecutive route computations. This allowed gPPD and gTED databases being fed with the new PLI and wavelength availability information through OSPF-TE, once the signaling phase of the related connection was completed. As highlighted in [PSA+11], however, this operation forces the NPOT to remain several seconds idle between successive computations, that is, the time it takes OSPF-TE to update, network wide, the new state of the involved resources plus the time needed to add the new active lightpath to the NPOT gPPD. In order to avoid the aforementioned idle periods in the centralized NPOT, which would likely lead to undesirably high restoration times, the NPOT-OCC protocol has been extended to enable a resource pre-reservation mechanism to prevent the use of resources assigned to lightpaths under establishment when processing new requests.

Aiming to avoid the gPPD and gTED inaccuracies between successive path computations, two alternative states are introduced for the lightpaths stored in the gPPD database, namely, *under establishment* and *active*. When a route is successfully computed for an incoming lightpath request, and before replying the source node that performed the request, the lightpath is stored in the gPPD and its state set to under establishment. Moreover, those resources supporting the lightpath are blocked, thus preventing them to be used in successive lightpath computations, while waiting to be eventually committed. Specifically, a timeout counter is set for every lightpath under establishment. In this way, if no commit request is received before the timeout counter expires, the lightpath is released and its respective resources unblocked. Finally, the computed route and wavelength for the lightpath, together with the identifier of the lightpath are sent back to the source node of the request.

Upon receiving the route and wavelength from the centralized NPOT, the source node of the connection starts the signaling procedures through standard RSVP-TE protocol. In case that the lightpath is successfully established, a commit message is sent from the source node to the centralized NPOT, containing the identifier of the lightpath whose state has to be changed to active, and its related network resources committed. Conversely, if the signaling of the lightpath fails for any reason, the source node sends a rollback

message, containing the identifier of the lightpath under establishment that has to be released and its resources unblocked.

FIGURE 5.27: Extended NPOT-OCC protocol: action message to commit or rollback the pre-reservation.

Figure 5.27 depicts the format of the new NPOT-OCC protocol message. As illustrated in section 5.5.3 (page 80), in the route request message (Figure 5.19, left), the identifiers of the source and destination nodes, together with the identifier of the new lightpath request (LSP Id) generated by the source node are included. In turn, the reply message (Figure 5.19, right) returns the result of the request (whether or not a route and a wavelength has been found), the assigned wavelength and, then, the succession of node and output interface identifiers throughout the path. Finally, the new action message (Figure 5.27) serves the purpose of informing the centralized NPOT about the success or failure of the signaling process. In such cases, the action to be performed is set to commit (action = 0) or rollback (action = 1), respectively. Note that, even though the proposed protocol is proprietary, it has been designed to tightly match the information needed by the standard control plane protocols, such as RSVP-TE.

5.7.3 Impairment-aware lightpath restoration example

Once the enhancements to the NPOT for fast restoration have been presented, this subsection shows an example with all the interactions among the control plane elements to achieve a successful restoration. For this purpose, the scenario depicted in Figure 5.28 is used, which represents the same network and m-trails design as in Figure 5.25 (page 87).

In the figure, assume that a lightpath is established between nodes 0 and 3, and a failure occurs on the link that connects nodes 1 and 3. This failure affects both the active lightpath, whose connectivity has to be restored as soon as possible, and the m-trails supported on the failed link, namely m-trails 1 and 2. Specifically, m-trail disruptions will be detected on the trail termination nodes 0 and 2, respectively, where monitors are located. These LoL events are notified to the corresponding OCCs. Detecting the failure, OCCs also inform the centralized NPOT about the failure state and the m-trail that has detected it. Therefore, the failure localization module in the NPOT can localize the failure by a simple alarm code table look-up. Indeed, looking at the alarm code table on the right of the figure, one can easily see that code 6 (i.e., T2=1, T1=1, T0=0) identifies the failed link (1,3). After localizing the failure, the module isolates the link

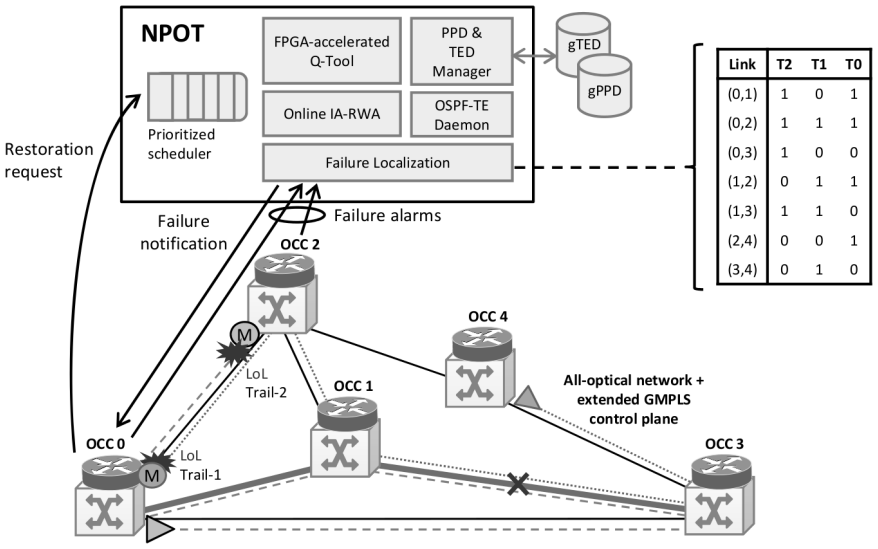


FIGURE 5.28: Impairment-aware restoration example: a bidirectional lightpath is established between nodes 0 and 3 (following the route 0-1-3), and a failure occurs on the link connecting nodes 1 and 3, thus disrupting m-trails 1 and 2.

by updating the gPPD and gTED databases accordingly. Next, it notifies the source nodes of all affected lightpaths, so that they can start the restoration procedures. In the implemented centralized restoration scheme, failure restoration is delegated to the GMPLS control plane, in particular, to the standard RSVP-TE protocol [RFC4872]. Therefore, as soon as the source node is notified about the failure, it requests a backup lightpath computation to the NPOT for restoration purposes. This request is initially stored in the NPOT scheduler, where it will be queued according to its priority, until served by the online IA-RWA module. If a feasible backup route is found, it is returned to the source node of the lightpath and the involved resources are pre-reserved in gPPD and gTED databases. Finally, the source OCC triggers the RSVP-TE signaling to establish the backup path. Upon successful backup lightpath establishment, every intermediate node will update the state of the allocated resources by means of the OSPF-TE protocol. Moreover, the source OCC sends a commit action message to the centralized NPOT in order to commit the pre-reserved resources and change the state of the lightpath to active. Conversely, if the backup lightpath establishment is unsuccessful, the source OCC informs the NPOT about this event by means of an action rollback message, which releases the lightpath under establishment and restores the availability of those pre-reserved resources that will not be finally used.

5.7.4 Experimental results

The performance of the proposed centralized lightpath restoration approach was validated on the CARISMA network. The test bed described the same topology as the 14-node DT network (Appendix A.3), assuming 10 bidirectional wavelengths per fiber link. Moreover, the centralized NPOT was deployed on an Intel Core i3 350 Linux server at 2.93 GHz.

This computer was directly connected to the FPGA board implementing the Q-Tool over a Gigabit Ethernet full-duplex link. The specific details of the used FPGA can be found in the previous section.

For evaluation, five different network scenarios with 10, 20, 30, 40 and 50 bidirectional lightpaths established between randomly selected node pairs with acceptable QoT were considered. A high or a low priority was assigned to these lightpaths according to a 25-75% high-low priority ratio. In each of these scenarios, the network was loaded until having the desired number of active lightpaths. Then, a failure was forced on a randomly selected link among those supporting lightpaths. Such operations were performed 100 times for each scenario, ensuring in this way a statistically relevant number of experiments.

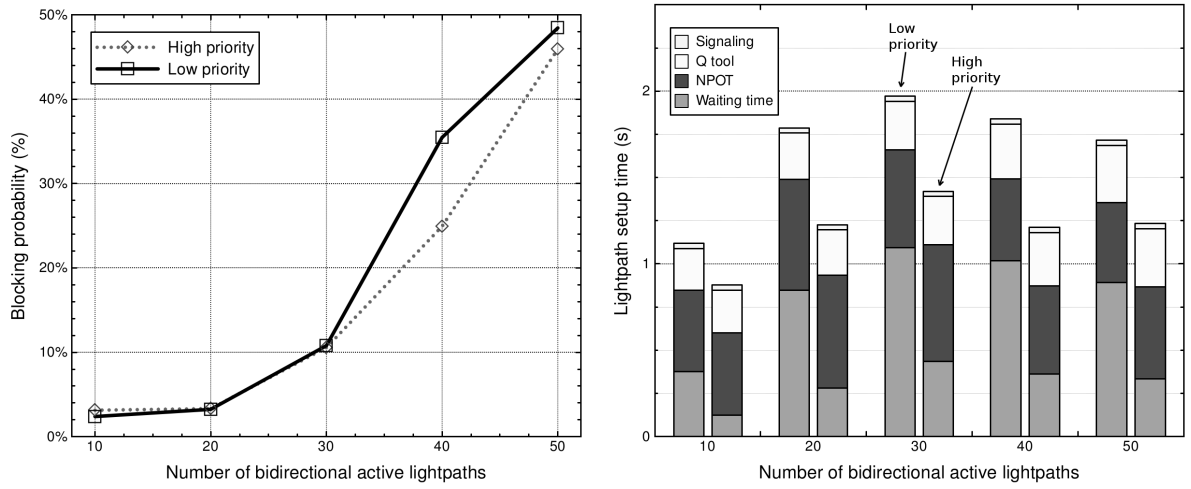


FIGURE 5.29: (a) Lightpath restoration BP as a function of the number of active lightpaths in the network for high and low priority traffic classes; (b) Average setup time breakdown as a function of the number of active lightpaths for high (right bars) and low (left bars) priority traffic. The waiting time in the scheduler, NPOT processing, Q-Tool computation and backup path signaling time contributions have been considered.

Figure 5.29 (a) depicts the observed restoration blocking probability in every evaluated scenario, that is, the probability that a restoration request is blocked either due to unacceptable QoT or resource unavailability. As can be observed, this probability increases with the number of active lightpaths in the network. This can be justified by the fact that a lower number of spare resources exists in the network. Hence, while some lightpath restoration requests cannot be completed due to resource unavailability, others also fail due to very long physical routes that do not match the QoT requirements. It is interesting to highlight that, in the scenario with 40 active lightpaths, high priority traffic experiences notably lower restoration blocking probability, since high priority restoration requests are served first using the spare network resources. With 50 active lightpaths, however, both curves get closer again because of the general resource scarcity, as only 10 wavelengths are available per fiber link. As a matter of fact, this resource scarcity is

the main cause of the increased restoration blocking probability observed in the highest loaded network scenarios.

Figure 5.29 (b) plots the average restoration time in the different load scenarios, which has been additionally split in their major contributions. Specifically, the average of the waiting time in the prioritized scheduler, the processing time in the NPOT, the Q-Tool computation time and the RSVP-TE backup path signaling time are considered. The processing time in the NPOT comprises the total amount of time needed to process the incoming route request once coming out of the prioritized scheduler, execute the IA-RWA algorithm, collect the resulting backup path and send it back to the source node OCC. It can be observed that restoration times start increasing with the number of lightpaths in the network. This is mainly caused by the increasing number of connections affected per failure that have to be queued in the prioritized scheduler until they can be processed by the NPOT online IA-RWA module. Indeed, while the other contributions remain quite steady, the waiting time incurred in the prioritized scheduler highly determines the restoration time behavior. This demonstrates the necessity of a resource pre-reservation strategy to ensure fast resource update between successive computations, together with priorities in the NPOT scheduler, guaranteeing low restoration times for the high priority critical traffic. Focusing on the restoration time differences between high and low priority traffic, they also increase with the number of active lightpaths. For instance, while waiting time differences of approximately 250 ms are observed in the scenario with 10 active lightpaths, these rise up to 650 ms in the scenario with 30 active ones. Towards higher loads, however, restoration times start decreasing because backup routes are more difficult to find. Thus, if no available route exists for a request, there is no need to invoke the NPOT Q-Tool, which reduces the queuing time of the subsequent requests still waiting to be served. This restoration time reduction is especially noticeable for low priority lightpaths, which are always the last ones to be served. The results in Figure 5.29 (b) also prove the scalability of the FPGA-accelerated Q-Tool that was integrated into the NPOT. In fact, Q-Tool time differences of merely tens of milliseconds are experienced when changing from the scenario with 10 active lightpaths to the one with 50.

To assess the benefits of the hardware-accelerated Q-Tool, Figure 5.30 compares the performance of the accelerated Q-Tool with a software-based Q-Tool (previously used in [APA+10a]) running along with the NPOT over the same hardware platform. In particular, the Q-Tool time improvement achieved with the FPGA-based hardware acceleration is depicted for the five different network scenarios considered so far. As shown, under light traffic loads, improvements remain quite low (around 20-30%), since the number of overlapping active lightpaths in the backup route, whose QoT feasibility has to be checked, is generally low. However, as more and more active lightpaths exist in the network, which can be supported on any of the backup route spans, QoT estimations become more computationally intensive. This makes software-based Q-Tool implementations inefficient, thus inappropriate for supporting critical network functionalities like dynamic impairment-aware lightpath restoration. For instance, in the scenario with 50 active lightpaths, FPGA-accelerated QoT estimations take around 330 ms, whereas they increase up

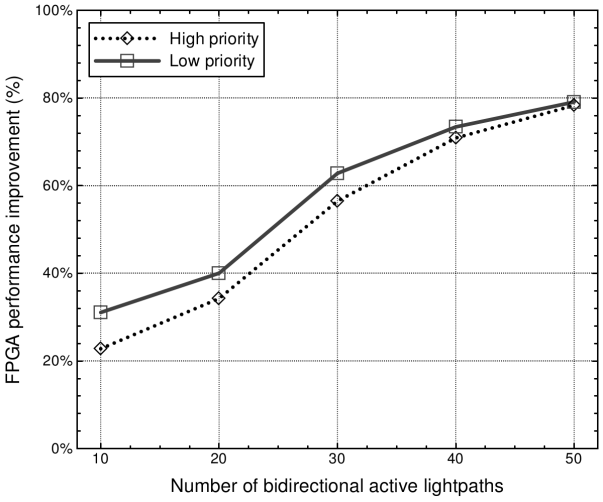


FIGURE 5.30: Relative FPGA QoT estimation time improvement as a function of the number of active lightpaths in the network compared to the case where a software-based Q-Tool is used.

to 1.5 seconds in the software-based Q-Tool, leading to an improvement around 78%. The slight Q-Tool time improvement differences between low and high priority restoration requests arise from the fact that, being the last ones to be served, low priority backup path computations face a sensibly increased lightpath overlap, as backup routes are typically longer than the primary ones. This increases the computational complexity in the Q-Tool, making the software-based Q-Tool to perform worse in such situations.

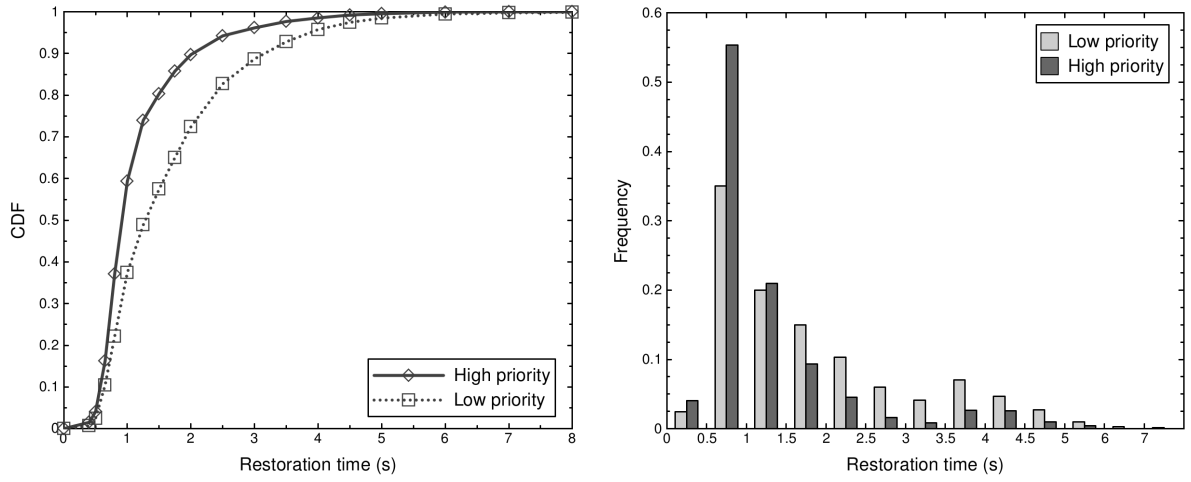


FIGURE 5.31: Cumulative Distribution Function (a) and Frequency (b) of the lightpath restoration time for high and low priority traffic classes.

In Figure 5.31 (a), the Cumulative Distribution Function (CDF) of the lightpath restoration time for both high and low priority traffic classes has been illustrated. For these results, the entire set of successful restorations in all network scenarios has been used. The CDF function represents the probability that a lightpath restoration takes a time

lower or equal to a certain value. As seen, the CDF of the high priority traffic grows significantly faster than the one of the low priority traffic, meaning faster restoration. For instance, it can be observed that more than 60% of the high priority restorations incur sub-second restoration times, whereas 90% of them are restored in less than 2 seconds. Moreover, focusing on the low priority traffic, even though it experiences longer restoration times, almost 90% of the low priority restorations are performed in less than 3 seconds, totally acceptable if we assign this class of service to the best-effort traffic on the network. For better illustration, Figure 5.31 (b) plots the lightpath restoration time frequency for both high and low priority traffic classes. From the figure, most of the high priority restorations (55%) are concentrated on the interval between 0.5 and 1 second, while none of them experience a restoration time higher than 5 seconds. In the case of the low priority traffic, it also experiences its peak (35%) on the same interval between 0.5 and 1 second, but with a longer and more pronounced tail that finally results in such slightly increased restoration times. Taking all these results into account, average restoration times of 1.16 and 1.64 seconds for the high and low priority traffic have been finally obtained, respectively. Such an outcome highlights the success of the resilience differentiation policies applied in the NPOT, so as to meet the diversity of resilience requirements posed by new emerging data applications. Moreover, the obtained values represent important restoration time improvements compared to the previous work reported in [APA+10a], where the NPOT implementation used a software-based Q-Tool, and the control plane lacked fast resource pre-reservation features for successive backup path computations.

5.8 PCE-based scheme for impairment-aware GMPLS-controlled dynamic optical networks

In previous chapters the Path Computation Element was presented as the main character for centralized route calculation in a GMPLS-controlled optical transport network. The work realized by the standardization bodies in this regard was presented as well. A step forward is done here, targeting at the standardization of the centralized scheme introduced in section 5.5. In brief, provided that the NPOT is essentially a route calculation entity, it can be encapsulated inside a PCE and integrated into the control plane by means the PCEP communication protocol.

Nonetheless, the transformation of the NPOT into a PCE as defined by the IETF is not straightforward due to the operational characteristics of the NPOT. The centralized NPOT follows a stateful approach, that is, it keeps track of both available network resources and active lightpaths in the network (stored in the gPPD database, as mentioned in section 5.5). Compared to the standard PCE, where a stateless implementation may be enough to compute end-to-end routes, based only on the network resource availability, statefulness is mandatory in the centralized NPOT, so as to account for the non-linear

PLIs (SPM, XPM and FWM) when assessing the feasibility of the incoming lightpath requests. In this regard, since [RFC4655] contemplates the possibility of using a stateful PCE, PCEP extensions are proposed in next section to enable the evolution of the NPOT towards a stateful PCE. It is worth to mention here that, although stateful PCE has not been a main goal in the scope of the IETF, some efforts have been recently dedicated to this issue (see [PCE-StatFul]). Furthermore, while the centralized NPOT performs both routing and wavelength assignment, the *PCRep* message, which is used by the PCE to send the route to the OCC for signaling, lacks an object to convey the wavelength assigned to the new lightpath. To address this, a *Wavelength* object is proposed to be added to the *PCRep* message in next section.

With these extensions, the lightpath establishment mechanism in the new scheme becomes basically the same as the one depicted in section 5.5.1 for the centralized approach but, in this case, the communication between the OCC and the PCE is done by means of the extended PCEP protocol. In this way, the route is requested using a *PCReq* message and the calculated route is sent back through a *PCRep* message. Once the lightpath has been signaled, the new *PARes* message (described in the section below) is used by the source OCC to notify the PCE about the successful or failed setup. Note that the extensions added to the PCEP protocol incorporate the functionalities formerly assigned to the previously defined NPOT-OCC protocol.

Figure 5.32 depicts this evolved impairment-aware centralized control plane architecture, where the NPOT becomes the routing engine of a stateful PCE. Note that a new NPOT-PCE interface (based on the former NPOT-OCC one, section 5.5.3) is used to facilitate the communication between the NPOT and PCE modules.

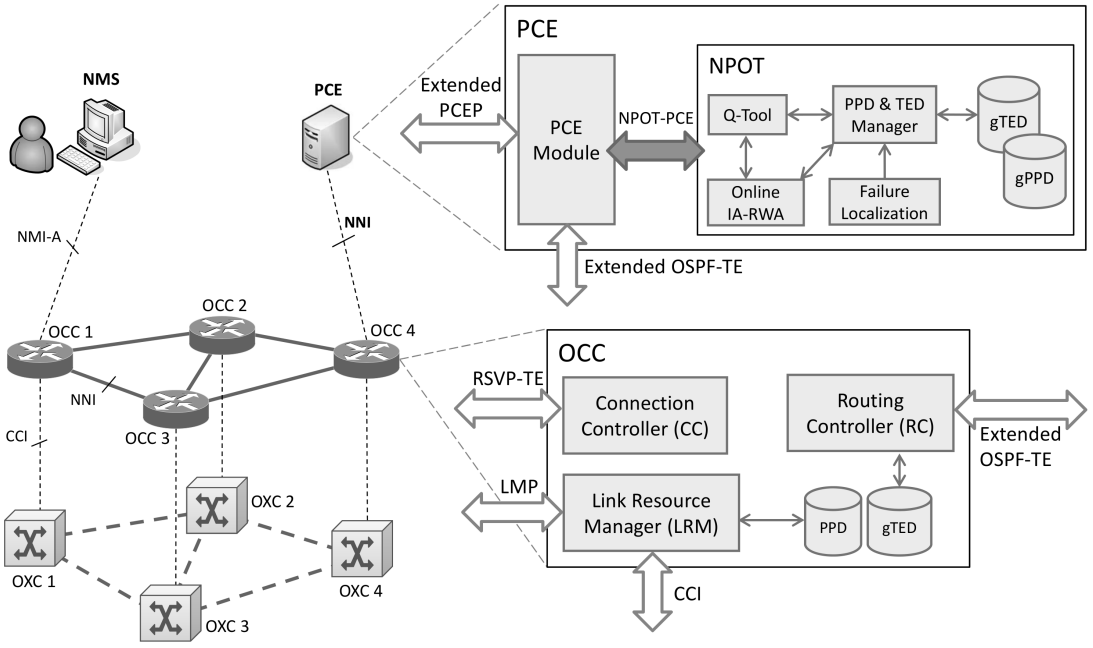


FIGURE 5.32: PCE-based centralized control plane integration scheme

5.8.1 PCEP protocol extensions

Three PCEP extensions have been defined to enable the integration of the NPOT within the PCE module. As said, such extensions comprise a new *Wavelength* object, which is added to the *PCRep* message and conveys the wavelength assigned to the new lightpath, and two new messages that allow the stateful PCE to keep track of the active lightpaths in the network.

5.8.1.1 Wavelength object

Figure 5.33 shows the specification of the *Path Computation Request* message once it has been extended to carry the new *Wavelength* object. This object contains the common object header defined in [RFC5440] whose assigned values for the object class and type are 21 and 1, respectively. Moreover, the object carries the wavelength assigned by the NPOT to the requested lightpath encoded as stated in [RFC6205] (right side of the figure). Hence, the *Wavelength* object is added to the path attributes list which, joined to the *ERO*, completes the RWA information needed by the control plane to establish the lightpath. The *RP* object contains the identifier of the lightpath to facilitate its management during operation.

```
 <PCRep Message> ::= <Common Header>
                     <response-list>

where:

 <response-list> ::= <response>[<response-list>]
 <response> ::= <RP>
                [<NO-PATH>]
                [<attribute-list>]
                [<path-list>]
 <path-list> ::= <path>[<path-list>]
 <path> ::= <ERO><attribute-list>

where:

 <attribute-list> ::= <Wavelength>
                      [<LSPA>]
                      [<BANDWIDTH>]
                      [<metric-list>]
                      [<IRO>]
 <metric-list> ::= <METRIC>[<metric-list>]
```

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Object-Class  |  OT   |Res|P|I|         Object Length         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid |  C.S. |    Reserved     |              n                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

FIGURE 5.33: Extended Path Computation Reply (PCRep) message and Wavelength object specification.

5.8.1.2 Path Allocation Result message

The proposed *Path Allocation Result* (*PARes*) message is used by the OCC to inform the PCE about the success or failure of the lightpath establishment process. As illustrated in Figure 5.34, upon a successful establishment, the message conveys the lightpath identifier, the signaled route, and the used wavelength by means of the *RP*, *RRO* and *Wavelength* objects, respectively. Then, the new lightpath information is stored in the NPOT's gPPD. Contrariwise, a *NO-PATH* object is sent within the message.

```
<PARes Message> ::= <Common Header>
                    <RP>
                    [<RRO><Wavelength>] or [<NO-PATH>]
```

FIGURE 5.34: Path Allocation Result message.

5.8.1.3 Path Tear-down Result message

Once a lightpath is torn down, the source OCC notifies the PCE through a *Path Tear-down Result* (*PTRes*) message (Figure 5.35). In this way, the lightpath is erased from the gPPD and not taken into account by the NPOT for further route computations.

```
 <PTRes Message> ::= <Common Header>
                     [<RP-list>]
where:

 <RP-list> ::= <RP>[<RP-list>]
```

FIGURE 5.35: Path Tear-down Result message.

5.9 Chapter summary

In this chapter transparent optical transport networks have been introduced as the most efficient solution to implement future core optical networks, since they promise to overcome bandwidth limitations also reducing the CAPEX. However, these improvements may be reached at a cost. In particular, transparent networks present two important drawbacks, related to lightpath QoT and failure management, that need to be tackled.

In this context, a cross-layer approach, relying on the collaboration between management, control and data planes, seems to be suitable to face both challenges. In this way, the monitoring of the data plane provides the control plane with information related to the status of the physical layer components which is then used for QoT assured lightpath establishment. Similarly, monitoring the physical layer allows failure detection. Here, alarms are sent to the control plane, which starts failure localization and lightpath

restoration mechanisms. This chapter has presented different proposals to implement this solution from the control plane perspective.

To motivate the subject, a brief explanation of the physical layer effects involved in all-optical data transmission has been done. Next, since the main contributions of this thesis are focused on the control plane, the most relevant architectural solutions (centralized and distributed) to implement an impairment-aware control plane have been presented. To conclude the introductory part of the chapter, the framework in which this work was conducted (i.e., the ICT FP7 DICONET project) has been described.

Section 4 has detailed the distributed approach proposed within the DICONET project to deploy a GMPLS-controlled impairment-aware optical transport network. This solution has been afterwards compared to the centralized scheme proposed in the framework of this thesis, which has been presented in section 5. Architectural and functional details have been provided for each approach in these sections.

In the evaluation conducted in section 6, the distributed approach has provided one fifth of the lightpath setup time than that of previously reported (centralized) alternatives found in the literature, also outperforming the centralized approach detailed in section 5 especially for high traffic loads. For low traffic loads, however, the proposed centralized approach results in reduced lightpath blocking ratio, and setup time delays similar to the distributed solution, thus being more appropriate in such scenarios. Nonetheless, the efforts devoted to reduce lightpath setup time by means of an enhanced version of the NPOT, provided with an FPGA-accelerated Q-Tool, have shown their results leading to setup time figures close to the distributed architecture while preserving reduced BP ratios. This new version of the centralized scheme takes advantage of an improved QoT estimation, which is the bottleneck in the IA-RWA process, to obtain reduced overall setup times.

Section 7 has reported the experimental demonstration of a dynamic impairment-aware restoration scheme implementing differentiated resilience support. The proposed scheme builds upon a centralized NPOT entity, which provides the network with online IA-RWA, as well as failure localization and isolation functionalities upon fiber link failures. First of all, the failure handling module of the NPOT has been described. Next, the enhancements applied to the NPOT for fast lightpath restoration have been described. Such enhancements are put together with the FPGA-accelerated Q-Tool introduced in section 6 and include the implementation of a prioritized scheduler in the NPOT so as to efficiently serve lightpaths with differentiated resilience requirements, and the implementation of a fast resource pre-reservation protocol allowing consecutive path computations without having to remain idle until the respective resource state updates. From the experimental evaluation, successful impairment-aware lightpath restoration times of around 1.16 and 1.64 seconds have been achieved for the high and low priority traffic, respectively. These results, totally permissible in real network environments, leverage dynamic restoration as a key element to provide the desired reliability levels to the future optical Internet.

Finally, section 8 has investigated the requirements to standardize the proposed centralized scheme. A PCE-based approach has been proposed here as the most appropriate solution, and the necessary functional and protocol extensions have been provided.

Chapter 6

PCE-based routing in GMPLS-controlled multi-domain optical transport networks

As current transport network infrastructures grow, they are commonly segmented into multiple domains due to administrative, technological, or scalability reasons. Furthermore, the interconnection of network infrastructures managed by different operators is a mandatory requirement to provide long distance connectivity. Here, confidentiality and reliability concerns become of paramount importance. Indeed, these particular interconnections usually require a certain degree of opacity when exchanging network information, since operators want to share the least information possible with their potential competitors. When computing end-to-end inter-domain paths, this is normally translated into a procedure in which each operator computes the path segment inside its administrative domain (intra-domain path). In the context of GMPLS-controlled optical networks, the IETF efforts are currently focused on the standardization of the Path Computation Element as the main entity for calculating end-to-end paths in both single-domain and multi-domain scenarios. While the former (i.e., intra-domain end-to-end path computation) has been studied extensively, inter-domain end-to-end path computation is still a work in progress.

This chapter proposes several mechanisms to provide end-to-end paths in a multi-domain scenario. In particular, the interconnection of several PCE-based GMPLS-controlled optical transport networks is assumed. Based on the configurations outlined in section 2.3.3, the architecture considered in this work is firstly detailed and motivated. Next, the current routing strategies utilized in the scenario under study, namely per-domain, BRPC and PCF, are reviewed in section 2. The main contributions of this chapter are presented in sections 3 and 4, which are devoted to studying the trade-off between scalability and performance of enhanced versions of the PCF and BRPC mechanisms, respectively. These contributions are summarized in the last section.

6.1 PCE-based multi-domain control plane architecture

Section 2.3.3 has introduced the different multi-domain architectural solutions proposed by standardization entities. In GMPLS-controlled optical networks, which is the focus of this thesis, the PCE arises as the most accepted solution for multi-domain route computation. Currently, two main PCE deployments are being studied. The first one contemplates the collaboration between peer PCEs each one controlling a single domain (flat model). In this model, each PCE computes the segment of the end-to-end multi-domain route that traverses its controlled domain. The second one relies on a hierarchical deployment where a parent PCE entity is placed over the (child) PCEs responsible for each domain. Here, the parent PCE uses information coming from child PCEs to compute end-to-end multi-domain routes. Although hierarchical PCE models are gaining momentum in the framework of the IETF (e.g., see [PCE-Hierarc, H-PCEP, GPCC11, CMM+11]), the contributions in this chapter are based on the peer PCE deployment scheme (Figure 6.1). In fact, the hierarchical model presents some practical problems such as who should take the responsibility to manage a parent PCE. This issue is specially clear in a scenario involving multiple operators.

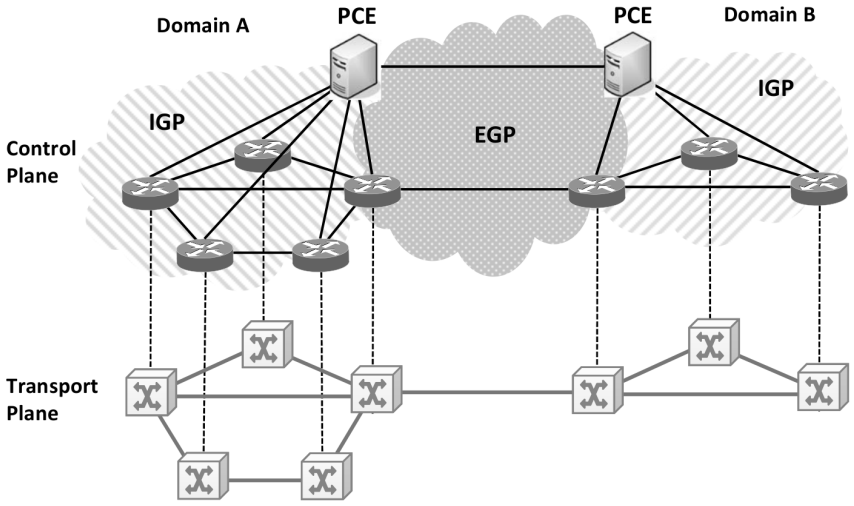


FIGURE 6.1: Collaborative PCE multi-domain architecture.

The figure above depicts the flat model where PCEs cooperate to provide end-to-end multi-domain paths. Therein, the interconnection between the DCNs of two GMPLS-based control planes is illustrated. While the connectivity among the network elements of the same domain is provided by an IGP (e.g., OSPFv2), an EGP (like BGP) is responsible for providing inter-domain connectivity. An equivalent scheme, based on IGP routing areas instead of IGP/EGP domains, can be found in [VAM+10].

These IP protocols enable the TE activity in the whole scenario, that is, TE routing and signaling, by means of OSPF-TE, PCEP and RSVP-TE protocols. It is noteworthy that

no TE exterior routing is assumed in this first approach to the architecture. Being this issue a key point in multi-domain, it will be discussed throughout the chapter.

6.2 PCE-based multi-domain routing strategies

This section summarizes the mechanisms proposed by the standardization bodies in the context of end-to-end inter-domain path computation. These mechanisms arise as a former solution to the problem but, as will be pointed out along this chapter, there are several open issues to be solved yet.

In the first approach PCEs are provided with the sequence of domains to be traversed from source to destination. Then, a *per-domain path computation* can be performed over this sequence [RFC5152]. In particular, the end-to-end inter-domain path is computed and provisioned on a per-domain basis so that every domain computes the path segment from its ingress node to the egress node connected to the following domain. Nonetheless, this concatenation of locally optimal-path segments likely leads to sub-optimal inter-domain paths.

Facing these limitations, the IETF defines a *backward recursive path computation* (BRPC) procedure in [RFC5441] where the peering PCEs in the pre-set domain sequence cooperate for computing an end-to-end path. Specifically, the responsible PCE in each domain in the pre-defined domain sequence generates a *virtual shortest path tree* (VSPT) with all the potential end-to-end paths crossing that sequence. Then, the PCE responsible for the source domain collects all this information and processes the complete tree to obtain the best end-to-end path. Besides this, in BRPC, PCEs can preserve the confidentiality across domains returning the VSPT as a loose path, or applying the *path-key* technique [RFC5441, RFC5520]. While the former may require a new path computation in the PCE of the domain to be traversed during the signaling process, in the latter, the path segments computed by each domain are encoded by means of a key when shared with the remainder, and decoded during the signaling phase.

As said, once the source domain PCE receives the VSPT, it selects the optimal candidate path. Therefore, the optimality of the computed path strongly depends on how the domain sequence is selected. In this context, the PCE standardization has not come up with any definitive solution to this issue, being still a work in progress within the IETF. As mentioned in [PCE-Hierarc], a *path computation flooding* (hereafter, PCF) approach, where all the possible domain sequences were explored, would allow BRPC to obtain optimal end-to-end inter-domain paths automatically without requiring any pre-defined domain sequence. However, due to its poor scalability and tremendous overhead, PCF is initially discarded since it is envisaged as unpractical for large multi-domain networks.

6.3 Scalable routing in PCE-based multi-domain networks

This section evaluates the performance of the approaches introduced in the previous one in terms of connection blocking probability, network control overhead, and path computation complexity. As will be shown, PCF drastically reduces the connection BP of the per-domain approach, but faces increased path computation complexity and control overhead. This motivates the proposal of a low overhead PCF (LoPCF) mechanism, whose performance is compared to PCF and BRPC solutions. Finally, in order to completely assess LoPCF, its feasibility on top of current PCE standardization is analyzed.

6.3.1 Low overhead path computation flooding mechanism

Working on the guidelines given in [PCE-Hierarc], next paragraphs introduce a PCF mechanism allowing BRPC to directly determine the optimal domain sequence in the path computation. To this end, the network is modeled as a graph $G = (V, E)$ where V and E represent the sets of nodes and links, respectively. This global graph joins D sub-graphs, $G^i = (V^i, E^i)$, where $1 \leq i \leq D$, as D independent domains. In particular, $V^i = \{v_1^i, ..., v_{\gamma^i}^i\}$ is the set of γ^i intra-domain nodes in G^i, so that $V^i \cap V^j = \emptyset$, $i \neq j$, $1 \leq i, j \leq D$, E^i is the set of intra-domain links, $E^i = \{e(v_w^i, v_k^i) \in E : v_w^i, v_k^i \in V^i, w \neq k\}$, and $\delta_{mn}^{ij} = \{e(v_m^i, v_n^j) \in E : v_m^i \in V^i,\ v_n^j \in V^j;\ 1 \leq m \leq \gamma^i,\ 1 \leq n \leq \gamma^j\}$ is the set of inter-domain links. Each domain G^i has a PCE^i responsible for the computation of paths inside it.

In the PCF mechanism, if a source node v_s^i needs an inter-domain path to a destination node v_d^j, it requests the end-to-end route to PCE^i using a *PCReq* message. PCE^i then forwards this *PCReq* message directly to PCE^j (i.e., the PCE in the destination domain). Upon receiving the message, PCE^j looks for every adjacent domain through which the source one can be reached, and sends a *PCRep* message to each PCE in those domains.

These *PCRep* messages contain the identifier of the current path computation request (*Request Id*) in addition to the computed VSPT to the specific neighboring domain (the path segments representing the best routes between the destination node and the border nodes connected to it). Each path segment in the VSPT is identified as $(BN_n^j, v_m^j, C_{nm}^j, K_{nm}^j)$, where BN_n^j is the n-th border node in G^j, v_m^j and C_{nm}^j are the destination node and the cost metric associated to the path segment, and K_{nm}^j is the path-key stored in PCE^j together with the computed intra-domain route from BN_n^j to v_m^j, which ensures the confidentiality between domains. Finally, although not considered in [PCE-Hierarc], the PCF operation is improved with a domain list in the *PCRep* messages containing the identifier of the PCEs that have processed this *PCRep*. In this way, a loop-free flooding mechanism is obtained.

A PCE receiving a *PCRep* message updates the included VSPT, adds its PCE Id to the domain list, and replicates the message to the PCEs of all potential upstream domains, except those appearing in the list. This is repeated at every transit domain. In the end, PCE^i will receive multiple *PCRep* messages with different VSPTs. These VSPTs will be used to create an N-ary path tree over which the optimal end-to-end route will be found and returned to v_s^i in a *PCRep* message. This reverse flooding allows PCF to consider every possible VSPT from source to destination but at expenses of increased overhead and path computation complexity as demonstrated in next subsection.

Aiming to improve the scalability of PCF, LoPCF is proposed as a low overhead PCF mechanism. In LoPCF, the source domain PCE (PCE^i) sends a direct *PCReq* message to the destination domain PCE (PCE^j). Then, PCE^j finds all possible upstream domains and sends a *PCRep* message to their PCEs. Such *PCRep* messages contain the *Request Id* and the VSPT to the specific upstream domain.

```
begin
    if (PCRep → SourceNode ∈ G^u) then
        Update (PCRep → VSPT)
        AddVSPTInformationToNaryTree ()
    else if (ProcessedRequestID (PCRep → RequestId) == true) then
        Discard (PCRep)
    else
        foreach upstream domain G^k do
            PCRepCopy = Copy (PCRep)
            Update (PCRepCopy → VSPT)
            Send (PCRepCopy, PCE^k)
        StoreProcessedRequestID (PCRep → RequestId)
        Discard (PCRep)
```

Algorithm 6.1: LoPCF: PCRep message processing at PCE^u.

When a neighboring PCE^u receives a *PCRep* message, it processes it as depicted in Algorithm 6.1. Firstly, it checks if the source node is contained in its domain. If so, it updates the contained VSPT and includes this information in the N-ary tree for that *Request Id*. Otherwise, it checks if any *PCRep* for that *Request Id* was already processed. If this is true, the message is directly discarded. In contrast to PCF, LoPCF only allows a given domain to be included in the VSPT of the first *PCRep* message received, which may lead to sub-optimal inter-domain paths. Nonetheless, LoPCF allows the network to be sufficiently explored while introducing lower control overhead. If that *Request Id* was not previously processed, a *PCRep* message is forwarded to every upstream domain PCE with the VSPT accordingly updated. Finally, the *PCRep* message *Request Id* is stored to permit discarding any further *PCRep* for the same path computation request. This operation itself assures loop-free VSPT flooding, making the domain list used in PCF not strictly necessary in LoPCF. Nonetheless, it demands the *Request Id* values to be globally unique and constant along the path computation, which is not a requisite in [RFC5441].

An important issue in both PCF and LoPCF is how the *PCRep* messages are collected at the source node to update the N-ary path tree. In this work, a timer is set to ensure that all the undiscarded *PCRep* messages are collected, which allows to quantify the scalability of PCF and LoPCF.

6.3.2 Performance evaluation

The performance of PCF and LoPCF was compared to standard BRPC by means of a self-developed C++ simulator describing the 9-domain transport network depicted in Figure 6.2 (Appendix A.4), where each link carries 8 bidirectional wavelengths. In this network, intra-domain nodes are all-optical (without conversion), while border nodes employ O/E/O conversion. Poisson connection requests were generated following 70/30% intra/inter-domain ratio. For inter-domain connections, source and destination domains were uniformly selected, as well as the source/destination nodes in the respective domains. For intra-domain connections, source and destination nodes were randomly chosen in the same domain. Different loads were obtained by fixing the connection HT to 200 seconds and varying the IAT accordingly (load = HT/IAT). Connection requests were assumed to demand a full wavelength. In the simulator, PCE and domain reachability were configured manually, although the latter could be alternatively provided by means of BGP. Note that PCEP messages are sent through the shortest route in terms of PCE hops.

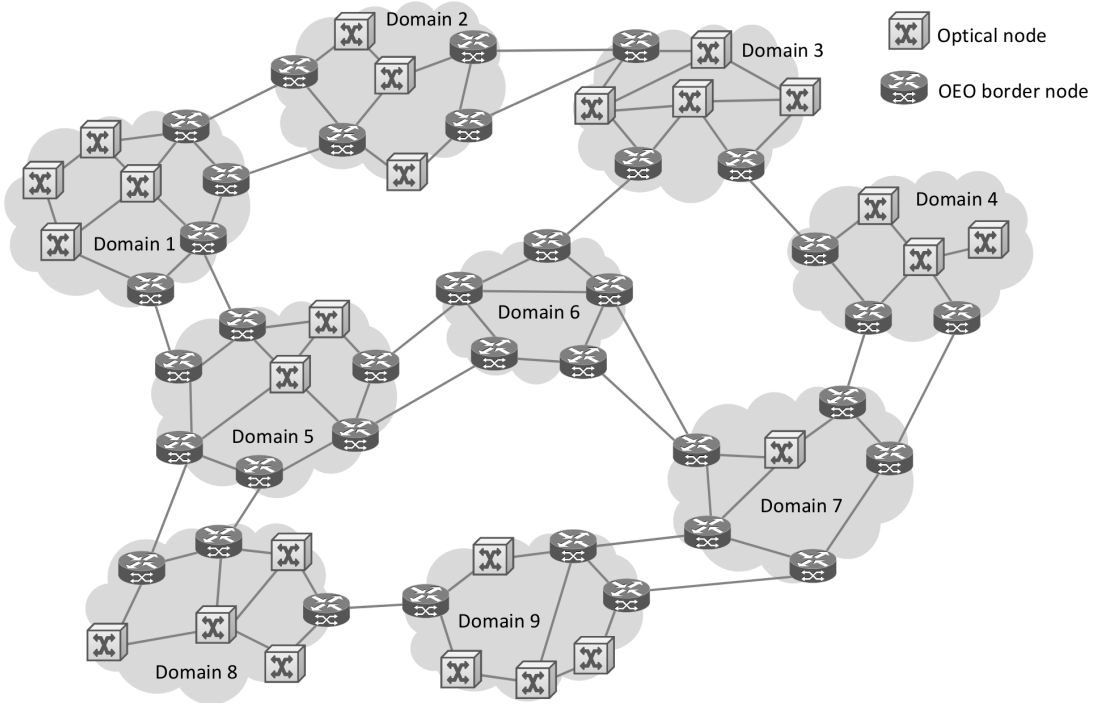
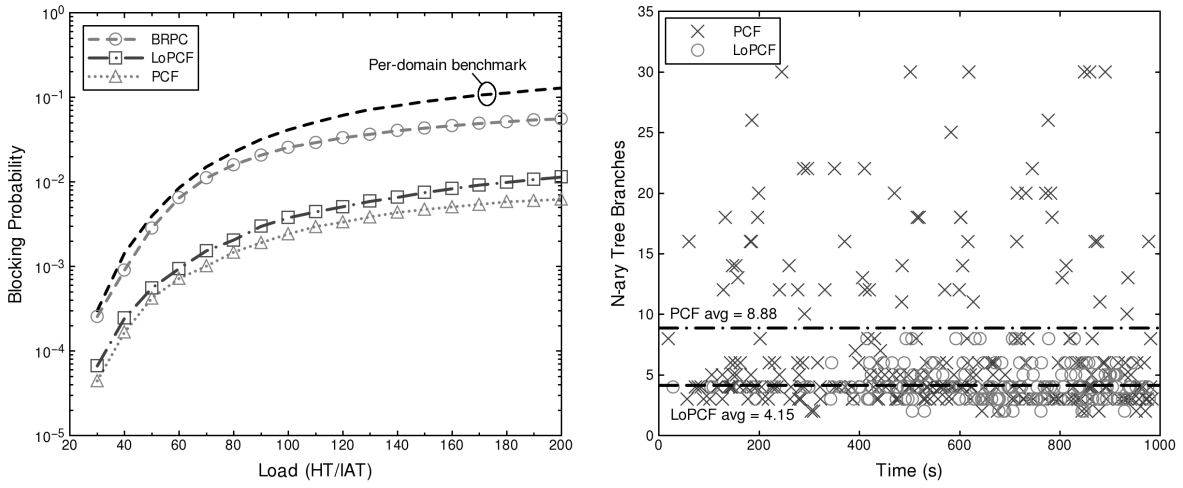


FIGURE 6.2: Nine-domain network composed of 61 nodes (including 19 inter-domain nodes) and 95 links.

Figure 6.3 (a) depicts the connection blocking probability achieved by standard BRPC, PCF and LoPCF. Besides, the per-domain approach is also shown as a benchmark. In

BRPC, the shortest domain sequence in terms of traversed domains is used. This leads to significantly increased blocking probability compared to PCF and LoPCF. For instance, fixing BP≈0.5%, the offered load to the network can be doubled if PCF or LoPCF is implemented (from 60 to 130-150). As seen, PCF yields the best BP figures, closely followed by LoPCF.



(a) (b)

FIGURE 6.3: Comparison of standard BRPC, PCF, LoPCF: (a) network blocking probability and (b) path computation complexity represented as the number of branches in the N-ary tree.

Figure 6.3 (b) compares the scalability of PCF and LoPCF in terms of the path computation complexity at source nodes. To this goal, for 150 Erlang (BP≈0.5%), the size of the N-ary tree to compute each requested inter-domain path was quantified during a 1000 seconds time interval. The high number of VSPTs collected by PCF even lead to 30 N-ary tree branches in some cases with very far off source/destination domains (averaging 8.88 in the whole time interval). In contrast, this number is reduced by 53% when LoPCF is applied, averaging 4.15 in the considered time interval. It shall be noted that LoPCF also results in much reduced control overhead in terms of PCEP messages/s to serve a certain set of intra- and inter-domain connections. For instance, for 150 Erlang, LoPCF reduces PCF control overhead by 51%, finally resulting in only 130% the overhead in BRPC.

The obtained low BP figures and scalable behavior allow to conclude on the appropriateness of LoPCF for PCE-based multi-domain networks. Moreover it would also fit the current PCE framework well. Looking at the PCEP standardization, no message format modification is required for LoPCF as the *Request Id* value can be carried in the Request-ID-number field of the *RP* object [RFC5440]. In fact, the most notable differences to standard BRPC and PCEP are found in the LoPCF behavior, although they should not entail critical implementation issues: firstly, the *PCReq* is sent directly to the destination instead of hop-by-hop through the domain sequence; secondly, the backward flooding (already introduced in PCF [PCE-Hierarc]) breaks the initial PCEP client/server model,

as *PCRep* messages are not sent in response of a *PCReq*; finally, the *Request Id* values must be globally unique, which is not an initial requisite in [RFC5441].

6.4 Enhanced PCE-based multi-domain end-to-end path computation

As multi-domain solutions catch the attention of both standardization entities and vendors, optimal end-to-end path computation in such a scenario arises as the main goal to be achieved. Here, route calculation agents have to deal with the limited topological information available due to the aforementioned confidentiality and scalability requirements.

In this context, the authors in [SM04] compare an algorithm with full network visibility (CombSPF), where all domains share full topological and link-state information, to a per-domain path computation approach (MSPF) over a WDM multi-domain optical network. In the latter, the BGP-based shortest *AS_PATH* [RFC4271] is used to obtain the domain sequence to be used. As expected, the CombSPF algorithm reduces the blocking probability in front the per-domain approach (MSPF). However, full topology dissemination is neither scalable nor matches the confidentiality requested by domain operators.

In order to enhance the inter-domain path computation without sharing complete internal domain information, different topology aggregation techniques have been also proposed (e.g., see [LKG+06] and [DLRRS08]). Such methods apply summarization mechanisms so that only certain information about the domain is shared with the other ones. For example, the authors in [LKG+06] evaluate two intra-domain aggregation methods, referred to as simple node abstraction, where each domain is represented as a single virtual node, and full mesh abstraction in which each domain is characterized by a group of virtual links that connect its border nodes. Then, aggregated topologies are shared amongst the different domains, finally resulting in a simplified global topology over which inter-domain routes are calculated. Even though topology aggregation mechanisms present better scalability for larger network scenarios, because the aggregated topologies usually comprise less information than the real ones, a considerable number of update messages need to be exchanged between domains to maintain the aggregated multi-domain topology updated. In order to alleviate the control overhead due to frequent dissemination processes, the authors in [CCJ+09] present a full mesh abstraction based on pre-reserved intra-domain resources for inter-domain connections. In this way, only those newly allocated or released inter-domain connections may trigger aggregated topology updates (i.e., modifications on intra-domain connections do not trigger any updates since they are not allowed to use resources assigned to inter-domain transit connections). From the obtained results, the authors conclude that this approach reduces the control overhead, but at expenses of a higher intra-domain connections blocking probability due to the lower resources available for their provisioning.

Rather than relying on route computations over aggregated topologies that may lead to sub-optimal solutions, other approaches focus on providing more adequate domain chains to the existing PCE-based mechanisms (such as BRPC) aiming to obtain improved end-to-end paths. In the proposed standard mechanisms (depicted in section 6.2), such domain sequences may be pre-configured, which results unfeasible as the multi-domain scenario is expanded, or obtained using inter-domain routing mechanisms. Existing routing mechanisms in IP networks, such as BGP, become unsuitable for QoS routing since no traffic engineering information is disseminated. Moreover, the domain sequence selection criteria used by the BGP, which is mainly supported on the shortest *AS_PATH* (user-defined policies can also be used), do not guarantee the optimal end-to-end route. To solve this, some TE extensions to BGP are proposed in [MRBD09]. However, the scalability of this approach may be questioned in dynamic scenarios where frequent changes occur due to the high number of lightpath setups and tear-downs. Although being a work in progress, the IETF is dedicating some extra efforts in this regard lately, as shown in [BGP-TE]. A more sophisticated domain sequence calculation is done in [WCZZ09]. There, aggregated topological information is used to compute a suitable domain chain which is then added to the BRPC procedure. Nevertheless, since a topology abstraction is maintained, this mechanism leads to poor scalability as well.

In line with the trends towards achieving better domain chains, next subsection describes a domain sequence selection procedure that, once added to the standard BRPC mechanism, makes a good trade-off between the introduced control overhead and the accuracy of the computed end-to-end route.

6.4.1 Disjoint domain sequences for enhanced end-to-end multi-domain path computation

As said, this subsection presents an alternative route computation mechanism named *k-Backward Recursive PCE-based Path Computation* (k-BRPC). This mechanism enhances the standard BRPC letting it explore multiple domain-disjoint sequences from the destination to the source domain and, consequently, obtain close to optimal end-to-end inter-domain paths using the collected routing information.

In order to introduce k-BRPC, the multi-domain network is modeled as a graph $G = (V, E)$. This global graph joins D sub-graphs corresponding to D independent domains. In particular, the domain G^i is defined as $G^i = (V^i, E^i)$, where V^i and E^i represent the sets of intra-domain nodes and links in G^i. Note that $V^i = \{v^i_1, ..., v^i_{\gamma^i}\}$, being γ^i the total number of nodes in that domain and $V^i = \{V^i \subset V : V^i \cap V^j = \emptyset, \forall i \neq j\}$, where $1 \leq i, j \leq D$. In turn, $E^i = \{e(w,k) \in E : v^i_w, v^i_k \in V^i; w \neq k\}$. Specifically, $e(w,k)$ describes a bidirectional intra-domain link connecting v^i_w and v^i_k. Inter-domain links are also bidirectional and defined as $\delta^{ij}_{mn} = \{e \in E, e(m,n) : v^i_m \in V^i, v^j_n \notin V^i \mid v^i_m \notin V^j, v^j_n \in$

$V^j; m \neq n; i \neq j\}$, where $1 \leq m \leq \gamma^i$ and $1 \leq n \leq \gamma^j$. Each domain G^i has a PCE^i responsible for the computation of paths inside it.

The standard BRPC calculates end-to-end paths assuming a single pre-configured sequence of domains. This may lead to sub-optimal routes if the selected domain sequence is not the most appropriate one according to the current network state. In light of this, k-BRPC enhances the end-to-end route calculation by means of the computation of k-shortest domain-disjoint sequences from the destination domain to the source one. In this way, the PCE in the source domain can obtain an extended set of domain-disjoint VSPTs and, then, decide the most appropriate path between a source node v_s^i and a destination node v_d^j. Like the standard BRPC protocol, the operation of k-BRPC is supported on the PCEP protocol, which allows the communication between a PCC and a PCE, or between different PCE peers. These interactions are implemented by means of *PCReq* and *PCRep* messages.

Algorithm 6.2 depicts the path computation procedure in k-BRPC. When an inter-domain connection between v_s^i and v_d^j is requested, v_s^i firstly sends a *PCReq* to its responsible PCE (PCE^i). After receiving this message, PCE^i forwards it directly to the destination domain PCE (PCE^j). As soon as the message reaches PCE^j, this one calculates k-shortest domain-disjoint sequences to the source domain using a multi-domain connectivity graph of the whole network. Figure 6.4 (right) shows an example of this simple graph, which can be manually created during the network bootstrap phase. In this graph, each vertex represents a domain of the network and each edge characterizes the adjacency between domains (i.e., two vertices are connected if the domains they represent have at least one inter-domain link). Note that this aggregation method does not provide information about the internal domain resources. Further details on the generation of this multi-domain connectivity graph are discussed in section 6.4.2.

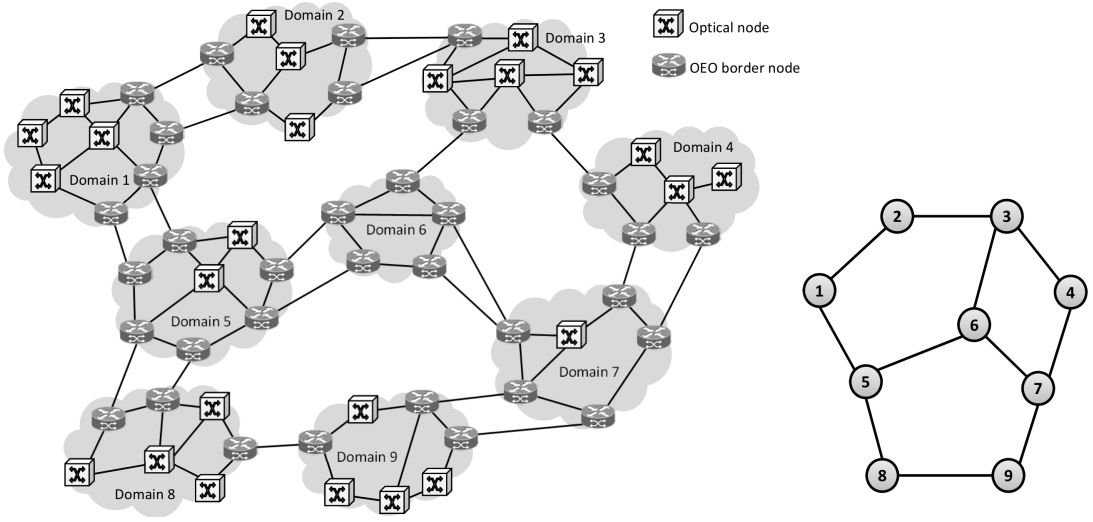


FIGURE 6.4: Nine-domain test network topology (Appendix A.4) and its resulting nine-domain connectivity graph (right).

```
begin
    if PCReq then
        if (PCReq → SourceNode ∈ G^u) then
            if (PCReq → DestinationNode ∈ G^u) then
                IntraDomainComputation (PCReq → DestinationNode)
            else
                DestPCE = findPCE (PCReq → DestinationNode)
                send (PCReq, DestPCE)
        else if (PCReq → DestinationNode ∈ G^u) then
            SourceDomain = findDomain (PCReq → SourceNode)
            KshortestRoutes (SourceDomain)
            foreach upstream domain G^k do
                PCRep = CreatePCRep ()
                FillDomainList (KshortestRoute^k)
                UpstreamDomain = findUpstreamDomain (KshortestRoute^k)
                PCRep → VSPT = ComputeVSPT(UpstreamDomain)
                UpstreamPCE = findPCE (UpstreamDomain)
                Send (PCRep, UpstreamPCE)
        else
            DiscardPCEPMsg ()
    else if PCRep then
        if (PCRep → SourceNode ∈ G^u) then
            PCRep → VSPT = ComputeVSPT (SourceDomain)
            AddVSPTInformationToNaryTree ()
        else
            NextDomain = CheckDomainList (PCRep)
            PCRep → VSPT = ComputeVSPT (NextDomain)
            NextPCE = findPCE (NextDomain)
            Send (PCRep, NextPCE)
    else
        DiscardPCEPMsg ()
```

ALGORITHM 6.2: k-BRPC procedure for PCE^u.

In this work, k takes the same value for all PCEs in the network. Nonetheless, it is worth mentioning here that the number of attainable domain-disjoint sequences is tightly related to the network topology and, in certain situations, there is no chance to obtain k domain-disjoint chains. Therefore, the k becomes an upper limit of the number of sequences to be obtained in each computation.

In k-BRPC, the k domain-disjoint sequences are treated as k separate pre-configured domain sequences. Hence, a *PCRep* message is sent over each one of these computed sequences. To do this, a domain list is included in each *PCRep* message with the information of the domain sequence to be explored. Moreover, PCE^j also fills the *PCRep* message with the VSPT to the upstream domain indicated by such sequence (i.e., the path segments representing the shortest route between v_d^j and the border nodes connected to the upstream domain). As will be detailed in section 6.4.2, each path segment in the VSPT is assigned a cost metric and is protected using a path-key (instead of sending the

explicit route), thus ensuring confidentiality. Any intermediate PCE receiving a *PCRep* message updates the included VSPT with its local domain route information. Right after, it forwards the message to the responsible upstream PCE according to the received domain list. This operation is repeated in every intermediate PCE until PCE^i is reached. As soon as PCE^i receives the k VSPTs, it builds an N-ary tree using all the gathered information. This N-ary tree is rooted at v_s^i and all branches have the same leaf which is v_d^j. Every branch is created using a concatenation of the accumulated path segments composing the received VSPTs. After processing the N-ary tree, PCE^i is able to respond the v_s^i request with the optimal end-to-end path to v_d^j constrained to any of the explored k domain-disjoint sequences.

6.4.2 k-BRPC mechanism in a GMPLS-controlled PCE-based control plane

This section describes the modifications that need to be applied to the standard PCE-based control plane in order to deploy the previously introduced k-BRPC mechanism. To this end, the architecture proposed by the IETF in [RFC4655] and the PCEP protocol defined in [RFC5440] are taken as starting points. In such standards, the IETF defines a client/server model for the interactions with the PCE module where every sent message has a reply. However, k-BRPC breaks this model in two ways. First, given that the *PCReq* message is sent from the source domain directly to the destination one instead of following the domain-by-domain basis proposed in [RFC5441], the backward messaging process from the destination domain to the source one does not respond to that client/server model. Second, since the *PCReq* message sent by the source PCE is translated into the reception of up to k *PCRep* messages, it is clear that the model is also broken.

A manual configuration of the multi-domain connectivity graph is assumed in this work. Nevertheless, since the information needed by the PCEs to build the graph is related to the reachability between domains, it could be provided by the standard version of the BGP protocol. This, along with the PCE discovery mechanisms proposed in [RFC4674], would allow a dynamic scenario where domains could be added and removed during network operation.

As detailed in 6.4.1, k-BRPC relies on a set of domain sequences calculated in the destination PCE and added to the *PCRep* messages to compute the path. In order to add a domain sequence to each *PCRep* the same strategy as in standard BRPC is followed. This strategy consists of including the standard optional *Include Route Object* (*IRO*, [RFC5440]) in the message to convey the calculated domain sequence. Different from BRPC, where the pre-calculated chain is first added (by means of the *IRO* object) to the *PCReq* message and then also conveyed in the *PCReq*; in k-BRPC the domain sequence is only included in *PCRep* messages since it is calculated by the PCE at the destination domain.

Confidentiality between domains has been presented in this thesis as an important issue to be taken into account. Hence, such a feature has to be kept in mind when thinking about deploying the proposed mechanism in a real scenario. In this regard, the k-BRPC strategy can take benefit from the mechanism proposed in [RFC5520] where the path segments calculated by each domain are substituted by a path-key, thus preserving the topological confidentiality between domains. In this mechanism, instead of containing a list of *ERO* subobjects composing the explicit route, the *ERO* object in the *PCRep* carries the ingress node of the path segment traversing the domain, a *Path-Key Subobject* (*PKS*) encoding the explicit route, and the egress node of that segment. Likewise, during the signaling process, the *ERO* object transported within the *RSVP-TE PATH* message contains the *PKS* subobjects as well. Once the *PATH* reaches the ingress node of the domain, the node requests to its responsible PCE to decode the path-key (by means of a *PCReq* message). The PCE, in turn, returns the explicit route to the node using a *PCRep*. The explicit route is then added to the *ERO* object in the *PATH* message, and the signaling process goes on. This process is repeated for each key to be decoded.

Finally, it is worth mentioning here that, even though a scenario where each domain has electrical terminations in its border routers (thus allowing wavelength conversion) is studied, an all-optical multi-domain network could be also assumed. In this case, the wavelength continuity constraint would have to be preserved throughout the end-to-end path. To this aim, the extensions proposed in [CMMG09] for the standard BRPC in a transparent multi-domain optical network would be applicable to the k-BRPC approach as well.

6.4.3 Performance evaluation

The performance of k-BRPC was compared to standard BRPC and PCF by means of simulation. The scenario used for this purpose is depicted in Figure 6.4 (left). The network topology is composed of 61 nodes and 95 links (19 inter-domain) carrying each one 8 bidirectional wavelengths per link. In this multi-carrier topology, border nodes provide optical-electrical-optical conversion while intra-domain nodes are all-optical (without O/E/O conversion). In all simulations, 10^5 connection requests were generated following a 70/30% intra/inter-domain ratio [LKG+06]. Source and destination nodes were randomly selected for intra-domain connections and all requests demanded a whole wavelength capacity. For inter-domain connections, source and destination domains were uniformly selected and the respective source and destination nodes were uniformly chosen. Mean holding time was set to 600 seconds and request inter-arrival time varied with the network offered load according to load=HT/IAT. Both times were exponentially distributed. Figure 6.4 (right) describes the multi-domain connectivity graph over which the set of k-shortest domain-disjoint sequences was computed.

Figure 6.5 depicts the connection blocking probability achieved by standard BRPC, PCF, and k-BRPC with k=1, 2, 3. Here, the BRPC implementation assumes as the pre-defined

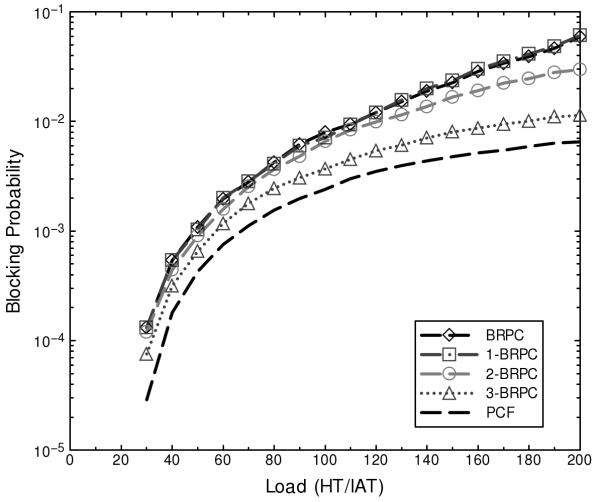


FIGURE 6.5: Connection blocking probability.

domain sequence the shortest one in terms of domain hops (as if it was provided by BGP). As explained in section 6.4.1, the k in k-BRPC represents the maximum number of domain sequences that can be explored between destination and source domains. Likewise, it represents the upper bound of the number of VSPTs provided to the source PCE over which the inter-domain end-to-end route is computed. Therefore, the best k value is closely related to the physical topology of the multi-domain network. Specifically, the number of adjacent domains restricts the number of explorable domain-disjoint sequences and, at the same time, it limits the number of possible end-to-end routes. Extensive simulations (not depicted in the figures) show that values of k higher than 3 do not decrease the BP in the simulated network topology. Moreover, as will be illustrated along this section, an upper bound of k=3 makes the best trade-off between blocking probability, introduced overhead and path computation complexity.

Looking at Figure 6.5, the mechanisms leading to the best and worst performance in terms of blocking are PCF and standard BRPC, respectively. As explained before, the flooding applied in PCF allows exploring any domain sequence from the destination to the source node. In this way, the source PCE is provided with complete information to select the optimal end-to-end inter-domain route in every case. However, as will be quantified later in this section, this operation becomes unscalable as the network size increases. On the contrary, a single fixed domain sequence is considered in standard BRPC. This leads to sub-optimal end-to-end routes in numerous cases, given that the provided shortest domain sequence does not assure the optimal end-to-end path. Furthermore, fixed domain chains lead to an inflexible operation since the network status is not taken into account in path computation, thus increasing BP. Finally, k-BRPC approaches lay between standard BRPC and PFC. It is clear that k=1 results in the same performance as standard BRPC, since only the shortest domain sequence is explored. Nonetheless, more information is gathered for the final inter-domain path computation as the value of k is increased. Hence, having a more accurate view of the current network status improves the route selection

and, in consequence, the blocking probability is reduced. Moreover, keeping the number of explored domain sequences bounded makes from k-BRPC a scalable solution.

The cost of obtaining optimal shortest end-to-end paths by means of PCF is twofold. First, the PCF mechanism is associated to a huge control overhead in terms of PCEP messages to be disseminated among PCEs. Second, the size of the N-ary tree resulting from the topological information collected during the flooding of PCEP messages increases the computational cost in the source PCE. To evaluate both statements Figure 6.6 (a) and (b) show the average number of PCEP messages per path computation request, and the number of branches of the resulting N-ary tree. In both figures, the standard mechanisms (i.e. BRPC and PCF) are taken as benchmarks and compared to the results obtained for k=1, 2, 3. These results have been obtained fixing the load at 200 Erlang (which implies a BP≈1%). The goal of this study is to measure the feasibility and scalability of the proposed mechanism in front of the standards.

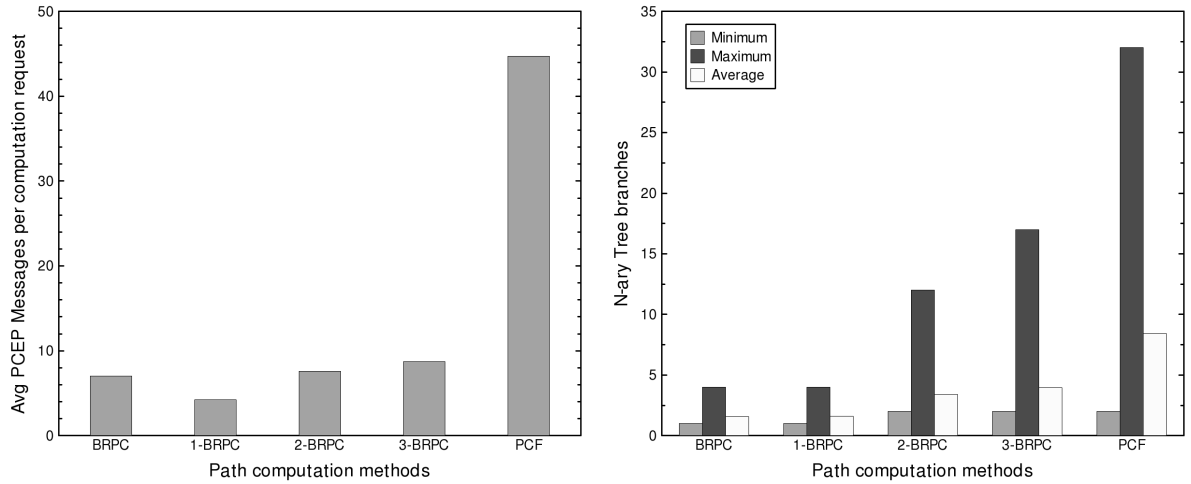


FIGURE 6.6: (a) Average number of PCEP messages per inter-domain path computation request; (b) path computation cost.

Figure 6.6 (a) shows that 1-BRPC halves the average number of PCEP messages per path computation used by standard BRPC. This is due to the fact that in k-BRPC the *PCReq* message is sent directly to the destination domain and not following a per-domain basis like in BRPC. As explained in section 6.4.2, this improvement is achieved at expenses of breaking the client/server model associated to the PCE architecture. As could be expected, 2-BRPC doubles the number of messages in front of 1-BRPC. Interestingly, for k=3 the number of messages barely exceeds the one obtained for k=2. This is given by the fact that in some cases only two disjoint domain sequences can be provided. As shown in the figure, PCF increases the number of PCEP messages dramatically (about a 500% versus 3-BRPC), thus making it unfeasible in large network scenarios.

Finally, the computational cost is plotted in Figure 6.6 (b). There, the size of the obtained N-ary tree that is processed in the source PCE is taken as a figure of merit. The

figure shows that standard BRPC and 1-BRPC, which explore the shortest domain sequence, result into the same computational cost averaging 1.7 branches per N-ary tree (i.e. per path computation). Unsurprisingly, 2-BRPC doubles this result by averaging 3.42 branches. Similar to the number of PCEP messages, 3-BRPC slightly increments the number of branches per tree in front of 2-BRPC (a 14%). PCF, however, shows again its weakest point yielding to the highest computational cost (8.41 branches per N-ary tree in average) which doubles the figures obtained in 3-BRPC.

The obtained results prove that k-BRPC outperforms the standard BRPC in terms of blocking probability, being close to the optimal values obtained by means of PCF with an appropriate value for the k. Moreover, the proposed strategy avoids the scalability problems associated to PCF keeping a reduced control overhead as well as an affordable computational cost. This, along with the low impact from the standardization point of view, detailed in section 6.4.2, makes from k-BRPC a valid strategy to be implemented in a GMPLS-enabled PCE-based control plane.

6.5 Chapter summary

This chapter has studied the performance and scalability of end-to-end path computation mechanisms in a PCE-based multi-domain scenario. To this end, section 1 has been devoted to the description of the architecture under study. Next, standard path computation mechanisms have been outlined in section 2.

Section 3 has presented an improved version of the PCF mechanism described by the IETF. LoPCF aims at reducing the path computation complexity introduced by PCF, thus increasing scalability, but also keeping PCF's optimal blocking probability figures. Taking as a measure of computation complexity the size of the N-ary tree containing the possible end-to-end routes between two given nodes, simulation results show a 50% reduction in the number of branches of the tree when using LoPCF. Furthermore, although yielding to sub-optimal end-to-end routes, LoPCF outperforms standard BRPC and per-domain schemes in terms of blocking probability.

Section 4 has studied a well-known lack associated to the BRPC mechanism, which is the domain sequence selection. In this context, a BRPC-based mechanism named k-BRPC has been presented and evaluated. This mechanism relies on a domain connectivity graph to dynamically compute a set of domain sequences aiming to provide close to optimal end-to-end paths in a scalable way. Results demonstrate that this mechanism provides a good trade-off between blocking probability, computational complexity and network control overhead. Moreover, slight changes are necessary in PCEP to implement the proposed mechanism. All this makes from k-BRPC an appropriate solution to be implemented in a PCE-based GMPLS-controlled multi-domain optical transport network.

Chapter 7

Conclusions and future work

Optical transport networks appear as the most promising solution to cope with current data traffic demands. While OPS and OBS models still present severe technological limitations, OCS networks rely on a mature transport technology, which can be managed by an intelligent control plane. Specifically, the standardized ASON/GMPLS architecture represents the most accepted model to be implemented in the short to mid term. The ASON/GMPLS architecture targets at improved transmission rates and more efficient operation than traditional IP/ATM/SDH/WDM networks, while still providing QoS and resilience. On this basis, the present thesis identifies and addresses several open issues in ASON/GMPLS networks from the control plane perspective. In brief, the main goals of this work are providing integrated multi-layer management, control plane support of physical layer impairments, and multi-domain end-to-end path computation. Most of these contributions are experimentally demonstrated.

To motivate the focus of this thesis, the evolution of optical transport networks was surveyed in chapter 2. This overview ranged from initial evolutionary stages, where only transmission is done in the optical domain, to the future all-optical networks. Special attention was paid to dynamic wavelength-routed optical transport networks and its main architectural proposal: ASON/GMPLS. The most important open issues presented by such an architecture, to be addressed in this thesis, were identified in this chapter as well. First, the efficiency problem due to the coarse bandwidth granularity offered by wavelength-routed optical transport networks was described. Second, the technological problems faced by all-optical data transmission and their impact on the connection provisioning process were introduced. Finally, the path computation problem in a multi-domain scenario was identified as the third challenge to be tackled.

Since most of the studies carried out in this thesis are experimental, chapter 3 was devoted to the description of the ASON/GMPLS CARISMA network, which is the test bed were the experimental contributions were implemented and evaluated. Besides, the CARISMA test bed was used as a driver to realize a thorough description of the ASON model and, specially, the GMPLS control plane architecture and protocols.

Chapter 4 concentrated on the efficient utilization of network resources in ASON/GMPLS. In order to overcome the bandwidth waste associated to wavelength-routed optical transport networks, the traffic grooming concept, which consists of aggregating high level data flows into the same wavelength for optical transmission, was presented and its implementation in GMPLS by means of the FA entity was detailed. In this context, the experimental development and validation of an integrated GMPLS-based multi-layer control plane was conducted. To this aim, the necessary GMPLS protocol and functional extensions were detailed. The experimental results exhibited dramatic benefits of traffic grooming in front of pure all-optical solutions in terms of blocking probability. Next in this chapter, a TE mechanism, based on an FA creation cost function, was presented and experimentally evaluated in front of the standard FA approach. An opaque and an all-optical scenarios were taken as benchmarks as well. The experimental evaluation demonstrated that the proposed mechanism achieves a network usage very close to the optimal, which is provided by opaque networks, but also providing an interesting reduction of the client O/E port usage of up to 20% in the scenarios under study. In conclusion, the multi-layer paradigm arises as the most valid mechanism to implement bandwidth-efficient and cost-effective optical transport networks.

Chapter 5 focused on the impairment-awareness in ASON/GMPLS networks. There, a centralized architectural solution was firstly presented to implement an impairment-aware GMPLS-based control plane capable of providing end-to-end lightpaths with guaranteed QoT. This scheme was compared to a distributed approach. The results of the experimental evaluation of both mechanisms revealed that the centralized approach obtains better blocking probability than the distributed one, thanks to the availability of full topological and PLI information for path computation. However, the costs of collecting and processing all that information leads to increased setup delays. Hence, the results conclude that the centralized scheme is more suitable for low loaded scenarios where request inter-arrival times are higher than connection setup times. Identifying the QoT estimation as the bottleneck in centralized IA-RWA, since it is a computationally intensive task, an FPGA-accelerated Q-Tool was deployed in the centralized NPOT. Therefore, reducing the QoT estimation times, the overall lightpath establishment time was automatically accelerated. To endorse this, the resulting figures exposed an improvement of the new centralized model in front of the former centralized approach in terms of setup time, thanks to the hardware-accelerated Q-Tool, getting very close to the setup delays obtained by the distributed scheme. Moreover, the new model keeps the good blocking probability numbers of the former centralized approach since it uses full topological and PLI information. Later on, being the resilience a key point in ASON/GMPLS networks, a centralized impairment-aware restoration scheme was studied in this chapter. Targeting at reduced restoration times, an enhanced version of the centralized control plane, provided with differentiated resilience support and a resource pre-reservation protocol, was implemented and experimentally tested. The obtained restoration times of 1.16 and 1.64 seconds for high and low priority traffic, respectively, make from this scheme a totally valid implementation for real network environments. Finally, a standardized version of

the centralized scheme, based on the PCE architecture was proposed. The functional and protocol requirements for such standardization were investigated.

Chapter 6 dealt with end-to-end path computation in GMPLS-controlled multi-domain optical networks. In particular, a PCE-based architecture was put under study. First, the scalability of the path computation schemes proposed by the IETF was studied and compared to the introduced LoPCF mechanism. As demonstrated by the obtained results, the proposed LoPCF and the optimal PCF are very close in terms of blocking probability. Nonetheless, LoPCF achieves a 50% reduction in computational complexity, which makes from this scheme a feasible solution to be implemented in the scenario under study. Similarly, the second contribution of this chapter was intended to improve the BRPC mechanism designed by the IETF. This improvement relied on the exploration of different domain sequences in order to obtain better end-to-end routes but, at the same time, keeping the overhead and, thus, the scalability under control. The k-BRPC scheme was proposed to this end and compared to the standard approaches, showing a better trade-off between blocking probability, control overhead and computational complexity than the others. This, in addition to the low impact of this novel mechanism on standardization, demonstrate that k-BRPC could be perfectly implemented in a real GMPLS-controlled multi-domain optical network.

Three major research lines result from this thesis and may be followed in future work. First, the contributions presented in chapter 6 concerning GMPLS-based multi-domain end-to-end path computation could be implemented and validated in the CARISMA network, as done with the multi-layer and impairment-aware architectures. As reviewed in chapter 6, some functional and architectural extensions should be necessary in the control plane, particularly affecting to the PCE and PCEP. In this subject, the emerging hierarchical PCE models [PCE-Hierarc, H-PCEP] could be explored as well. The second and probably the most ambitious goal for future work would be the integration of the three main elements addressed throughout this thesis, namely multi-layer, impairment-awareness and multi-domain, in a unified GMPLS-enabled control plane for ASON. Some intermediate integration stages should be defined in this process due to its complexity. For example, a multi-domain impairment-aware GMPLS-controlled optical transport network might be the first milestone in such integration. Nonetheless, it is clear that this would be a long term project. Finally, the last future research line would be focused on improving the network energy efficiency, which has been a growing concern in the last years. Indeed, ICT consumes about an 8% of the current total electricity over the world, and this amount may double over the next five years [ZCTM10]. Hence, power requirement will be a major constraint of next generation networks. In light of this, the control plane has to evolve in tune with the transport technology to provide energy efficient services.

Appendix A

Reference transport network scenarios

This appendix introduces the scenarios that have been used for the experimental evaluation of the mechanisms proposed throughout this thesis. For convenience, cross-references are also provided to easily obtain the sections and subsections where each network has been used.

A.1 9-Node German transport network

To validate integrated routing algorithms in multi-layer optical networks, the authors in [NGB03] defined a fictitious 9-Node topology that deployed 13 links interconnecting some of the major German cities. Departing from this topology, two links were removed to obtain a scenario with 9 nodes interconnected by 11 links, resulting in an average ND of 2.44 (Figure A.1). This topology has been used in multi-layer experimental results presented in subsections 4.2.4, 4.3.1.1 and 4.3.2.1.

A.2 16-Node EON core topology

This network scenario is the result of modifying the number of nodes of the 28-Node EON basic topology. Being one of the main achievements of the COST 266 and Lion European Projects [IKM+03, MCL+03], the 28-Node EON has been quite a *de facto* choice in last years optical networking scientific literature. The resulting topology (Figure A.2), which only comprises the core part of the 28-Node EON basic network, contains 16 nodes interconnected by 23 links, leading to an average ND of 2.88. This topology has been used to experimentally evaluate the traffic grooming capabilities in subsection 4.3.2.

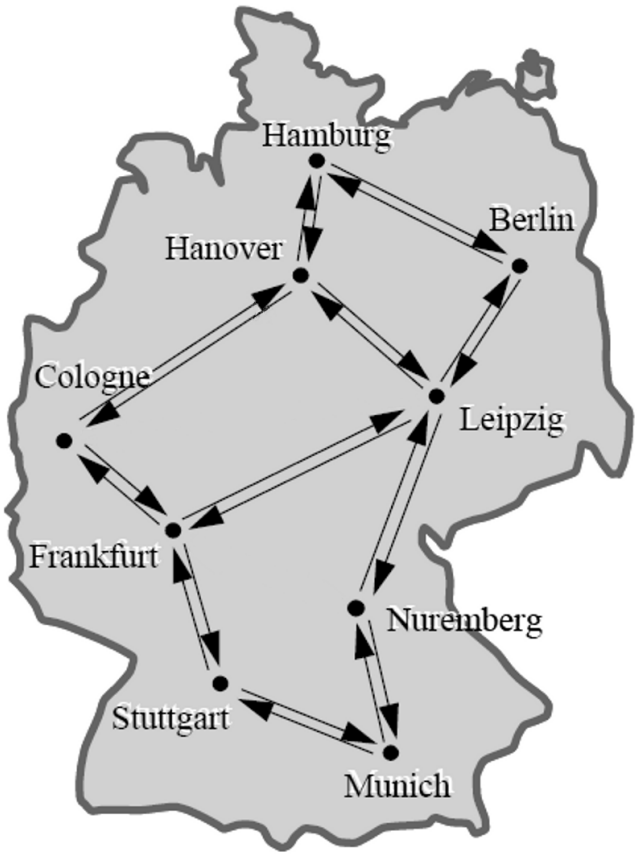


FIGURE A.1: 9-Node fictitious German network

FIGURE A.2: 16-Node EON core topology

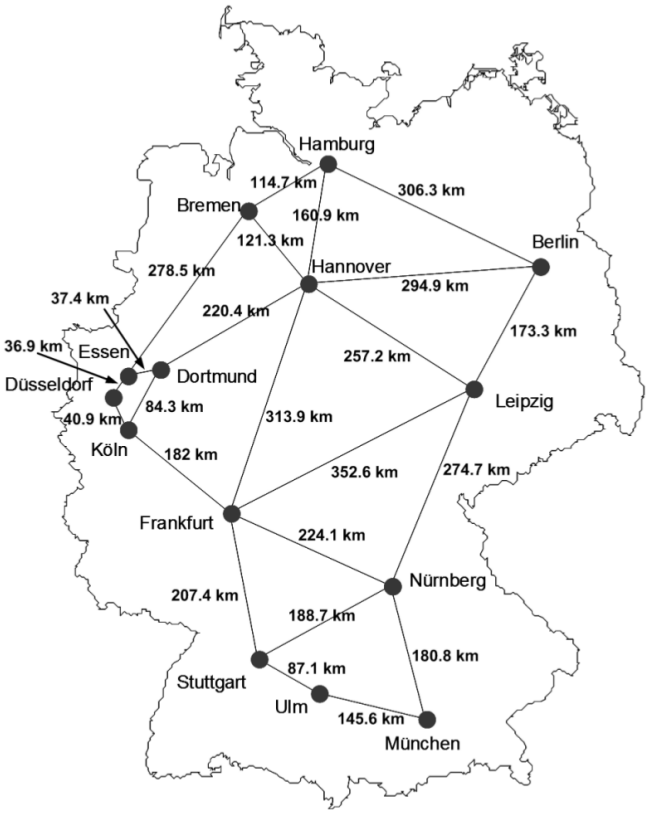


FIGURE A.3: 14-Node Deutsche Telekom topology

A.3 14-Node Deutsche Telekom (DTAG) network

The 14-Node Deutsche Telekom network was proposed as one of the reference networks to be used within the DICONET project. Provided by Deustsche Telekom AG itself, this network is composed of 14 nodes and 23 links deployed as shown in Figure A.3, resulting in an average ND of 3.29.

A.4 9-Domain multi-domain test network topology

Extracted from [LKG+06] and used in different multi-domain publications, this network provides good inter-domain connectivity, averaging 4.22 inter-domain links per domain. This network topology has been used in the studies conducted in chapter 6 and its topology is depicted in Figure 6.2 (page 106).

Appendix B

Physical layer parameters

This annex contains the specification of the physical layer information that is used during the operation of the different impairment-aware control plane architectures illustrated in chapter 5. In the distributed approach described in section 5.4, the information is conveyed by means of the extended RSVP-TE protocol (section 5.4.2) and used in the Q-check process. Furthermore, this information is flooded by means of the extended version of the OSPF-TE protocol (section 5.5.2), and used during the operation of the centralized impairment-aware control plane architecture depicted in section 5.5, and its PCE-based version illustrated in section 5.8.

Component	**Parameter**	**Wavelength related**
Transmitter	Bitrate	Yes
	Power	Yes
	Center wavelength	Yes
	Ref wavelength	No
	Extinction ratio	Yes
	Modulation format	Yes
	Type of FEC	Yes

TABLE B.1: Parameters of the Transmitter component.

Component	**Parameter**	**Wavelength related**
Fiber	Type	No
	Length	No
	Dispersion parameter	No
	Dispersion slope	No
	Attenuation	No
	Non-linear parameter	No
	Effective core area	No
	PMD	No
	Insertion Loss	No

Table B.2: Parameters of the Fiber component.

Component	**Parameter**	**Wavelength related**
Attenuator	Attenuation	Yes

Table B.3: Parameters of the Attenuator component.

Component	**Parameter**	**Wavelength related**
Receiver	Responsitivity	Yes
	Absolute threshold	Yes
	Thermal noise density	No
	Power	Yes
	PMD	No
	Residual CD	No
	Q value	Yes
	BER	Yes
	Insertion Loss	No

Table B.4: Parameters of the Receiver component.

Component	**Parameter**	**Wavelength related**
Amplifier	Spontaneous emission	No
	Gain	No
	Insertion Loss	No
	Input power	No
	Output power	No

Table B.5: Parameters of the Amplifier component.

Component	**Parameter**	**Wavelength related**
Filter	Type	No
	Order	No
	NEB factor	No
	Bandwidth	Yes
	Wavelength	Yes

Table B.6: Parameters of the Filter component.

Component	**Parameter**	**Wavelength related**
Physical node	Type	No
	Order	No
	NEB factor	No
	Bandwidth	Yes
	Wavelength	Yes

Table B.7: Parameters of the Physical node component.

Appendix C

Publication List

C.1 Publications in Journals

1. S. Azodolmolky, J. Perelló, M. Angelou, **F. Agraz**, L. Velasco, S. Spadaro, Y. Pointurier, A. Francescon, C.V. Saradhi, P. Kokkinos, E. Varvarigos, S. Al Zahr, M. Gagnaire, M. Gunkel, D. Klonidis and I. Tomkos. Experimental Demonstration of an Impairment Aware Network Planning and Operation Tool for Transparent/ Translucent Optical Networks. *Journal of Lightwave Technology*, 29(4): 439-448, February 2011.

2. G. Hernández-Sola, J. Perelló, **F. Agraz**, L. Velasco, S. Spadaro and G. Junyent. Enhanced domain disjoint backward recursive TE path computation for PCE-based multi-domain networks. *Photonic Network Communications*, 21(2): 141-151, April 2011.

3. **F. Agraz**, L. Velasco, J. Perelló, M. Ruiz, S. Spadaro, G. Junyent and J. Comellas. Deployment and Validation of GMPLS-Controlled Multi-layer Integrated Routing over the ASON/GMPLS CARISMA Test-bed. *Journal of Networks*, 5(11): 1321-1327, November 2010.

4. J. Perelló, G. Hernández-Sola, **F. Agraz**, S. Spadaro, and J. Comellas. Scalable Path Computation Flooding Approach for PCE-Based Multi-domain Networks. *ETRI Journal*, 32(4): 622-625, August 2010.

5. L. Velasco, **F. Agraz**, R. Martínez, R. Casellas, S. Spadaro, R. Muñoz and G. Junyent. GMPLS-based Multidomain Restoration: Analysis, Strategies, Policies and Experimental Assessment. *IEEE/OSA Journal of Optical Communications and Networking*, 2(7): 427-441, July 2010.

6. J. Perelló, **F. Agraz**, S. Spadaro, J. Comellas and G. Junyent. Using Updated Neighbor State Information for Efficient Contention Avoidance in OBS Networks. *Elsevier Computer Communications*, 33(1): 65-72, January 2010.

7. **F. Agraz**, L. Velasco, J. Perelló, M. Ruiz, S. Spadaro, G. Junyent, and J. Comellas. Design and Implementation of a GMPLS-Controlled Grooming-capable Optical Transport Network. *IEEE/OSA Journal of Optical Communications and Networking*, 1(2): A258-A269, July 2009.

C.2 Publications in Conferences

1. A. Morea, J. Perelló, **F. Agraz** and S. Spadaro. Demonstration of GMPLS-Controlled Device Power Management for Next Generation Green Optical Networks. In *Proceedings of the Optical Fiber Communication Conference (OFC 2012)*, Los Angeles (USA), March 2012.

2. A. Morea, S. Spadaro, O. Rival, J. Perelló, **F. Agraz** and D. Verchere. Power Management of Optoelectronic Interfaces for Dynamic Optical Networks. In *Proceedings of 37th European Conference on Optical Communications (ECOC 2011)*, Geneva (Switzerland), September 2011.

3. S. Spadaro, J. Perelló, **F. Agraz**, M. Angelou, S. Azodolmolky, Y. Qin, R. Nejabati, D. Simeonidou and I. Tomkos. Dynamic Impairment-Aware Optical Networking: Some Experimental Results of the EU DICONET Project. In *Proceedings of 13th International Conference on Transparent Optical Networks (ICTON 2011)*, Stockholm (Sweden), June 2011.

4. M. Angelou, **F. Agraz** , P. Kokkinos, J. Perelló, S. Azodolmolky, E. Varvarigos, S. Spadaro and I. Tomkos. Experimental Comparison of Impairment-Aware RWA Algorithms in a GMPLS-controlled Dynamic Optical Network. In *Proceedings of Future Network and Mobile Summit 2011*, Warsaw (Poland), June 2011.

5. S. Spadaro, J. Perelló, **F. Agraz**, S. Azodolmolky, M. Angelou, Y. Qin, R. Nejabati, D. Simeonidou, P. Kokkinos, E. Varvarigos, Y. Ye and I. Tomkos. Experimental Demonstration of an Enhanced Impairment-Aware Path Computation Element. In *Proceedings of the Optical Fiber Communication Conference (OFC 2011)*, Los Angeles (USA), March 2011.

6. J. Perelló, S. Spadaro, **F. Agraz**, M. Angelou, S. Azodolmolky, Y. Qin, R. Nejabati, D. Simeonidou, P. Kokkinos, E. Varvarigos, S. Al Zahr, M. Gagnaire and I. Tomkos. Experimental Evaluation of Centralized Failure Restoration in a Dynamic Impairment-Aware All-Optical Network. In *Proceedings of the Optical Fiber Communication Conference (OFC 2011)*, Los Angeles (USA), March 2011.

7. **F. Agraz**, J. Perelló, M. Angelou, S. Azodolmolky, L. Velasco, S. Spadaro, P. Kokkinos, E. Varvarigos and I. Tomkos. Experimental Evaluation of Path Restoration for a Centralised Impairment-Aware GMPLS-Controlled All-Optical Network.

In *Proceedings of 36th European Conference on Optical Communications (ECOC 2010)*, Torino (Italy), September 2010.

8. **F. Agraz**, S. Azodolmolky, M. Angelou, J. Perelló, L. Velasco, S. Spadaro, A. Francescon, C.V. Saradhi, Y. Pointurier, P. Kokkinos, E. Varvarigos, M. Gunkel and I. Tomkos. Experimental Demonstration of Centralized and Distributed Impairment-Aware Control Plane Schemes for Dynamic Transparent Optical Networks. In *Proceedings of the Optical Fiber Communication Conference (OFC 2010)*, San Diego (USA), March 2010.

9. R. Martinez, R. Casellas, L. Velasco, **F. Agraz**, R. Muñoz and S. Spadaro. Experimental Validation of End-to-End GMPLS-Enabled Restoration in Multi-Domain Transparent WSON. In *Proceedings of the Optical Fiber Communication Conference (OFC 2010)*, San Diego (USA), March 2010.

10. J. Perelló, L. Velasco, **F. Agraz**, S. Spadaro, J. Comellas and G. Junyent. Demonstration of GMPLS-controlled integrated IP/WDM routing over a grooming-capable ASON/GMPLS network test-bed. In *Proceedings of 35th European Conference on Optical Communication (ECOC 2009)*, Vienna (Austria), September 2009.

11. S. Spadaro, L. Velasco, J. Perelló, **F. Agraz**, J. Comellas and G. Junyent. Some open issues in multi-domain/multi-operator/multi-granular ASON/GMPLS networks. In *Proceedings of 11th International Conference on Transparent Optical Networks (ICTON 2009)*, Island of Sao Miguel, Azores (Portugal), June 2009.

12. **F. Agraz**, L. Velasco, J. Perelló, M. Ruiz, S. Spadaro, G. Junyent and J. Comellas. An experimental GMPLS-controlled network test-bed enabling sub-wavelength connection provisioning. In *Proceedings of 11th International Conference on Transparent Optical Networks (ICTON 2009)*, Island of Sao Miguel, Azores (Portugal), June 2009.

13. J. Perelló, L. Velasco, **F. Agraz**, S. Spadaro, G. Junyent, and J. Comellas. A comparison of in-fiber and out-of-fiber GMPLS-based control plane configurations: Benefits, drawbacks and solutions. In *Proceedings of 10th International Conference on Transparent Optical Networks (ICTON 2008)*, Athens (Greece), June 2008.

14. L. Velasco, R. Romeral, **F. Agraz**, S. Spadaro, J. Comellas, G. Junyent and D. Larrabeiti. On the design of MPLS-ASON/GMPLS Interconnection Mechanisms. In *Proceedings of 7th Workshop in G/MPLS networks, co-located with 12th International Conference on Optical Networking Design and Modeling (ONDM 2008)*, Vilanova i la Geltrú (Spain), March 2008.

15. R. Muñoz, R. Martínez, F. Galán, R. Morro, H.M. Foisel, S. Szuppa, J. Jiménez, O. González, H. Dentler, E. Escalona, **F.Agraz**, S. Spadaro and B. Berde. Experimental interconnection and interworking of the multi-domain (ASON-GMPLS) and

multi-layer (TDM-LSC) NOBEL2 Test-beds. In *Proceedings of 33rd European Conference on Optical Communications (ECOC 2007)*, Berlin (Germany), September 2007.

16. J. Comellas, J. Perelló, **F. Agraz**, S. Spadaro, and G. Junyent. Grooming Strategies for GMPLS Controlled WDM Networks. In *Proceedings of 9th International Conference on Transparent Optical Networks (ICTON 2007)*, Rome (Italy), June 2007.

17. S. Spadaro, J. Perelló, E. Escalona, L. Velasco, **F. Agraz**, J. Comellas, and G. Junyent. The CARISMA ASON/GMPLS Network: Overview and Open Issues. In *Proceedings of 9th International Conference on Transparent Optical Networks (ICTON 2007)*, Rome (Italy), June 2007.

18. J. Perelló, **F. Agraz**, S. Spadaro, J. Comellas, and G. Junyent. Assessment of LMP-based Recovery Mechanisms for GMPLS Control Planes. In *Proceedings of Workshop on GMPLS Performance: Control Plane Resilience (co-located with IEEE ICC 2007)*, Glasgow (Scotland), June 2007.

19. J. Perelló, E. Escalona, S. Spadaro, **F. Agraz**, J. Comellas, and G. Junyent. Control Plane Protection Using Link Management Protocol (LMP) in the ASON/GMPLS CARISMA Network. In *Proceedings of 5th International Networking Conference (NETWORKING 2006)*, Coimbra (Portugal), May 2006.

20. J. Perelló, A. Sánchez, **F. Agraz**, E. Escalona, S. Spadaro, J. Comellas, and G. Junyent. Experimental Link Resource Manager implementation for the ASON/GMPLS CARISMA network. In *Proceedings of 5th Workshop in G/MPLS Networks*, Girona (Spain), March 2006.

Bibliography

[Agr02] G. P. Agrawal. *Fiber-optic communications systems*. John Wiley & Sons, Inc., 2002.

[AKM+09] S. Azodolmolky, M. Klinkowski, E. Marin, D. Careglio, J. Solé-Pareta, and I. Tomkos. A survey on physical layer impairments aware routing and wavelength assignment algorithms in optical networks. *Elsevier Computer Networks*, 53(7):926–944, May 2009.

[Ala04] W. Alanqar. Optical networking evolution in ITU-T and IETF: A reality check. In *Proceedings of the Optical Fiber Communication Conference (OFC 2004)*, Los Angeles (USA), February 2004.

[Ang12] M. Angelou. *Cross-layer optimization for optical networks*. PhD Dissertation, Universitat Politècnica de Catalunya (UPC), 2012.

[APA+09] S. Azodolmolky, Y. Pointurier, M. Angelou, J. Solé-Pareta, and I. Tomkos. An offline impairment aware RWA algorithm with dedicated path protection consideration. In *Proceedings of the Optical Fiber Communication Conference (OFC 2009)*, San Diego (USA), March 2009.

[APA+10a] F. Agraz, J. Perello, M. Angelou, S. Azodolmolky, L. Velasco, S. Spadaro, P. Kokkinos, E. Varvarigos, and I. Tomkos. Experimental evaluation of path restoration for a centralised impairment-aware GMPLS-controlled all-optical network. In *Proceedings of 36th European Conference on Optical Communications (ECOC 2010)*, Torino (Italy), September 2010.

[APA+10b] S. Azodolmolky, Y. Pointurier, M. Angelou, J. Solé-Pareta, and I. Tomkos. Routing and wavelength assignment for transparent optical networks with QoT estimation inaccuracies. In *Proceedings of the Optical Fiber Communication Conference (OFC 2010)*, San Diego (USA), March 2010.

[APMS07] V. Anagnostopoulos, C. T. Politi, C. Matrakidis, and A. Stavdas. Physical layer impairment aware wavelength routing algorithms based on analytically calculated constraints. *Elsevier Optics Communications*, 270(2):247–254, 2007.

[AT02] M. Ali and L. Tancevski. Impact of polarization-mode dispersion on the design of wavelength-routed networks. *IEEE Photonics Technology Letters*, 14(5):720–722, May 2002.

[Azo10] S. Azodolmolky. *Physical impairments aware planning and operation of transparent optical networks*. PhD Dissertation, Universitat Politècnica de Catalunya (UPC), 2010.

[BCM07] G. Bogliolo, V. Curri, and M. Mellia. Considering transmission impairments in RWA problem: Greedy and metaheuristic solutions. In *Proceedings of the Optical Fiber Communication Conference (OFC 2007)*, Anaheim (USA), March 2007.

[BDL+01] A. Banerjee, L. Drake, L. Lang, B. Turner, D. Awduche, L. Berger, K. Kompella, and Y. Rekhter. Generalized multiprotocol label switching: an overview of signaling enhancements and recovery techniques. *IEEE Communications Magazine*, 39(7):144–151, July 2001.

[BGP-TE] H. Gredler, A. Farrel, J. Medved, and S. Previdi. North-Bound Distribution of Link-State and TE Information using BGP. Internet Draft, work in progress: draft-gredler-idr-ls-distribution-01, Internet Engineering Task Force, March 2012.

[BSBS08] J. Berthold, A. Saleh, L. Blair, and J. Simmons. Optical networking: Past, present, and future. *Journal of Lightwave Technology*, 26(9):1104–1118, May 2008.

[Car99] A. V. T. Cartaxo. Cross-phase modulation in intensity modulation-direct detection WDM systems with multiple optical amplifiers and dispersion compensators. *Journal of Lightwave Technology*, 17(2):178–190, February 1999.

[CCJ+09] M. Chamania, X. Chen, A. Jukan, F. Rambach, and M. Hoffmann. An adaptive inter-domain PCE framework to improve resource utilization and reduce inter-domain signaling. *Elsevier Optical Switching and Networking*, 6(4):259–267, December 2009.

[CCM05] R. Cardillo, V. Curri, and M. Mellia. Considering transmission impairments in wavelength routed networks. In *Proceedings of 9th International Conference on Optical Network Design and Modeling (ONDM 2005)*, Milan (Italy), 2005.

[CCV+07] P. Castoldi, F. Cugini, L. Valcarenghi, N. Sambo, E. Le Rouzic, M. J. Poirrier, N. Andriolli, F. Paolucci, and A. Giorgetti. Centralized vs. distributed approaches for encompassing physical impairments in transparent optical networks. In *Proceedings of 11th International Conference on*

Optical Network Design and Modeling (ONDM 2007), Athens (Greece), 2007.

[CFH+06] I. Cerutti, A. Fumagalli, R. Hui, P. Monti, A. Paradisi, and M. Tacca. Plug and play networking with optical nodes. In *Proceedings of the 8th International Conference on Transparent Optical Networks (ICTON 2006)*, Nottingham (United Kingdom), June 2006.

[CMM+11] R. Casellas, R. Martínez, R. Muñoz, L. Liu, T. Tsuritani, I. Morita, and M. Tsurusawa. Dynamic virtual link mesh topology aggregation in multi-domain translucent WSON with hierarchical-PCE. *OSA Optics Express Journal*, 19(26):B611–B620, December 2011.

[CMMG09] R. Casellas, R. Martinez, R. Muñoz, and S. Gunreben. Enhanced backwards recursive path computation for multi-area wavelength switched optical networks under wavelength continuity constraint. *Journal of Optical Communications and Networking*, 1(2):A180–A193, July 2009.

[CMP+03] J. Comellas, R. Martinez, J. Prat, V. Sales, and G. Junyent. Integrated IP/WDM routing in GMPLS-based optical networks. *IEEE Network*, 17(2):22–27, March/April 2003.

[CPVC07] F. Cugini, F. Paolucci, L. Valcarenghi, and P. Castoldi. Implementing a path computation element (pce) to encompass physical impairments in transparent networks. In *Proceedings of the Optical Fiber Communication Conference (OFC 2007)*, Anaheim (USA), March 2007.

[CSA+08] F. Cugini, N. Sambo, N. Andriolli, A. Giorgetti, L. Valcarenghi, P. Castoldi, E. Le Rouzic, and J. Poirrier. Enhancing GMPLS signaling protocol for encompassing quality of transmission (QoT) in all-optical networks. *Journal of Lightwave Technology*, 26(19):3318–3328, October 2008.

[DLRRS08] R. Douville, J.-L. Le Roux, J.-L. Rougier, and S. Secci. A service plane over the PCE architecture for automatic multidomain connection-oriented services. *IEEE Communications Magazine*, 46(6):94–102, June 2008.

[DR02] R. Dutta and G. Rouskas. Traffic grooming in WDM networks: past and future. *IEEE Network*, 16(6):46–56, November/December 2002.

[DS05] T. Deng and S. Subramaniam. Adaptive QoS routing in dynamic wavelength-routed optical networks. In *Proceedings of 2nd IEEE International Conference on Broadband Communications, Networks and Systems (BROADNETS 2005)*, Boston (USA), October 2005.

[DVP04] E. Dotaro, M. Vigoureux, and D. Papadimitriou. Multi-region networks: generalized multi-protocol label switching (GMPLS) as enabler for vertical integration. In *Proceedings of Global Telecommunications Conference (GLOBECOM 2004)*, pages 374–379, Dallas (USA), November 2004.

[ELW+04] G. Ellinas, J.-F. Labourdette, J. Walker, S. Chaudhuri, L. Lin, E. Goldstein, and K. Bala. Network control and management challenges in opaque networks utilizing transparent optical switches. *IEEE Communications Magazine*, 42(2):S16–S24, February 2004.

[G.707] ITU. Rec. G.707/Y.1322: Network node interface for the synchronous digital hierarchy (SDH), International Telecommunication Union, ITU-T, August 2003.

[G.7712] ITU. Rec. G.7712/Y.1703: Architecture and specification of data communication network, International Telecommunication Union, ITU-T, March 2003.

[G.7713.2] ITU. Rec. G.7713.2/Y.1704.2: Distributed call and connection management: Signalling mechanism using GMPLS RSVP-TE, International Telecommunication Union, ITU-T, March 2003.

[G.7715] ITU. Rec. G.7715/Y.1706: Architecture and requirements for routing in the automatically switched optical networks, International Telecommunication Union, ITU-T, June 2002.

[G.7715.2] ITU. Rec. G.7715.2/Y.1706.2: ASON routing architecture and requirements for remote route query, International Telecommunication Union, ITU-T, February 2007.

[G.805] ITU. Rec. G.805: Generic functional architecture of transport networks, International Telecommunication Union, ITU-T, March 2000.

[G.807] ITU. Rec. G.807/Y.1302: Requirements for automatic switched transport networks (ASTN), International Telecommunication Union, ITU-T, July 2001.

[G.8080] ITU. Rec. G.8080/Y.1304: Architecture for the automatically switched optical network (ASON), International Telecommunication Union, ITU-T, November 2001.

[G.872] ITU. Rec. G.872: Architecture of optical transport networks, International Telecommunication Union, ITU-T, November 2001.

[GPCC11] A. Giorgetti, F. Paolucci, F. Cugini, and P. Castoldi. Hierarchical PCE in GMPLS-based multi-domain wavelength switched optical networks. In *Proceedings of the National Fiber Optic Engineers Conference (NFOEC 2011)*, Los Angeles (USA), March 2011.

[GR08] S. Gunreben and F. Rambach. Assessment and performance evaluation of pce-based inter-layer traffic engineering. In *Proceedings of 12th International Conference on Optical Network Design and Modeling (ONDM 2008)*, Vilanova i la Geltrú (Spain), March 2008.

[Gro04] W. Grover. *Mesh-based survivable networks: options and strategies for optical, MPLS, SONET, and ATM Networking*. Prentice Hall PTR, 2004.

[H-PCEP] F. Zhang, Q. Zhao, O. Gonzalez de Dios, R. Casellas, and D. King. Extensions to Path Computation Element Communication Protocol (PCEP) for Hierarchical Path Computation Elements (PCE). Internet Draft, work in progress: draft-zhang-pce-hierarchy-extensions-01, Internet Engineering Task Force, October 2011.

[HBP06a] J. He and M. Brandt-Pearce. RWA using wavelength ordering for crosstalk limited networks. In *Proceedings of the Optical Fiber Communication Conference (OFC 2006)*, Anaheim (USA), March 2006.

[HBP06b] J. He and M. Brandt-Pearce. Dynamic wavelength assignment using wavelength spectrum separation for crosstalk limited networks. In *Proceedings of 3rd IEEE International Conference on Broadband Communications, Networks and Systems (BROADNETS 2006)*, San José (USA), October 2006.

[HBPP+07] J. He, M. Brandt-Pearce, Y. Pointurier, C. Brown, and S. Subramaniam. Adaptive wavelength assignment using wavelength spectrum separation for distributed optical networks. In *Proceedings of IEEE International Conference on Communications (ICC 2007)*, Glasgow (Scotland), June 2007.

[HBPPS07] J. He, M. Brandt-Pearce, Y. Pointurier, and S. Subramaniam. QoT-aware routing in impairment-constrained optical networks. In *Proceedings of Global Telecommunications Conference (GLOBECOM 2007)*, Washington (USA), November 2007.

[HDG10] A. Haddad, E. Doumith, and M. Gagnaire. A meta-heuristic approach for monitoring trail assignment in wdm optical networks. In *Proceedings of IFIP/IEEE Reliable Networks Design and Modeling (RNDM 2010)*, Moscow (Russia), October 2010.

[HHM05] Y. Huang, J. Heritage, and B. Mukherjee. Connection provisioning with transmission impairment consideration in optical WDM networks with high-speed channels. *Journal of Lightwave Technology*, 23(3):982–993, March 2005.

[IKM+03] R. Inkret, A. Kuchar, B. Mikac, C. Gauger, and M. Köhn. Advanced Infrastructures for photonic Networks - Extended Final Report of COST Action 266, 2003.

[INOT94] K. Inoue, K. Nakanishi, K. Oda, and H. Toba. Crosstalk and power penalty due to fiber four-wave mixing in multichannel transmissions. *Journal of Lightwave Technology*, 12(8):1423–1439, August 1994.

[Jai05] A. Jaiszczyk. Automatically switched optical networks: benefits and requirements. *IEEE Communications Magazine*, 43(2):S10–S15, February 2005.

[KCMV09] P. Kokkinos, K. Christodoulopoulos, K. Manousakis, and E. A. Varvarigos. Multi-parametric online RWA based on impairment generating sources. In *Proceedings of Global Telecommunications Conference (GLOBECOM 2009)*, Honolulu (USA), December 2009.

[KLXF08] D. King, Y. Lee, H. Xu, and A. Farrel. Path computation architectures overview in multi-domain optical networks based on ITU-T ASON and IETF PCE. In *Proceedings of IEEE/IFIP Network Operations and Management Symposium (NOMS 2008)*, Salvador da Bahia (Brazil), April 2008.

[KTMT05] P. Kulkarni, A. Tzanakaki, C. Machuka, and I. Tomkos. Benefits of Q-factor based routing in WDM metro networks. In *Proceedings of 31st European Conference on Optical Communications (ECOC 2005)*, volume 4, Glasgow (United Kingdom), September 2005.

[LHDF07] J. Li, K. Hinton, S. Dods, and P. Farrell. Enabling ASON routing via novel signal quality metrics. In *Proceedings of the Optical Fiber Communication Conference (OFC 2007)*, Anaheim (USA), March 2007.

[LKG+06] Q. Liu, M. Kök, N. Ghani, V. Muthalaly, and M. Wang. Hierarchical inter-domain routing in optical DWDM networks. In *Proceedings of 25th IEEE International Conference on Computer Communications (INFOCOM 2006)*, Barcelona (Spain), April 2006.

[LKGG06] Q. Liu, M. Kök, N. Ghani, and A. Gumaste. Hierarchical routing in multi-domain optical networks. *Elsevier Computer Communications*, 30(1):122–131, 2006.

[MBL+08] A. Morea, N. Brogard, F. Leplingard, J.-C. Antona, T. Zami, B. Lavigne, and D. Bayart. QoT function and A∗ routing: an optimized combination for connection search in translucent networks. *OSA Journal of Optical Networking*, 7(1):42–61, January 2008.

[MCK+09] K. Manousakis, K. Christodoulopoulos, E. Kamitsas, I. Tomkos, and E. A. Varvarigos. Offline impairment-aware routing and wavelength assignment algorithms in translucent WDM optical networks. *Journal of Lightwave Technology*, 27(12):1866–1877, June 2009.

[MCL+03] S. D. Maesschalck, D. Colle, I. Lievens, M. Pickavet, P. Demeester, C. Mauz, M. Jaeger, R. Inkret, B. Mikac, and J. Derkacz. Pan-European optical transport networks: An availability-based comparison. *Photonic Network Communications*, 5(3):203–225, May 2003.

[MKH+05] T. Miyamura, T. Kurimoto, R. Hayashi, I. Inoue, K. Shiomoto, and S. Urushidani. Demonstration of PCE-controlled dynamic traffic engineering for GMPLS-based multilayer service network. In *Proceedings of 31st European Conference on Optical Communications (ECOC 2005)*, Glasgow (United Kingdom), September 2005.

[MPC+06] R. Martinez, C. Pinart, F. Cugini, N. Andriolli, L. Valcarenghi, P. Castoldi, L. Wosinska, J. Comellas, and G. Junyent. Challenges and requirements for introducing impairment-awareness into the management and control planes of ASON/GMPLS WDM networks. *IEEE Communications Magazine*, 44(12):76–85, December 2006.

[MRBD09] A. Manolova, S. Ruepp, J. Buron, and L. Dittmann. On the efficiency of BGP-TE extensions for GMPLS multi-domain routing. In *Proceedings of 13th International Conference on Optical Network Design and Modeling (ONDM 2009)*, Braunschweig (Germany), February 2009.

[MTT05] C. Mas, I. Tomkos, and O. Tonguz. Failure location algorithm for transparent optical networks. *IEEE Journal on Selected Areas in Communications*, 23(8):1508–1519, August 2005.

[Muk00] B. Mukherjee. WDM optical communication networks: progress and challenges. *IEEE Journal on Selected Areas in Communications*, 18(10):1810–1824, October 2000.

[NGB03] M. Necker, C. Gauger, and S. Bodamer. A new efficient integrated routing scheme for SDH/SONET-WDM multilayer networks. In *Proceedings of the Optical Fiber Communication Conference (OFC 2003)*, Atlanta (USA), March 2003.

[OIF11] OIF. External Network-Network Interface (E-NNI) OSPF-based Routing - 2.0 (Intra-Carrier) Implementation Agreement, Optical Internetworking Forum, OIF, July 2011.

[OOI+05] S. Okamoto, T. Otani, W. Imajuku, D. Shimazaki, M. Hayashi, K. Ogaki, M. Miyazawa, I. Nishioka, M. Nanba, K. Morita, S. Kano, S. Seno,

K. Sagara, N. Arai, and H. Otsuki. Nationwide GMPLS field trial using different types (MPLS/TDM/lambda) of switching capable equipment from multiple vendors. In *Proceedings of the Optical Fiber Communication Conference (OFC 2005)*, Anaheim (USA), March 2005.

[OOO08] S. Okamoto, H. Otsuki, and T. Otani. Multi-ASON and GMPLS network domain interworking challenges. *IEEE Communications Magazine*, 46(6):88–93, June 2008.

[OZW05] Y. Ouyang, Q. Zeng, and W. Wei. Dynamic lightpath provisioning with signal quality guarantees in survivable translucent optical networks. *Optic Express*, 13(26):10457–10468, December 2005.

[PCE-Hierarc] D. King and A. Farrel. The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS. Internet Draft, work in progress: draft-ietf-pce-hierarchy-fwk-02, Internet Engineering Task Force, May 2012.

[PCE-StatFul] E. Crabbe, J. Medved, R. Varga, and I. Minei. PCEP Extensions for Stateful PCE. Internet Draft, work in progress: draft-crabbe-pce-stateful-pce-02, Internet Engineering Task Force, January 2012.

[PH06] C. Pinart and H. Harai. On using optical-layer link information parameters in distributed impairment constraint-based routing. In *Proceedings of the Optical Fiber Communication Conference (OFC 2006)*, Anaheim (USA), March 2006.

[PPK08] S. Pachnicke, T. Paschenda, and P. Krummrich. Assessment of a constraint-based routing algorithm for translucent 10 Gbits/s DWDM networks considering fiber nonlinearities. *OSA Journal of Optical Networking*, 7(4):365–377, April 2008.

[PRSV06] S. Pachnicke, J. Reichert, S. Spalter, and E. Voges. Fast analytical assessment of the signal quality in transparent optical networks. *Journal of Lightwave Technology*, 24(2):815–824, February 2006.

[PSA+11] J. Perello, S. Spadaro, F. Agraz, M. Angelou, S. Azodolmolky, Y. Qin, R. Nejabati, D. Simeonidou, P. Kokkinos, E. Varvarigos, S. Al Zahr, M. Gagnaire, and I. Tomkos. Experimental evaluation of centralized failure restoration in a dynamic impairment-aware all-optical network. In *Proceedings of the Optical Fiber Communication Conference (OFC 2011)*, Los Angeles (USA), March 2011.

[PSLR+11] C. Pinart, N. Sambo, E. Le Rouzic, F. Cugini, and P. Castoldi. Probe schemes for Quality-of-Transmission-aware wavelength provisioning. *Journal of Optical Communications and Networking*, 3(1):87–94, January 2011.

[QCT+10] Y. Qin, K. Cheng, J. Triay, E. Escalona, G. Zervas, G. Zarris, N. Amaya-Gonzalez, C. Cervello-Pastor, R. Nejabati, and D. Simeonidou. Demonstration of C/S based hardware accelerated qot estimation tool in dynamic impairment-aware optical network. In *Proceedings of 36th European Conference on Optical Communications (ECOC 2010)*, Torino (Italy), September 2010.

[QY99] C. Qiao and M. Yoo. Optical Burst Switching (OBS) - a New Paradigm for an Optical Internet. *Journal of High Speed Networks*, 8(1):69–84, December 1999.

[RDF+99] B. Ramamurthy, D. Datta, H. Feng, J. Heritage, and B. Mukherjee. Impact of transmission impairments on the teletraffic performance of wavelength-routed optical networks. *Journal of Lightwave Technology*, 17(10):1713–1723, October 1999.

[RFC2225] M. Laubach and J. Halpern. Classical IP and ARP over ATM. RFC 2225 (Proposed Standard), Internet Engineering Task Force, April 1998. Updated by RFC 5494.

[RFC2328] J. Moy. OSPF Version 2. RFC 2328 (Standard), Internet Engineering Task Force, April 1998. Updated by RFC 5709.

[RFC2370] R. Coltun. The OSPF Opaque LSA Option. RFC 2370 (Proposed Standard), Internet Engineering Task Force, July 1998. Obsoleted by RFC 5250, updated by RFC 3630.

[RFC3209] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow. RSVP-TE: Extensions to RSVP for LSP Tunnels. RFC 3209 (Proposed Standard), Internet Engineering Task Force, December 2001. Updated by RFCs 3936, 4420, 4874, 5151, 5420, 5711.

[RFC3471] L. Berger. Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description. RFC 3471 (Proposed Standard), Internet Engineering Task Force, January 2003. Updated by RFCs 4201, 4328, 4872, 6002, 6003, 6205.

[RFC3472] P. Ashwood-Smith and L. Berger. Generalized Multi-Protocol Label Switching (GMPLS) Signaling Constraint-based Routed Label Distribution Protocol (CR-LDP) Extensions. RFC 3472 (Proposed Standard), Internet Engineering Task Force, January 2003. Updated by RFCs 3468, 4201.

[RFC3473] L. Berger. Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions. RFC 3473 (Proposed Standard), Internet Engineering Task

Force, January 2003. Updated by RFCs 4003, 4201, 4420, 4783, 4874, 4873, 4974, 5063, 5151, 5420, 6002, 6003.

[RFC3477] K. Kompella and Y. Rekhter. Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE). RFC 3477 (Proposed Standard), Internet Engineering Task Force, January 2003. Updated by RFC 6107.

[RFC3630] D. Katz, K. Kompella, and D. Yeung. Traffic Engineering (TE) Extensions to OSPF Version 2. RFC 3630 (Proposed Standard), Internet Engineering Task Force, September 2003. Updated by RFCs 4203, 5786.

[RFC3717] B. Rajagopalan, J. Luciani, and D. Awduche. IP over Optical Networks: A Framework. RFC 3717 (Informational), Internet Engineering Task Force, March 2004.

[RFC3784] H. Smit and T. Li. Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE). RFC 3784 (Informational), Internet Engineering Task Force, June 2004. Obsoleted by RFC 5305, updated by RFC 4205.

[RFC3945] E. Mannie. Generalized Multi-Protocol Label Switching (GMPLS) Architecture. RFC 3945 (Proposed Standard), Internet Engineering Task Force, October 2004. Updated by RFC 6002.

[RFC4054] J. Strand and A. Chiu. Impairments and Other Constraints on Optical Layer Routing. RFC 4054 (Informational), Internet Engineering Task Force, May 2005.

[RFC4139] D. Papadimitriou, J. Drake, J. Ash, A. Farrel, and L. Ong. Requirements for Generalized MPLS (GMPLS) Signaling Usage and Extensions for Automatically Switched Optical Network (ASON). RFC 4139 (Informational), Internet Engineering Task Force, July 2005.

[RFC4202] K. Kompella and Y. Rekhter. Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS). RFC 4202 (Proposed Standard), Internet Engineering Task Force, October 2005. Updated by RFCs 6001, 6002.

[RFC4203] K. Kompella and Y. Rekhter. OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS). RFC 4203 (Proposed Standard), Internet Engineering Task Force, October 2005. Updated by RFCs 6001, 6002.

[RFC4204] J. Lang. Link Management Protocol (LMP). RFC 4204 (Proposed Standard), Internet Engineering Task Force, October 2005.

[RFC4206] K. Kompella and Y. Rekhter. Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE). RFC 4206 (Proposed Standard), Internet Engineering Task Force, October 2005. Updated by RFCs 6001, 6107.

[RFC4271] Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol 4 (BGP-4). RFC 4271 (Draft Standard), Internet Engineering Task Force, January 2006. Updated by RFC 6286.

[RFC4394] D. Fedyk, O. Aboul-Magd, D. Brungard, J. Lang, and D. Papadimitriou. A Transport Network View of the Link Management Protocol (LMP). RFC 4394 (Informational), Internet Engineering Task Force, February 2006.

[RFC4397] I. Bryskin and A. Farrel. A Lexicography for the Interpretation of Generalized Multiprotocol Label Switching (GMPLS) Terminology within the Context of the ITU-T's Automatically Switched Optical Network (ASON) Architecture. RFC 4397 (Informational), Internet Engineering Task Force, February 2006.

[RFC4655] A. Farrel, J.-P. Vasseur, and J. Ash. A Path Computation Element (PCE)-Based Architecture. RFC 4655 (Informational), Internet Engineering Task Force, August 2006.

[RFC4674] J. L. Roux. Requirements for Path Computation Element (PCE) Discovery. RFC 4674 (Informational), Internet Engineering Task Force, October 2006.

[RFC4872] J. Lang, Y. Rekhter, and D. Papadimitriou. RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery. RFC 4872 (Proposed Standard), Internet Engineering Task Force, May 2007. Updated by RFC 4873.

[RFC5152] J. Vasseur, A. Ayyangar, and R. Zhang. A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs). RFC 5152 (Proposed Standard), Internet Engineering Task Force, February 2008.

[RFC5212] K. Shiomoto, D. Papadimitriou, J. L. Roux, M. Vigoureux, and D. Brungard. Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN). RFC 5212 (Informational), Internet Engineering Task Force, July 2008.

[RFC5440] J. Vasseur and J. L. Roux. Path Computation Element (PCE) Communication Protocol (PCEP). RFC 5440 (Proposed Standard), Internet Engineering Task Force, March 2009.

[RFC5441] J. Vasseur, R. Zhang, N. Bitar, and J. L. Roux. A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths. RFC 5441 (Proposed Standard), Internet Engineering Task Force, April 2009.

[RFC5520] R. Bradford, J. Vasseur, and A. Farrel. Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism. RFC 5520 (Proposed Standard), Internet Engineering Task Force, April 2009.

[RFC5787] D. Papadimitriou. OSPFv2 Routing Protocols Extensions for Automatically Switched Optical Network (ASON) Routing. RFC 5787 (Experimental), Internet Engineering Task Force, March 2010.

[RFC6163] Y. Lee, G. Bernstein, and W. Imajuku. Framework for GMPLS and Path Computation Element (PCE) Control of Wavelength Switched Optical Networks (WSONs). RFC 6163 (Informational), Internet Engineering Task Force, April 2011.

[RFC6205] T. Otani and D. Li. Generalized Labels for Lambda-Switch-Capable (LSC) Label Switching Routers. RFC 6205 (Proposed Standard), Internet Engineering Task Force, March 2011.

[RSM09] S. Rai, C. F. Su, and B. Mukherjee. On provisioning in all-optical networks: An impairment-aware approach. *IEEE/ACM Transactions on Networking*, 17(6):1989–2001, December 2009.

[SIL+98] R. Sabella, E. Iannone, M. Listanti, M. Berdusco, and S. Binetti. Impact of transmission performance on path routing in all-optical transport networks. *Journal of Lightwave Technology*, 16(11):1965–1972, November 1998.

[SLMBMT+05] S. Sanchez-Lopez, X. Masip-Bruin, E. Marin-Tordera, J. Solé-Pareta, and J. Domingo-Pascual. A hierarchical routing approach for GMPLS based control plane for ASON. In *Proceedings of IEEE International Conference on Communications (ICC 2005)*, Seoul (South Korea), May 2005.

[SM04] T. Saad and H. Mouftah. Inter-domain wavelength routing in optical WDM networks. In *Proceedings of 11th International Telecommunications Network Strategy and Planning Symposium (Networks 2004)*, Vienna (Austria), June 2004.

[SS09] C. Saradhi and S. Subramaniam. Physical layer impairment aware routing (PLIAR) in WDM optical networks: issues and challenges. *IEEE Communications Surveys & Tutorials*, 11(4):109–130, Fourth quarter 2009.

[SS11] A. Saleh and J. Simmons. Technology and architecture to enable the explosive growth of the internet. *IEEE Communications Magazine*, 49(1):126–132, January 2011.

[SSC+07] S. Stanic, G. Sahin, H. Choi, S. Subramaniam, and H.-A. Choi. Monitoring and alarm management in transparent optical networks. In *Proceedings of 4th IEEE International Conference on Broadband Communications, Networks and Systems (BROADNETS 2007)*, Raleigh (USA), September 2007.

[SSZ+09a] C. Saradhi, E. Salvadori, A. Zanardi, S. Dalsass, and R. Piesiewicz. Hybrid control plane architecture for dynamic impairment-aware routing in transparent optical networks. In *Proceedings of the International Conference on Photonics in Switching (PS 2009)*, Pisa (Italy), September 2009.

[SSZ+09b] C. Saradhi, E. Salvadori, A. Zanardi, S. Dalsass, R. Piesiewicz, and I. Tomkos. Control plane issues in cross-layer optimized dynamic optical networks. In *Proceedings of the 11th International Conference on Transparent Optical Networks (ICTON 2009)*, Azores (Portugal), July 2009.

[SYS+09] E. Salvadori, Y. Ye, C. Saradhi, A. Zanardi, H. Woesner, M. Carcagni, G. Galimberti, G. Martinelli, A. Tanzi, and D. La Fauci. Distributed optical control plane architectures for handling transmission impairments in transparent optical networks. *Journal of Lightwave Technology*, 27(13):2224–2239, July 2009.

[SYT+02] K. Sato, N. Yamanaka, Y. Takigawa, M. Koga, S. Okamoto, K. Shiomoto, E. Oki, and W. Imajuku. GMPLS-based photonic multilayer router (Hikari router) architecture: an overview of traffic engineering and signaling technology. *IEEE Communications Magazine*, 40(3):96–101, March 2002.

[TMKO08] T. Tsuritani, M. Miyazawa, S. Kashihara, and T. Otani. Optical path computation element interworking with network management system for transparent mesh networks. In *Proceedings of the Optical Fiber Communication Conference (OFC 2008)*, San Diego (USA), February 2008.

[TWH09] J. Tapolcai, B. Wu, and P. H. Ho. On monitoring and failure localization in mesh all-optical networks. In *Proceedings of 28th IEEE International Conference on Computer Communications (INFOCOM 2009)*, Rio de Janeiro (Brazil), April 2009.

[VAM+10] L. Velasco, F. Agraz, R. Martínez, R. Casellas, S. Spadaro, R. Muñoz, and G. Junyent. GMPLS-based multidomain restoration: Analysis,

strategies, policies and experimental assessment. *Journal of Optical Communications and Networking*, 2(7):427–441, July 2010.

[WCZZ09] X. Wan, Y. Chen, H. Zhang, and X. Zheng. Dynamic domain-sequencing scheme for inter-domain path computation in WDM networks. In *Proceedings of Asia Communications and Photonics Conference and Exhibition (ACP 2009)*, Shanghai (China), November 2009.

[WHY09] B. Wu, P. Ho, and K. Yeung. Monitoring trail: On fast link failure localization in all-optical WDM mesh networks. *Journal of Lightwave Technology*, 27(18):4175–4185, September 2009.

[WSON-Enc] G. Bernstein, Y. Lee, D. Li, and W. Imajuku. Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks. Internet Draft, work in progress: draft-ietf-ccamp-rwa-wson-encode-13, Internet Engineering Task Force, October 2011.

[WW04] M. Wu and W. I. Way. Fiber nonlinearity limitations in ultra-dense WDM systems. *Journal of Lightwave Technology*, 22(6):1483–1498, June 2004.

[YADA01] Y. Ye, C. Assi, S. Dixit, and M. Ali. A simple dynamic integrated provisioning/protection scheme in IP over WDM networks. *IEEE Communications Magazine*, 39(11):174182, November 2001.

[YAZG10a] M. Youssef, S. Al Zahr, and M. Gagnaire. Cross optimization for RWA and regenerator placement in translucent WDM networks. In *Proceedings of 14th International Conference on Optical Network Design and Modeling (ONDM 2010)*, Kyoto (Japan), February 2010.

[YAZG10b] M. Youssef, S. Al Zahr, and M. Gagnaire. Traffic-driven vs. topology-driven strategies for regeneration sites placement. In *Proceedings of IEEE International Conference on Communications (ICC 2010)*, Cape Town (South Africa), May 2010.

[YMYD03] S. Yao, B. Mukherjee, S. Yoo, and S. Dixit. A unified study of contention-resolution schemes in optical packet-switched networks. *Journal of Lightwave Technology*, 21(3):672–683, March 2003.

[Yoo06] S. Yoo. Optical Packet and Burst Switching Technologies for the Future Photonic Internet. *Journal of Lightwave Technology*, 24(12):4468–4492, December 2006.

[YR05a] X. Yang and B. Ramamurthy. Dynamic routing in translucent WDM optical networks: the intradomain case. *Journal of Lightwave Technology*, 23(3):955–971, March 2005.

[YR05b] W. Yao and B. Ramamurthy. A link bundled auxiliary graph model for constrained dynamic traffic grooming in WDM mesh networks. *IEEE Journal on Selected Areas in Communications*, 23(8):1542–1555, August 2005.

[YS99] J. Y. Youe and S. W. Seo. An algorithm for virtual topology design in WDM optical networks under physical constraints. In *Proceedings of IEEE International Conference on Communications (ICC 2010)*, Vancouver (Canada), June 1999.

[ZCTM10] Y. Zhang, P. Chowdhury, M. Tornatore, and B. Mukherjee. Energy efficiency in telecom optical networks. *IEEE Communications Surveys & Tutorials*, 12(4):441–458, Fourth quarter 2010.

[ZH04] H. Zeng and C. Huang. Fault detection and path performance monitoring in meshed all-optical networks. In *Proceedings of Global Telecommunications Conference (GLOBECOM 2004)*, Dallas (USA), November/December 2004.

[ZHLL04] D. Zhemin, M. Hamdi, J. Lee, and V. Li. Integrated routing and grooming in GMPLS-based optical networks. In *Proceedings of IEEE International Conference on Communications (ICC 2004)*, Paris (France), June 2004.

[ZJM00] H. Zang, J. P. Jue, and B. Mukherjee. A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks. *Optical Networks Magazine*, 1(1):47–60, January 2000.

[ZLS+06] Y. R. Zhou, A. Lord, S. Santoni, D. Setti, T. Fischer, G. Lehmann, H. Bulow, H. Haunstein, and A. Schinabeck. PMD rules for physical constraint-based routing in all optical networks. In *Proceedings of the Optical Fiber Communication Conference (OFC 2006)*, Anaheim (USA), March 2006.

[ZM02] K. Zhu and B. Mukherjee. Traffic grooming in an optical WDM mesh network. *IEEE Journal on Selected Areas in Communications*, 20(1):122–133, January 2002.