Optical Networks
Series Editor: Biswanath Mukherjee

Suresh Subramaniam
Maïté Brandt-Pearce
Piet Demeester
Chava Vijaya Saradhi *Editors*

Cross-Layer Design in Optical Networks

Springer

Optical Networks

For further volumes:
http://www.springer.com/series/6976

Suresh Subramaniam • Maïté Brandt-Pearce
Piet Demeester • Chava Vijaya Saradhi
Editors

Cross-Layer Design in Optical Networks

Springer

Editors
Suresh Subramaniam
Department of Electrical
and Computer Engineering
George Washington University
Washington, DC, USA

Maïté Brandt-Pearce
Department of Electrical
and Computer Engineering
University of Virginia
Charlottesville, VA, USA

Piet Demeester
Department of Information Technology
Ghent University – iMinds
Ghent, Belgium

Chava Vijaya Saradhi
Create-Net
Networks and Services Department
Povo-Trento, Italy

ISSN 1935-3839 ISSN 1935-3847 (electronic)
ISBN 978-1-4614-5670-4 ISBN 978-1-4614-5671-1 (eBook)
DOI 10.1007/978-1-4614-5671-1
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013931993

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Chapter 1
Introduction

Suresh Subramaniam, Maïté Brandt-Pearce, Piet Demeester, and Chava Vijaya Saradhi

Optical networking based on Wavelength Division Multiplexing (WDM) has truly established itself as the preferred technology for ultra-long haul communications and networking. Tremendous strides in optical communications technology have been made over the past few decades: wavelength spacings have decreased to 25 GHz; wavelength counts have increased to more than 100; advanced modulation schemes have increased wavelength rates to 100 Gbps; and distributed Raman amplification has increased the optical reach to more than 1,000 km. Meanwhile, optical switching is becoming more and more widespread with the development of Reconfigurable Optical Add-Drop Multiplexers (ROADMs).

Despite the rapid strides in technology, there are still some fundamental challenges in need of satisfactory solutions. The Open Systems Interconnection (OSI) layered model of networks (and networking techniques subsequently developed to suit it) was based on small, low-speed networks, by current optical standards. To fully exploit the capability of optical technology, one must stray from the layered network approach, and into the realm of so-called cross-layer approaches. The purpose of this book is to present a collection of chapters on the topic of cross-layer design written by leading researchers in the field.

S. Subramaniam (✉)
George Washington University, Washington, DC, USA
e-mail: suresh@gwu.edu

M. Brandt-Pearce
University of Virginia, Charlottesville, VA, USA
e-mail: mb-p@virginia.edu

P. Demeester
Department of Information Technology, Ghent University – iMinds, Ghent, Belgium
e-mail: piet.demeester@intec.ugent.be

C.V. Saradhi
Nanyang Technological University, Singapore, Singapore
e-mail: saradhi@ntu.edu.sg

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*,
Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_1,

One challenge brought on by the desire for ever-increasing speeds and distances, and difficult to fully address through the layered approach, is the presence of significant physical layer impairments. Optical signals that traverse long distances are subject to a number of impairments such as amplifier noise, dispersion, crosstalk, four-wave mixing, self- and cross-phase modulation, etc. While improvements in the underlying technology can certainly help in reducing the severity of the impairments, a powerful way of mitigating the remaining effects is through cross-layer design and management, i.e., implementing physical-layer aware networking, and network-layer (or lambda-layer) aware physical layer design. Cross-layer design is routine in wireless communications and networking, but has only recently become a mainstream subject of study in optical networking, partly because it has just now become necessary due to the increase in wavelength data rates, reduced wavelength spacings, and increased optical switching possibilities resulting in ultra-long lightpaths.

There has also been a tremendous amount of research carried out in the context of the lambda-layer and the layer above, which can be called the service layer. Lightpaths at the lambda-layer serve as logical links to carry service-layer traffic, and the joint design and optimization of these layers has been known to improve performance significantly. In this area, traffic grooming is a significant research topic with many advancements achieved over the years. This topic has been covered extensively in other venues. In this book, we focus on other cross-layer topics that have been insufficiently addressed in the literature.

Since the topic of cross-layer design is currently an actively researched topic, we realize that it is impossible to keep abreast of all the progress being made on this topic. Nevertheless, there does not exist a single source for a collection of up-to-date work on the topic. Our aim in putting together this collection of chapters is to ensure that new researchers in the field can have a handy single source of reference to start from. The book can also serve as a guide to researchers who are already working on the topic.

The audience that the book is intended for consists of students in academia learning about and doing research on cross-layer optical network design, industrial practitioners working on optical networks, industry researchers designing next-generation networks plagued by some of the problems addressed in this book, and faculty members and researchers in academia wishing to teach an advanced course on optical networks, or conducting research in the area of cross-layer design. Our hope is that the reader gains a deep understanding of the ideas behind optical cross-layer network design and is inspired to go forward and find solutions for the remaining challenges, or better solutions for already-solved problems.

The book is organized by subject area. Chapters 2–9 focus on physical-layer aware network design, whereas chapters 10–16 deal with service-layer aware network design. A brief overview of each chapter follows.

Chapter 2 by Politi, Matrakidis, and Stavdas presents a tutorial on physical layer impairments that new researchers to the area will find invaluable. Chapter 3 by

Angelou, Azodolmolky, and Tomkos describes the latest work on the problem of online impairment-aware routing and wavelength assignment. As optical networks grow and evolve, various fiber types and line rates are expected to coexist within the same network. Chapter 4 by Nag, Tornatore, Liu, and Mukherjee presents issues and challenges in such mixed line rate networks. In Chapter 5, Christodoulopoulos and Varvarigos discuss the problem of off-line network design considering both linear and non-linear impairments. Both transparent networks without optical–electrical–optical (OEO) regenerators and translucent networks are considered. Chapter 6, by Maier et al., investigates the effects of physical parameter uncertainty on the network dimensioning problem. Much of the work in impairment-aware networking has to rely on simulations. Pointurier and He present some recent efforts on analytically modeling the performance of physically-impaired optical networks. Chapter 8, by Salvadori, Zanardi, and Saradhi, surveys the work on control plane architectures for handling physical layer impairments. Chapter 9 by Brandt-Pearce and Subramaniam present approaches for designing survivable networks under physical layer impairments.

The subject area service-layer aware network design starts off with a chapter by Puype et al. on energy-efficient traffic engineering. First, approaches for energy-efficient lightpath routing to carry higher layer IP traffic are presented, and then routing of lightpaths to optimize energy consumption through optical bypass is considered. Chapters 11 and 12 explore the same topic using different approaches. Qin and Gurusamy consider a multi-protocol label switching (MPLS)-over-WDM network and present integrated routing algorithms for protecting Label Switched Paths (LSPs) in Chapter 11. In the following chapter by Lee, Lee, and Modiano, the impact of a single physical link on multiple logical links (lightpaths) is examined through the introduction of new metrics and a quantitative analysis. In Chapter 13, Zervas and Abosi present multi-layer architectures for supporting grid and cloud services. Figuerola et al. present architectural concepts in the evolution of the user-controlled lightpath paradigm and Infrastructure-as-a-Service, or IaaS, for optical networks in Chapter 14. In Chapter 15, by Jinno and Tsukishima, the authors present a cross-layer network design and control testbed developed by NTT. In the final chapter of the book, Chan presents a comprehensive cross-layer design approach for free-space optical networks.

All the articles in this book have been written by leading researchers who have made significant contributions to this field over the past decade and longer. The articles were closely reviewed by the editors and others in their groups. The chapters published are revised versions based on the reviewer comments. Although every effort has been made to eliminate discrepancies and correct errors, some will inevitably remain, and for those we apologize.

We would like to thank all of the contributing authors for their time and effort in providing the excellent articles that make up this book. We would also like to acknowledge the help of Juzi Zhao and Farshad Ahdi at George Washington University, Houbing Song, Yi Tang, Xu Wang, and Mohammad Noshad at the University of Virginia, and Didier Colle, Wouter Tavernier Marc De Leenheer,

Pieter Audenaert, Bart Puype, and Dimitri Staessens at Ghent University – iMinds who helped in reviewing the articles. We are also grateful to Alex Greene at Springer for his help and patience, without which this book would not have been possible. Finally, we would like to thank Prof. Biswanath Mukherjee of the University of California at Davis, for suggesting the topic of the book to us and guiding us through the various stages of the book in his capacity as Springer Optical Networks Book Series Editor.

Chapter 2
A Tutorial on Physical-Layer Impairments in Optical Networks

C.T. Politi, C. Matrakidis, and A. Stavdas

1 Introduction

Exploiting the wavelength domain in order to utilise the abundant fiber spectrum has been one of the cornerstones of optical communications. By transmitting different channels on different wavelengths, one could multiply the transmitted capacity without the need of advanced multiplexing subsystems. The history of the WDM technique began once laser and receiver technology became widely available at room temperature; hence, the transmission of two signals simultaneously, one at the 1.3 μm and one at the 1.55 μm, was possible in the 1980s. Meanwhile the advent of fiber amplifiers in the late 1980s together with improved optical subsystems alleviated the need for a number of optoelectronic regenerators and brought about the breakthrough of the WDM technology in mid-1990s resulting in optical networks that soon looked like the one shown in Fig. 2.1. Each signal modulates one of the tens of wavelengths which are now multiplexed passively onto the same fiber. Many optical channels can now share the same optical components (fibers, amplifiers etc.), leading to an enormous cost reduction. Further enhancement of the overall capacity was achieved by technology improvements that could bring WDM channels closer. The next generation of WDM systems was based on higher channel bit rates per wavelength. 40 Gb/s made techno-economic sense only when efficient modulation formats combined with high-speed digital signal processing could actually replace adequately their 10Gb/s counterparts.

Nevertheless optical communications have been aiming at cost reduction through efficient utilisation of common resources. By looking into the history of optical communication systems [1, 2], both experimentally achieved and commercially available, it is evident that there are two ways that their capacity has been

C.T. Politi (✉) • C. Matrakidis • A. Stavdas
Department of Telecommunications Science and Technology,
University of Peloponnese, Tripolis, Greece
e-mail: tpoliti@uop.gv

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*,
Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_2,

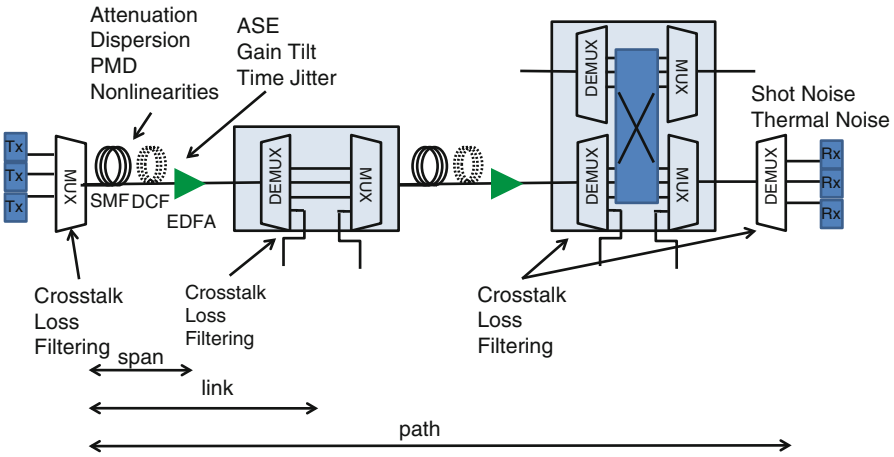


Fig. 2.1 WDM system that comprises a number of dispersion compensated and amplified spans and also ROADMs and OXCs. The figure indicates the impairments and their place to the system that will be thoroughly discussed in this chapter

evolved throughout the years: the one concerns efficient *multiplexing* and the other concerns efficient data *modulation* for spectral efficiency. Meanwhile optical networking has been based on the notion of *statistical multiplexing* for temporally efficient resource sharing.

2 Wavelength Division Multiplexing Transmission Systems

Multiplication of the transported capacity over a single fiber has been historically achieved through packing of individually transmitted channels into the same medium, with the least possible impact on each signal transmission, that is, multiplexing. We can then assume that the overall capacity of a wavelength domain multiplexed system is determined by the useable bandwidth, the spectral efficiency hence the number of channels and the bit rate per channel. Meanwhile the overall transmission reach is an important parameter for the overall cost of an optical WDM system. Evidently there is a trade-off between the above parameters in order to achieve the highest possible capacity together with the longest possible span. Achieving a high capacity times length product ($B \times L$) has been the main target of all different WDM system designs and is being compromised by all the physical degradations that are discussed in this chapter.

The maximum usable bandwidth in a WDM system is usually determined by the amplifier bandwidth and/or the demultiplexer bandwidth. In the most advanced WDM transmission systems in order to maximise the $B \times L$ product, both the C and L transmission bands are ulitised with EDFAs, but a further upgrade using, for instance, Raman amplifiers is suggested extending useable fiber bandwidth to the long-wavelength window of the fiber. The impressive achieved capacity of above

64 Tbit/s that has been reported in [3] is of the order of magnitude of the fiber raw bandwidth available in the spectral region between 1.3 μm and 1.7 μm. Recently other "domains" have been utilised in order to extend the multiplexing factor of WDM systems. Polarisation [4], code division [5] and spatial division multiplexing [6] have all been proposed as means to achieve orthogonal multiplexing of signals, with the latter being possibly the record capacity achieved so far in a single fiber. Recently due to its superior scalability, frequency domain has been also used, and coherent Orthogonal Frequency Division Multiplexing is well positioned to be an attractive choice for >100 Gb/s transmission [7]. Since then, there has been extensive innovation towards developing various forms of optical OFDM. Combination of all these multiplexing techniques with WDM is the trend in contemporary hero experiments that seek to achieve record B × L [1–3, 5–10].

Meanwhile the main way to increase the capacity of a WDM transmission system is the utilisation of the bandwidth more efficiently with spectrally efficient modulation formats. The idea is that other than modulating the amplitude of the complex signal A, as in *amplitude-shift keying* (ASK) modulation formats, one can let the phase, or its time derivative, carry the information, as in *phase-shift keying* (PSK) and *frequency-shift keying* (FSK), respectively [4]. PSK has better tolerance to noise at the receiver than ASK formats; however, *coherent receivers* required for phase detection became readily available only recently when digital signal processing allowed their advent. Quadrature modulation doubles the number of transmitted bits per symbol. Differential PSK modulation formats where the phase of a bit is modulated with respect to that of the previous one relaxed the need of complex receivers. Furthermore, a combination of more than one signal characteristics can be modulated, for example, the amplitude *and* the phase of the signal to represent different information symbols, as in *amplitude and phase-shift keying* (APSK). With respect to the number of states, one refers to *binary transmission* if only two states are allowed for A, representing two symbols 0 and 1, whereas in *M-ary transmission*, M states are allowed for A, representing M symbols. Although $\log_2 M$ times information is carried by symbols in M-ary transmission, which translates into higher spectral efficiency, transmitter and receiver complexity is higher and tolerance to noise decreases, like in duobinary or quaternary formats [4].

In [1–3] all hero experiments with respect to their overall capacity and spectral efficiency until 2010 are reviewed. Since then, the list has grown significantly mainly due to the advent of the new multiplexing techniques [6–10]. If we define spectral efficiency as the ratio of information rate per channel over the channel spacing, a bandwidth efficiency of 1 bit/s/Hz has been achieved in an 80 × 107 Gb/s system [1] or in a 16 × 112 Gbit/s system [2] up to 9.3 bit/s/Hz reported in [11]. In [12] 1.28 Tbit/s in a single-channel experiment has been reported. In [13] aggregate transmitted capacity of 25.6 Tbit/s has been reported in 2007, 64 Tbit/s in 2011 [3] and 102.3 Tbit/s [10] and 305Tbit/s in [6] in 2012. The transmission lengths are also impressive with 10.608 km reported in the experiment of [9]. All those experiments that combine advanced modulation formats with advanced communication techniques like coherent detection manage also to enhance the length of the B × L as they might exhibit improved robustness to the fiber impairments and

achieve increased receiver sensitivity [14]. In this chapter we seek to understand the signal degradations that arise when a multiwavelength high-bit-rate channel propagates into a system like the one shown in Fig. 2.1.

3 Wavelength Switched Optical Networks

WDM technology has been the key enabler in the evolution of optical networks from high-capacity point-to-point links towards flexible meshed optical networks with dynamic resource allocation utilising two-way reservations. Different ways have been used to describe optical networks. The concept of reconfigurable optical networks is not new – in fact, it has been around for at least two decades [15, 16]. In WSON, transportation from a source node to the destination node is completed in a transparent way by setting up a lightpath that comprises a number of spans (Fig. 2.1). This network supports an optical layer utilising multi-degree reconfigurable optical add/drop (ROADM) and cross-connect nodes (OXC) and provides traffic allocation, routing and management of the optical bandwidth. ROADMs although have emerged in the market before 2000 due to the downturn of telecommunications market following that year have only managed to be part of the major vendors portfolio around the globe only in 2002. Long-haul DWDM networks have nearly all been constructed with ROADMs in the last years. Hence, other than the transmission links, the principal building blocks of optical networks are those optical nodes shown in Fig. 2.1. To support flexible path provisioning and network resilience, OXCs utilise a switch fabric to enable routing of any incoming wavelength channel to the appropriate output port. Similarly ROADMs utilise switch fabrics in order to add/drop locally a number of wavelengths out of the WDM comb. First-generation ROADMs were of degree two and supported ring or line architectures. Currently ROADMs support high-degree nodes, and hence, they are featured as colourless and directionless. Colourless means that add/drop ports are not wavelength specific, and directionless feature enables any transponder to be connected to any degree.

Several designs have been proposed for robust ROADMs and OXCs based on different switching technologies and different architectures [17–22]. In Fig. 2.2 a ROADM and a OXC in a typical configuration are shown [23, 24]. Depending on the switching technology used, OXC designs are commonly divided into opaque and transparent [23]. Opaque OXCs incorporate either an electrical switch fabric or optical ones with OEO interfaces. They support sub-wavelength switching granularities and offer inherent regeneration, wavelength conversion and bit-level monitoring; multi-casting is possible if required. Switching times however are limited to ~ μsec if electronic switching is used. In opaque OXCs with an optical switch fabric, signal monitoring and regeneration can still be implemented but with added complexity, bit-rate limitation and increased power consumption. The complexity and power consumption are related to the processing that takes place in the

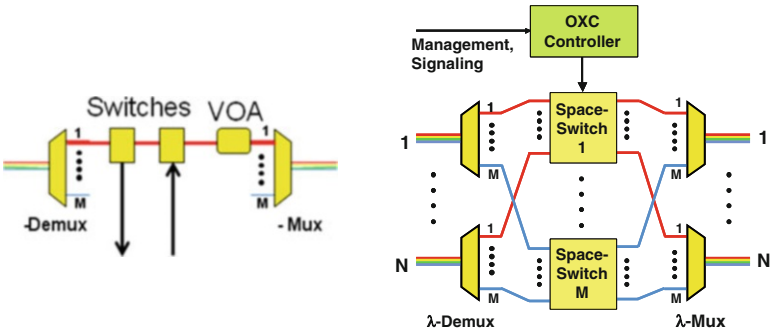


Fig. 2.2 A typical configuration of a ROADM [17] and an OXC architecture [16]. For the wavelength selective OXC architecture, the switching fabric is segmented, so following the wavelength demultiplexing stage, the incoming wavelength channels are directed to discrete switches each supporting a single wavelength

transponders. In [18] optoelectronic conversion is used as the means to perform regeneration without the implications of electronic signal processing.

For transparent OXCs a variety of optical switch fabrics have been developed [25, 26] that exhibit different physical performance. Switching time is a very significant feature and sets the switching time of the node.

When discussing physical impairments in optical networks, it is imperative to discuss them in the context of the switch architecture. To build a high-port-count node, the simplest solution is based on a central switch fabric, like the crossbar switch. However today, guided-wave technology can achieve small to moderate port counts (less than about 128 $\times$ 128) with moderate to high insertion loss and rapid switching speeds. Free-space technologies are more likely to achieve larger port counts (256 and higher) with low loss and slower switching speeds [19, 26]. Hence large crossbar switches are not a feasible architecture beyond a certain port count. Therefore, various multistage optical switch structures have been suggested like the three-stage Clos, the wavelength selective switch (WSS) and the λ-S-λ architectures [27].

4 Physical Impairments in Optical Networks

A typical WDM optical network path is illustrated in Fig. 2.1 with the physical representation of transmitter and receiver equipment, the transmission system that comprises the fiber spans and amplifiers, and also switching equipment like a ROADM and OXC. The figure indicates the place, where different degradation mechanisms occur that are going to be discussed in this chapter. The optical signals are generated by the modulation of N electronic signals on different wavelengths which are in turn multiplexed on the same fiber using a passive multiplexer.

They are then copropagating into the same fiber spans where they suffer different linear and nonlinear effects. Each fiber span is assumed to comprise a transmission fiber together with a dispersion compensating element and an optical amplifier to compensate for the span losses. Evidently different WDM designs exist. After propagating over a series of transmission spans and periodically spaced amplifiers, the WDM channels may be separated again in order to be switched at an OXC or dropped locally at a ROADM where it will be received by the receiver.

The following most important degradation mechanisms can be identified:

- Transmitter (Tx) and receiver (Rx)-related impairments
- Transmission-related impairments that are either linear effects (dispersion, polarisation mode dispersion) or nonlinear ones (Kerr effect related and scattering effects)
- Amplifier-related noise (amplified spontaneous emission generated in optical amplifiers)
- ROADM- and OXC-related distortion, that is, crosstalk (in multiplexers and switches) and nonideal filter characteristics and filter concatenation

In this chapter we will seek to understand the physical impairments that occur in such systems and give some analysis in most cases of how they can be computed and what impact they seem to have on the B $\times$ L product of an optical network.

4.1 Optical Signal Generation and Detection

Transmitters and receivers in WDM systems may induce their own impairments. The type and significance of transmitter/receiver impairments are related to the modulation format and type of detection [4, 28, 29].

Transmitters comprise optical source and modulators. Single-mode lasers with very low-side-mode suppression ratio are desirable as they can be spectrally spaced too closely combined with modulators with high extinction ratio that do not degrade the receiver sensitivity [4]. Furthermore phase noise could be detrimental, especially in PSK modulation formats.

Receivers on the other hand inherently induce shot noise and thermal noise during optical signal detection [4, 29] while depending on the detection technique beat noise may manifest itself due to the local oscillator. All these terms are discussed in the next chapter and are shortly defined as follows:

- Shot noise is related to the quantum nature of photons in the sense that random fluctuations of photons during the duration of a bit are translated into fluctuation of photocurrent electrons. The variance of this noise term is directly related to the receiver power; hence, it may be variable according to the signal modulation and/or according to the local oscillator amplitude.
- Thermal noise is associated with the random move of electrons in the electronic part of the receiver due to finite temperature.

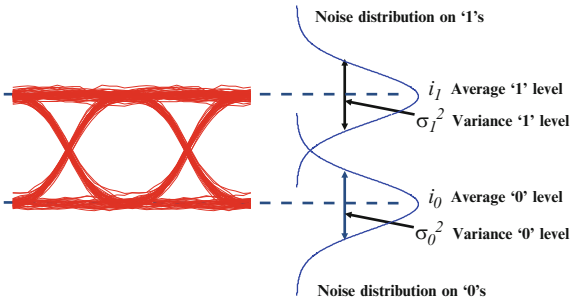


Fig. 2.3 Fluctuating signal received by the decision circuit and Gaussian probability densities of 1 and 0 together with the corresponding eye diagram

- Beat noise in coherent systems refers to the "beating" of more than one optical field at the photodetector.

Evidently the relative strength of the above factors strongly is associated on the receiver design, signal power and detection scheme. Furthermore their relative values with respect to the other system noise terms are important when designing a high-capacity WDM system.

4.2 Eye Diagram, Q Factor and BER

In any communication system, noise and distortion in the signal result in errors in the recovered signal. Here, we will briefly explain some of the main figures utilised in order to compare performance of different systems in optical communications. The ultimate measure of a system's performance in digital communications is the bit error rate (BER). This is defined as probability of faulty detected bits. BER calculations in a system are usually modelled by Monte Carlo simulations, Gaussian approximation or a deterministic approach [30]. The impairments discussed in this chapter are assumed to have Gaussian distribution.

Figure 2.3 shows schematically the fluctuating signal received by the decision circuit at the receiver. The corresponding "eye diagram" is illustrated which represents a repetitively sampled signal, superimposed in a way to provide representation of the noise behaviour. Characteristic degradations occur to the eye diagram that indicates imperfect transmission.

If i_D is the decision threshold, i_1 the photocurrent when the signal bit is "one", i_0 the photocurrent when the signal bit is "zero" and σ_0 and σ_1 are the photocurrent variances for 0 and 1 bits, the formula that describes the BER for an ASK signal after detection is therefore given by [29]:

$$\mathrm{BER} = \frac{1}{2}\frac{1}{2}\mathrm{erfc}\left(\frac{1}{\sqrt{2}}\frac{i_1 - i_D}{\sigma_1}\right) + \frac{1}{2}\frac{1}{2}\mathrm{erfc}\left(\frac{1}{\sqrt{2}}\frac{i_D - i_0}{\sigma_0}\right) \tag{2.1}$$

The optimum threshold is that for which the BER is minimised, and it is approximately:

$$i_D = \frac{\sigma_0 i_1 + \sigma_1 i_0}{\sigma_1 + \sigma_0}$$

by substituting this into (2.1) we get:

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

where Q is the Q factor defined as:

$$Q = \frac{|i_1 - i_0|}{\sigma_1 + \sigma_0} \tag{2.2}$$

It is evident that the Q factor gives an indication of the signal power level with respect to the noise induced. Even in the case of a simple system in the power level falls to a level that is of the order to the noise, the Q factor is affected and the signal is irreversibly destroyed; hence, it cannot be detected.

The BER improves, when the Q factor increases and takes values lower than 10^{-12} for $Q > 7$. For a typical communication system, the minimum acceptable BER would be 10^{-9} ($Q \sim 6$). The presence of Forward Error Correction (FEC) however allows for a better margin. With the introduction of FEC in optical communication systems in the mid-1990s, transmission experiments started to include a coding overhead (typically 7 % increase of the bit rate), to allow the operation of the system at a higher BER value. This is typically 10^{-5} so that at the output of a hard decision FEC module, the BER is less than 10^{-12} but can reach values as high as 10^{-3} depending on the codes, overheads and system under investigation.

For a given system, the receiver sensitivity typically corresponds to the average optical power for which a BER = 10^{-9}, or any other prespecified BER value. Equation (2.2) can assist in the determination of receiver sensitivity, for a back-to-back system, where only receiver-induced noise is affecting system performance. In order to calculate the receiver sensitivity, in a direct detection ASK system, we assume that $P_0 = 0$ (no power is carried on the 0 bits). The power carried on signal bit *one* is $P_1 = \mathrm{i}_1/R_{\mathrm{D}}$, where R_{D} is the receiver responsivity. The average received power will be $P_{\mathrm{s}} = (P_1 + P_0)/2 = P_1/2$. The RMS noise currents σ_1 and σ_0 include contributions from shot noise and thermal noise from the receiver and thus can be written as

$$\sigma_0 = \sigma_{\mathrm{T}} \quad \text{and} \quad \sigma_1 = \sqrt{\sigma_{\mathrm{T}}^2 + \sigma_{\mathrm{s}}^2}$$

If we neglect the dark current for the PIN receiver with Δf being the bandwidth of the receiver, we get the following expression from which P_s can be calculated:

$$Q = \frac{i_1}{\sqrt{\sigma_s^2 + \sigma_T^2} + \sigma_T} = \frac{R_D P_1}{\sqrt{2qR_D P_1 \Delta f + \sigma_T^2} + \sigma_T} \tag{2.3}$$

In binary phase modulation (BPSK) systems with coherent detection, a local oscillator (laser) is mixed with the detected signal before the photodiode. Typically a differential receiver is employed where the mixed signal is fed into two photodiodes out of phase and their outputs subtracted. The result is that the direct detection terms of the local oscillator (P_{LO}) and signal (P_s) are cancelled leaving only the mixing terms. Therefore, we have

$$i_1 = 2R_D\sqrt{P_{LO}P_s}$$
$$i_0 = -2R_D\sqrt{P_{LO}P_s} = -i_1$$

and

$$i_D = 0$$

Similarly, the RMS noise currents are equal for both transmitted symbols, and we have

$$\sigma_1 = \sigma_0 = \sqrt{\sigma_T^2 + \sigma_s^2}$$

for an unamplified back-to-back system. The shot noise term is now dominated by P_{LO} that is typically higher than the received signal P_s, so we have

$$\sigma_s^2 \approx 2qR_D P_{LO}\Delta f$$

The outcome of this analysis is that we can calculate the Q factor for the back-to-back coherent binary phase modulation receiver as

$$Q = \frac{2R_D P_{LO} P_s}{\sqrt{\sigma_T^2 + 2qR_D P_{LO}\Delta f}} \tag{2.4}$$

For quaternary (QPSK) modulated systems, there are two receivers, one for the in phase and one of the quadrature orthogonal components, with the above analysis holding for each one of them.

Now, the impact of the transmission channel on the signal degradation is manifested via the increase in the power required by the receiver to achieve the same BER. All the effects that will be discussed in the remaining of this chapter

degrade either the relative levels of i_1 and i_0, for example, effects like dispersion that act as intersymbol interference effects, or they increase the noise terms in the Q factor, hence degrading receiver sensitivity. To give an example of how the impact of the transmission channel on the signal Q factor degradation can be modelled analytically, let us assume a dispersion-compensated amplified WDM system that operates at 10 Gb/s with direct detection ASK signals like the one described in [31]. Here, one can argue that cross-phase modulation (XPM), four-wave mixing (FWM) and amplifier noise are the significant effects. All these effects can be modelled as Gaussian random variables. If we assume that all the above effects are included into the current fluctuations at the receiver, the standard deviation of the latter will be given by

$$\begin{aligned}\sigma_0 &= \sqrt{\sigma_{\mathrm{T}}^2 + \sigma_{\mathrm{sp-sp}}^2} \\ \sigma_1 &= \sqrt{\sigma_{\mathrm{T}}^2 + \sigma_{\mathrm{S}}^2 + \sigma_{\mathrm{s-sp}}^2 + \sigma_{\mathrm{sp-sp}}^2 + \sigma_{\mathrm{XPM}}^2 + \sigma_{\mathrm{FWM}}^2}\end{aligned} \tag{2.5}$$

where σ_{T} is the thermal and σ_{S} the shot noise of the receiver. The ASE-related spontaneous-spontaneous $\sigma_{\mathrm{sp\text{-}sp}}$ noise and signal-spontaneous $\sigma_{\mathrm{s\text{-}sp}}$ noise are calculated as in [32] for the whole amplifier chain and are explained in the second part of this chapter. σ_{XPM} and σ_{FWM} are the standard deviations of the XPM and FWM generated fluctuations, respectively, and are explained in the second part of the chapter. Correspondingly if coherently detected phase modulation is deployed, the standard deviation terms above should be appropriately modified with the inclusion of P_{LO}.

4.3 Optical Signal Propagation Through Optical Fibers and Amplifiers

4.3.1 Attenuation and Losses

In their quest to achieve a high capacity times length product, telecommunication transmission systems have introduced optical fiber as the main transmission medium, especially in long-haul systems. This is because attenuation in optical fiber is of the order of 0.36 dB/km at 1300 nm and 0.2 dB/km at 1550 nm. The effects that comprise the attenuation are mainly absorption, Rayleigh scattering and loss due to geometric effects. Absorption consists of intrinsic IR and UV absorption, natural property of the glass itself, that contributes to the absorption of very short and very long wavelengths [33, 34]. Meanwhile, impurities are a major source of loss due to manufacturing procedures in fibers. Two types of impurities are particularly bothersome to minimise in glass: metal ions and OH ions; these are important impurities as they create a peak at the attenuation. Finally, Rayleigh scattering acts as a theoretical boundary as it is the scattering by the small inhomogeneities in the material.

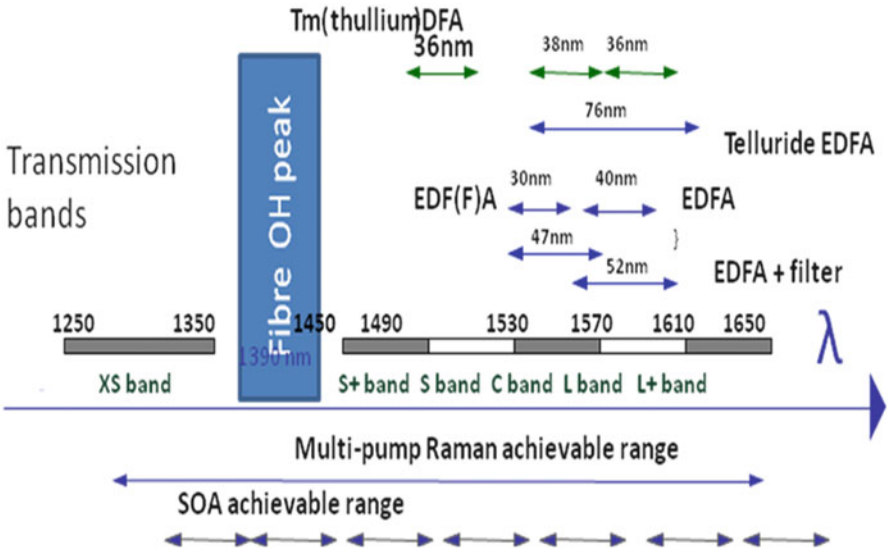


Fig. 2.4 WDM transmission bands with respect to the wavelength and different amplifier technologies that cover those wavelengths [35, 36]

4.3.2 Amplifiers

After having propagated through substantially long distances as well as through different optical networking elements, the signal power falls well beyond the levels that can be detected by the receiver. The advent of the optical amplifier and specifically the EDFA made the compensation of losses feasible simultaneously for a number of WDM channels, as long as these wavelengths were all confined in the gain bandwidth of the device. All optical signals are then optically amplified without the need of separate power consuming optoelectronic conversion. Different optical amplifiers have been proposed in the literature and have been developed with different characteristics as far as the operation principle, the material and operational characteristics are concerned. More specifically in order to achieve the widest useable bandwidth possible, different technologies may be devised as can be seen in Fig. 2.4.

Here we will discuss some general concepts of optical amplifications. In most cases it is based on the principle of stimulated emission, as in the case of laser. Without the use of optical feedback, however, only population inversion is required in gain medium for optical amplification to occur. Inserted photons are amplified by the gain medium along the length of the amplifier depending on the wavelength and the input signal power. Meanwhile spontaneously emitted photons of various wavelength, phase and direction are travelling along the amplifier and invoke amplification in the same amplifying medium. Hence, amplified photons exit the device together with amplified spontaneous emission (ASE) that acts as noise to the signal.

In most cases amplifiers are modelled ignoring the influence of noise, and in many cases they can be modelled as gain element that compensates exactly the losses from previous spans. This approximation is valid when strong saturation conditions are applied. An amplifier model that accounts for the self-saturating effects needs to be used as in [37]:

$$G = \frac{P_{\text{sat}}\text{productlog}\left(\frac{e^{\frac{P_{\text{in}}}{P_{\text{sat}}}} G_{\text{ss}} P_{\text{in}}}{P_{\text{sat}}}\right) + BhfNF}{P_{\text{in}} + BhfNF} \tag{2.6}$$

where P_{sat} is the saturation power of the amplifier, P_{in} is the input power, B is the bandwidth, f the frequency, NF the noise figure of the amplifier and the function productlog(x) is the solution of the differential equation $dy/dx = y/(x(1 + y))$.

In order to understand the effect of ASE on the Q factor, we have to assume an NRZ ASK signal. If we assume that the OSNR is measured in a specific $\Delta\nu_1$ optical bandwidth, then we can write [29]

$$\text{OSNR} = \frac{P_{\text{ave}}}{S_{\text{v}} \triangle \nu_1} \tag{2.7}$$

where S_{v} is the power spectral density of the ASE. $\Delta\nu_{\text{opt}}$ is the optical filter bandwidth after the amplifier, and $P_{\text{sp}} = \Delta\nu_{\text{opt}}\, S_{\text{v}}$ is the spontaneous emission noise power that enters the receiver. It is evident that when the signal propagates through a number of concatenated amplifiers, the spontaneous emission will build up together with the signal, and the overall OSNR will degrade.

At a direct detection receiver, the ASE-induced current noise has its origin in the beating of the signal electric field with the spontaneous emission noise but also of the beating of the spontaneous emission with itself. So the total variance of the current fluctuations is now going to have four terms [32]:

$$\sigma^2 = \sigma_{\text{T}}^2 + \sigma_{\text{s}}^2 + \sigma_{\text{s-sp}}^2 + \sigma_{\text{s-sp}}^2 \tag{2.8}$$

Correspondingly in (2.4) the ASE-related terms are the following:

$$\sigma_{\text{s}}^2 = 2qR(P_1 + P_{\text{sp}})\Delta f$$
$$\sigma_{\text{s-sp}}^2 = 4R^2 P_1 S_{\text{v}} \Delta f$$
$$\sigma_{\text{sp-sp}}^2 = 4R^2 S_v^2 \Delta f \Delta v_{\text{opt}}$$

Evidently P_1 accounts for the amplified power of 1, so both P_1 and S_{V} account for G [29]. By using the equations above, the Q factor can be calculated in (2.5). Note that the gain is assumed equal for the whole WDM comb although in reality gain tilting may affect the design of a WDM system. Furthermore ASE can also

induce time jitter in a bit sequence by shifting the optical pulses from their original time position randomly.

The modulation format plays an important role in the propagation of the signal through a chain of optical amplifiers or a generally amplified system. Although in order to compare different modulation formats directly one has to make several assumptions about the optical filtering and electronic hardware implementations that sometimes lead to nonoptimal solutions for specific cases, it is generally well recognised that RZ formats perform well among OOK modulation formats also due to the higher peak power. Especially RZ AMI exhibits an enhanced performance [14]. Among the PSK formats DPSK shows as expected a very good performance, if balanced detection is used, which allows in principle to double the noise-limited transmission distance. Evidently all the above concern a typical amplifier that reamplifies only the amplitude of the signal. Phase-sensitive amplifiers (PSA) have been reported [38] and may have a sub-quantum-limited noise figure with respect to the conventional erbium-doped fiber amplifier that has an associated inherent minimum noise figure of approximately 3 dB. All the above should be taken into account when/if advanced modulation formats are considered.

4.4 Dispersion

4.4.1 Chromatic Dispersion

Dispersion of light propagating into a fiber is the phenomenon where the group velocity of a propagating pulse depends on the wavelength. Typically pulses propagating in a single-mode fiber are hardly monochromatic; hence, chromatic dispersion broadens the pulse duration while causing intersymbol interference [4, 28, 29]. Although dispersion may severely affect the performance of a system based on single-mode fiber, it is a linear phenomenon; hence, it can be compensated completely if the reverse dispersion is applied. Mathematically, chromatic dispersion is defined either by the second derivative of the propagation constant β which is denoted as β_2 or by the so-called group velocity dispersion (GVD) coefficient normally denoted by D. Two are the components of the dispersion in a fiber:

Material dispersion: The silica dielectric constant, and, therefore, the refractive index, depends on the transmitted frequency.

Waveguide dispersion: The effective propagation constant related to the waveguiding nature of the fiber depends on the wavelength even if the core and cladding indices are constant.

Evidently to optimise the transmission of a multichannel WDM signal over a long fiber length is not a trivial task. The interplay of dispersion and nonlinearities plays an important role. Dispersion management has been introduced into systems that require high local dispersion for the sake of nonlinearities and negligible total accumulated dispersion (dispersion $\times$ length). The details of a transmission system

with dispersion management are shown in Fig. 2.1, where special dispersion-compensating fiber modules are utilised for each fiber span that retain high local values for the reduction of the effect of nonlinearities and overall dispersion remains zero. Different dispersion management schemes exist [4]. Besides the best choice of the amplifier and compensation module spans, the maximum and minimum powers and the exact dispersion map are important. As it is cost-efficient to co-place amplifiers and dispersion compensator system, the intra-amplifier spacing dictates the cost. Although the longer the spacing the lower the cost, the effect on the total span length achievable should also be considered.

With respect to the demands on dispersion management, the single-channel bit rate is essential. It is noted that the impact of dispersion measured as an eye-opening penalty increases with the bit rate and specifically scales quadratically. For 2.5 Gb/s per channel, no dispersion compensation is needed, and the signals have only to be amplified after some fiber span. For 10 Gb/s per channel long-haul systems, dispersion management is necessary, but fixed though properly adjusted compensators over the whole WDM bandwidth are sufficient. Furthermore, today advanced digital signal processing can be performed to complement the system tolerance to dispersion at those rates. For 40 Gb/s and above tuneable per channel, compensators have to complement the fixed compensators in order to address residual and time-varying dispersion and or advanced equalisation techniques.

Meanwhile the effect of modulation format has to be taken under consideration as spectrally narrow formats may yield significantly better dispersion tolerance. Duobinary and DQPSK formats have shown especially high-dispersion tolerance; however, the effect has to be discussed in relation to other system characteristics like, for example, filtering and/or fiber nonlinearity. As mentioned above it is the combination of format and bit rate that dictates the optimum dispersion compensation scheme. For a typical NRZ-ASK WDM transmission system, full dispersion compensation by optical means can be assumed to be feasible for the whole WDM comb.

4.4.2 Polarisation Mode Dispersion

Other than chromatic dispersion high-bit-rate transmission systems suffer from polarisation mode dispersion (PMD). The effect is related to the birefringent nature of the fiber and the two polarisations of light that copropagate in. So, when an optical pulse is injected in a fiber, at the end of the fiber, the pulse is split up in two pulses which have orthogonal polarisations and a delay against each other. At the receiver the two pulses are taken as a single broadened pulse, and this effect is called PMD [39]. For 10 Gb/s systems, PMD compensation is not necessary. For long-haul systems with 40 Gb/s and above, some compensation of PMD is necessary. Whereas dispersion compensation can be achieved using fixed, passive schemes, polarisation mitigation has to be adjusted to the actual state of PMD, and it has to be tuneable.

PMD affects the eye opening hence the Q factor of the channels in a statistical way. The effect is obviously related to the modulation format other than the channel bit rate. As in the case of chromatic dispersion, the tolerance of a specific format to the effect is related to the waveform and the filter; hence, direct comparison is out of the scope of this chapter. As a rule, however, one could assume that the first-order PMD sensitivity scales linearly with the symbol duration; hence, DQPSK is expected to be more tolerant than binary modulation formats although RZ formats are more resilient than NRZ due to their resilience to intersymbol interference [14].

Because of the possibly rare manifestation of the phenomenon instead of worst system design, specific margins are allowed in the operation of a system with predefined outage probability. To give an example, let us assume a system with k spans of SMF and DCF fiber and specific PMD. For a given fiber, the DGD is a random variable with a Maxwellian distribution. Following the analysis in [40], the distribution of the eye-opening penalty is derived using the statistics of the PMD vector which results in the pdf of the first- and higher-order PMD. As a result the pdf of the PMD-affected Q factor can be calculated. Now due to the statistical nature of the effect, the lowest acceptable achieved Q factor can be calculated for a specific outage probability (OP) derived by integration of the pdf (Q). By setting the outage probability equal to 0.00018, the achieved Q as a function of the Q factor without PMD (Q_{wopmd}) can be derived: $Q = Q_{\text{wopmd}}\text{OP}^{(\log 10/10n)}$ where $n = 16/\text{m}$ and $m = \text{A}\pi\text{k}(\text{DB})^2$ where D is related to the lengths and dispersion parameters of the single-mode fiber and the dispersion-shifted fiber of the span.

4.4.3 Nonlinear Effects

Nonlinear effects in optical fibers are related to the Kerr effect and scattering effects. The first is related to the variation of the fiber refractive index with incident optical power. In multichannel WDM transmission systems the optical power confined in the core of the fiber induces changes to the refractive index and hence to the phase change of the electromagnetic fields. Scattering effects are of two types Raman and Brillouin and can be generally overcome in WDM systems so will only be discussed for the sake of completeness.

Fiber nonlinearities are summarised in Fig. 2.5 after [14]. They are further divided in two large categories depending whether they occur as consequence of the interaction of the pulses of the same WDM channel or between the one channel and the ASE noise (intra-channel nonlinearities) or as a result of the interactions of two or more WDM channels (interchannel nonlinearities).

As far as the signal–signal interaction is concerned, effects like cross-phase modulation (XPM) and four-wave mixing (FWM) occur between WDM channels but also between individual pulses and lead to the phenomena of intra-channel XPM (IXPM) and intra-channel FWM (IFWM). The nonlinear interaction of a channel or a pulse with itself is referred to as self-phase modulation (SPM) which is a single-channel effect. Regarding signal noise interactions the dominant optical source of noise in a transmission line is typically the ASE generated by optical amplification.

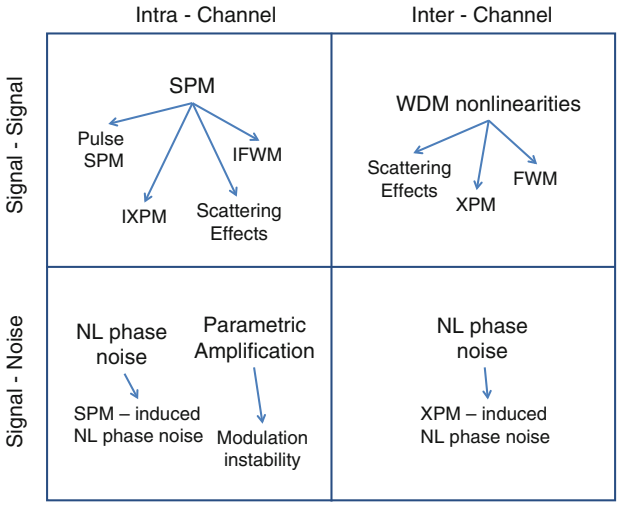


Fig. 2.5 Classification of fiber nonlinearities after [14]. SPM is self-phase modulation, XPM is cross-phase modulation, and FWM is the four-wave mixing. NL describes nonlinear. The I in front indicates the effect intra-channel

The dominant form of nonlinear signal–noise interactions manifests itself in noise-induced nonlinear distortions of specific signal characteristics (e.g. amplitude, timing, phase or frequency [41]). An example of signal–noise nonlinear interaction that affects PSK formats is nonlinear phase noise, which is sometimes referred to as the Gordon–Mollenauer effect [42].

The significance of each type of nonlinearities is related to the per-channel bit rate. In [14] the effects are thoroughly discussed, and it is argued that interchannel effects play important role at WDM systems where per-channel bit rates are of 10 Gb/s and below, whereas intra-channel nonlinearities affect systems most strongly when bit rates are above 10 Gb/s. The exact impact of each type of fiber nonlinearity also depends on the local fiber dispersion. Evidently lower dispersion fibers have stronger interchannel effects than fibers with high local dispersion.

Furthermore the modulation formats play a very important role, and we will discuss in each section the significance of the modulation format. Although a detailed understanding of the limitations that fiber nonlinearities have on advanced modulation formats is still a research topic, some numerical and experimental investigations suggest the types of nonlinear effects that are typically limiting transmission: Intensity-modulated formats are limited by all nonlinear effects but only above 40 Gb/s intra-channel effects start getting significant. For PSK formats, at 2.5 and 10 Gb/s, the main limitations come from nonlinear phase noise. At 40 Gb/s and above, intra-channel nonlinearities dominate, although nonlinear phase noise can become a limitation at low OSNR.

Since the effects of fiber nonlinearity depend on many system parameters, the comparison of modulation formats with respect to their resilience to such nonlinear

phenomena may depend on the system design and operating conditions. Hence when describing the impact of nonlinearities on an advanced modulation format, system characteristics should be specified.

Additional comparisons of various advanced modulation formats for nonlinear transmission can be found in [43–45].

Interchannel Effects

XPM

At the same time the intensity modulation of one channel effectively modulates the optical phase of all other WDM channels. The additional spectral components that may appear due to this induced phase modulation will generally lead to pulse distortion when operating in high-dispersion regime. Depending on the operating conditions, however, it has been widely accepted that higher dispersion can lead to an averaging of the effect, as each channel is affected by the impact of many pulses travelling with different speed.

XPM is an important source of performance degradation in multichannel WDM systems. In the following paragraph the analysis presented in [31] is adopted. By treating XPM as noise inducing effect, one can utilise equation (2.5) to analyse the XPM-induced Q factor degradation. In Fig. 2.6 we have modelled a transmission system that comprises 20 spans like in Fig. 2.1. 40 WDM channels modulated with ASK-NRZ copropagate with 1 mW of power per channel and 50 GHz channel spacing. The rest of the system parameters are modelled as in [32]. One of the curves shows the Q factor calculated with XPM as the only nonlinear effect included. Evidently XPM severely affects the Q factor of the specific system. Furthermore the effect on different channels implies that some are more severely affected than others which adds an extra degree of complexity in the optical network physical design. It is evident that XPM can be a limiting factor in a high-bit-rate WDM system with small channel spacing; however, it can be neglected if 100 GHz spacing is used.

Four-Wave Mixing

In a nonlinear medium where more than one electromagnetic wave propagate, mixing effects arise when the beating of two waves at their difference frequency drives material excitations. The coherent output to a new frequency is then the result of diffraction of the third wave from this material excitation. FWM is one of these phenomena which is severe in WDM systems with uniform channel spacing as the new products fall on neighbouring channels.

Performance degradation is caused in two ways:

- The generation of new components at different frequencies represents a loss of signal energy.

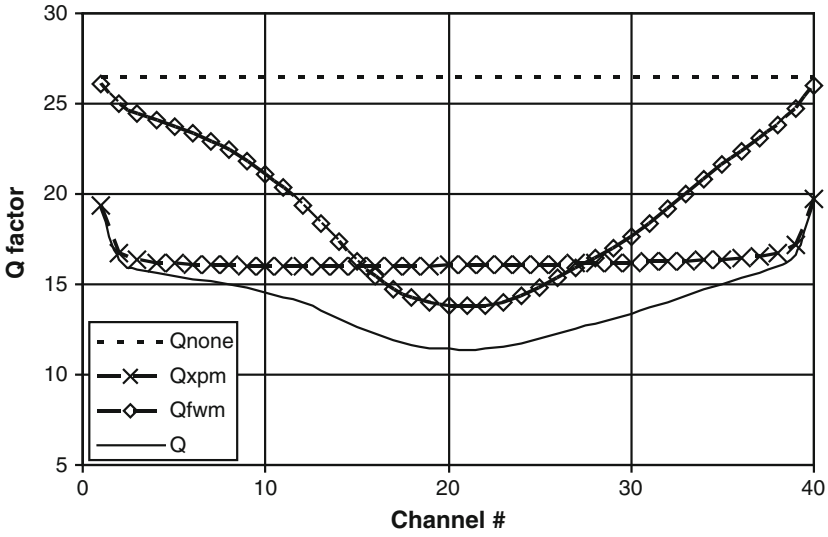


Fig. 2.6 Example of Q factor calculation with (solid line Q) and without nonlinear effects (dashed line Q_{none}) after [31] and [32]. Equation (2.5) is used to include XPM effects as in [46] and FWM effects as in [47]. All figures include ASE-related effects through (2.9)

- In WDM system using equally spaced channels the new components fall on frequencies allocated to other channels, causing severe crosstalk.

In a WDM system with N channels with equal channel spacing, the total time averaged FWM power generated at channel k at the end of the Mth link, assuming that the same input power per channel for all channels, can be calculated as in [47]. Like all nonlinearities, a moderate power level per channel is one possible solution for FWM. If fiber with uniform dispersion characteristics for all WDM channels is used, FWM may be tolerable for a WDM system. If the system is such that one channel falls exactly at zero dispersion wavelength, then only unequal channel spacing could be used to mitigate FWM as the products will fall out of the band of the signals. By assuming that FWM is a degradation-inducing effect that can be described as Gaussian noise, one can use equation 2.5 in order to model the effect of the FWM on the Q factor. In Fig. 2.6 we have modelled a transmission system that comprises 20 spans similar to the one in Fig. 2.1. One of the curves shows the value of the Q factor that is calculated when FWM is the only nonlinear effect included. It is obvious that FWM significantly affects the performance of some channels under the specific conditions.

Scattering Effects: Stimulated Raman and Stimulated Brillouin Scattering

Brillouin scattering is related to the Brownian motion of fiber molecules. Part of the light travelling through the fiber is backscattered by the proper component of the acoustic noise. This backscattered light, called Stokes wave, interferes

with the propagating light that acts as a pump. A stationary wave is induced, and a coherent acoustic wave is created that stimulates the Brillouin scattering, which in turn reinforces the acoustic wave. This feedback process is called stimulated Brillouin scattering (SBS). The SBS process can be summarised as an energy transfer from the pump wave to the Stokes wave; however, in source linewidths like today's WDM networks, it is not considered a limitation.

Stimulated Raman scattering (SRS) is a nonlinear effect appearing in systems also involving high-power sources. Light in the fiber interacts with molecular vibrations, and scattered light is generated at a wavelength longer than the incident light. Another signal co- or contra-propagating in the fiber will undergo amplification providing its wavelength correspond to the up-conversion. In wide WDM systems loss of energy and crosstalk between channels take place.

SRS may limit the performance of a WDM system by depleting the lower wavelength channels while adding crosstalk to the higher wavelength ones. In [31] an exact analytical solution for SRS is given, and equation (2.5) can be used to describe the effect of SRS in the Q factor of a WDM system.

In [4] it is argued that for a 40-channel WDM system with 100 GHz channel spacing, the SRS-related penalty can be reduced below 0.5 if the power per channel drops below 3 mW. Hence in today's networks by keeping moderate power levels, one can reduce the effect of this phenomenon.

Intra-channel Effects

SPM

As far as the SPM is concerned, which is a major single-channel nonlinear effect that affects all systems, the modulation of the signal power gives rise to a temporal variation of the optical phase of its signal channel which combined with the local dispersion may lead to pulse spreading not to mention the distortion in phase-modulated signals. It is actually the interplay between SPM and fiber dispersion that makes SPM an especially important nonlinear effect since it distorts the received waveform, degrades receiver sensitivity and limits transmission distance and/or optical amplifier output power.

In [48] the SPM is treated like a phenomenon that leads to frequency chirping of optical pulses. The SPM-induced chirp depends both on the variation of power and the pulse shape; hence, analytical treatment is not straightforward. In general spectral broadening of the pulse induced increases the signal bandwidth considerably and limits the performance of the system.

Other Intra-channel Effects

Due to the high fiber dispersion in dispersion managed systems, pulses within each channel tend to overlap during a significant part of transmission span, and as a result they interact through the fiber nonlinearity.

Together with SPM all pulses generated a contribution to the nonlinear phase shift that depend now on the power of the pulses and is called IXPM. Similarly to the XPM this effect causes different data pulses to experience time shifts causing time jitter. However, since locations where real acceleration of pulses may take place are related to the dispersion, IXPM can be reduced by launch position optimisation [49].

In a similar fashion IFWM is the result of a sum of three different pulses in the time domain that overlap due to the dispersion broadening, and a fourth pulse is generated. This will lead to a ghost pulse on a zero bit or amplitude jitter on a "one" bit. Different methods have been proposed for the suppression of IFWM leading to higher complexity receivers and transmitters. Utilising alternate polarisation among neighbouring bits reduces IFWM efficiency [4]. Sub-channel multiplexing in time domain with slightly different wavelengths has been suggested as means to detune the IFWM product out of the signal band.

4.5 OXC and ROADM Physical Impairments

Each OXC and ROADM introduces physical impairments that, similarly to other effects described so far, also limit the abundant fiber bandwidth in several ways. Hence, although transparency is a great asset for optical networks, in some cases the term has been misleading as the real offered transparency depends on many different system characteristics. Let us assume that by transparent, we characterise a system where no electronic processing takes place. The combination of transparency and fast reconfigurability gives the OXCs and ROADMs great flexibility.

The OXC and ROADM systems affect the physical performance in three different ways: introducing loss to the system, introducing crosstalk terms and imperfect filtering characteristics. In order to investigate the physical performance of such an optical networking element, the scalability versus cascadeability performance of such a switch should be evaluated. This is the physical performance of the element and how it is compromised by the increase in the number of wavelengths and fibers and the number of concatenated nodes.

Like all physical elements, OXCs and ROADMs introduce loss to the propagating signals which can be seen as an aggregate loss of all the elements that comprise the system (see Fig. 2.1), that is, multiplexers, demultiplexers and switch fabrics together with other required components like filters or attenuators. The main degradation source, however, is the loss of the switch fabric, mainly due to the high number of the devices that are required when high-port node architectures are designed. Insertion loss depends on the switching technology, the switch architecture and its size. Switch fabrics with large loss require power-consuming transceivers and demand more optical amplifiers to compensate for loss. Both can be really complex and power consuming when advanced modulation formats are utilised.

However, there is more into optimising the loss performance of an OXC than having low insertion loss. The variation of loss among different paths across the fabric must also be as low as possible. For example, this is one of the main disadvantages that may arise in multistage architectures. Although semiconductor optical amplifiers (SOAs) have been proposed as gates (on/off switches) also because of their induced gain, they still affect the OSNR of the bypass signals. It is also desirable that the performance of the optical switch be wavelength and polarisation independent. Polarisation-dependent loss (PDL) and polarisation mode dispersion (PMD) should be as low as possible.

However, the main physical-layer impairment introduced by the optical elements is crosstalk. Crosstalk terms arise due to the imperfect gating of signals or at demultiplexers, multiplexers and filters. It originates from the possible power leakage of imperfect devices but also due to the limited extinction ratio of the switches and can be either at the same wavelength with the interfering signal or at different wavelengths. The higher the extinction ratio of the switching device, the better it is. Residual optical power coming from a device when it is at the off state can be the source of crosstalk in a switching system. Crosstalk terms at the same wavelength as the main signal produce interferometric noise, which significantly compromises system operation. The phase noise in the interfering terms is converted to intensity noise when they, are converted to electrical form by a square law detector. These current fluctuations manifest themselves as relative intensity noise (RIN) and add to those resulting from shot noise and thermal noise in equations (2.5). In [50] an analytical treatment of the RIN is reported. As a result of this extra parameter, the Q factor is now reduced in the presence of intensity noise, and in order to maintain the same Q factor, the received power must be increased. Crosstalk terms that are at different wavelength appear as power addition crosstalk which is also expected to reduce the Q factor.

Other than the residual power that is due to imperfect filtering, filter shapes dictate specific performance to optical signals that propagates through them. In optical networks optical signals propagate through a number of mux/demux and transparent OXCs. The concatenation of filters narrows the overall optical bandwidth of the devices, and the propagating signal may be affected by passband misalignments. This evidently is true for advanced modulation formats for which careful filter design considerations are sought. Hence for some modulation formats, tight filtering is optimum as ASE is suppressed (e.g. duobinary ASK formats), while others (NRZ) which are susceptible to ISI may not be ideal following considerable concatenation of filters.

5 Conclusions

Wavelength Switched Optical Networks (WSONs) have arisen as a natural continuation of the success of WDM systems. Although the introduction of fiber in the 1970s was to demonstrate the delivery of a humble capacity of 6 Mbps, it was clear that the optical

channel could offer much more than that. However, at the time it seemed that attenuation was the only factor that could limit the promising prospects of the newly developed transmission medium. Since then, optical communication history has proven that physical impairments are numerous and always related to the context of the achievable B $\times$ L. Today that achievable capacities reach the available raw fiber bandwidths and spectral efficiency is starting to become the main issue, cross-channel effects that were discussed in the context of WDM seemed already as a trivial design issue, and intra-channel effects are taking over as the bothersome obstacles in system development. Meanwhile crosstalk and filtering effects that arise from the optical networking elements are adding some complexity to the optimisation of the physical layer. In the context of this chapter we discussed all the impairments that impact the performance of the system and are related to the transmitter and receiver, the fiber nonlinear effects and linear effects, the amplifier-induced noise and the ROADM filtering and crosstalk effects. However, it was made clear that the significance of the various impairments strongly depends on multiplexing scheme, modulation format and detection mode, while their interplay between linear and nonlinear impairments complicates things. So in low-bit-rate amplitude modulation, FWM and XPM are the main nonlinear impairments; however, as data rates increase and formats become more spectrally efficient, intra-channel effects dominate. For low-bit-rate phase-modulated formats, the main limitations on nonlinear transmission generally come from nonlinear phase noise. At 40 Gb/s and above, intra-channel nonlinearities dominate.

However, exact conclusions on relative impairment impact are system design specific. This is one of the main reasons behind the versatility of methods for combating linear physical impairments has not been an easy task. Advances in optical enabling technologies, such as dispersion compensation modules, and in high-speed electronics, such as feedforward equalisers, have made possible the design of spectrally efficient high-capacity transmission system where impairments are mitigated via well-studied digital communication and signal processing techniques like modulation, coding and equalisation. Specifically as data rates move beyond 100 Gb/s per channel, modulation formats and line coding are used to mitigate linear and nonlinear impairments explained in the chapter while achieving high spectral efficiencies in optical network environments. Meanwhile multi-level modulation is being applied as means to apply electronic pre- and post-processing by using lower rate digital electronics hardware.

At the same time optical component technology has been researched as means to convey the abundant fiber bandwidth to other functionalities in the network which up to now have been occupied by electronic technology. Optical subsystems are appealing as alternatives to electronic counterparts only when substituting many components with one subsystem for efficient resource sharing (fiber, amplifier, dispersion compensation). As a result, optical technology is more attractive as a genuinely cost effective and recently energy efficient suggestion albeit unsuitable for the intelligent manipulation and processing of bits. Towards this direction the focus of optical intelligence today has moved from devising optical bit-rate tailored subsystems like 2R regenerators towards impairment resilient modulation formats, for example. Combating physical-layer impairments with optical means while

utilising the available bandwidth efficiently seems to be the next cost-efficiency achievement of optical technology that will probably win over the progressive advancement of electronic processing.

Acknowledgments This work was carried out with the support of IP STRONGEST and Network of Excellence ICT BONE both funded by the European Commission.

References

1. Essiambre RJ et al (2010) Capacity limits of optical fiber networks. J Lightwave Technol 28 (4):662–700
2. Essiambre RJ et al (2009) Capacity limits of fiber-optic communication systems. March 2009, San Diego, OFC 2009
3. Zhou X et al (2011) 64-Tb/s, 8 b/s/Hz, PDM-36QAM transmission over 320 km using both pre- and post-transmission digital signal processing. J Lightwave Technol 29(4):571–577
4. Stavdas A (2010) Core and metro networks. Wiley, pp 382–438
5. Prucnal PR (2005) Optical code division multiple access: fundamentals and applications. Optical science and engineering. CRC press
6. Sakaguchi J et al (2012) 19-Core fiber transmission of 19x100x172-Gb/s SDM-WDM-PDM-QPSK signals at 305Tb/s. Post deadline paper OFC 2012, PDP5C.1, Los Angeles, USA
7. Lowery AJ, Du LB (2011) Optical orthogonal division multiplexing for long haul optical communications: a review of the last five years. Optical Fiber Telecommunications 17:421–438 (Invited review article)
8. Schuh K, Lach E, Junginger B, Veith G, Renaudier J, Charlet G, Tran P (2009) 8 Tb/s (80 X 107 Gb/s) DWDM NRZ-VSB transmission over 510 km NZDSF with 1 bit/s/Hz spectral efficiency. Bell Labs Tech J 14(1):89–104, Alcatel-Lucent. Published by Wiley Periodicals, 2009
9. Cai J et al (2010) Transmission of 96x100G pre-filtered PDM-RZ-QPSK channels with 300% spectral efficiency over 10.608 km and 400% spectral efficiency over 4,368km, San Diego, California, March 21, 2010, Postdeadline paper, Optical Fiber Communication Conference OFC 2010
10. Sano A et al (2012) 102.3-Tb/s (224 x 548-Gb/s) C- and extended L-band all-Raman transmission over 240 km Using PDM-64QAM single carrier FDM with digital pilot tone, post deadline paper OFC 2012
11. Nzakawa M (2008) Challenges of FDM-QAM coherent transmission with ultrahigh spectral efficiency. In: Proc. ECOC 2008, Paper, Th3.4.4
12. Gual P et al (2010) 1.28 Tbit/s/channel single-polarization DQPSK transmission over 525 km using ultrafast time-domain optical Fourier transformation, We.6.C.3, ECOC 2010, 19–23 Sep 2010, Torino, Italy
13. Guank AH et al. 25.6 Tb/sec C+L band of polarisation multiplexed RZQPSK signals. In: Proc OFC 007, post deadline paper PDP 19
14. Winzer PJ, Essiambre R-J (2006) Advanced modulation formats for high-capacity optical transport networks. J Lightwave Technol 24(12):4711–4728
15. Hill G (1988) A wavelength routing approach to optical communication networks. Br Telecom Technol J 6(3):24–31
16. Stavdas A, Politi C(T), Orphanoudakis T, Drakos A (2008) Optical packet routers: how they can efficiently and cost-effectively scale to petabits per second [Invited]. J Opt Netw 7(10):867
17. Wilfong G et al (1999) WDM cross-connect architectures with reduced complexity. J Lightwave Technol 17(10):1732–1741
18. Iannone E, Sabella R (1996) Optical path technologies: a comparison among different cross-connect architectures. J Lightwave Technol 14(10):2184–2196

19. Chu PB, Lee S-S, Park S (2002) MEMS: the path to large optical cross-connects. IEEE Commun Mag 40(3):80–87
20. De Dobbelaere P et al (2002) Digital MEMS for optical switching. IEEE Commun Mag 40(3):88–95
21. Ramaswami R. Using all-optical crossconnects in the transport network (invited), WZ1-I, OFC2001
22. El-Bawab TS (2006) Optical switching. Springer, ISBN 978-0387-26141-6, 2006
23. Ben Yoo SJ (2006) Optical packet and burst switching technologies for the future photonic internet. J Lightwave Technol 24(12):4468
24. Grobe K (2006) Applications of ROADMs and control planes in metro and regional networks. NTuC1, OFC 2006
25. Papadimitriou GI, Papazoglou C, Pomportsi CAS (2003) Optical switching: switch fabrics, techniques, and architectures. J Lightwave Technol 21(2):384–405
26. Zheng X et al (2003) Three-dimensional MEMS photonic cross-connect switch design and performance. IEEE J Sel Top Quantum Electron 9(2):571–578
27. Stavdas A, Bianco A, Pattavina A, Raffaelli C, Matrakidis C, Piglione C, Politi C(T), Savi M, Zanzottera R (2012) Performance evaluation of large capacity broadcast-and-select optical crossconnects. Opt Switch Netw 9(1):13–24
28. Ramaswami R, Sivarajan K (1998) Optical networks. Morgan Kauffman Publishers, Burlington
29. Agrawal P (1997) Fibre optic communication systems. Wiley, New York
30. Jeruchim MC, Balaban P, Shanmugan KS (1992) Simulation of communication systems. Plenum Press, New York
31. Djordjevic I, Stavdas A, Skoufis C, Sygletos S, Matrakidis C (2003) Analytical modelling of fibre non-linearities in amplified dispersion compensated WDM systems. Int J Model Simul 23(4):226–233
32. Anagnostopoulos V, Politi C(T), Matrakidis C, Stavdas A (2007) Physical layer impairment aware wavelength routing algorithms based on analytically calculated constraints. Opt Commun 270(2):247–254
33. Palais JC (1998) Fiber optic communications. Prentice-Hall, Upper Saddle River
34. Kazovsky L, Benedetto S, Willner A (1996) Optical fiber communication systems. Artech house, London
35. Nakagawa K. Dep. EIE, Yamagata University, Progress in optical amplifiers and the future of optical communications systems, OAA'1999 Nara
36. Bayart D, Baniel P, Bergonzo A, Boniort JY, Bousselet P, Gasca L, Hamoir D, Leplingard F, LeSauze F, Nouchi P, Roy F, Sillard P (2000) Broadband optical fiber amplification over 17.7 THz range. Electron Lett 36(18):1569–1571
37. Stavdas A, Sygletos S, O'Mahoney M, Lee H, Matrakidis C, Dupas A (2003) IST-DAVID: concept presentation and physical layer modelling of the metropolitan area network. IEEE J Lightwave Technol 21(2):372
38. Tong Z, Bogris A, Karlsson M, Andrekson PA (2010) Full characterization of the signal and idler noise figure spectra in single-pumped fiber optical parametric amplifiers. Opt Express 18:2884–2893
39. Bülow H, Lanne S. Optical and electronic PMD compensation. OFC 2003; Tutorial Notes; pp 175–199
40. Pachnicke S, Gravemann T, Windmann M, Voges E (2006) Physically constrained routing in 10-Gb/s DWDM networks including fiber nonlinearities and polarization effects. IEEE J Lightwave Technol 24(9):3418
41. Haus HA, Lai Y (1990) Quantum theory of soliton squeezing: a linearized approach. J Opt Soc Am B, Opt Phys 7(3):386–392
42. Gordon JP, Mollenauer LF (1990) Phase noise in photonic communications systems using linear amplifiers. Opt Lett 15(23):1351–1353

43. Dahan D, Eisenstein G (2002) Numerical comparison between distributed and discrete amplification in a point-to-point 40-Gb/s 40-WDM-based transmission system with three different modulation formats. J Lightwave Technol 20(3):379–388
44. Hodži´c A, Konrad B, Petermann K (2002) Alternative modulation formats in n 40 Gb/s WDM standard fiber RZ-transmission systems. J Lightwave Technol 20(4):598–607
45. Gnauck AH, Liu X, Wei X, Gill DM, Burrows EC (2004) Comparison of modulation formats for 42.7-Gb/s single-channel transmission through 1980 km of SSMF. IEEE Photonics Technol Lett 16(3):909–911
46. Cartaxo AVT (1999) Cross-phase modulation in intensity modulation-direct detection WDM systems with multiple optical amplifiers and dispersion compensators. J Lightwave Technol 17 (2):178–190
47. Zeiler W et al (1996) Modeling of four-wave mixing and gain peaking in amplified WDM optical communication systems and networks. J Lightwave Technol 14(9):1933–1942
48. Martensson J, Westlund M, Berntson A (2000) Intra channel pulse interactions in 40Gbpse dispersion managed RZ transmission systems. Electron Lett 36(3):244–246
49. Djordjevic I, Vasic Constrained B (2006) Coding techniques for the suppression of intra-channel nonlinear effects in high –speed optical transmission. J Lightwave Technol 24 (1):411–419
50. Takanishi H, Oda K, Toda H (1996) Impact of crosstalk in an arrayed-waveguide multiplexer on NxN optical interconnection. J Lightwave Technol 14(6):1097–1105

Chapter 3
Dynamic Impairment-Aware Routing and Wavelength Assignment

Marianna Angelou, Siamak Azodolmolky, and Ioannis Tomkos

The deployment of the first fiber-based networks in the 1980s was a decisive moment in telecom history as it provided the ground for the important upcoming developments. Nevertheless, it was the emergence of wavelength-division multiplexing (WDM) a decade later that allowed a big step forward offering significant benefits in terms of capacity and spectral efficiency. Multiplexing optical channels over a single fiber-optic strand enabled the intensity-modulated communication systems to transmit 10^{-2} to 10^{-1} bit/s/Hz [1] as opposed to the single-channel systems that were then limited to 10^{-6} to 10^{-7} bit/s/Hz. As a consequence, core networks managed to significantly increase the offered capacity in a cost-effective manner.

From that point on, core networks followed a path towards continuously higher capacities and improved economics with denser WDM systems, higher bit rates [2] and advanced modulation formats but also intelligent optical network nodes. Indeed in the early days of WDM, core networks consisted of a set of point-to-point optical links. A signal that propagated across the network had to be converted to the electrical domain and then back to optical at every intermediate node in order to travel long distances. The idea of avoiding all these costly optical–electronic–optical (OEO) conversions triggered the realization of optical add–drop multiplexers (OADMs) and optical cross-connects (OXCs) that allowed the optical signals to pass through or get switched at a node all-optically. Thus, a network, depending on its scale, either is *transparent* denoting that there are no OEO interfaces throughout the nodes or has few optical regeneration sites that form islands of transparency inside the network graph [3]. The term *translucent* is used to describe a network in the latter case. Due to the physical-layer effects that affect the quality of transmission (QoT) of an optical signal, the optical reach is finite and depends on the topological, physical and traffic parameters that regulate the network [4]. In order to establish connections that are longer than the maximum optical reach, regenerators are necessary to re-amplify, reshape and retime (i.e., 3R) the optical signal and maintain the acceptable QoT [5].

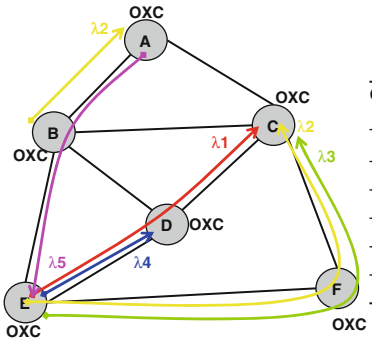
M. Angelou (✉) • S. Azodolmolky • I. Tomkos
Athens Information Technology Center, 19.5 km Markopoulou Avenue, 19002 Peania, Greece
e-mail: mang@ait.edu.gr

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*,
Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_3,

1 Impairment-Aware RWA

In a core WDM network the traffic that travels from node to node is carried over a set of optical channels effectively dividing the bandwidth of the fiber medium into a number of nonoverlapping wavelengths. Given a connection demand between two optical nodes, a *lightpath* is established between the destination and the source node forming a (semi)-permanent circuit. An LP is essentially an optical signal running on a specific wavelength travelling possibly via multiple fiber links until it reaches the destination node. To serve a connection demand, an LP is realized by determining the sequence of nodes and allocating to it one of the available optical channels. The traffic matrix (or demand set) that is active in a wavelength-routed network is served by an equal number of LPs, as depicted in Fig. 3.1. An OXC serves either as a switching element when it is an intermediate node or as an add/drop element when it is the source/destination node, respectively. All links carry equal numbers of wavelengths and typically consist of one fiber per direction; link A $\rightarrow$ B is different than B $\rightarrow$ A.

The selection of the route and the wavelength is an optimization problem known in the literature as routing and wavelength assignment (RWA). RWA is indeed an important process for the efficient operation of a WDM network. Given a network topology and a number of available wavelengths per link, an RWA algorithm returns an optimal solution when it manages to accommodate the connection demands with minimum rejection (blocking). Blocking may occur due to lack of resources or other performance metric. The valuable resource here is the available wavelengths. In the case where the allocated wavelength of a LP remains the same along the traversed links of the path, it is said to satisfy the *wavelength continuity* constraint (see Fig. 3.1). If the optical cross-connects (OXCs) are equipped with wavelength conversion capabilities, then the problem is relaxed to typical routing in a circuit-switched network where the only constraint is the number of channels per link. RWA was introduced more than a decade ago, and numerous works have proposed different solutions to address it as summarized in [6].



Connection Demand	Node Sequence	Wavelength
B→A	B→A	λ2
A→E	A→B→E	λ5
E→C	E→D→C	λ1
E→D	E→D	λ4
E→C	E→F→C	λ3
E→C	E→F→C	λ2

Fig. 3.1 Lightpaths established in a typical WDM network

Most of these works though assume that the optical fiber is a non-impairing transmission medium, and therefore, all RWA solutions are feasible. In reality, as the optical signal propagates inside the fiber, it experiences certain physical effects that attenuate and distort the optical pulses, limiting the reach. Therefore, a selected LP that does not yield acceptable physical-layer performance is not really a valid solution. In the end this LP would not be established as it would fail to be correctly detected at the receiver point because of the high bit-error rate (BER). Hence, to overcome the limitations imposed by the physical-layer impairments (PLIs), QoT is introduced in the RWA process as an additional constraint. To achieve this, typically the impact of the dominant PLIs is mapped to a single figure of merit (e.g. Q-factor) that determines the QoT of a LP, which in turn is considered in the routing and wavelength allocation decision. The term commonly used in the literature to describe the RWA process that always takes into account the PLIs that degrade the quality of the signal is *IA-RWA*. IA-RWA has received great attention by the research community as evident in [7, 8]. The authors in [7] studied and classified more than 100 publications dealing with the IA-RWA problem.

2 Dynamic Core WDM Networks

Besides the first OADMs and OXCs that triggered the vision of transparency, *reconfigurable* add/drop and switching elements were later developed to achieve a higher degree of flexibility. Indeed the early optical nodes were restricted to a fixed configuration that only allowed predetermined LPs to enter or leave a node on predetermined network ports. In the case of a traffic variation, a manual intervention was required to reconfigure the equipment on-site. A core WDM network equipped with reconfigurable OADMs and OXCs may adapt remotely and on demand to the potential traffic changes while eliminating the associated operational expenses [9]. Reconfigurable nodes are rapidly adopted due to their inherent ability to support the dynamic traffic evolution caused by the new bandwidth-intensive multimedia applications. Dynamic optical core networks present a clear business case since the operators do not have to overprovision their network with equipment meant to serve future variations in traffic.

IA-RWA is a resource optimization process that employs a cross-layer technique to minimize the resources and improve the overall network performance. During the *planning* phase of a core WDM network and given an initial traffic matrix, IA-RWA algorithms are used to route all the predetermined connection demands and essentially to dimension the network. IA-RWA applies, nonetheless, also during the *operational* phase of a reconfigurable core network where the demands arrive and depart in a dynamic fashion.

Figure 3.2 illustrates abstractly the two phases of a core WDM network, that is, the planning and the operational phase. During the planning phase, it is assumed that the topology is in place meaning that the operator is aware of the characteristics of the network such as distances, node connectivity, link design and amplifier spans.

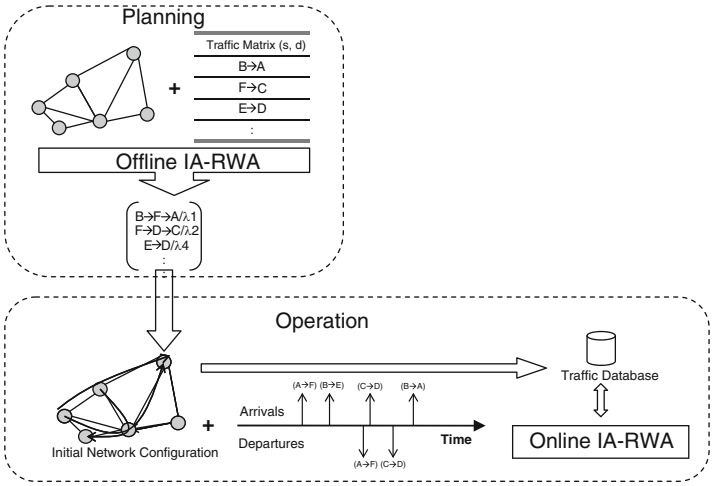


Fig. 3.2 Planning and operational phases of a WDM core network. In the planning phase, the traffic is considered static and the routing is handled by an offline IA-RWA algorithm, whereas in the operational phase, demands arrive and depart at random time instants. Specially designed online dynamic IA-RWA algorithms compute the route and wavelength of each demand on a one-by-one basis as they arrive

Then for a set of connection demands that are known in advance of the actual operation of the network, IA-RWA algorithms compute the paths and allocate the optical channels considering the impact of the PLIs. These IA-RWA algorithms, usually referred to as *offline,* assume that the input traffic is static and optimize the RWA process for the entire set of LPs. In the case where the WDM network is also equipped with regenerators, the appropriate offline IA-RWA is used that also minimizes their use. Since this phase is indeed "offline", it gives the operator the opportunity to revisit their infrastructure needs, as the objective here is to maximize the traffic that is served for a given number of wavelengths or minimize the necessary wavelengths (resources) for a given initial traffic matrix. All in all, the offline case of IA-RWA is a global design and optimization problem, and as it takes place before the operation of the network, the processing time is not an important constraint.

Following the planning phase, the initial configuration of the network is set having established all the predetermined LPs. During the operational phase though, additional demands request connection at random time instances. These connection demands are served as soon as they arrive and in a sequential way. This type of LP provisioning is known as *online* (*or dynamic*) and considers the case where during the network operation demands arrive at any point in time without a priori knowledge. In a dynamic traffic scenario, demands get established upon their arrival, remain established for a finite time period and get released at random points in time. In a core network with reconfigurable optical nodes, dynamic traffic gets served on the fly and remotely. The online dynamic IA-RWA, as opposed to the

offline case, computes a LP on a one-by-one basis taking into account the traffic that is already established in the network but also the current physical-layer parameters. The objective here is to minimize over time the blocking rate with the resources that are available at each moment. In a dynamic core network, existing LPs should not get disrupted by the connections established or torn down over time, unless a physical failure occurs. In that case restoration mechanisms take over and again online IA-RWA is required to compute new routes and channels for the connections that failed.

3 Dynamic IA-RWA

In the context of planning, traffic is considered static, and therein, offline IA-RWA is applied. During operation, LPs have a finite lifetime [10] and may be distinguished in two variants: scheduled dynamic demands and ad hoc dynamic demands (Fig. 3.3). In the case of scheduled demands, the arrival time and lifetime are known in advance and therefore can be pre-planned accordingly. Provisioning of layer 1 can support this type of demands through virtual private networks (VPNs). The second variant includes the demands where their arrival time and duration are not known a priori. Indeed these two parameters are used to describe the traffic load induced in a network over the network's lifetime. Ad hoc dynamic demands are the subject of dynamic IA-RWA. Dynamic IA-RWA may be generally defined as the process that takes into account the current state of the network to compute LPs for demands requesting connections at arbitrary time instants. In what follows, the physical-layer aspect of dynamic IA-RWA is discussed. Next, we examine the algorithmic approach that can be followed to solve the dynamic IA-RWA problem through an overview of related published works. This section closes with the metrics that are used to assess the performance and suitability of dynamic IA-RWA algorithms.

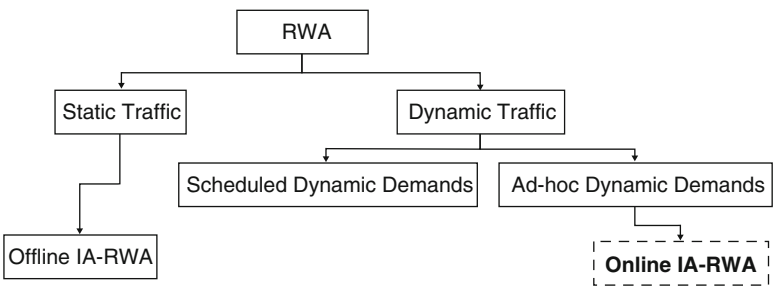


Fig. 3.3 RWA and traffic categories. Dynamic or online IA-RWA deals with traffic demands that arrive at random time instances while the network is already in operation

3.1 Physical-Layer Assessment

In a core optical network it is essential to consider the degradation effects that affect the transmission due to potentially very long paths and high WDM channel count. Indeed IA-RWA incorporates the impact of the PLIs on the QoT and uses this information to make sophisticated decisions to optimize the network in terms of resources and quality of service (QoS). Particularly for online IA-RWA, the corresponding algorithm requires fast yet accurate estimation of the physical degradation, as the process has to be performed on the fly.

The most important PLIs that affect the optical signal as they propagate through the optical fiber medium include the following: amplifier spontaneous emission (ASE) noise, chromatic dispersion (CD), polarization mode dispersion, polarization-dependent loss (PDL), crosstalk (XT), filter concatenation (FC), self-phase modulation (SPM), cross-phase modulation (XPM), four-wave mixing (FWM), stimulated Brillouin scattering (SBS) and stimulated Raman scattering (SRS). These PLIs can be classified into linear (ASE, CD, PDL, XT, FC) and non-linear (SPM, XPM, FWM, SBS, SRS). In the literature, IA-RWA works take into account some of these PLIs, ideally the most dominant in an effort to assess accurately the physical-layer degradations. These works either consider only linear impairments [11–21] or both linear and non-linear [22–26].

Nevertheless, when it comes to dynamic IA-RWA, it is necessary to identify the impairments that are either *single channel or multichannel*. Single-channel effects refer to the physical phenomena that affect every LP individually; LPs suffer also degradations caused by their neighbours due to the multichannel effects present in a WDM system, such as XPM, FWM and XT (Fig. 3.4). This denotes that in a dynamic network environment, a single change in the optical paths (e.g. a new request for optical path establishment) will not only require the performance evaluation of the newly appearing LP in the presence of other active channels but also the performance "re-evaluation" (i.e. recalculation) of all the existing paths in order to determine the impact of the new one on them. This cross-channel dependence sets demanding requirements in terms of computational and processing power.

Hence, the way impairments are quantified plays an important role in the total computation time, and the selection of the appropriate technique needs to provide fast processing for dynamic variations yet also acceptable accuracy. Various approaches are used in literature including analytical modelling, numerical simulations, real-time optical monitoring measurements and experimental data. In the absence of monitors, the impact of random effects that introduces fluctuations at the amplitude levels of the optical pulse such as ASE, XPM and FWM is typically quantified via analytical models [18, 27] that provide individual fast estimations. Deterministic effects, on the other hand, impose distortions that affect the eye opening and may be computed using numerical simulations [28, 29]. In general, feeding the dynamic IA-RWA process with real-time information about the various physical-layer parameters (e.g. power, wavelength utilization) improves the accuracy of the QoT assessment.

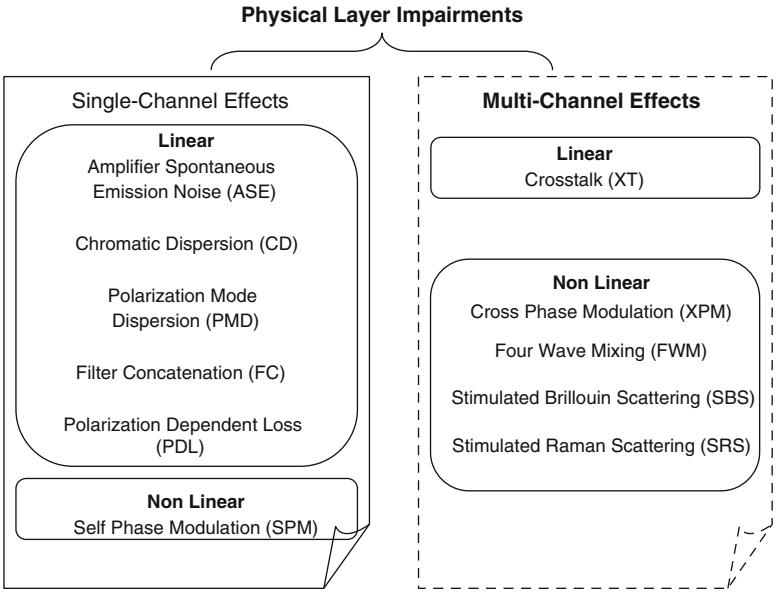


Fig. 3.4 Single and multichannel physical-layer impairment. In dynamic IA-RWA, multichannel effects are particularly important as these determine the impact of a new connection demand on the already established traffic

Following the modelling of all the dominant impairments, an IA-RWA algorithm typically incorporates the PLIs using a formula that combines all of them into a single figure of merit, such as the Q-factor or the optical signal-to-noise ratio (OSNR). Q-factor reflects the quality of a received eye diagram, while OSNR quantifies the impact of all the noise-like impairments. Q-factor seems suitable for the comprehensive assessment of the signal QoT due to its direct relation with bit-error rate (BER) through $\text{BER} = 0.5\ \text{erfc}\left(Q/\sqrt{2}\right)$. Indeed, many works have selected Q-factor [15, 25, 30–35] or its equivalent BER [18, 22, 36–38] as the QoT metric that defines the signal performance in the RWA process. OSNR has been also utilized by several dynamic IA-RWA works to determine the QoT level of a prospect LP [13, 39–41]. Other less common QoT metrics (Table 3.1) include the optical power level [42] and the so-called equivalent length that maps a transparent network element into an equivalent length of fiber based on its losses or contributing noise [16, 19, 26].

3.2 *Algorithmic Approach*

Following the definition of the problem and its conditions, it is important to investigate how dynamic IA-RWA is addressed from an algorithmic point of view. In general, IA-RWA algorithms can be divided into two broad categories:

Table 3.1 QoT metrics used for the assessment of the QoT of a prospective lightpath

QoT metric	References
Q-factor	[15, 25, 30–35]
BER	[18, 22, 36–38]
OSNR	[13, 39–41]
Power level	[42]
Equivalent length	[16, 19, 26]

the algorithms that follow a *sequential* approach for solving the problem and the ones that follow a *combinatorial* approach. The sequential approach employs heuristic or meta-heuristic algorithms that usually provide suboptimal solutions, whereas the combinatorial approach seeks an optimal solution.

The pure RWA problem, that is, without considering the physical-layer constraints, is NP-complete. Therefore, the optimal solution cannot be found in polynomial time. On the other hand, IA-RWA is also constrained by the physical-layer parameters becoming a problem with additional computational complexity. To overcome this issue, the RWA problem may be decomposed into two smaller problems, and each subproblem solved separately: one dedicated to the computation of the route and the other to the appropriate selection of the wavelength. For each of these subproblems, a two-step process may be followed. The first step determines the set of candidate routes/wavelengths, and the second decides the optimum among the given set. As already mentioned, online IA-RWA has to consider only the resources that are available at a given moment. Therefore, the set of candidate routes/wavelengths cannot contain the LPs that are already established. Alternatively the use of a heuristic or meta-heuristic for the solution of each of the two subproblems lacks optimization capability yet offers advantages in terms of computation time. Finding a solution in reasonable amount of time is of utmost importance in online IA-RWA.

Indeed because of additional complexity imposed by the models of the PLIs, online IA-RWA algorithms most frequently resort to simple *heuristic* methods [22, 35, 43–45]. The various heuristic algorithms proposed in the literature typically base their solution to the routing subproblem on a shortest path algorithm (e.g. Dijkstra algorithm). Then, among those, two classes of algorithms can be identified, the ones that follow a *single-path* method and the ones that follow a *multipath* method, also known as *k*-shortest path. Each network link in a shortest path-based algorithm is characterized by a cost (or weight) parameter that is used to find a minimum cost solution. The link cost may refer to the physical length of the link, or it may correspond to any parameter related to the link state including physical-layer information.

From the algorithms based on single-path routing, a number of them follow the minimum hop shortest path approach as reported in [18, 22, 35, 46]. On the other hand, other works use a PLI-related cost to compute the shortest path, such as a function of the link residual dispersion in [46], crosstalk in [24] or the noise variance in [47]. In this way the concept of impairment awareness is incorporated into the path computation process. The algorithms that follow the multipath approach compute a set of candidate paths, usually the *k*-shortest, and similar to

Table 3.2 Heuristic algorithms for the solution of the dynamic IA-RWA problem

Subproblem	References
Routing	Single path [18, 22, 24, 35, 46, 47] Multipath [10, 20, 25, 36, 48]
Wavelength assignment	First Fit (FF) [10, 15, 20, 23, 36, 44, 46, 47, 50] Best fit [24]
Joint routing + wavelength assignment	[39, 45, 51, 52]

the single-path case use either the hop count as a cost metric [10, 20] or a physical-layer-related parameter [48]. Distance may be also considered as a simple PLI parameter, as it indirectly refers to the impairments that are proportional to the physical length of the path (e.g. ASE, PMD). The authors in [25, 36, 48] apply a more comprehensive link cost, employing a Q-factor penalty approach.

The selection of an appropriate path among the set of candidates may be performed in a *sequential* [20, 49] or a *parallel* manner [44]. The former implies that a sequence of attempts is performed and the first available path that meets the given performance requirements is selected. In the parallel case, all the candidate paths are examined, and the optimal according to a certain criterion is chosen.

Following the solution of the routing subproblem, the list of available wavelengths needs to be determined. This list may be ordered randomly or according to a certain policy. In [22] the authors initially considered the wavelengths that are most separated in the frequency domain among the available ones, and then these were ordered so as to maximize the frequency separation. Similar to the path selection phase, the appropriate wavelength is selected either sequentially or in parallel. In the sequential approach, the first available wavelength is selected according to the decided physical-layer or network-layer criteria. The well-known first-fit (FF) method has been employed in several IA-RWA algorithms [10, 15, 20, 23, 36, 44, 46, 47, 50]. Alternatively an IA-RWA algorithm can perform an exhaustive search among the candidate wavelengths to find the best fit. A best fit may be decided according to PLI-related criteria. Indeed in one of the algorithms presented in [24], the wavelength that minimizes the FWM interference is selected.

Apart from solving separately the routing and wavelength assignment problems, there are heuristics where the IA-RWA problem is addressed jointly (Table 3.2). Such an example is the dynamic IA-RWA algorithm presented in [45]. Therein, a shortest path algorithm is used to find the candidate routes where the cost is the link distance. For each candidate route, the number of crosstalk components along the route is calculated. Among all the candidate routes, it chooses the route and the wavelength with the minimum crosstalk intensity. Other works considering a joint IA-RWA process include the following [39, 51, 52].

In general the IA-RWA methods that were discussed up to this point use a single parameter to represent the link cost in the path computation process. This cost may be a function of more arguments, yet it is a scalar. These algorithms are referred to as *single cost*. In the *multi-cost* approach, a vector of different cost parameters is utilized and assigned to each link. Multi-cost routing algorithms were proposed as a

solution to the limited performance of single-cost algorithms with respect to general cost functions and their inability to support QoS differentiation. The cost vector in a multi-cost approach may include PLI and other network parameters that uniquely characterize the link, allowing thus the PLIs to be handled differentially and efficiently. The authors in [31] developed a multi-cost online IA-RWA algorithm that directly accounts for the PLIs. The vector parameters used by the algorithm are the variances of the random PLIs on a per-link basis that can be combined to estimate the QoT of a LP in a fast and efficient manner. Another multi-cost online algorithm in [32] considers indirectly the PLIs such as the path length, the hop count (filtering effects) and the number of adjacent channels (multichannel effects).

Besides the heuristic-based algorithms that are particularly appropriate for the routing of traffic during the operational phase of a network, there is a class of IA-RWA in literature that employs *meta-heuristic* methods. Popular meta-heuristic approaches are based on genetic algorithms (typically for static traffic [53, 54]) and ant colony optimization [42], whose common principle is the convergence to an optimum solution through successive iterations.

A certain issue particularly important in the online case of RWA is the way the existing connections are taken into account in terms of the effect they have on a new connection but also the effect the new connection has on them. As mentioned already in Sect. 3.3.1, the multichannel PLIs are responsible for this cross-channel dependence. In [43, 45], this issue is addressed by considering the crosstalk due to the already established LPs. The approach proposed in [22] utilizes the BER of the candidate and affected connections to allow the LP establishment. In a similar manner, the multi-cost algorithms in [31, 32] estimate the Q-factor of the candidate and existing LPs and check whether all pass a predetermined Q-factor threshold. If not, then the candidate LP is not granted admission, and another LP has to be computed.

To summarize the different approaches used in literature to account for the PLIs in the online IA-RWA problem, three different cases can be identified: (a) route computation and wavelength assignment without considering PLIs (pure RWA) and in the end verification of the QoT of the LPs (new and existing) to proceed to LP establishment, (b) consider PLIs in the routing and/or the wavelength assignment processes and (c) consider PLIs in the routing and/or the wavelength assignment processes and in the end also perform QoT verification (Fig. 3.5).

Examples of the first case are proposed in [10, 35, 44]. Cases (b) and (c) may be further divided to sub-cases where either both the routing and wavelength assignment processes account for the PLI constraints or only one of the two accounts for the PLI constraints. For instance, in case (b), [46] follows a PLI-constrained path computation process, yet the wavelength is decided without considering the PLIs. Zhai et al. [45] proposes an approach where the wavelength is decided taking into account the crosstalk intensity, and in [21, 36, 43], PLIs are incorporated in both the routing and wavelength assignment phases. The third group of algorithms is a combination of the first two cases. Again here the physical-layer constraints are incorporated in the routing phase or the wavelength assignment phase or both but include also a final QoT verification step that allows the algorithm to seek for an alternative LP [22, 47, 48].

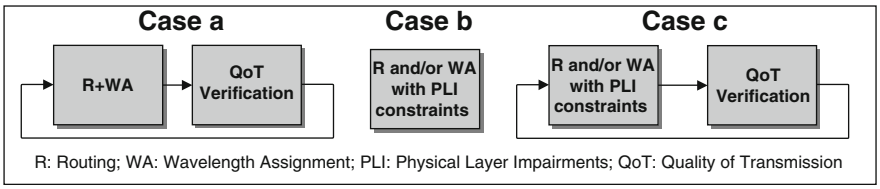


Fig. 3.5 Three cases are identified when considering the physical-layer impairments in the algorithmic approach

3.3 Formulation of an Online IA-RWA Algorithm

Following the discussion about the various algorithmic approaches, we present here a multi-constraint online IA-RWA algorithm which uses a single mixed cost metric per link and employs a typical source–destination shortest path routing algorithm, such as Dijkstra, to select the connection's path [30].

In a multi-constraint case, each network link has multiple weights, where the multi-constraint path (MCP) problem can be defined as a routing problem that seeks a path that satisfies a number of (additive) constraints. However, in the multi-constraint IA-RWA algorithm presented here, an MCP engine with a single mixed metric (SMM) is used, instead of dealing with multiple link weights. Using a SMM reduces the algorithmic complexity that is valuable for real-time operation.

A network can be represented by a directed or undirected graph $G = (V, E)$, where V is the set of nodes (vertices) and E is the set of links (edges). It is assumed that the network graph is connected, meaning that at least one path exists between each pair of nodes in the network. For a certain source S and destination D, we let Π_{SD} be the set of all paths between S and D. The MCP problem can then be formulated as follows.

Consider a network topology $G = (V, E)$, a source node S and a destination node D. Each link $e \in E$ is characterized by M additive non-negative weights, $w_m(e),\ m = 1, 2, \ldots, M$. Given constraints $C_m,\ m = 1, 2, \ldots, M$, find a path p such that

$$\sum_{e \in p} w_m(e) < C_m; \quad m = 1, 2, \ldots, M. \tag{3.1}$$

The cost metric that is used here and that transforms the MCP problem into a problem with a SMM is

$$\mathrm{SMM}_d(e) = \mu_d(e)\ [\Delta_d(e) + \varepsilon]; \quad 0<\varepsilon<1, \tag{3.2}$$

where

$$\mu_d(e) = \frac{1}{M} \sum_{m=1}^{M} \left(\frac{w_m(e)}{C_m} \right)^d; \quad d \geq 1, \tag{3.3}$$

$$\Delta_d(e) = \sum_{m=1}^{M} \left[\left(\frac{w_m(e)}{C_m} \right)^d - \mu_d(e) \right]^2. \tag{3.4}$$

This relation considers the impact of both mean and variance of normalized weights in a single mixed metric. The contribution of mean [as defined in Eq. (3.3)] is controlled with parameter e. The novelty of this approach is the exploitation of the single-cost metric for k-shortest path or diverse k-shortest path algorithms. Indeed authors in [55, 56] modified the Dijkstra algorithm to find a single path between the source and destination node, while this algorithm is using the MCP framework to compute a set of candidate paths between source and destination nodes for a dynamic demand.

Regarding the link cost vector, there is no particular limit for the number of constraints in this proposed algorithm. Various constraints may be considered as the link cost parameters. For instance, the physical link length could be one of them as it indirectly accounts for the physical impairments whose impact is proportional to the physical distance. Another constraint may be related to the PMD through the average squared differential group delay. PMD management requires the time-average differential time delay Δt between the two orthogonal states of the polarization to be less than a fraction α of the bit duration $T = (1/B)$, where B is the bit rate. A typical value that can be used for the fraction α is 0.1 (10 %). Suppose that the transparent segment consists of M fiber spans, where the kth span has length $L(k)$, and fiber PMD parameter $D_{\mathrm{PMD}}(k)$. Then the constraint on the average differential delay can be expressed as $B\sqrt{\sum_{k=1}^{M} D_{\mathrm{PMD}}^2(k)L(k)} < \alpha$. Another element may be the "static" Q-factor estimation. The term static refers to the single-channel impairments that do not depend on the utilization of the network (Sect. 3.3.1). The main idea here is to consider the static physical impairments inside the MCP computation engine in order to exclude paths that do not satisfy the minimum QoT requirements (as far as these physical impairments are concerned).

Having introduced the problem definition and the way constraints may be considered, the structure of the algorithm is presented hereafter (Fig. 3.6). Upon the arrival of a connection request, the current network topology is decomposed into W layers (wavelength planes), where W is the total number of available wavelengths in each fiber. For each wavelength plane, the algorithm computes the candidate paths that satisfy the multiple constraints using the MCP framework. These candidate paths correspond to a set of candidate LPs, based on the wavelength planes each path belongs to. This way the candidate LPs also conform to the wavelength availability constraint.

The next step is to construct another set of LPs, the "usable" LPs. In particular, each candidate LP is temporarily added to the set of currently established LPs, and the impact of this addition on the overall QoT is computed (i.e. the QoT of all established LPs at the moment the new connection demand arrived). If all QoT values are above a certain threshold, the candidate LP under consideration will be

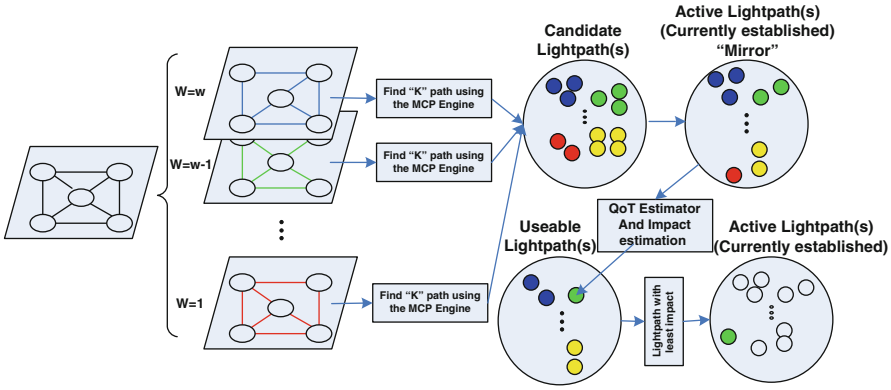


Fig. 3.6 The structure of the online IA-RWA algorithm in [30]. The algorithm is based on a multi-constraint framework that utilizes a single mixed cost metric

moved to the "useable" LP set. The final step of the algorithm is to select the best useable LP. In order to find this LP, we select the one that introduces the minimum impact on the currently established LPs. In the case the "useable" set is empty, the demand will be blocked.

3.4 Performance Metrics

In order to evaluate a typical IA-RWA, whether offline or online, the metric commonly used is the number of blocked connections. However, in a dynamic scenario it is useful to compute the *blocking ratio* with respect to the induced traffic load, whereas in a static scenario the blocking ratio is studied with respect to the amount of the required resources (i.e. wavelengths, regenerators, fibers). Besides, during the operational phase, the network is already planned, and therefore, the objective is to minimize the blocked connections over an infinite amount of time.

Furthermore, to assess the performance of an IA-RWA algorithm in the context of online operation, it is essential to compute the average *execution time* [20] of a given connection demand as this will contribute to the total LP establishment delay. Since the physical-layer awareness implies complex models or formulations, the execution time should be taken into consideration when designing a dynamic IA-RWA algorithm. A case in point is [31] where the authors show the average execution time per connection with respect to the traffic load, and in [33] the average execution time is reported as a function of the number of available wavelengths for fixed traffic load.

Figures 3.7 and 3.8 illustrate the results of an evaluation of the multi-constraint algorithm briefly presented in Sect. 3.3.3 and the k-SP-Q algorithm [57] in terms of the blocking rate and the average execution time per connection. The k-SP-Q algorithm computes the k-shortest available routes between each source and destination pair and utilizes the first-fit wavelength assignment scheme to allocate a

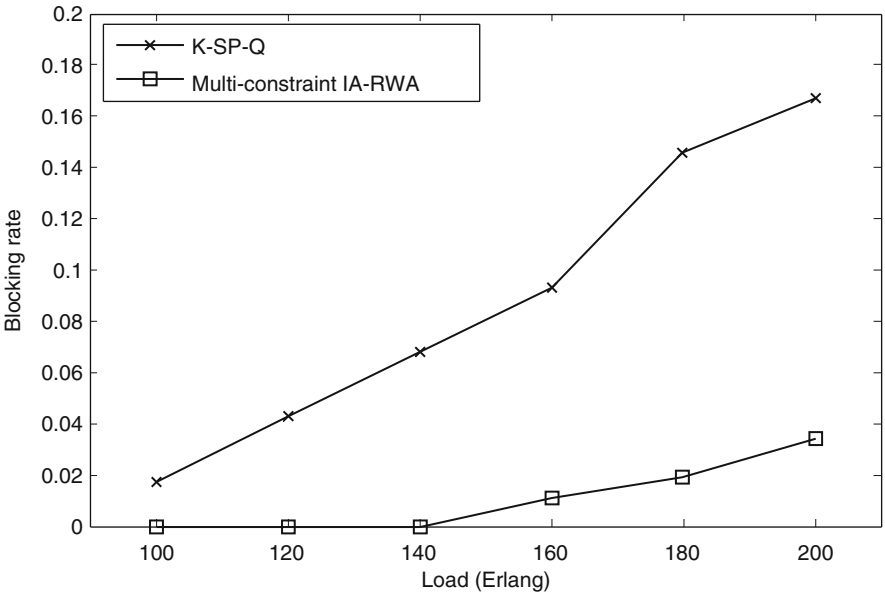


Fig. 3.7 Blocking rate vs. traffic load for fixed number of wavelengths

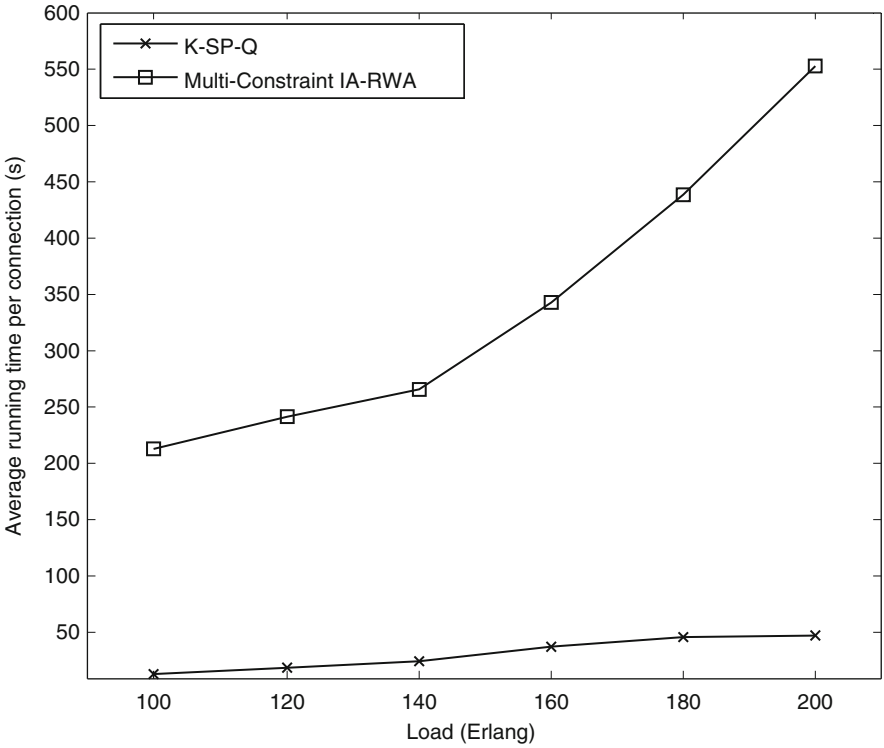


Fig. 3.8 Execution time vs. traffic load

wavelength to the shortest available route among the k; a verification step checks the QoT value of the decided LP. For the performance evaluation, the case where $k = 5$ was simulated. The two online algorithms were studied for dynamic traffic scenarios with "arrivals" and "departures". One thousand requests for 10 Gbps connections were generated according to a Poisson process with a rate of λ requests/time unit. The source and destination of the connection requests were uniformly selected among the network nodes. The lifetime of a connection was given by a random variable with exponential distribution and average value $1/\mu$. The ratio λ/μ expresses the total traffic load (in Erlangs) that in this study varied from 100 to 200.

Figure 3.7 depicts the performance of the two algorithms for increasing values of the traffic load and a fixed number of wavelengths W. As expected, increasing the load leads to higher blocking rate for both algorithms, yet for the case of the multi-constraint algorithm, it remains under 4 % even for the highest assumed traffic load. The incorporation of the physical-layer parameters in the LP computation certainly allowed the established LPs to be less prone to blocking due to low QoT. On the other hand, it affected the computation complexity of the algorithm. As shown in Fig. 3.8, the average running time (in absolute value on a specific computer platform) of the multi-constraint algorithm is significantly longer than the average running time of k-SP-Q.

Hence, as demonstrated in this exercise, both the execution time and the blocking rate are crucial for the selection of a suitable online IA-RWA algorithm. Focusing on one or the other is not enough to make an educated decision.

4 IA-RWA and Control Plane

Dynamic IA-RWA relies heavily on a control plane that together with the algorithm enables the impairment-aware LP provisioning. The goal of this section is not to describe the various control plane approaches but to discuss their role in correlation with the IA-RWA in an operational network.

Recently, the adoption of the GMPLS framework developed by the Internet Engineering Task Force (IETF) seems to prevail as an effective solution for controlling an optical network. Although GMPLS offers standardized protocols for signalling, routing and resource management, it lacks the ability to disseminate throughout the network the physical-layer-related information that is necessary for the impairment-aware networking. As a consequence, the idea of encompassing the PLIs into the control plane functionalities has attracted the attention of the research community [28, 58–60].

Mainly two different approaches may be followed to incorporate the IA-RWA into the control plane, centralized or distributed. A *centralized* solution implies the availability of a central point of control that is aware of the complete network topology, the available resources and the PLI-related information [10, 25, 33, 46]. This element is responsible for collecting and updating all this information and is accessible by all network nodes. It performs the IA-RWA and due to the centralized

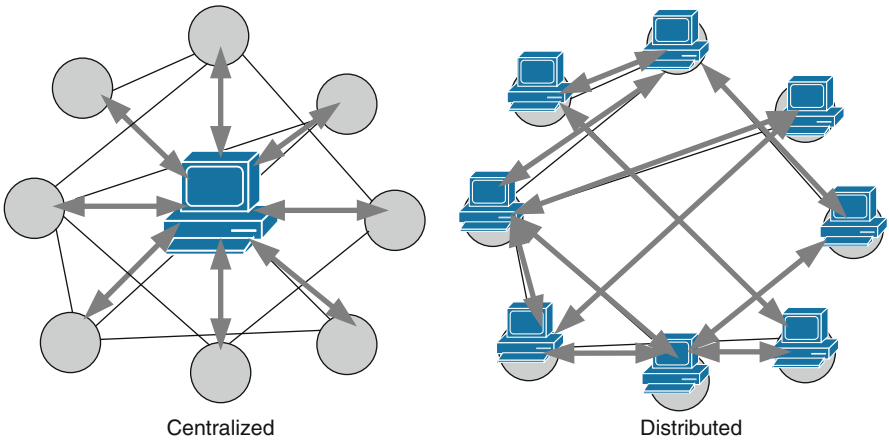


Fig. 3.9 Centralized and distributed approaches for dynamic IA-RWA. Dynamic IA-RWA relies heavily on a control plane that has to be properly enhanced to support the physical-layer impairments

architecture ensures that the LP computation considers all up-to-date information. In a centralized solution, the routing protocol needs to be properly extended to carry the PLI information. The central element may be either a Path Computation Element or the Network Management System [60].

In the *distributed* approach each node is responsible for computing, establishing and maintaining LPs [21–23, 35, 42]. Nevertheless, the nodes are deprived of global data knowledge. They collect information on the available resources through the routing protocol, run the RWA process and establish a new connection using the signalling protocol (Fig. 3.9). Extensions to the signalling and/or the routing protocol are necessary here to introduce the impairment-awareness feature [59, 61].

In [28, 58, 60] the authors studied and compared both configurations. Particularly in [28] a multiplane experimental test-bed was utilized to evaluate the performance of the two schemes under the same dynamic traffic conditions. In the centralized case, the central element (PCE) used a specially designed online IA-RWA algorithm that incorporated PLI constraints into the RWA step and also QoT verification in the end (case c in Sect. 3.3.2) to perform the path computation. On the contrary, the distributed scheme utilized a method of case (a) where the source node of a given connection demand computes k-candidate routes based on the available resources. The QoT verification step is performed by the destination node and all the destination nodes of potentially affected LPs. The PLI information is collected from source to destination along a candidate route using a properly extended signalling GMPLS protocol. The authors measured the overall LP setup delay that included both the RWA process and also the explicit steps of the establishment process. The distributed scheme yielded shorter setup times taking benefit of the parallel LP establishments. However, the centralized scheme demonstrated lower blocking ratio that is justified by the centralized nature of

Table 3.3 Employed control plane schemes

Control plane approach	References
Centralized	[10, 25, 33, 46]
Distributed	[21–23, 35, 42]

this configuration since the IA-RWA can have a complete picture of the traffic and physical-layer parameters (Table 3.3).

5 Summary

This chapter dealt with the topic of dynamic IA-RWA. Dynamic IA-RWA applies in core optical WDM networks, computing and assigning LPs to the connection demands that arrive during the operational phase. The dynamic (or online) counterpart of RWA implies that requests for traffic arrive in the network at random time instants. New demands are served as soon as they arrive and one at a time. Upon their arrival, traffic already established in the network must not be disrupted.

IA-RWA takes into account the QoT of the LPs in an effort to minimize the blocked connections and ensure quality of service. Besides, in WDM systems multichannel physical effects impose additional complexity in the solution of the RWA problem. These effects essentially stem from one optical signal and affect their neighbours. Therefore, the establishment of a new connection may deteriorate the QoT of the existing LPs to an unacceptable level. The way to assess the PLIs was discussed in Sect. 3.3.1

Following the definition and the conditions of the problem, the algorithmic approach that can be followed in IA-RWA algorithms was discussed through a comprehensive survey of related works in Sect. 3.3.2. Heuristics or meta-heuristics, single shortest path or multiple shortest paths and single cost or multi-cost are only few of the questions to be answered when developing an online IA-RWA algorithm. Furthermore, IA-RWA algorithms that have been proposed in the literature utilize different techniques to incorporate the impairments. What is critical for the online application of IA-RWA is to minimize the execution time needed per connection that is largely dependent on the way PLIs are considered. Online IA-RWA has to be able to assign LPs "on the fly" but at the same time ensure acceptable QoT in order to minimize the blocking ratio.

Finally, in an operational core network, IA-RWA is supported by a properly enhanced control plane that essentially enables the dynamic impairment-aware networking. An IA-RWA algorithm may be implemented either for a centralized or a distributed control plane whose protocols are extended to disseminate the valuable physical-layer information.

References

1. Nosu K, Iwashita K (1987) Coherent FDM transmission techniques. In: Optical fiber communication, 1987 OSA Technical Digest Series, Optical Society of America, paper TUG2
2. Berthold J, Saleh AAM, Blair L, Simmons JM (2008) Optical networking: past, present, and future. J Lightwave Technol 26:1104–1118
3. Saleh A (1998) Islands of transparency—an emerging reality in multiwavelength optical networking. Presented at the IEEE/LEOS summer topical meeting broadband optical networks technologies, Monterey, July 1998
4. Simmons JM (2005) On determining the optimal optical reach for a long-haul network. J Lightwave Technol 23(3):1039–1048
5. Sygletos S, Tomkos I, Leuthold J (2008) Technological challenges on the road toward transparent networking. IEEE J Opt Netw 7(4):321–350
6. Zang H, Jue JP, Mukherjee B (2000) A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks. Opt Netw Mag 1:47–59
7. Azodolmolky S, Klinkowski M, Marin E, Careglio D, Solé-Pareta J, Tomkos I (2009) A survey on physical layer impairments aware routing and wavelength assignment algorithms in optical networks. Comput Netw 53(7):926–944
8. Saradhi CV, Subramaniam S (2009) Physical layer impairment aware routing (PLIAR) in WDM optical networks: issues and challenges. IEEE Commun Surv Tutor 11(4):109–130
9. Ruffini M et al (2008) Cost study of dynamically transparent networks. In: Proceedings of the OFC/NFOEC, paper OMG2
10. Ali Ezzahdi M, Al Zahr S, Koubaa M, Puech N, Gagnaire M (2006) LERP: a quality of transmission dependent heuristic for routing and wavelength assignment in hybrid WDM networks. In: Proceedings of the ICCCN, Arlington, October 2006, pp 125–136
11. He J, Brandt-Pearce M, Subramaniam S (2009) QoS-aware wavelength assignment with BER and latency constraints for all-optical networks. J Lightwave Technol 27:462–474
12. He J, Brandt-Pearce M, Subramaniam S (2007) QoS-aware wavelength assignment with BER and Latency Guarantees for Crosstalk Limited Networks. In: Proceedings of the IEEE international conference on communications, Glasgow, June 2007
13. Huang Y, Wen W, Heritage J, Mukherjee B (2003) Signal-quality consideration for dynamic connection provisioning in all-optical wavelength-routed networks. In: Proceedings of the SPIE OptiComm: optical networking and communications
14. Pointurier Y, Brandt-Pearce M, Subramaniam S (2007) Analysis of blocking probability in noise and crosstalk impaired all-optical networks. In: Proceedings of the IEEE INFOCOM, Anchorage, May 2007, pp 2486–2490
15. Pointurier Y, Brandt-Pearce M, Subramaniam S, Xu B (2008) Crosslayer adaptive routing and wavelength assignment in all-optical networks. IEEE J Sel Areas Commun 26(6):32–44
16. Pinart C, Sambo N, Rouzic E, Cugini F, Castoldi P (2011) Probe schemes for quality-of-transmission-aware wavelength provisioning. J Opt Commun Netw 3(1):87–94
17. Rahbar AG (2010) Dynamic impairment-aware RWA in multifiber wavelength-routed all-optical networks supporting class-based traffic. J Opt Commun Netw 2(11):915–927
18. Ramamurthy B, Datta D, Feng H, Heritage JP, Mukherjee B (1999) Impact of transmission impairments on the teletraffic performance of wavelength-routed optical networks. J Lightwave Technol 17(10):1713–1723
19. Sambo N, Pinart C, Le Rouzic E, Cugini F, Valcarenghi L, Castoldi P (2009) Signaling and multi-layer probe-based schemes for guaranteeing QoT in GMPLS transparent networks. In: Proceedings of the IEEE OFC
20. Yang X, Ramamurthy B (2005) Dynamic routing in translucent WDM optical networks the intradomain case. J Lightwave Technol 23(3):955–971
21. Zulkifli N, Okonkwo C, Guild K (2006) Dispersion optimized impairment constraint based routing and wavelength assignment algorithms for all-optical networks. In: Proceedings of the ICTON, vol 3, Nottingham, June 2006, pp 177–180

22. He J, Brandt-Pearce M, Pointurier Y, Subramaniam S (2007) QoT-aware routing in impairment-constrained optical networks. In: Proceedings of the IEEE GLOBECOM, Washington, November 2007
23. Lin W, Hahn T, Wolff RS, Mumey B (2011) A distributed impairment aware QoS framework for all-optical networks. Opt Switch Netw 8(1):56–67
24. Marsden A, Maruta A, Kitayama K (2008) Routing and wavelength assignment encompassing FWM in WDM lightpath networks. In: Proceedings of the IFIP ONDM, Vilanova i la Geltrú, March 2008
25. Markidis G, Sygletos S, Tzanakaki A, Tomkos I (2007) Impairment-constraint-based routing in ultralong-haul optical networks with 2R regeneration. IEEE Photon Technol Lett 19 (6):420–422
26. Sambo N, Cugini F, Cerutti I, Valcarenghi L, Castoldi P, Poirrier J, Le Rouzic E, Pinart C (2008) Probe-based schemes to guarantee lightpath Quality of Transmission (QoT) in transparent optical networks. In: European conference and exhibition on optical communication (ECOC), Brussels
27. Pachnicke S, Reichert J, Spalter S, Voges E (2006) Fast analytical assessment of the signal quality in transparent optical networks. J Lightwave Technol 24(2):815
28. Azodolmolky S, Perello J, Angelou M, Agraz F, Velasco L, Spadaro S, Pointurier Y, Francescon A, Vijaya Saradhi C, Kokkinos P, Varvarigos E, Al Zahr S, Gagnaire M, Gunkel M, Klonidis D, Tomkos I (2011) Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks. J Lightwave Technol 29:439–448
29. Leibrich J, Rosenkranz W (2003) Efficient numerical simulation of multichannel WDM transmission systems limited by XPM. IEEE Photon Technol Lett 15(3):395–397
30. Azodolmolky S, Pointurier Y, Angelou M, Careglio D, Sole-Pareta J, Tomkos I (2011) A novel impairment aware RWA algorithm with consideration for QoT estimation inaccuracy. J Opt Commun Netw 3:290–299
31. Christodoulopoulos K, Manousakis K, Varvarigos E, Angelou M, Tomkos I (2009) A multicost approach to online impairment-aware RWA. In: IEEE international conference on communications
32. Kokkinos P, Christodoulopoulos K, Manousakis K, Varvarigos E (2009) Multi-parametric online RWA based on impairment generating sources. In: Proceedings of the IEEE Globecom, Honolulu
33. Manousakis K, Kokkinos P, Christodoulopoulos K, Varvarigos E (2010) Joint online routing, wavelength assignment and regenerator allocation in translucent optical networks. J Lightwave Technol 28(8):1152–1163
34. Pointurier Y, Brandt-Pearce M, Deng T, Subramaniam S (2006) Fair QoS-aware adaptive routing and wavelength assignment in all optical networks. In: Proceedings of the IEEE international conference on communications, Istanbul, pp 2433–2438
35. Pachnicke S, Paschenda T, Krummrich PM (2008) Physical impairment based regenerator placement and routing in translucent optical networks. In: Proceedings of the OFC/NFOEC, San Diego, February 2008
36. Deng T, Subramaniam S (2005) Adaptive QoS routing in dynamic wavelength-routed optical networks. In: Proceedings of the 2nd international conference on broadband networks, vol 1, October 2005, Boston, MA, USA, pp 184–193
37. Huang Y, Gencata A, Heritage J, Mukherjee B (2002) Routing and wavelength assignment with quality-of-signal constraints in WDM networks. In: European conference on optical communications, Copenhagen
38. Pinart C, Le Rouzic E, Martnez I (2007) Physical-layer considerations for the realistic deployment of impairment-aware connection provisioning. In: Proceedings of the IEEE international conference on transparent optical networks, Rome, pp 134–137
39. Cardillo R, Curri V, Mellia M (2005) Considering transmission impairments in wavelength routed networks. In: Proceedings of the ONDM, Milan, pp 421–429

40. Huang Y, Heritage J, Mukherjee B (2005) Connection provisioning with transmission impairment consideration in optical WDM networks with high-speed channels. J Lightwave Technol 23(3):982–993
41. Maranhao J, Soares A, Waldman H (2010) Wavelength assignment in optical networks considering physical impairments. In: Proceedings of the IEEE ICTON, Munich
42. Pavani GS, Zuliani LG, Waldman H, Magalhães MF (2008) Distributed approaches for impairment-aware routing and wavelength assignment algorithms in GMPLS networks. J Comput Netw 52(10):1905–1915
43. Pointurier Y, Brandt-Pearce M (2006) Fair routing and wavelength assignment in all optical networks. In: Proceedings of the OFC/NFOEC, Anaheim, March 2006
44. Salvadori E, Ye Y, Zanardi A, Woesner H, Carcagni M, Galimberti G, Martinelli G, Tanzi A, La Fauci D (2007) Signalling-based architectures for impairment-aware lightpath set-up in GMPLS networks. In: Proceedings of the IEEE GLOBECOM, Washington, November 2007, pp 2263–2268
45. Zhai Y, Pointurier Y, Subramaniam S, Brandt-Pearce M (2007) QoS-aware RWA algorithms for path-protected DWDM Networks. In: Proceedings of the OFC/NFOEC, Anaheim, March 2007
46. Carpenter TJ, Menendez RC, Shallcross DF, Gannett JW, Jackel J, Von Lehmen AC (2004) Cost-conscious impairment-aware routing. In: Proceedings of the IEEE/OSA OFC, vol 1, San Diego, February 2004
47. He J, Brandt-Pearce M, Pointurier Y, Subramaniam S (2007) Adaptive wavelength assignment using wavelength spectrum separation for distributed optical networks. In: Proceedings of the IEEE international conference on communication, Glasgow, June 2007, pp 24–28
48. Pachnicke S, Paschenda T, Krummrich P (2008) Assessment of a constraint based routing algorithm for translucent 10 Gbits/s DWDM networks considering fibre nonlinearities. J Optical Netw 7(4):365
49. Cugini F, Andriolli N, Valcarenghi L, Castoldi P (2004) A novel signaling approach to encompass physical impairments in GMPLS networks. In: Proceedings of the IEEE GLOBECOM workshops, Dallas, November 2004, pp 369–373
50. Wang L, Zhang J, Gao G, Chen X, Gu W (2009) Noise-aware wavelength assignment for wavelength switched optical networks. In: Proceedings of the IEEE international conference on communication, Dresden
51. Cardillo R, Curri V, Mellia M (2006) Considering transmission impairments in configuring wavelength routed optical networks. In: Proceedings of the IEEE/OSA OFC/NFOEC, Anaheim
52. Gurzi P, Steenhaut K, Nowe A (2011) Minimum cost flow based R&WA algorithm for dispersion and OSNR limited all-optical networks. In: Optical network design and modeling (ONDM), Bologna
53. Monoyios D, Vlachos K (2011) Multiobjective genetic algorithms for solving the impairment-aware routing and wavelength assignment problem. J Opt Commun Netw 3:40–47
54. Lima M, Cesar A, Araujo A (2003) Optical network optimization with transmission impairments based on genetic algorithm. In: Proceedings of the IEEE international microwave and optoelectronics conference, Brazil, pp 361–365
55. Khadavi P, Samavi S, Todd TD, Saidi H (2004) Multi-constraint QoS routing using a new single mixed metric. In: Proceedings of the IEEE international conference on communication, vol 4, 20–24 June 2004, pp 2042–2046
56. Khadavi P, Samavi S, Todd TD (2008) Multi-constraint QoS routing using a new single mixed metric. J Netw Comp Appl 4:656–676
57. Yang X, Shen L, Ramamurthy B (2005) Survivable lightpath provisioning in WDM mesh networks under shared path protection and signal quality constraints. J Lightwave Technol 23 (4):1556–1567
58. Castoldi P, Cugini F, Valcarenghi L, Sambo N, Le Rouzic E, Poirrier MJ, Adriolli N, Paolucci F, Giorgetti A (2007) Centralized vs. distributed approaches for encompassing physical

impairments in transparent optical networks. In: Tomkos I, Neri F, Solé-Pareta J, Masip-Bruin X, Sánchez-López S (eds) Optical network design and modeling, Lecture Notes in Computer Science, vol 4534. Springer, Berlin
59. Cugini F, Sambo N, Andriolli N, Giorgetti A, Valcarenghi L, Castoldi P, Le Rouzic E, Poirrier J (2008) Enhancing GMPLS signaling protocol for encompassing quality of transmission (QoT) in all-optical networks. J Lightwave Technol 26:3318–3328
60. Martinez R, Pinart C, Cugini F, Andriolli N, Valcarenghi L, Castoldi P, Wosinska L, Comellas J, Junyent G (2006) Challenges and requirements for introducing impairment-awareness into the management and control planes of ASON-GMPLS WDM networks. IEEE Commun Mag 44(12):76–85
61. Cugini F, Andriolli N, Valcarenghi L, Castoldi P (2005) Physical impairment aware signalling for dynamic lightpath set up. In: Proceedings of the ECOC, vol 4, Glasgow, September 2005, pp 979–980

Chapter 4
Routing and Wavelength Assignment in WDM Networks with Mixed Line Rates

Avishek Nag, Massimo Tornatore, Menglin Liu, and Biswanath Mukherjee

1 Introduction

Telecommunication networks are under constant change over the past few years to support the need for enormous bandwidth. Network operators and service providers are upgrading the networks in all its hierarchies, viz., the backbone, the metro and the access networks. The bandwidth demand especially in backbone networks is huge as it is the "big ocean" where internet traffic from different sources flow. Therefore to support such large volume of traffic, the capacities of information carrying channels (which is essentially a wavelength channel in an optical fiber) in the backbone networks are getting raised from 10 to 40 Gbps and also to 100 Gbps. But the question is: Is 100 Gbps always better? Surely, such a high data rate is desirable because it can carry a huge amount of traffic, but at such high bit rates, the signal impairments significantly limit the reach of the regenerator-free optical distance. So, in the context of wavelength-division multiplexed (WDM) backbone telecommunication networks, increasing the capacity of a wavelength to 100 Gbps presents a tradeoff between capacity and reach. In fact, using 100 Gbps everywhere in a backbone network will require expensive equipment to have increased reach, i.e., either expensive transmission equipment will be required or more expenditure on regenerators should be required [1]. Moreover, traffic demands across the network may not require 100 Gbps capacity everywhere. So, a cost-effective network design needs to exploit the possibility of mixed line rates (MLRs) [2–6].

In such a network, the wavelength channels can have a variety of capacities (10/40/100 Gbps), and the high-bit-rate transceivers (viz., 40/100 Gbps) can exploit

A. Nag (✉) • M. Liu • B. Mukherjee
University of California, Davis, USA
e-mail: anag@ucdavis.edu; mlliu@ucdavis.edu; mukherje@cs.ucdavis.edu

M. Tornatore
Politecnico di Milano, Milano, Italy
e-mail: tornator@elet.polimi.it

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*,
Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_4,

the volume discount[1] of transceivers for large traffic demands. However, due to accumulating signal impairments, and nonlinear interaction between wavelength channels with different bit rates, the reach of some of the high-bit-rate lightpaths could be limited, based on some threshold signal quality [e.g., bit error rate (BER)]. In fact, depending on certain physical layer parameters, viz., launch power, modulation format, dispersion map, etc. different line rates will have different reach. Thus, in a network design (or upgrade) problem, there remains a tradeoff between assignment of line rates to wavelengths and the all-optical reach of the line rates. So, to ensure that the total traffic is supported with minimum capital expenditure (CapEx) one needs to have more of the higher line rates with increased optical reach. This interplay between capacity and reach is thus the basic challenge for planning a MLR optical network.

This chapter will present various aspects and challenges for cost-efficient cross-layer design of MLR networks. The main issues in cross-layer design of MLR networks are twofold. One is from the service layer to the network layer, i.e., traffic grooming[2] in asymmetrical wavelength capacities, and the other is from the physical layer to the network layer where the signal impairments play a more significant role compared to a single-line-rate (SLR) case. More details on these issues will be discussed in Sects. 4.3 and 4.4. Section 4.2 will present some of the research works on MLR, that have been done so far. Section 4.5 summarizes the chapter.

2 Related Work

Operators of networks carrying 10-Gbps WDM systems are evaluating how and when to upgrade their networks. There is now a range of vendors providing 40-Gbps (per wavelength) transmission and switching systems, and based on announcements from various leading vendors, 100 Gbps per wavelength is also available. But the topic of MLR network design needs further research.

Migration options to higher rates are being investigated [5–9]. In [7], an operational expenditure (OpEx) analysis for such a migration is given. References [6, 8–11] describe the transition from low- to high-bit-rate services, focusing on

[1] Volume discount means that the cost of a resource increases at a rate that is lower than the linear increase of the rate, e.g., the cost of a 40-Gbps transceiver can eventually be 2.5 times that of a 10-Gbps transceiver under steady-state conditions, which is less than the rate increase of four times. This scale-up factor is independent of modulation scheme, i.e., with same modulation scheme the scale-up ratio in cost is 2.5 times from 10 to 40-Gbps and 1.5 times from 40 to 100-Gbps transceivers. However, if different modulation schemes are used for different line rates, the scale-up factors will change depending on the different hardware complexities associated with different modulation schemes.

[2] Traffic grooming means to pack different sub-wavelength traffic demands (i.e., demands which are smaller than the capacity of one wavelength channel) into the spare capacity of a wavelength channel.

the physical requirements and reducing signal impairments for higher bit rates. In [5], the authors study the optimization of routing and aggregation in terms of overall capital expenditure (CapEx), using multi-period planning approaches [12, 13]. A comparison between two multi-period planning approaches, viz., all-period planning and incremental planning, is presented. Using these two approaches, the network CapEx is optimized with bit-rate migration in view. The work in [14] also compares network-upgrade options from 10 to 40 Gbps.

Considering a network planning viewpoint, the MLR network poses a number of research challenges in order to include the effects of physical-layer impairments in the design phase. A body of research has dealt with networking issues related to physical-layer-impairment-aware (PLIA) routing and design. To be concise, we note two recent surveys on this topic [15, 16]. Most of this work deals with single-line-rate (SLR) networks, while MLR and PLIA routing and design was a largely unexplored problem until the authors in [4] proposed an impairment-aware design methodology for a transparent[3] optical network using MLR. In this work, it is shown how MLR network design can save more CapEx compared to SLR networks. The idea behind this MLR design is to assign line rates to the lightpaths such that the optical reach of the line rate supports the maximum route length of the path. Thus, the routing and wavelength assignment (RWA) problem gets extended to routing, wavelength, and rate assignment (RWRA). In [17], a translucent MLR network design is proposed, where it is shown that the CapEx can be further reduced by placing some regenerators in some of the intermediate nodes such that a large number of high-bit-rate paths can be bypassed over long distances. By using multiple modulation formats for different line rates, the reach of the high-bit-rate lightpaths can also be increased and this can be another method to reduce the cost of a MLR network [18].

Other physical layer parameters such as WDM channel spacing and launch optical power can also play sensitive role in MLR network design and planning. In [19], it is shown how static MLR network planning with various combinations of launch power can affect the CapEx, whereas [20] provides a dynamic launch power control algorithm which optimizes the blocking performance of 10 Gbps on–off keyed (OOK) line rates and 40 Gbps differential quadrature phase-shift keyed (DQPSK) line rates. The dynamic provisioning with impairment awareness in MLR networks is yet to be explored to the best of our knowledge. Some dynamic lightpath provisioning algorithms considering fixed and tunable node architectures appear in [21], but impairments are not considered as all nodes are assumed to be opaque enabled with O–E–O conversion.

Some more research works on MLR networks especially on survivability have also been reported in the literature. In [22], the authors have dealt with transparent

[3] In *transparent* networks the end-to-end lightpaths are all-optical and they do not undergo any regeneration at intermediate nodes. Optical networks can be also classified as *translucent* and *opaque* networks. In a *translucent* network, an optical signal travels from the source node through the network "as far as possible" before its quality degrades and it needs to be regenerated through optical-to-electrical-to-optical (O–E–O) conversion at an intermediate node. *Opaque* networks include signal regenerators at every intermediate node.

MLR network design with dedicated protection (where every primary lightpath carrying data, has a dedicated backup lightpath such that if the primary path fails, the affected traffic can be switched on to the backup path), demonstrating the beneficial interaction among MLR and protection. While dedicated protection has fast protection switching, shared protection (considered in [22]) is more resource efficient due to backup sharing. In [23], a more-effective shared-protection scheme based on pre-deploying subconnections is proposed (essentially pre-lit lightpaths) [24]. A study on the energy efficiency of the MLR networks has also been reported [25]. It shows that MLR networks are more energy efficient compared to SLR networks.

3 Role of Physical Impairments in MLR Network Design

Physical impairments play an important role in the design and planning of an optical network. In MLR networks, the role of physical impairments becomes more crucial because the same network infrastructure is used by different line rates. The effects of the impairments are different on different line rates and also their co-existence creates interference effects on each other which are more pronounced and hard to manage or mitigate. Especially with the legacy 10-Gbps system using a particular optimized dispersion map over the fiber links, installing new line rates on the same links creates unoptimized transmission performances because of the interplay between dispersion and fiber nonlinearities. Let us summarize the major impairments that specifically affect a MLR network.

(a) *Dispersion*: Dispersion in optical fibers affects the performance of a propagating signal based on which wavelength it is launched as well as its bit rate. Thus, dispersion becomes a vital issue in a MLR system. Dispersion can be compensated effectively for a given line rate but management of dispersion in a multi-rate scenario is difficult. In a MLR network, dispersion management can reduce the signal degradations due to fiber nonlinearities to extend the transmission reach. Dispersion management chooses the right amount of chromatic dispersion (CD) at the amplifier huts[4] (by deploying dispersion compensating fiber (DCF)) in order to reduce the non-linear effects. Right amount of CD per fiber span (known as residual dispersion per span (RDPS)) can be optimized during system design and can reduce the cross-phase modulation (XPM)[5]

[4] An amplifier hut is a location along the fiber routes where the optical amplifiers are installed along with gain flattening filters and dispersion compensating fibers. These huts are spaced about 80–100 km from each other.

[5] Cross-phase modulation (XPM) is a nonlinear effect where the phase of an optical signal changes due to the intensity fluctuations of other signals in adjacent WDM channels. This happens because when the fiber operates in the nonlinear regime, its refractive index becomes a function of the intensity of the signal.

penalty on co-propagating wavelengths in a dispersion-managed link [11]. Choosing the right RDPS optimizes the dispersion map (DM) for a fiber plant. Legacy networks are dispersion-minimized for 10-Gbps signals, i.e., 10-Gbps signals can be received at the receiver without any pre/post dispersion compensation. This means the residual dispersion after transmission is within the dispersion tolerance limits of 10 Gbps. However, 40 Gbps, being shorter in time window, has a lower dispersion tolerance compared to 10-Gbps signals. So, the reach of 40 Gbps will be effected by the 10-Gbps DM [11]. Same applies for 100-Gbps signals on a 10-Gbps dispersion-minimized fiber (DMF). Therefore, a transmission system dispersion-minimized for 10-Gbps signals is not essentially optimal for 40-Gbps or 100-Gbps signals [10]. Pre/post dispersion compensation is needed for 40-Gbps/100-Gbps signals while they are transmitted over a 10-Gbps DMF [10, 11].

In future, as networks will be enriched with 40-Gbps and 100-Gbps line rates, one can expect to have fiber plants dispersion-minimized for 40-Gbps and 100-Gbps line rates also. Hence, the network will have heterogeneous dispersion-minimized transmission links (where some links are 10-Gbps DMFs, some are 40 Gbps DMFs, and some are 100 Gbps DMFs). When a 40-Gbps wavelength goes through a 10-Gbps DMF along with other 10-Gbps wavelengths, one can expect to have a shorter reach of the 40-Gbps wavelength compared to that of the 40-Gbps wavelength through 40-Gbps DMF link. Similarly, a 10-Gbps wavelength should experience shorter reach in a 40-Gbps DMF link. Same argument applies for 100-Gbps wavelengths and 100-Gbps DMF links. This will affect the network design and upgrade in a significant way [26].

(b) *Nonlinearities*: As mentioned above, an optimum dispersion map helps to control the nonlinear effects such as four-wave mixing (FWM),[6] cross-phase modulation (XPM), and self-phase modulation (SPM).[7] But then, in MLR systems, it is hard to optimize a dispersion map for all the line rates, and thus nonlinear effects will impair the signals in different channels by different amounts based on their bit rates.

In order to retrofit high-bit-rate channels in the existing 50-GHz WDM grid, advanced modulation formats such as DQPSK and polarization-multiplexed QPSK seem to be good choices [6]. These phase-modulated systems are highly susceptible to nonlinear effects such as XPM and SPM that affect the phase of

[6] Four-wave mixing (FWM) is a nonlinear effect where three optical signals in the WDM band interact to produce a fourth signal with a frequency that falls within the frequency band of the three interacting signals but of less power compared to the interacting signals. Thus the fourth signal exhibits as an in-band crosstalk signal.

[7] Self-phase modulation (SPM) is another nonlinear effect where the phase of the signal gets modulated due to change of refractive index of the fiber with intensity variations. The difference between XPM, as defined earlier, and SPM lies in the fact that the phase variations in a signal are induced due to intensity variations in the same signal (SPM) or other co-propagating signals (XPM). More details about the nonlinear effects can be found in [27] and [28].

the propagating optical signal. In particular, XPM is highly detrimental while DQPSK modulated 40-Gbps signals are transmitted along with on-off-keyed 10-Gbps signals [11]. Depending on how the bit rates are placed in the wavelength grid, their optical reach gets affected due to XPM accordingly. For example, if 10-Gbps and 100-Gbps signals are in adjacent channels, then the signal degradation due to XPM is more severe than in the cases where 10-Gbps and 40-Gbps are adjacent or 40-Gbps and 100-Gbps are adjacent. This is because, due to the difference in group velocities of the optical signal pulses, there is a walk off between the two pulses, i.e., if they start together, they will separate as they propagate through the medium. Nonlinear interaction takes place as long as they physically overlap in the medium. Smaller the dispersion, smaller will be the difference in group velocities (assuming closely-spaced wavelengths) and the longer they will overlap. This would lead to stronger XPM effects. So obviously, the combination of 10 Gbps and 100 Gbps on adjacent wavelengths is the worst combination as far as relative walk off is concerned as 10-Gbps pulses are broadest temporally whereas the 100-Gbps pulses are narrowest temporally. Therefore, in MLR systems, the wavelength assignment to different lightpaths should be done with utmost care to minimize the effect of these nonlinearity-induced interference.

(c) *Optical Filter Concatenation*: Optical filters are present in the nodes for multiplexing and demultiplexing of wavelengths. For a long lightpath, the optical signal has to travel through multiple hops and hence has to pass through several such optical filters. The filters being tuned to 50-GHz bandwidth ideally (which happens to be the channel spacing) may not be exactly aligned to their center wavelengths. Some of them may be shifted to the left and some may be shifted to the right of their center wavelengths. Hence, the effective bandwidth for a chain of such filters will be less than 50 GHz. A 10-Gbps signal spectrum with binary on–off keying may well fit into that bandwidth, but for higher bit rates, especially 100 Gbps, it requires advanced modulation formats to narrow down the spectrum so that it fits into the 50-GHz channel spacing. So the narrowing down of the effective bandwidth for a chain of optical filters affects the higher bit rates more. Due to nonidealities in the laser sources, the signal spectrum may be offset from its center frequency as well, and the signal gets impaired due to bandlimiting. It also manifests as crosstalk on the neighboring channels. So, while designing a MLR network, an effective channel management may be required.

The physical-layer performance summary for a MLR scenario with preferred modulation formats are summarized below in Table 4.1.

4 MLR Network Design

As discussed so far in this chapter, it may be convenient for next-generation networks to set up lightpaths with different bit rates over different wavelengths and/or over different links. Equipping the network with different bit rates (1) adds

Table 4.1 Physical-layer performance summary [6, 29, 30]

Line Rate	10 Gbps	40 Gbps	100 Gbps
Preferred modulation format	NRZ-OOK	50% RZ-DQPSK	Pol-Mux DQPSK
CD tolerance	1,500 ps/nm	219 ps/nm	140 ps/nm
PMD tolerance	10 ps	8 ps	2.5 ps
Penalty due to 50 GHz channel spacing	Low	Low	Medium
XPM Penalty	On 10 Gbps: low On 40 Gbps: medium On 100 Gbps: high	On 10 Gbps: low On 40 Gbps: low On 100 Gbps: medium	On 10 Gbps: low On 40 Gbps: medium On 100 Gbps: low

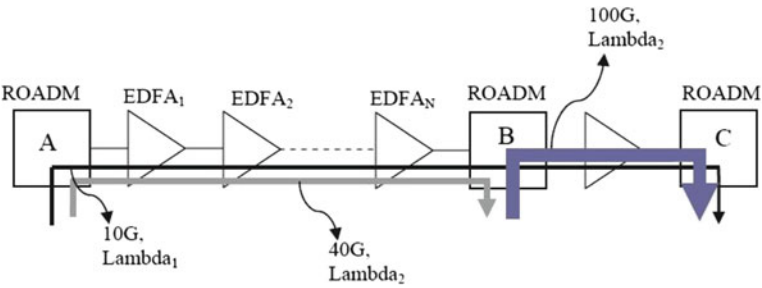


Fig. 4.1 Basic scheme of MLR networks. *EDFA*: Erbium-Doped Fiber Amplifier; *OXC*: Optical Cross-Connect; lightpaths between OXCs (switching nodes) can be set up at different rates

flexibility in the network design by avoiding low-bandwidth connections over high-capacity lightpaths, (2) supports multi-rate transport protocols and hence avoids complex multiplexing schemes, and (3) uses the optimal combination (number/rate) of wavelengths on each link which addresses both traffic and network asymmetry. Thus, MLR networks are becoming a new paradigm. The MLR idea also matches with recent progress in carrier-grade Ethernet, where native Ethernet frames are transported on Ethernet tunnels over WDM channels (Ethernet-over-WDM). Ethernet provides cost-efficient interfaces operating at different line rates to deal with traffic heterogeneity, which can be well supported by MLR networks.

In a MLR network, high-bit-rate transceivers (viz., 40/100 Gbps) can exploit the volume discount principle. Due to accumulated signal impairments, the reach of some high-bit-rate lightpaths could be limited, based on a threshold signal quality (e.g., bit-error rate (BER)). Figure 4.1 shows some basic properties of a MLR network in which (1) an optical node may contain optical switches, called optical cross-connects (OXCs); (2) a fiber link between two optical nodes may have several

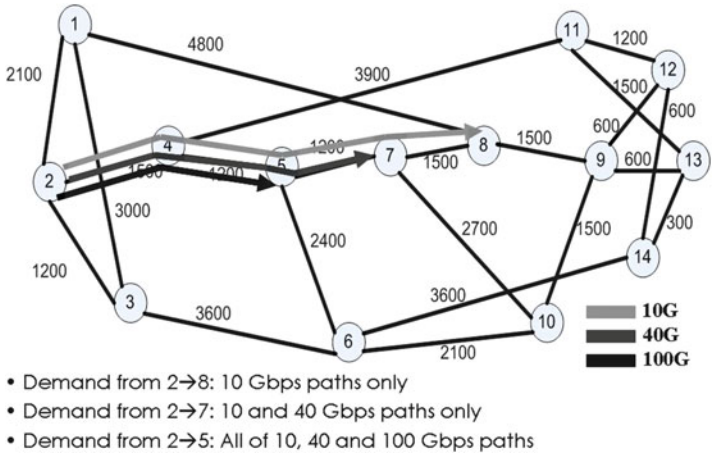


Fig. 4.2 Reach of lightpaths at different line rates in a US-wide network (Link lengths in km.)

Erbium-Doped Fiber Amplifiers (EDFAs) approximately every 80 km or so apart in practice for (analog) signal amplification; and (3) most importantly, different line rates may co-exist on the same fiber on different wavelengths, e.g., a 10-Gbps lightpath on λ_1 between nodes A and C; a 40-Gbps lightpath on λ_2 between nodes A and B; and a 100-Gbps lightpath also on λ_2 between nodes B and C. Figure 4.2 shows that a lower-rate lightpath can have a longer reach (regenerator-free distance) than a higher-rate lightpath because of signal-quality constraint. So, in a MLR network, there is a tradeoff between capacity and reach.

A MLR network, because of its efficient utilization of capacity, is more cost-effective than a single-line-rate (SLR) network [4]. To further reduce the cost of a MLR network, the network planner can exploit the heterogeneity in the transmission systems, e.g., some long-distance high-bit-rate paths could be lit up with an *improved modulation format* having higher bandwidth-distance product and less susceptibility to impairments [30]. Note that, using a basic modulation format, these paths could be infeasible if the signal quality at the receiver was poor (BER constraint). Thus, the reach of some high-bit-rate paths can be increased and one can save on CapEx by exploiting more volume discount [18]. Another approach to reduce the cost of a MLR network is to place *3R regenerators* (or just regenerators, for short, which is the term used in the rest of this document)[8] in the network at select locations such that the congested parts of the network can be served by high-bit-rate paths [18]. The idea is to have considerable amount of volume discount by spending a little "extra" on regenerators. Below, we discuss some holistic design approaches which consider the role of physical impairments in the design of a MLR network.

[8] Regeneration can be classified as 1R, 2R, and 3R, where 1R = Reamplification (which is provided by optical amplifiers such as EDFA); 2R = Reamplification and Reshaping; and 3R = Reamplification, Reshaping, and Retiming. Note that 2R and 3R apply to digital signals, which are considered here and in most networking studies, while 1R is an analog operation.

4.1 Transparent Optical Network Design with Mixed Line Rates

The design of a transparent MLR network is a complex holistic network design problem. As an example, a two-step approach can be employed. First, the BER of lightpaths over a set of pre-computed paths (using fixed routing algorithms such as shortest path or disjoint k-shortest path) are calculated over each possible line rate (10/40/100 Gbps). The calculated BER is compared with a threshold BER (typically 10^{-3} is an acceptable BER with modern modulation formats), and thus the feasibility of a lightpath over a particular line rate is determined (those combinations of path, line rate, and wavelength, for which the BER is higher than the threshold BER, are considered to be infeasible). Next, knowing the feasible lightpaths, an optimization problem can be posed such that a minimum-cost MLR network design is obtained, as shown below.

BER Estimation: To estimate the signal quality (BER), let us first consider static impairments in a lightwave system, viz.: (1) dispersion (chromatic and polarization-mode), (2) optical amplifier noise (amplified spontaneous emission (ASE) noise in fiber amplifiers), (3) crosstalk, (4) optical filter concatenation, (5) laser frequency offset, and (6) receiver noises. Dynamic impairments such as four-wave mixing (FWM), cross-phase modulation (XPM), and self-phase modulation (SPM) become intractable to estimate for network design problems, so they are accounted for by using a typical safety margin in the Q-factor, say 1 dB [31]. There are ways in which one can still capture the effects of nonlinearities in a static network design problem and they are discussed later in the chapter.

The signal component of the received lightwave is given by:

$$e_s(t) = E_s d_s(t) \cos[2\pi(f_s + \Delta f_s)t + \phi_s(t) + \theta_s] \tag{4.1}$$

where $E_s = 2\sqrt{P_R}$ is the lightwave amplitude, Δf_s is laser frequency misalignment, $\phi_s(t)$ is the laser phase noise term, θ_s is the random epoch, and $d_s(t)$ is the encoded data pulse (NRZ, duobinary, etc.). Bandlimiting due to filter concatenation occurs in optical switches at intermediate nodes and in the receiver at the destination node [18, 32].

The signal dropped at the destination passes through a bandlimiting optical filter onto the photodetector and then through a low-pass filter (LPF) followed by a decision circuit. The current at the photodetector output is given by:

$$i_p(t) = R_\lambda \langle e_R^2(t) \rangle + i_{sh}(t) + i_{th}(t) \tag{4.2}$$

where $R_\lambda \langle e_R^2(t) \rangle$ = square-and-average response of the photodetector, $e_R(t)$ = incident lightwave, R_λ = photodetector responsivity, $i_{sh}(t)$ = shot noise produced by the incident lightwave, and $i_{th}(t)$ = receiver thermal noise. The incident lightwave consists of signal, crosstalk, and spontaneous lightwaves; and these three will

produce beating terms at the output of the receiver and contribute to noise. Evaluation of BER requires the statistics of the electrical noise at the receiver output which follows Gaussian statistics whose variances depend on the network parameters. Noise variances for "0" and "1" reception consist of variances of thermal noise, shot noise (signal dependent), and other beat-noise terms between signal (will be absent for "0" reception), spontaneous, and crosstalk lightwaves. Noise variances may be calculated easily [33, 34].

The effect of chromatic dispersion is determined by convolving the input pulse with the fiber transfer function and then calculating the power penalty due to pulse broadening [35]. Polarization-mode dispersion (PMD) can be taken care of by a power penalty factor [36]. At the receiver, the decision circuit following the LPF compares the sampled LPF output with a threshold current given by:

$$I_T = \frac{\sigma_0 I_{s1} + \sigma_1 I_{s0}}{\sigma_0 + \sigma_1} \tag{4.3}$$

where I_{s1} and I_{s0} are signal current due to "1" and "0" reception, respectively. The BER at the destination receiver is thus given by:

$$BER = \frac{1}{4}\left[erfc\left(\frac{I_T}{\sqrt{2}\sigma_0}\right) + erfc\left(\frac{I_{s1} - I_T}{\sqrt{2}\sigma_1}\right)\right] \tag{4.4}$$

The BER calculation described above is a simplified one as mentioned, without the modeling of the nonlinear impairments. As described in Sect. 4.3, some of the nonlinear impairments are crucial for MLR network design. So, other physical layer models can be devised which can capture the effect of nonlinearities as well. An approach to do so can be to modify the Nonlinear Schrödinger's Equation (NLSE) for signal propagation in an optical fiber for MLR scenario and then solving it. A similar simulation based on NLSE can be found in [37].

Design of Minimum-Cost MLR Networks: Based on the BER model, one can identify the set of admissible paths on each line rate. The next step is to develop a design strategy for MLR networks. Design methods, based both on exact mathematical methods, such as Integer Linear Program (ILP) formulations and heuristic approaches, can be devised to solve the problem. For illustration purposes, hereafter let us consider a network with single fiber links and a fixed number of wavelengths per link. In this case, the main contribution to the network cost would be the cost of the transponders to be installed to support a given set of traffic demands. The minimal-cost network design can be stated as follows.

Input Parameters:

- $G(V, E)$: Physical topology of the network with V nodes and E links.
- $T = [\Lambda_{sd}]$: Traffic matrix with aggregate demands Λ_{sd} in Gbps between a source-destination (s–d) pair.
- $R = \{r_1, r_2, \ldots, r_k\}$: Set of available channel rates.

- D_k: Cost of a transponder with rate r_k.
- L_{ij}: Length of the lightpath between a *s–d* pair *ij*(in km).
- l_{mn}: Physical link between nodes *m-n*.
- W: Maximum number of wavelengths supported on a link, $\lambda \in \{1, 2, \ldots, W\}$.
- B: Threshold BER: a lightpath with a higher BER will be rejected.
- $BER_{ijk\lambda}$: BER for the lightpath between a *s–d* pair *ij* at rate r_k and wavelength λ.
- $\alpha_{ijk\lambda} = \begin{cases} 1 & \text{if } BER_{ijk\lambda} \leq B \\ 0 & \text{otherwise} \end{cases} \quad \forall\ (i,j), k, \lambda$
- P_{mn}: Set of lightpaths passing through link l_{mn}.

Variables:

- $X_{ijk\lambda}$: Number of lightpaths at rate r_k and wavelength λ between nodes *ij*.
- f_{ij}^{sd}: Traffic from source s to destination d routed on lightpath between nodes *ij*.

Problem Formulation:

$$\text{Minimize}: \sum_{\lambda} \sum_{ij} \sum_{k} X_{ijk\lambda} \cdot D_k \tag{4.5}$$

Subject to:

Capacity Constraint:

$$\sum_{\lambda} \sum_{k} r_k \cdot X_{ijk\lambda} \cdot \alpha_{ijk\lambda} \geq \sum_{s,d} f_{ij}^{sd} \quad \forall\ (i,j) \tag{4.6}$$

Unique Wavelength Constraint:

$$\sum_{i,j \in P_{mn}} \sum_{k} X_{ijk\lambda} \cdot \alpha_{ijk\lambda} \leq 1 \quad \forall\ (m,n),\ \forall\ \lambda \tag{4.7}$$

Solenoidality Constraint:

$$\sum_{i} f_{ij}^{sd} - \sum_{i} f_{ji}^{sd} = \begin{cases} \Lambda_{sd} & \text{if } s = j \\ -\Lambda_{sd} & \text{if } d = j \\ 0 & \text{otherwise} \end{cases} \quad \forall\ (i,j) \tag{4.8}$$

The objective function in Eq. (4.5) computes the overall cost due to transponders at various bit rates. The $\alpha_{ijk\lambda}$'s determine whether a lightpath between a pair of nodes is feasible over a particular wavelength and a bit rate (i.e., respects the BER threshold), and are calculated off-line for each possible combination of i, j, k, λ over the physical routes. In Eqs. (4.6) and (4.7), multiplication of $X_{ijk\lambda}$ with $\alpha_{ijk\lambda}$, refers to the fact that only those variables are present in the summations for which the combination of i, j, k, and λ yields $\alpha_{ijk\lambda} = 1$.

Equation (4.6) sets a capacity constraint on the traffic demands routed over the lightpaths. The wavelength-continuity (i.e., optical transparency) constraint is taken care of by Eq. (4.7), which implicitly also enforces a constraint on the available capacity (number of channels W) over a physical link. Equation (4.8) satisfies the flow-conservation constraints. The output of this ILP is the number of lightpaths over different rates and wavelengths. This formulation can be downgraded for a SLR network by forcing the value of r_k to a predefined bit rate.

Different Variations of the Transparent Mixed-Line-Rate Network Design:

1. *Splittable vs. Non-Splittable Flows*. Note that, call splitting at the source is allowed in this formulation by defining the variables f_{ij}^{sd} to be integers. An alternative approach where call splitting is not allowed, can be easily formulated by forcing the variables f_{ij}^{sd} to be binary, replacing Λ_{sd} 's in Eq. (4.8) with 1's, and multiplying Λ_{sd} with f_{ij}^{sd} in Eq. (4.6).
2. *Adjustments for Non-Linear Impairments*. The 1-dB margin in the Q-factor that is mentioned before in order to account for the nonlinear impairments, is one way of approaching the cross-layer design of MLR networks. Another approach could be to incorporate and hence limit the effects of inter-channel crosstalk due to nonlinearities directly in the ILP model as described below. Let N_{ij} be the set of intermediate nodes in the path connecting the nodes *i–j* and let l_{ij} be the set of intermediate links in the path connecting nodes *i–j*. Let $s^2_{XT_k,n}$ be the intra-channel crosstalk noise variance that affects a lightpath on line rate k crossing node n and $s^2_{XPM_k,l}$ be the 1st order cross-phase modulation variance on link l due to line rate k. Then the constraint which limits the interference that accumulates on a candidate lightpath due to intra-channel crosstalk, cross-phase modulation and four-wave mixing is given by:

$$\sum_{n\in N_{ij}}\left(\sum_{i'j'|n\in N_{i'j'}}\sum_{k} s^2_{XT_k,n}X_{i'j'k\lambda}+\sum_{i'j'|l\in L_{i'j'}}\sum_{k} s^2_{XPM_k,l}\left(X_{i'j'k\lambda-1}+X_{i'j'k\lambda+1}\right)\right)$$
$$+\, c_{FWM}+BX_{ijk\lambda}-S\le \sigma^2_{ijk\lambda(max)}+B \quad \forall i,j,\lambda \tag{4.9}$$

where, $X_{ijk\lambda}$ is the desired lightpath and $X_{i'j'k\lambda}$ denotes the interfering lightpath; c_{FWM} is the worst case four-wave mixing (FWM) contribution [38], $\sigma^2_{ijk\lambda(max)}$ is the maximum noise variance bound contributed due to impairments that do not depend on the interference between lightpaths [38], B is a large positive constant and S is a positive slack variable[9] which should be added to the objective function as well. The worst case FWM contribution c_{FWM} is computed assuming

[9] Slack variables are used in optimization problems to convert constraints which are in the form of inequalities, into equalities. Here the purpose of the slack variable is to account for any surplus impairments that might occur due to nonlinearities and these surplus inequalities are minimized by adding the variable S in the objective function.

all wavelengths in all links are fully loaded. The XPM and crosstalk variances in Eq. (4.9) are evaluated using analytical tools presented in [39] and [40]. The use of the large positive constant B is as follows. Suppose, the variable $X_{ijk\lambda}$ takes value one, then the constraint in Eq. (4.9) becomes:

$$\sum_{n\in N_{ij}}\left(\sum_{i'j'|n\in N_{i'j'}}\sum_{k}s^2_{XT_k,n}X_{i'j'k\lambda}+\sum_{i'j'|l\in L_{i'j'}}\sum_{k}s^2_{XPM_k,l}\left(X_{i'j'k\lambda-1}+X_{i'j'k\lambda+1}\right)\right) + c_{FWM}-S\le\sigma^2_{ijk\lambda(max)}\quad\forall i,j,\lambda \tag{4.10}$$

which holds true as $\sigma^2_{ijk\lambda(max)}$ is the maximum noise variance bound. Again, if $X_{ijk\lambda}$ takes value zero, then Eq. (4.9) becomes:

$$\sum_{n\in N_{ij}}\left(\sum_{i'j'|n\in N_{i'j'}}\sum_{k}s^2_{XT_k,n}X_{i'j'k\lambda}+\sum_{i'j'|l\in L_{i'j'}}\sum_{k}s^2_{XPM_k,l}\left(X_{i'j'k\lambda-1}+X_{i'j'k\lambda+1}\right)\right) + c_{FWM}-S\le\sigma^2_{ijk\lambda(max)}+B\quad\forall i,j,\lambda \tag{4.11}$$

which also holds true since B is a large positive variable. So, modifying the ILP formulation given by Eqs. (4.5) to (4.8) by adding constraint given in Eq. (4.9), the effect of nonlinear impairments can be limited to a certain extent. For more details please see [19] and [38].

3. *k-Paths*. As mentioned before, the pre-computed set of feasible paths can be estimated using either shortest path algorithm or disjoint k-shortest path algorithm. In case of k-shortest path algorithm, the different paths can be denoted by having another subscript to the variable $X_{ijk\lambda}$ as $X_{ijk\lambda}p$, where p denotes the path dimension. The summations in the above formulation that employ $X_{ijk\lambda}p$, will then have another summation over all p i.e., all paths.
4. *Minimizing the Routing Cost along with Transponder Cost*. In the optimization model only the equipment costs are considered. A more enhanced version of this model can be implemented by considering the routing cost as well. This can be accomplished by defining a variable H_k for the number of hops a lightpath can have on a particular line rate r_k. Then the objective function in Eq. (4.5) can be rewritten as:

$$\text{Minimize:}\quad\sum_{\lambda}\sum_{ij}\sum_{k}X_{ijk\lambda}\cdot D_k+\varepsilon\sum_{\lambda}\sum_{ij}\sum_{k}H_k\cdot X_{ijk\lambda} \tag{4.12}$$

where ε is a weighting factor which puts appropriate weightage for the routing optimization. The second term in Eq. (4.12), enables to choose a lesser hop path within the transmission range requirement.

Solving the optimization problem (stated through Eqs. (4.5) to (4.8)) for a typical US-wide network such as the one in Fig. 4.2, with 80 wavelengths per

fiber and a total traffic volume of 1 Tbps, has very high computation time (approx. 10–12h on ILOG CPLEX 9.0 running on an Intel Pentium-4 CPU, with 3.20 GHz processor speed and 1 GByte of random-access memory (RAM)), and the problem becomes practically unsolvable for larger networks. So, the MLR design problem needs to be supplemented with proper heuristic optimization methods.

Heuristic Solution: The inputs to the heuristic are (1) for each line rate, the reachability information between the various node pairs (which we refer as the logical connectivity pattern), (2) traffic demand matrix, (3) set of rates along with their costs, and (4) a set of possible physical routes (lightpaths) for each link in the logical connectivity pattern and, for each of these routes, the set of rates (viz., 10 Gbps; or 10/40 Gbps; or 10/40/100 Gbps) that can be installed. The output of this algorithm is the set of lightpaths for different *s*–*d* pairs over different rates. Demands are sorted between each $s-d$ pair in descending order, and then the demands are served sequentially. In each iteration, we check the residual capacity of the lightpaths set up in the previous iteration to see if all the current demand or part of it can be routed through the lightpaths having spare capacity. If the entire current demand can "fit in" to spare capacities, the network cost does not increase. If only part of the current demand can be "fit in", then for the remaining demand, we set up new lightpath(s) with an integer combination of 10, 40, and 100 Gbps line rates such that the cost of the additional capacity is minimum.

The run time of this heuristic is small (on the order of seconds for the same problem size and hardware platform as earlier) compared to the ILP formulation. A final note on the wavelength assignment: a first-fit wavelength assignment is applied here, but effective MLR-aware wavelength-assignment policies need to be studied: e.g., wavelengths that are less prone to static impairments can be reserved for higher bit rates, or to minimize the dynamic impairments, an opportunistic wavelength assignment can help by significantly reducing impairments such as XPM and cross-talk.

Illustrative Results: Some illustrative results obtained by an ILP for MLR design are presented here, and compared with SLR network design. The network topology used in our study is shown in Fig. 4.2. The traffic matrix is shown in Table 4.2 and it represents a total traffic of 1 Tbps, which is multiplied by different factors to represent a range of loads. The costs of 10 Gbps, 40 Gbps, and 100 Gbps transponders are, respectively, $1\times, 2.5\times$ and $3.75\times$ [41]. Thus, higher-rate transponders provide volume discount, i.e., the cost of capacity does not scale up linearly as capacity increases; the higher the capacity per wavelength, the lower is the cost per unit of capacity.

The modulation scheme considered is optical duobinary which has least transmitter and receiver complexity among advanced modulation formats [30]. BER threshold is taken as $B = 10^{-3}$ (worst possible value correctible using error-correcting codes) [30]. $W = 80$ wavelengths per link. Transmitted power is 0 dBm, switch crosstalk is -35 dB, channel spacing is 50 GHz, chromatic dispersion coefficient is 17 ps/(nm.km), and PMD coefficient is 0.06 ps/$\sqrt{\text{km}}$. Other physical-layer parameters are chosen as in [33].

Table 4.2 Traffic matrix

Node	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	2	1	1	1	4	1	1	2	1	1	1	1	1
2	2	0	2	1	8	2	1	5	3	5	1	5	1	4
3	1	2	0	2	3	2	11	20	5	2	1	1	1	2
4	1	1	2	0	1	1	2	1	2	2	1	2	1	2
5	1	8	3	1	0	3	3	7	3	3	1	5	2	5
6	4	2	2	1	3	0	2	1	2	2	1	1	1	2
7	1	1	11	2	3	2	0	9	4	20	1	8	1	4
8	1	5	20	1	7	1	9	0	27	7	2	3	2	4
9	2	3	5	2	3	2	4	27	0	75	2	9	3	1
10	1	5	2	2	3	2	20	7	75	0	1	1	2	1
11	1	1	1	1	1	1	1	2	2	1	0	2	1	61
12	1	5	1	2	5	1	8	3	9	1	2	0	1	81
13	1	1	1	1	2	1	1	2	3	2	1	1	0	2
14	1	4	2	2	5	2	4	4	1	1	61	81	2	0

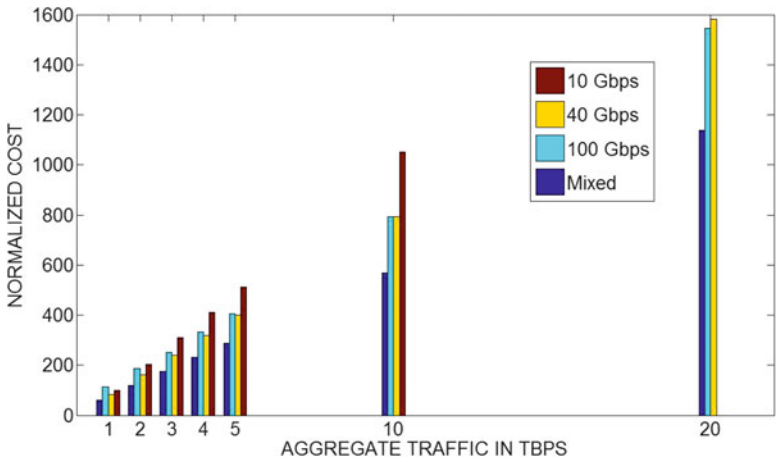


Fig. 4.3 Transponder costs for SLR and MLR networks

Figure 4.3 reports the cost of the network in Fig. 4.2 from ILP formulation, in terms of line cards, for the following four scenarios: MLR network and three SLR networks, each equipped with either 10 Gbps, or 40 Gbps, or 100 Gbps line cards, considering the BER constraint. As expected, an MLR network with BER constraints has always the least cost (because all SLR solutions are part of the MLR solution space). For SLR cases, interesting observations are: (1) at 10 Gbps, the cost increases almost linearly with traffic[10]; (2) at 40 Gbps, the cost is less than

[10] Note that, in this example, a 10-Gbps SLR network cannot support 20 Tbps of traffic with 80 wavelengths. So, no cost is reported in the results for 20 Tbps in case of 10-Gbps SLR network.

Table 4.3 Normalized cost for ILP and heuristic

Traffic (Tbps)	ILP	Heuristic with Grooming	Heuristic without Grooming
1	58	115.5	219.5
2	115	191.5	249.5
3	173	255	319
5	289	353	379
10	568	610	615
20	1,138	1,152	1,152

that of 10 Gbps because of volume discount; and finally, (3) at 100 Gbps, due to the BER constraint, only a few 100 Gbps end-to-end lightpaths can be used, so many more multi-hop connections[11] have to be established. So the volume discount that 100 Gbps can provide, if all paths were feasible, is lost. These results confirm that an MLR network is able to support traffic heterogeneity achieving relevant cost savings compared to SLR networks.Next, the performance of the heuristic compared to the ILP formulation isshown in Table 4.3, where, the fourth column presents results for provisioningof separate end-to-end lightpaths for all the demands without anygrooming.[12]These results show that the MLR network design heuristic performs in betweenthe two extremes, viz., ILP formulation, where the most efficient packing oflightpaths occurs as it tries to globally minimize the total transponder cost, andthe case where no grooming takes place. Our algorithm tends to work better astraffic load increases, but the large gap with the optimal solution motivatesfurther research on improved algorithms: e.g., rerouting of traffic overlower-bit-rate connections to fill residual capacity in higher-bit-rate lightpathsand/or aggregation of lower-bit-rate connections on higher-bit-rate routes,whenever possible.

4.2 Transparent vs. Translucent Optical Network Design with Mixed LineRates

Another way to improve the efficiency of MLR network design is to embed regeneration capabilities in the network so that high-bit-rate paths have increased reach and they can serve the congested parts of the network, thereby reducing network cost through volume discount. Note that, based on current technology, regenerators can apply only on a single line rate, so different regenerators have to be installed to support different line rates. Since regenerators are costly, placement of regenerators

[11] Multi-hop connections refer to those source-destination connections which are established over a sequence of lightpaths compared to a single-hop connection which is established on a single end-to-end lightpath.

[12] By grooming, it is referred to source grooming where the sub-wavelength demands are packed at the source of a lightpath to fill up a wavelength's capacity. This is different from grooming at intermediate nodes as in an opaque network.

has to be carefully planned, minimizing their number. This approach is termed as translucent MLR network design, as opposed to the previous "transparent" design.

Here, we sketch a design method for translucent MLR networks as an extension of the earlier method for transparent MLR networks. As before, BER calculation comes first. Then, the design is divided into a two-step problem. In the first step, a location problem is solved. This places a *fixed* number of regenerators optimally such that the number of high-bit-rate paths is maximized. While maximizing the number of high-bit-rate paths, priority is given to the bit rates according to their capacity. The output of this step provides the location of the regenerators for various bit rates. The ILP formulation for the regenerator placement problem is presented below.

Input Parameters:

- $G(V, E)$: Physical topology of the network with V nodes and E links.
- $T = [\Lambda_{sd}]$: Traffic matrix with aggregate demands Λ_{sd} in Gbps between a *s–d* pair.
- $R = \{r_1, r_2, \ldots, r_k\}$: Set of available channel rates.
- $D_k(C_k)$: Cost of a transponder(regenerator) with rate r_k.
- l_{ijk}: Lightpath between a node pair *i–j* on rate k.
- W: Maximum number of wavelengths supported on a link, $\lambda \in \{1, 2, \ldots, W\}$.
- B: Threshold BER: a lightpath with a higher BER will be rejected.
- $BER_{ijk\lambda}$: BER for the lightpath between a *s–d* pair *ij* at rate r_k and wavelength λ.
- $\alpha_{ijk\lambda} = \begin{cases} 1 & \text{if } BER_{ijk\lambda} \leq B \\ 0 & \text{otherwise} \end{cases} \quad \forall\ (i,j), k, \lambda$
- P_{mn}: Set of lightpaths passing through physical link $m - n$.
- $L = \{l_{ijk} : \alpha_{ijk} = 0 \forall ijk\}$: Denotes set of infeasible l_{ijk}-s based on threshold BER.
- K: Maximum number of regenerators that can be placed in the network.

Variables:

- X_{ijk}: Binary variable, takes value 1 if lightpath l_{ijk} is feasible due to regenerator placement.
- $Y_{ek(ij)}$: Binary variable, takes value 1 if an intermediate node e along the path connecting end nodes *i–j* enables regeneration for rate r_k.
- $Z_{ek(ij)}$: Binary variable, takes value 1 if an intermediate node e for l_{ijk} exists, where l_{ijk} can be fragmented into two feasible lightpaths.
- $W_{ek(ij)}$: Binary variable, takes value 1 if both $Y_{ek(ij)}$ and $Z_{ek(ij)}$ take value 1.
- N_{ek}: number of regenerators at node e of line rate r_k.

Problem Formulation:

$$\text{Maximize} \sum_{i,j,k} X_{ijk} \tag{4.13}$$

subject to:

Logical AND of α_{iek} and α_{ejk}:

$$Z_{ek(ij)} \geq \alpha_{iek} + \alpha_{ejk} - 1 \quad \forall(e,i,j,k) \tag{4.14}$$

$$Z_{ek(ij)} \leq \alpha_{iek} \quad \forall(e,i,j,k) \tag{4.15}$$

$$Z_{ek(ij)} \leq \alpha_{ejk} \quad \forall(e,i,j,k) \tag{4.16}$$

Logical AND of $Z_{ek(ij)}$ and $Y_{ek(ij)}$:

$$W_{ek(ij)} \geq Z_{ek(ij)} + Y_{ek(ij)} - 1 \quad \forall(e,i,j,k) \tag{4.17}$$

$$W_{ek(ij)} \leq Z_{ek(ij)} \quad \forall(e,i,j,k) \tag{4.18}$$

$$W_{ek(ij)} \leq Y_{ek(ij)} \quad \forall(e,i,j,k) \tag{4.19}$$

$$X_{ijk} \leq W_{ek(ij)} \quad \forall(e,i,j,k) \tag{4.20}$$

$$N_{ek} \geq \sum_{ij} Y_{ek(ij)} \quad \forall(e,k) \tag{4.21}$$

$$\sum_{e,k} N_{ek} \leq K \tag{4.22}$$

The objective function in Eq. (4.13) maximizes the number of paths which were infeasible according to the threshold BER requirement but become feasible due to the regenerator placement. The summation in Eq. (4.13) is actually over the set of infeasible paths with default values to be zero and by maximizing that sum, the ILP is essentially enforcing as many X_{ijk}'s to be equal to one as possible subject to the constraints given by Eqs. (4.14) to (4.22). Equations (4.14), (4.15), and (4.16) assign value equal to 1 to variable Z, on a node e and rate k, if the placement of a regenerator in e would allow to enable the lightpath l_{ijk} (i.e., if the two fragments l_{iek} and l_{ejk} are feasible). Equations (4.17), (4.18), and (4.19) assign value equal to 1 to variable W, on a node e and rate k, if $Z = 1$ and $Y = 1$ (Y assumes value equal to 1 when a regenerator is actually deployed in node e). Equation (4.20) allows discovering the newly-enabled lightpaths according to the regenerator placement. Equation (4.21) evaluates the number of regenerators placed in each node e for each line rate k. Equation (4.22) imposes a maximal value on the number of regenerators.

Next, considering that lightpaths passing through nodes with regenerators have increased reach, a similar ILP formulation as in Sect. 4.4.1 can be solved. Let Q be the set of new lightpaths that become feasible after the placement of regenerators

and let S_e be the set of lightpaths that bypass through node e. Then the new ILP formulation considering regenerator placement will be as follows:

$$\text{Minimize}: \sum_{\lambda}\sum_{ij}\sum_{k} X_{ijk\lambda}\cdot D_k + \sum_{\lambda}\sum_{ij\in Q}\sum_{k} X_{ijk\lambda}\cdot C_k \tag{4.23}$$

Subject to:

Capacity Constraint:

$$\sum_{\lambda}\sum_{k} r_k \cdot X_{ijk\lambda}\cdot \alpha_{ijk\lambda} \geq \sum_{s,d} f_{ij}^{sd} \quad \forall\ (i,j) \tag{4.24}$$

Unique Wavelength Constraint:

$$\sum_{i,j\in P_{mn}}\sum_{k} X_{ijk\lambda}\cdot \alpha_{ijk\lambda} \leq 1 \quad \forall\ (m,n),\ \forall\ \lambda \tag{4.25}$$

Regeneration Constraint:

$$\sum_{i,j\in Q\cap S_e}\sum_{\lambda} X_{ijk\lambda} \leq N_{ek} \quad \forall e,k \tag{4.26}$$

Solenoidality Constraint:

$$\sum_{i} f_{ij}^{sd} - \sum_{i} f_{ji}^{sd} = \begin{cases} \Lambda_{sd} & \text{if } s=j \\ -\Lambda_{sd} & \text{if } d=j \\ 0 & \text{otherwise} \end{cases} \quad \forall\ (i,j) \tag{4.27}$$

The new objective function accommodates also the cost of regenerators (e.g., 1.4x the cost of a transponder in the illustrative examples below, but can be a parameter in general). Note that, while the first step gives an upper bound on the number of regenerators on select nodes based on a certain total maximum budget on regenerators, the second step shows how the network cost can be minimized using that information. The exact number of regenerators for a particular line rate in a regenerator-enabled node can be found by counting the number of bypass paths through those nodes obtained from the second-step ILP.

Illustrative Results: Table 4.4 shows the distribution of regenerators of different line rates at different nodes of the network in Fig. 4.2, for different values of K. Using the same setup as before, a cost comparison between a MLR network with regenerators (translucent) and a MLR network without regenerators (transparent) is conducted. Results are reported in Table 4.5, where K is the maximum number of regenerators allowed in the network. Costs for translucent cases are always less than the costs for the transparent case. So, using MLR, the network cost can be reduced if one chooses translucent design with effective regenerator placement. This is

Table 4.4 Distribution of regenerators

Node ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$K = 10$	–	–	40 $G = 1$	40 $G = 2$	–	40 $G = 2$ 100 $G = 1$	–	40 $G = 2$	–	–	–	–	–	–
$K = 20$	–	40 $G = 1$ 100 $G = 2$	40 $G = 1$ 100 $G = 1$	40 $G = 4$ 100 $G = 2$	–	40 $G = 2$ 100 $G = 1$	–	40 $G = 5$	–	–	–	–	40 $G = 1$	–
$K = 40$	–	40 $G = 3$ 100 $G = 2$	40 $G = 1$ 100 $G = 1$	40 $G = 6$ 100 $G = 4$	100 $G = 3$	40 $G = 2$ 100 $G = 2$	40 $G = 1$ 100 $G = 5$	40 $G = 5$ 100 $G = 1$	100 $G = 2$	–	–	–	40 $G = 1$	100 $G = 1$

Table 4.5 Transparent vs. translucent MLR network: cost comparison

Total Traffic	1 Tbps	2 Tbps	3 Tbps	4 Tbps	5 Tbps	6 Tbps
$K = 0$ (transp.)	58	115	173	231	289	347
$K = 10$	57	114	172	229	286	343
$K = 20$	57	114	172	229	285	342

because regeneration increases the possibility of more high-bit-rate paths and lowers the transponder costs through volume discount. The extra cost of regenerators (which is 1.4x the cost of a transponder with same line rate here) is compensated by the decrease in the transponder cost. For our example, we do not need many regenerators (after $K = 10$, there is no significant improvement).

Note that, the regenerator placement and the RWA problem for SLR networks can be solved as a single-step optimization problem using connectivity graphs as proposed in [42]. A similar approach with appropriate modifications can also be applied in case of MLR networks. But, for scalability purposes a two-step approach may be more appropriate. Moreover, the formulation for regenerator placement described above, accounts for regeneration at one intermediate node for a given lightpath. This works fine for set of pre-computed feasible and non-feasible lightpaths for the given network in Fig. 4.2. For networks with larger physical routes where a lightpath may need more than one intermediate regenerations, the formulation can be modified accordingly by modifying Eqs. (4.14), (4.15), and (4.16). Replace $Z_{ek(ij)}$ which is a binary variable that denotes a logical AND between α_{iek} and α_{ekj}, with $Z_{e'e''(ij)}$ which is a logical AND of $\alpha_{ie'k}$, $\alpha_{e'e''k}$, and $\alpha_{e''jk}$. Then Eqs. (4.14), (4.15), and (4.16) can be replaced by:

$$Z_{e'e''(ij)} \geq \alpha_{ie'k} + \alpha_{e'e''k} + \alpha_{e''jk} - 1 \quad \forall(e', e'', i, j, k) \tag{4.28}$$

$$Z_{e'e''(ij)} \leq \alpha_{ie'k} \quad \forall(e', e'', i, j, k) \tag{4.29}$$

$$Z_{e'e''(ij)} \leq \alpha_{e'e''k} \quad \forall(e', e'', i, j, k) \tag{4.30}$$

$$Z_{e'e''(ij)} \leq \alpha_{e''jk} \quad \forall(e', e'', i, j, k) \tag{4.31}$$

This signifies the lightpath between nodes ij can now have two regenerations at intermediate nodes e' and e''. Similar extensions can be made for three or more intermediate regenerations.

4.3 MLR Design with Multiple Modulation Formats

As discussed earlier, the set of feasible lightpaths can be populated with more high-bit-rate paths if advanced modulation formats (such as differential quadrature phase-shift keying (DQPSK) [18]) are used to increase their transmission reach.

Table 4.6 Relative cost comparison for MLR with single and multiple modulation formats

Traffic (Tbps)	MLR with Duobinary	MLR-MMF (Duobinary + DQPSK)
1	79	64
2	153	129
3	226	194
4	299	259
5	371	323
10	734	646
20	1,482	1,292

Using mixed modulation formats (MMF) in the network is another approach which, similar to the translucent design in Sect. 4.4.2, can further reduce the cost of the MLR network compared to the transparent MLR network design. A preliminary study [18], assuming that a DQPSK transceiver is 1.5 times costlier than a duobinary transceiver [43], shows that the network cost can be significantly reduced using a mixed (MLR/MMF) transmission system.

Table 4.6 shows the network design cost, calculated as the sum of the transceivers' cost for different scenarios. These results are obtained from a similar ILP model as described before in Sect. 4.4.1 with appropriate modifications made due to MMF. Details can be found in [18]. Here, regeneration is not considered and grooming, as mentioned before, is at the source only. We observe that, for low traffic load, the difference between the total cost of a MLR-MMF network and a MLR network with duobinary is not much, since most long-distance high-bit-rate paths on DQPSK do not need to be lit up. But, when traffic load becomes higher, the fraction of traffic routed over long-distance routes increases and the usage of high-bit-rate DQPSK-modulated paths becomes convenient for the MMF (duobinary + DQPSK) case. Thus, for higher traffic, MMF design achieves lower total cost compared to the case with only duobinary modulation format. (Note: the analysis is based on recent data available on the relative cost of the transceivers for the two different modulation formats [43]). Thus, the network cost can be reduced by having a mixed transmission system.

Another approach for multiple modulation formats in MLR networks is to use the basic on–off keying (OOK) for the 10 Gbps line rate, DQPSK for 40 Gbps line rate, and dual-polarized QPSK (DP-DQPSK) for the 100 Gbps line rate [6]. This would make all the line rates fit well without much bandlimiting, in the existing 50-GHz optical grid that already exists for the legacy systems.

5 Summary

Mixed line rate (MLR) optical networks seem to be the next paradigm for addressing the heterogeneity and increasing capacity of telecom backbone networks. Managing such different capacities in the physical layer is a challenging

task and hence the cross-layer design of MLR networks is gaining importance as a research problem. In this chapter, several efforts towards MLR optical network research have been reported and some minimum-cost design approaches have been described for transparent and translucent MLR networks.

References

1. Simmons JM (2005) On determining the optimal optical reach for a long-haul network. IEEE/OSA J Lightwave Tech 23(3):1039–1048
2. Berthold J, Saleh AAM, Blair L, Simmons JM (2008) Optical networking: past, present, and future. IEEE/OSA J Lightwave Tech 26(9):1104–1118
3. Batayneh M, Schupke DA, Hoffmann M, Kirstaedter A, Mukherjee B (2008) Optical network design for a multiline-rate carrier-grade Ethernet under transmission-range constraints. IEEE/OSA J Lightwave Tech 26(1):121–130
4. Nag A, Tornatore M (2008) Transparent optical network design with mixed line rates. In: Proceedings of IEEE advanced networks and telecommunication systems (ANTS), Mumbai, India
5. Meusburger C, Schupke D (2009) Optimizing the migration of channels with higher bit-rates. In: Proceedings of optical fiber communication conference (OFC), Postdeadline Papers
6. Wuth T, Chbat M, Kamalov V (2008) Multi-rate (100G/40G/10G) transport over deployed optical networks. In: Proceedings of national fiber optic engineers conference, OSA Tech. Dig. (CD), San Diego, CA
7. Shayani D, Machuca CM, Jager M, Gladisch A (2008) Cost analysis of the service migration problem between communication platforms. In: Proceedings of NOMS 2008, pp. 734–737
8. Schmidt T, Malouin C, Saunders R, Hong J, Maroccia R (2008) Mitigating channel impairments in high capacity serial 40G and 100G WDM transmission systems. In: Proc. IEEE/LEOS Summer Topical Meetings, pp. 141–142, July 2008
9. Zulkifli N, Guild K (2007) Moving towards upgradeable all-optical networks through impairment-aware RWA algorithms. In: Proc. Optical Fiber Communication Conference (OFC), March 2007
10. Bissessur H, Hugbart A, Ruggeri S, Bastide C (2005) 40G over 10G infrastructure - dispersion management issues. In: Proc. Optical Fiber Communication Conference (OFC), March 2005
11. Chandrasekhar S, Liu X (2007) Impact of channel plan and dispersion map on Hybrid DWDM transmission of 42.7-Gb/s DQPSK and 10.7-Gb/s OOK on 50-GHz grid. IEEE Photonic Tech Lett 19(22):1801–1803
12. Meusburger C, Schupke DA, Eberspcher J (2008) Multiperiod planning for optical networks - approaches based on cost optimization and limited budget. In: Proc. ICC08, Beijing, May 2008
13. Strauss S, Kirstaedter A, Schupke DA (2006) Multi-period planning of WDM-networks: comparison of incremental and EoL approaches. In: Proc. 2nd IEEE/IFIP International Conference in central Asia on Internet (ICI), pp. 1–7, Sept. 2006
14. Casier K, Colle D, Verbrugge S, Demeester P (2007) Impact of reliability constraints and operational expenses on equipment planning. In: Proc. DRCN, La Rochelle, France, Oct. 2007
15. Azodolmolky S, Klinkowski M, Marn-Tordera E, Careglio D, Sole-Pareta J, Tomkos I (2009) A survey on physical layer impairments aware routing and wavelength assignment algorithms in optical networks. Comput Netw 53(6):926–944
16. Saradhi CV, Subramaniam S (2009) Physical Layer Impairment Aware Routing (PLIAR) in WDM optical networks: issues and challenges. IEEE Comm Surv Tutorials 11(4):109–130, Fourth Quarter
17. Nag A, Tornatore M (2009) Transparent vs. translucent optical network design with mixed line rates. In: Proc. Optical Fiber Communication Conference (OFC), March, 2009

18. Nag A, Tornatore M, Mukherjee B (2010) Optical network design with mixed line rates and multiple modulation formats. IEEE/OSA J Lightwave Tech 28(4):466–475
19. Nag A, Tornatore M, Mukherjee B (2010) Power management in mixed line rate optical networks. In: Proc. Photonics in Switching, OSA Technical Digest (CD), Optical Society of America, Karlsruhe, Germany, 2010
20. Gao G, Zhang J, Gu W, Feng Z, Ye Y (2010) Dynamic power control for mixed line rate transparent wavelength switched optical networks. In: Proc. ECOC 2010, Torino, Italy, Sept. 2010
21. Chen Y, Hua N, Zheng X, Zhang H (2010) Dynamic connection provisioning in mixed-line-rate optical networks. In: Proc. Communications and Photonics Conference and Exhibition (ACP), 2010 Asia, Shanghai, China, Dec. 2010
22. Liu M, Tornatore M, Mukherjee B (2010) New and improved strategies for optical protection in mixed-line-rate WDM networks. In: Proc. IEEE/OSA OFC 2010, San Diego, CA, March 2010
23. Liu M, Tornatore M, Mukherjee B (2011) Efficient shared subconnection protection in mixed-line-rate optical WDM networks. In: Proc. Optical Fiber Communication Conference (OFC), Los Angeles, CA, March 2011
24. Simmons JM (2007) Cost vs. capacity tradeoff with shared mesh protection in optical-bypass-enabled backbone networks. In: Proc. Optical Fiber Communication Conference (OFC), March 2007
25. Chowdhury P, Tornatore M, Mukherjee B (2010) On the energy efficiency of mixed-line-rate networks. In: Proc. Optical Fiber Communication Conference (OFC), March 2010
26. Chowdhury P, Nag A, Tornatore M, Ip E, Wang T, Mukherjee B (2011) Mixed-Line-Rate (MLR) optical network design considering heterogeneous fiber dispersion maps. In: Proc. 37th European conference and exposition on optical communications (ECOC), OSA Technical Digest (CD)
27. Ramaswamy R, Sivarajan K (2002) Optical networks-a practical perspective," 2nd edn. Morgan Kaufman, Burlington, MA
28. Agarwal GP (2002) Fiber-optic communication systems, 3rd edn. Wiley, New York
29. Saunders R, Nicoll G, Wollenweber K, Schmidt T (2007) Can 100 Gb/s wavelengths be deployed using 10 Gb/s engineering rules. http://www.opnext.com/applications/papers/100G_Optics_East_Saunders_Nicholl_OE07IT402-15v2.pdf
30. Winzer PJ, Essiambre R-J (2006) Advanced modulation formats for high-capacity optical transport networks. IEEE/OSA J Lightwave Tech 24(12):4711–4728
31. Leplingard F, Morea A, Zami T, Brogard N (2009) Interest of an adaptive margin for the quality of transmission estimation for lightpath establishment. In: Proc. Optical Fiber Communication Conference (OFC), Post-Deadline Papers, March 2009
32. Roudas I, Antoniades N, Otani T, Stern TE, Wagner RE, Chowdhury DQ (2002) Accurate modeling of optical multiplexer/demultiplexer concatenation in transparent multiwavelength optical networks. IEEE/OSA J Lightwave Tech 20(6):921–936
33. Ramamurthy B, Datta D, Feng H, Heritage J, Mukherjee B (1999) Impact of transmission impairments on teletraffic performance of wavelength-routed optical networks. IEEE/OSA J Lightwave Tech 17(10):1713–1723
34. Datta D, Ramamurthy B, Feng H, Heritage J, Mukherjee B (1998) BER-based call admission in wavelength-routed optical networks. In: Proc. Optical Fiber Communication Conference (OFC), Feb. 1998
35. Forestieri E, Prati G (2001) Novel optical line codes tolerant to fiber chromatic dispersion. IEEE/OSA J Lightwave Tech 11(19):1675–1684
36. Cantrell CD (2003) Transparent optical metropolitan area network. In: Proc. 16th Annual Meeting IEEE/Laser Electro Optics Society, Tucson, AZ, 2003
37. Ip E (2010) Nonlinear compensation using backpropagation for polarization-multiplexed transmission. IEEE/OSA J Lightwave Tech 28(6):939–951

38. Christodoulopoulos K, Manousakis K, Varvarigos E (2010) Offline routing and wavelength assignment in transparent WDM networks. IEEE/OSA J Lightwave Tech 18(5):1326–1334
39. Pachnicke S, Reichert J, Spalter S, Voges E (2006) Fast analytical assessment of the signal quality in transparent optical networks. IEEE/OSA J Lightwave Tech 24(2):815–824
40. Sambo N, Secondini M, Cugini F, Bottari G, Iovana P, Cavaliere F, Castoldi P (2011) Modeling and distributed provisioning in 10-40-100-Gb/s multirate wavelength switched optical networks. IEEE/OSA J Lightwave Tech 29(9):1248–1257
41. Simmons JM (2008) Optical network design and planning. Springer, New York
42. Rizzelli G, Musumeci F, Tornatore M, Maier G, Pattavina A (2011) Wavelength-aware translucent network design. In: Proc. Optical Fiber Communication Conference (OFC), March 2011
43. Lemus A (2008) Optical digital communications technology and modulation formats. http://www.nsc.liu.se/nsc08/pres/lemus.pdf, October, 2008

Chapter 5
Considering Linear and Nonlinear Impairments in Planning WDM Networks

Konstantinos Christodoulopoulos and Emmanouel Varvarigos

1 Introduction

In a wavelength division multiplexed (WDM) network, each fiber link carries high-rate data on several different wavelengths, thus creating multiple channels within a single fiber. The most common architecture utilized for establishing communication in WDM optical networks is wavelength routing, where optical pulse trains are transmitted through *lightpaths*, that is, all-optical WDM channels that may span multiple consecutive fibers [1–3]. Since the lightpaths are the basic switched entities of a wavelength-routed WDM network, their effective establishment and usage is crucial. Thus, it is important to find efficient algorithms to select the routes for the requested connections and to assign wavelengths on each of the links along these routes, so as to serve the traffic and optimize a certain performance metric. This is known as the *routing and wavelength assignment* (RWA) problem. The constraints are that paths that share common links are not assigned the same wavelength (*distinct wavelength assignment*). Also a lightpath, in the absence of wavelength converters, must be assigned a common wavelength on all the links it traverses (*wavelength continuity constraint*).

The RWA problem is usually considered under two alternative traffic models. When the set of connection requests is known in advance (e.g., given in the form of a traffic matrix), the problem is referred to as *offline* or *static* RWA and is encountered in the planning phase of a WDM network, while when the connection requests arrive at random times, over an infinite time horizon, and are served on a one-by-one basis, the problem is referred to as *online* or *dynamic* RWA and is encountered in the operation phase. In this chapter we will focus on offline RWA, which is known to be a NP-hard problem [4]. Offline RWA is generally considered more difficult than online RWA, since it aims at jointly optimizing the lightpaths

K. Christodoulopoulos (✉) • E. Varvarigos
Computer Engineering and Informatics Department, University of Patras, Patras, Greece
e-mail: {kchristodou,manos}@ceid.upatras.gr

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*,
Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_5,

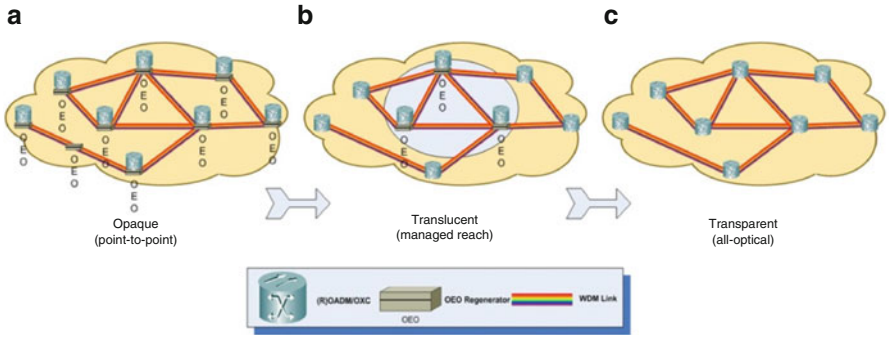


Fig. 5.1 Evolution of optical networks

used by the connections, in the same way that the multicommodity flow problem is more difficult than the shortest path problem in general networks.

Optical core networks in the past were mainly point-to-point (opaque) networks, with signal regeneration taking place at each intermediate node via optical–electronic–optical (OEO) conversion. As the size of opaque networks increased, network designers had to consider more electronic-terminating and switching equipment, increasing capital costs (CAPEX), heat dissipation, power consumption, physical space requirements, and operation and maintenance costs (OPEX). The current trend clearly shows an evolution towards more transparent networks that use fewer OEO transponders (Fig. 5.1). In such WDM networks, where lightpaths are not regenerated at each hop but remain in the optical domain for more than one link, physical limitations of fibers and optical components affect the quality of transmission (QoT) of the lightpaths. We will refer to such phenomena as physical-layer impairments (PLIs). Because of PLIs, signal quality may degrade to the extent that the bit error rate (BER) at the receiver is so high that signal detection is infeasible. For the remaining of this chapter, we will refer to such a phenomenon as *physical-layer blocking*, as opposed to the *network-layer blocking* that arises from the unavailability of an adequate number of wavelengths.

At present, and in the foreseeable future, the only satisfactory method to overcome these impairments is regeneration through OEO conversion (reamplifying, reshaping, and retiming of optical pulses—referred to as 3R regeneration). The network cost of an opaque network could be reduced by employing regenerators only at specific nodes instead of all the nodes. When regenerators are available, a lengthy end-to-end connection that needs regeneration at some intermediate node(s) is set up in a multi-segment manner so that it is served by two or more consecutive lightpath segments. The regenerator at the end of each lightpath segment serves as a "refueling station" that restores signal quality. This type of optical networks, where some lightpaths are routed transparently, while others have to go through a sequence of 3R regenerators, is referred to as a *translucent* optical networks. The ultimate goal is the development of an all-optical *transparent*

network, where the data signal remains in the optical domain for the entire path [5]. Since this may not be feasible for large-sized network, translucent networks seem a more appropriate solution at present and in the near future.

Clearly, PLIs limit the set of paths that can be used for routing. This interdependence between the physical and the network layers makes the RWA problem in the presence of impairments a cross-layer optimization problem. To address this problem, a number of approaches are emerging, usually referred to as impairment-aware (IA)-RWA algorithms. An important distinction is how the IA-RWA algorithms define the interaction between the network layer and the physical layer and if they jointly optimize the solutions over these two layers. Because of interference-related impairments, routing decisions made for one lightpath affect and are affected by the decisions made for other lightpaths. This interdependence is particularly difficult to formulate in offline IA-RWA where there are no already established connections and the utilization of the lightpaths are the variables of the problem. It is because of this difficulty that the majority of offline IA-RWA algorithms proposed to date do not handle interference-related impairments. In this chapter, we present offline IA-RWA algorithms for transparent and translucent networks that account for interference physical effects and jointly optimize the solution over the network and the physical layers.

In this chapter we will present an algorithm for solving the "impairment-unaware" (IU)-RWA problem (i.e., without considering physical impairments) that uses path-related variables and is based on a linear programming (LP)-relaxation formulation [6, 7]. To obtain good integrality performance, the presented formulation uses a piecewise linear cost function and a random perturbation technique. If, even with these techniques, the returned solution is not integer, we use iterative fixing and rounding techniques to obtain an integer solution.

The IU-RWA algorithm is next extended so as to also handle physical-layer impairments in a transparent WDM network, that is, a network that does not employ regeneration at any intermediate node. We present the Sigma-Bound IA-RWA algorithm that considers *directly* the physical impairments and performs a cross-layer optimization between the network and the physical layers [7]. For each candidate lightpath inserted in the formulation, we calculate a noise variance bound based on the impairments that do not depend on the interference among lightpaths. We use this bound and noise variance parameters to define appropriate constraints that limit the total interference noise accumulated on the lightpath. If the selected lightpaths satisfy these constraints, they have, by definition, acceptable quality of transmission. Our goal is to provide practical IA-RWA algorithms that can be used in real network and traffic scenarios. So, although the formulation presented can be solved optimally by ILP methods for small problems, we focus on LP relaxation combined with appropriate techniques for obtaining integer solutions that scale well and can give near-optimal solutions to problems of large sizes, within acceptable time.

We then turn our attention to translucent WDM networks. In such networks apart from the RWA problem, there is also the additional problem of choosing the connections that have to be served using regenerators and the sequence of regenerators

these connections should use. We present a solution based on decomposition: we formulate the problem of regenerator placement as a virtual topology design problem and address it using various algorithms, ranging from integer linear programming (ILP) formulations to simple greedy heuristics [8]. Once the sequence of regenerators to be used by the nontransparent connections has been determined, we transform the initial traffic matrix by replacing nontransparent connection requests with a sequence of transparent connection requests that terminate and begin at the specified 3R intermediate nodes. Using the transformed matrix, we then apply the transparent Sigma-Bound IA-RWA algorithm, outlined above, to route the traffic. We also present simulation results to evaluate the efficiency and applicability of the presented algorithms under realistic scenarios.

The rest of this chapter is organized as follows. In Sect. 2 we overview previous work on RWA and IA-RWA. In Sect. 3, we discuss ways in which PLIs can be incorporated in the RWA problem. In Sect. 4 we give our IU-RWA formulation. We then extend it and present in Sect. 5 the Sigma-Bound IA-RWA algorithm for transparent networks. In Sect. 6, we present a number of algorithms for planning translucent networks, as opposed to transparent networks of Sect. 5. Simulation results are given in Sect. 7. Our conclusions follow in Sect. 8.

2 Previous Work

Routing and wavelength assignment (RWA) has been extensively studied in the literature. The offline version of RWA is known to be a NP-complete problem [4]. To make computations tractable, a common approach is to decouple the RWA into its constituent routing (R) and wavelength assignment (WA) subproblems by first finding routes for all requested connections and then searching for an appropriate wavelength assignment [1, 9]. Note that both subproblems are NP-hard: routing a set of requested connections corresponds to a multicommodity flow problem, while wavelength assignment corresponds to a graph coloring problem. Various efficient heuristics have been developed for each subproblem, which may be combined and produce solutions for the joint RWA problem. However, such decomposition suffers from the drawback that the optimal solution of the (joint) RWA problem might not be included in the solutions provided for the two subproblems.

RWA integer linear programming (ILP) formulations were initially proposed in [10] and [11]. Since the associated ILPs are hard to solve, the corresponding relaxed linear programs (LP) have been used to get bounds on the optimal value. A review on various offline and online RWA algorithms can be found in [9]. A few newer and more sophisticated RWA algorithms are presented in [6, 12, 13]. The LP-relaxation formulation proposed in [12] and in [6] is able to produce exact RWA solutions in many cases, despite the absence of integrality constraints. This is the approach that we adopt and extend in our formulations.

Recently, RWA algorithms that consider the impact of physical-layer impairments have been the subject of intense research, referred to in this chapter as IA-RWA. Most of these studies consider the online (dynamic) version of

the problem [14–19]. Among these online algorithms, there are approaches that consider the quality of transmission (QoT) problem separately from the RWA problem by first solving the RWA problem and then considering PLI effects through the evaluation of the feasibility of the selected lightpaths in a separate module [14–16]. This approach may not yield a solution of acceptable quality, and iterations are usually performed to improve physical-layer blocking performance. Other online approaches incorporate PLIs into the cost function of the optimization problem and also consider the interference among the lightpaths [17–19].

In the dynamic traffic case, where connections are served one by one, the employed algorithm must examine the feasibility of a lightpath that is about to be established. This can be done by calculating (through appropriate models) or measuring (through performance monitors) the interference caused by the already established lightpaths to the candidate lightpath. However, this cannot be done in the static RWA case, where there are no existing connections, and the utilization of lightpaths are the variables of the problem. For this reason, the majority of offline RWA algorithms proposed to date do not consider inter-lightpath interference. In [20], the authors solve the offline impairment-unaware (IU)-RWA problem and then evaluate the feasibility of the chosen lightpaths in a post-processing phase. For requests whose lightpaths do not have acceptable transmission quality, new solutions are found by excluding the ones previously considered. An offline impairment-aware (IA)-RWA algorithm that assigns Q-factor costs to links before solving the problem is proposed in [21]. However, the proposed algorithm does not account for the actual interference among lightpaths and assumes a worst-case interference scenario. In [22], the authors include the optical power in the RWA formulation so as to ensure that the power level at the beginning of each optical amplifier and at the end of each fiber is above a certain threshold.

All the work discussed in the preceding paragraphs refers to the case of transparent networks. The RWA problem for translucent optical networks has also been addressed before by dividing a large-scale optical network into several *islands of transparency* or *optically transparent domains*. Within the same island, a lightpath can transparently reach any other node without intermediate signal regeneration. For communication across islands, 3R regeneration is performed at nodes at the island boundaries [23]. An alternative approach, which is the one taken in this chapter, is called *sparse placement of regenerators*, where some selected nodes in the network are used for regeneration and there is no explicit notion of transparent domains. The majority of RWA algorithms proposed so far for translucent networks assume a dynamic traffic scenario, that is, they address the online version of the problem. [24] presents a two-dimensional Dijkstra RWA algorithm for translucent optical network that takes into account the placement of regenerators and a constraint on the maximum transparent distance. When the length of a lightpath exceeds the maximum transparent distance bound, the lightpath is blocked. A different approach for dynamic resource allocation and routing is considered in [25] and [26], where spare transceivers (transmitter–receiver pairs connected back-to-back) at the nodes are used to regenerate signals. In this kind of network, every node with a *spare transceiver* is a *potential regenerator*. A *Max-spare algorithm* for selecting the

regeneration nodes for a lightpath is proposed in [27] and compared to a *greedy* algorithm used in conjunction with a wavelength-weighted and a length-weighted RWA algorithm. In [28], two online RWA algorithms for translucent networks with sparse regenerator placement are presented that assume either worst-case physical transmission penalties corresponding to a fully loaded system or that take into account the current network status and the actual number of active channels.

The work in [29] formulates the problem of maximizing the number of connections served, under a constraint on the maximum transparent length, as a *mixed-integer linear program* (MILP). Since MILP does not scale a heuristic algorithm to establish connections is also proposed. A heuristic for placing the fewest regenerator nodes to reach a given blocking probability for dynamic traffic, based on the ranked frequency of shortest routes, is given in [30]. In [31] the authors address the problem of translucent network design by proposing several regenerator placement algorithms based on different knowledge of future network traffic patterns. A QoT-based heuristic algorithm for IA-RWA in translucent networks is presented in [32]. In the first phase of that algorithm, a random search heuristic RWA algorithm is used, and in the second phase, regeneration placement is performed after estimating the BER of the lightpaths obtained in the first phase.

In the following sections we present a series of offline IA-RWA algorithms that solve the planning problem of heterogeneous transparent and translucent WDM networks. The algorithms presented take into account not only impairments that depend on the chosen lightpath but also impairments that depend on the interference among lightpaths as additional constraints on RWA, so as to perform a cross-layer optimization of the solution over the physical and the network layers.

3 Considering Physical-Layer Impairments in RWA

In transparent and translucent WDM networks, the signal quality of transmission (QoT) degrades due to the nonideal physical layer [2, 3]. Several criteria can be used to evaluate the signal quality of a lightpath. Among a number of measurable optical transmission quality attributes, such as optical power, optical signal-to-noise ratio (OSNR), chromatic dispersion (CD), and polarization mode dispersion (PMD), the Q-factor seems to be more suitable as a metric to be integrated in an RWA algorithm because of its immediate relation to the bit error rate (BER). In this chapter we look only at on–off keying (OOK) modulation, but similar expressions can be derived for higher-order modulations. The Q-factor is the electrical signal-to-noise ratio at the input of the decision circuit in the receiver's terminal. Assuming Gaussian-shaped noise, the Q-factor of a lightpath (p, w), that is, wavelength w on path p, is given by

$$Q(p,w) = \frac{I_{1}(p,w) - I_{0}(p,w)}{\sigma_{1}(p,w) + \sigma_{0}(p,w)},$$

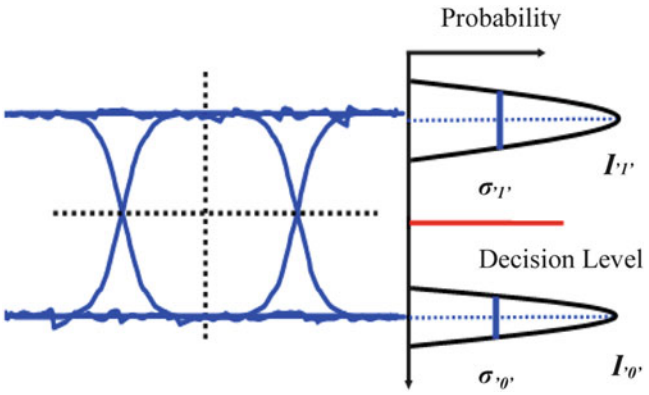


Fig. 5.2 Eye diagram and Q-factor

where $I_{,1,}$ and $I_{,0,}$ are the mean values of electrical voltage of signal 1 and 0, respectively, and $\sigma_{,0,}$ and $\sigma_{,1,}$ are their standard deviations, at the input of the decision circuit at the destination, which in this case is the end of path p. Figure 5.2 illustrates the relation between an eye diagram and the Q-factor. The higher the Q-factor, the smaller the BER is and the better the quality of the signal. Generally, a path has acceptable QoT when the Q-factor at the destination is higher than 15.5 dB. When forward error correction (FEC) is employed, a connection can be accepted with even smaller Q values.

Physical-layer impairments (PLIs) are usually categorized to linear and nonlinear, according to their dependence on the power. However, when considering IA-RWA, it is useful to categorize the PLIs into those that affect the same lightpath that generated them and those that are due the presence of other lightpaths, resulting in the following two classes for the most important PLIs:

- *Class 1–impairments that affect the same lightpath*: amplified spontaneous emission noise (ASE), polarization mode dispersion (PMD), chromatic dispersion (CD), filter concatenation (FC), self-phase modulation (SPM)
- *Class 2–impairments that are generated by other lightpaths*: crosstalk (XT) (intra-channel and interchannel crosstalk), cross-phase modulation (XPM), four-wave mixing (FWM)

We now discuss the ways in which PLIs can be incorporated in the RWA problem. PLIs of class 1 depend only on the selected lightpath and can be accounted for quite easily. We will assume that we are using a typical RWA algorithm that takes as input a set of candidate lightpaths and selects an appropriate subset of lightpaths to serve the connections. Then, for each candidate lightpath, we can pre-calculate the effects of the PLIs of class 1, using, for example, analytical models, and discard those that have unacceptable QoT performance. In this way, the RWA algorithm can be fed with candidate solutions that are acceptable at least for the impairments of class 1. PLIs of class 2 are more difficult to consider in

offline algorithms since they make decisions taken for one lightpath effect and be affected by decisions taken for other lightpaths. An obvious simplification is to consider a "worst-case scenario," by assuming that all wavelengths on all links are active, and calculating the worst-case interference on each lightpath. Lightpaths that do not have acceptable QoT under this worst-case assumption can be discarded before the RWA algorithm is executed, ensuring that the solution is feasible, irrespective of the final selection of lightpaths.

This worst-case approach does not optimize the problem for the given traffic but solves it as if the network was fully utilized, which will never happen. In practice, wavelength continuity limits the achievable network utilization, except for the degenerate case where all connections are between adjacent nodes. The worst-case assumption results in discarding many candidate lightpaths that are not really infeasible. The feasibility or not of these lightpaths depends on the lightpaths finally selected in the solution, which are not known before the RWA execution. By formulating the interference among lightpaths, we may be able to use these lightpaths, letting the algorithm choose the optimal among all feasible solutions.

In the following, we quantify through a realistic example the reduction in the routing solution space when PLIs are considered. We assume the Deutsche Telekom network topology (DTnet), shown in Fig. 5.3, with physical-layer parameters chosen to have realistic values. We have also used a quality of transmission evaluation module (Q-Tool) developed within the DICONET project [33] that uses analytical models to account for the most important PLIs (and in particular, the ones listed in the above classification). We assume an offline RWA algorithm that takes as input a set of candidate paths and selects a subset of them to satisfy the demands. In this example, we assume that all source–destination pairs require a connection of one wavelength for a total number of connection demands equal to 182. Initially, we calculate the k-shortest length paths for different values of the parameter k, and then we prune the set of candidate paths using the Q-Tool. In doing so, we either assume an empty network, discarding lightpaths that are infeasible due to impairments of class 1, or we assume a fully utilized network, discarding lightpaths that are infeasible due to impairments of class 1 and of class 2 under the worst-case interference scenario.

In Table 5.1 we see the reduction in the routing solution space after considering the PLIs. As we increase the initial path population size, by increasing parameter k, the percentage of paths that are discarded increases. This was expected since the path length largely determines the physical-layer effects, and thus as we move from the shortest to more lengthy paths, an increasing percentage of these paths turns out to be infeasible. Eventually, beyond some value of k, all paths will turn out to be infeasible, and the population reported in columns (b) and (c) will stabilize. Table 5.1 shows that the path population obtained after eliminating candidate paths due to the impairments of class 1 (column (b)) is considerably larger than when we use the worst-case interference assumption for the impairments of class 2 (column (c)). An IA-RWA algorithm that accounts for the actual interference among lightpaths can use as input the paths that correspond to column (b), while an algorithm that assumes a worst-case interference scenario will use as input the paths

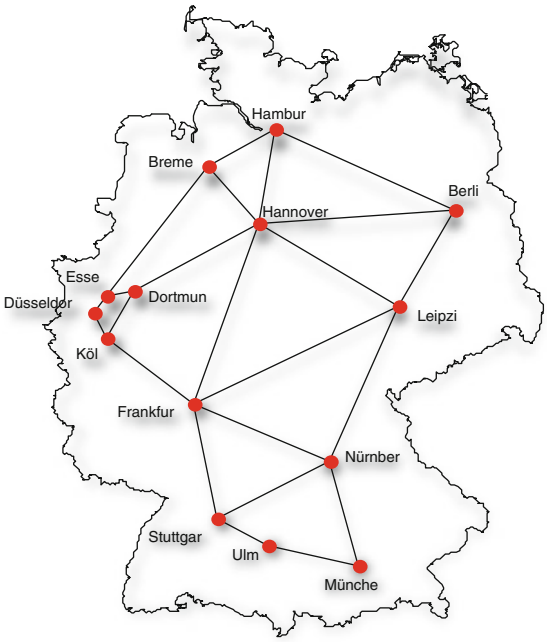


Fig. 5.3 Generic DT network topology, consisting of 14 nodes and 46 unidirectional links

Table 5.1 The reduction in the solution space due to the PLIs of class 1 and 2 (under the worst-case interference assumption) for the case of the DTnet and the reference traffic matrix

	(a) Initial population (k-shortest length paths)	(b) Population after discarding paths due to impairments of class 1	(c) Population after discarding paths due to impairments of class 1 and 2–assuming worst-case interference
$k = 1$	182	182	182
$k = 2$	364	359	333
$k = 3$	546	528	427
$k = 4$	728	653	479
$k = 5$	910	751	506
$k = 6$	1092	817	513

of column (c). This reduction in the solution space unnecessarily restricts the choices of the algorithm and leads to a deterioration of its performance.

Based on the above discussion, an IA-RWA algorithm that accounts for the actual inter-lightpath interference and performs a cross-layer optimization over the physical and the network layers will require fewer wavelengths to serve the given traffic than an IA-RWA algorithm that assumes a worst-case interference scenario. We will come back and quantify this performance difference later in this chapter. The reduction in the solution space due to the physical effects depends on the topology and the traffic profile. For example, in a small network that consists of short-length links, PLIs are negligible, and the difference between the population sizes corresponding to columns (a), (b), and (c) will be small, even for high values of k. Thus, the above discussion is valid when PLIs are significant enough, which is the case of interest in this chapter.

A frequently used argument in favor of the worst-case interference scenario is that a network configured using this assumption is more reliable, since future connections will not render the already established lightpaths infeasible. In current operational WDM networks, changes in the traffic matrix are usually slow and incremental. The increase of the average traffic between two end nodes is eventually going to require an additional wavelength, but the average traffic increase is small compared to the granularity of the wavelength, and the frequency of such incremental changes is low, of the order of a couple of new connections per year. For slowly changing traffic, an online IA-RWA algorithm is executed to adjust the previous solution to the new traffic demands, and/or an offline algorithm can be periodically re-executed after a certain time or a certain number of new established lightpaths to optimize the solution to the current traffic. It stands to reason that the establishment of new lightpaths under the worst-case interference assumption would continue making inefficient utilization of the available wavelengths, and the network will reach the point where it cannot serve more connections earlier than in the case where an IA-RWA that accounts for the actual inter-lightpath interference and performs a cross-layer optimization is used.

4 Impairment-Unaware RWA Algorithm

In this section we present an algorithm for solving the RWA problem without considering physical-layer impairments. In the subsequent section, we extend this algorithm in order to take into account physical-layer impairments and address the planning problem of transparent and translucent WDM networks.

The network is represented by a connected graph $G = (V,E)$. V denotes the set of nodes, assumed not to be equipped with wavelength converters. E denotes the set of single-fiber links. Each fiber supports a common set $C = \{1,2,\ldots,W\}$ of distinct wavelengths. The static version of RWA assumes an a priori known traffic scenario given in the form of a nonnegative integer matrix Λ, called the traffic matrix. Then, Λ_{sd} denotes the number of requested wavelengths from source s to destination d.

Note that there may be multiple lightpath requests for a given source–destination pair (s,d) and they can be routed over different paths.

The algorithm takes as input an RWA instance, that is, a network topology, a set of available wavelengths, and a traffic matrix. It returns the RWA solution, in the form of routed lightpaths and assigned wavelengths, and the blocking probability, in case the connections cannot all be served for the given set of wavelengths.

The "impairment-unaware" (IU)-RWA algorithm consists of four phases [6]. The first (preprocessing) phase computes a set of candidate paths to route the requested connections. RWA algorithms that do not use any set of predefined paths but allow routing over any feasible path (using multicommodity flow formulations) have also been proposed in the literature. These algorithms are bound to give at least as good solutions as the algorithms that use pre-calculated paths, such as the one proposed here, but use a much higher number of variables and constraints and do not scale well. In any case, the optimal solution can be also found with a RWA algorithm that uses pre-calculate paths, given a large enough set of paths. The second phase of the proposed algorithm utilizes Simplex to solve the LP that formulates the given RWA instance. If the solution returned by Simplex is not integer, the third phase uses iterative fixing and rounding techniques to obtain an integer solution. Non-integer solutions are not acceptable, since a connection is not allowed to bifurcate over alternative paths or wavelength channels. Finally, phase 4 handles the infeasible instances so that most (if all is not possible) requested connections are established.

Phase 1: In this phase, k candidate paths are calculated for each connection request using a variation of the k-shortest path algorithm: at each step, a shortest path is selected, and its link costs are increased (doubled in our experiments) so as to be avoided by the paths found in subsequent steps. Paths thus obtained tend to use different edges and be more representative of the path solution space. Other k-shortest path algorithms are also applicable. By selecting a sufficiently large value for k, the solution space is expected to contain an optimal RWA solution with large probability. After a set P_{sd} of candidate paths for each s–d pair is computed, the total set $P = \cup_{s-d} P_{sd}$ is inserted to the next phase. The preprocessing phase clearly takes polynomial time.

Phase 2: Taking into account the network topology, the number of available wavelengths, the traffic matrix, and the set of paths identified in phase 1, phase 2 formulates the given RWA instance as a linear program (LP). The LP formulation used is presented in Sect. 4.1. This LP is solved using the Simplex algorithm that is generally considered efficient for the great majority of all possible inputs and has additional advantages, as we will see, for the problem at hand. If the instance is feasible and the solution is integer, the algorithm terminates by returning the optimal solution and zero blocking. If the instance is feasible but the solution is non-integer, we proceed to phase 3. If the instance is infeasible, meaning that it cannot be solved for the given number of wavelengths, we proceed to phase 4.

Phase 3: In case of a fractional (non-integer) solution, phase 3 involves iterative fixing and rounding methods, as presented in Sect. 4.3, in order to obtain an integer solution. The maximum number of such iterations is the number of connection

requests, which is polynomial on the size of the input. Rounding can turn the problem infeasible, and then we proceed to phase 4. If we find a feasible solution, the algorithm terminates and outputs the RWA solution and zero blocking.

Phase 4: This phase is used when the LP instance is infeasible for the given number of wavelengths W. Infeasibility is overcome by progressively increasing the number of available wavelengths and re-executing phases 2 and 3 until a feasible solution is obtained. Then, at the end of phase 4, we have to select which connections should be blocked to reduce the number of wavelengths to the given W. The wavelengths removed are those occupied by the smallest number of lightpaths, so as to minimize blocking. The algorithm terminates and outputs the RWA solution, along with the blocking probability, which is in that case strictly greater than zero.

4.1 RWA LP Formulation

The proposed LP formulation aims at minimizing the maximum resource usage in terms of wavelengths used on network links. Let $F_l = f(w_l)$ denote the flow cost function, an increasing function on the number of lightpaths w_l traversing link l (the used formula is presented in the next subsection). The LP objective is to minimize the sum of all F_l values. The following variables are used:

- x_{pw}: an indicator variable, equal to 1 if path p occupies wavelength w, that is, if lightpath (p,w) is activated, and equal to 0, otherwise
- F_l: the flow cost function value of link l

$$\text{Minimize} \quad \sum_{l} F_l$$

subject to the following constraints:

- Distinct wavelength assignment constraints: $\sum_{\{p|l\in p\}} x_{pw} \leq 1$, for all $l \in E$ and all $w \in C$
- Incoming traffic constraints: $\sum_{p\in P_{sd}} \sum_{w} x_{pw} = \Lambda_{sd}$ for all (s,d) pairs
- Flow cost function constraints: $F_l \geq f(w_l) = f\left(\sum_{\{p|l\in p\}} \sum_{w} x_{pw}\right)$ for all $l \in E$
- The integrality constraints are relaxed to $0 \leq x_{pw} \leq 1$ for all $p \in P$ and all $w \in C$

The wavelength continuity constraints are implicitly taken into account by the definition of the path-related variables. Note that using inequalities for the flow cost function constraints in the above formulation is equivalent to using equalities, since these constraints will hold as equalities at the optimal solution. Using inequalities is convenient for employing a piecewise linear cost function f, as will be presented in

the next subsection. In Sect. 5, we will extend the above LP formulation so as to take into consideration the physical-layer impairments.

4.2 Flow Cost Function

The variable F_l expresses the cost of congestion on link l, for a specific selection of the routes, and is chosen to be an increasing function $f(w_l)$ of the number of lightpaths $w_l = \sum_{\{p|l\in p\}} \sum_w x_{pw}$ crossing link l. $F_l = f(w_l)$ is chosen to also be strictly convex (instead of, e.g., linear), implying a greater degree of "undesirability," when a link becomes congested. This is because it is preferable for overall network performance to serve an additional unit of flow using several low-congested links than using a link that is close to saturation. In particular, we utilize the following flow cost function:

$$F_l = f(w_l) = \frac{w_l}{W + 1 - w_l},\ 0 \le w_l \le W.$$

The above (nonlinear) function is inserted to the LP formulation in the approximate form of a piecewise linear function, consisting of W consecutive linear parts. The piecewise linear approximation is constructed as follows: we begin with $F_l(0) = 0$, and iteratively set, for $i = 1,2,\ldots,W$, $F_l{}^i(w_l) = a_i \cdot w_l + \beta_i$, $i - 1 \le w_l \le i$, where $a_i = F_l(i) - F_l(i-1)$ and $\beta_i = (i-1) \cdot F_l(i) - i \cdot F_l(i-1)$. We insert in the LP formulation W linear constraints of the form

$$F_l{}^i(w_l) = a_i \cdot w_l + \beta_i \le F_l,\ i = 1, 2, \ldots, W,$$

defined by the corresponding a_i and β_i values. Since the LP objective is to minimize the cost $\sum_l F_l$, for a specific value of w_l, one of these W linear cost functions, and in particular the one that yields the highest $F_l(w_l)$, is satisfied with equality at the optimal solution of the LP. All the remaining linear functions are deactivated, that is, they are satisfied as strict inequalities at the optimal solution (Fig. 5.4).

This piecewise linear function is equal to the nonlinear function $f(w_l)$ at integer values ($w_l = 1,2,\ldots,W$) and greater than that at non-integer values. Inserting such a piecewise linear function to the LP results in the identification of integer optimal solutions by Simplex in most cases [12]. This is because the vertices of the polyhedron defined by the constraints tend to correspond to the corner points of the piecewise linear function and tend to consist of integer components. Since Simplex moves from vertex to vertex of that polyhedron [34], there is a higher probability of obtaining integer solutions than with other (e.g., interior point) methods. The results in Sect. 7.1 show that this is the case in most problem instances.

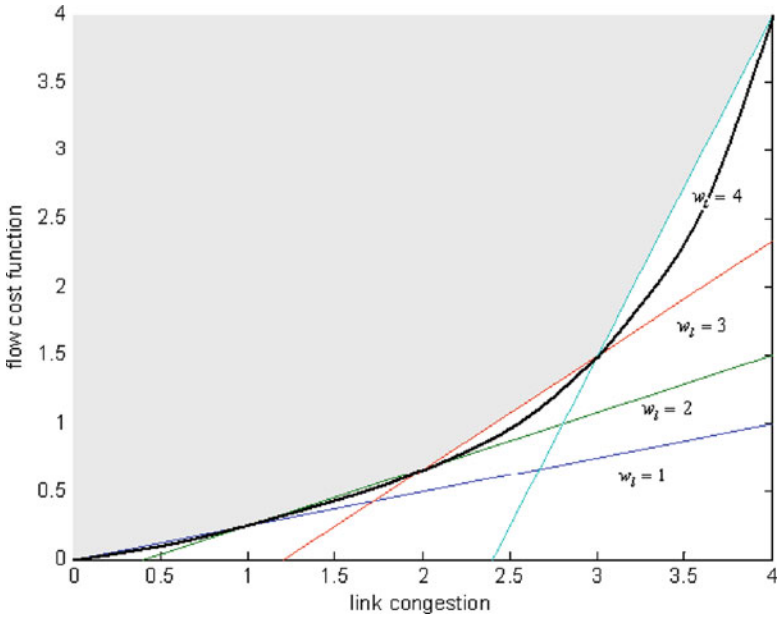


Fig. 5.4 The set of linear constraints inserted in the LP formulation. We use inequality constraints to limit our search in the colored area. Since the objective is to minimize the flow cost, we search for solutions only at its lower bounds, which identify the piecewise linear approximation of the flow cost function $F_l = f(w_l)$ (black line)

4.3 Random Perturbation Technique

Although the piecewise linear cost function presented above is designed so as to yield good integrality characteristics, there are still cases where some of the solution variables of the LP relaxation turn out to be non-integer. Recall that non-integer solutions for the flow variables are unacceptable. To increase the number of integer solutions obtained, we use the following random perturbation technique. In the general multicommodity flow problem, given an optimal fractional solution, a flow that is served by more than one path has equal sum of first derivatives of the costs of the links comprising these paths [35]. The reason is that if they were not equal, one could shift some small flow δ from one path to the other, reducing the total cost, which would mean that it is not an optimal solution. The objective function used in our RWA formulation sums the flow costs of all the links, and thus a request that is served by more than one lightpaths has equal sums of first derivatives over the links of these lightpaths. Note that the derivative of the cost on a specific link is given by the slope of the linear or piecewise linear flow cost function used. To make the situations where two lightpaths have equal first derivative lengths over the links that comprise them less probable, we multiply the slopes on each link with a random number that differs from 1.0 in the sixth decimal digit (see Fig. 5.5).

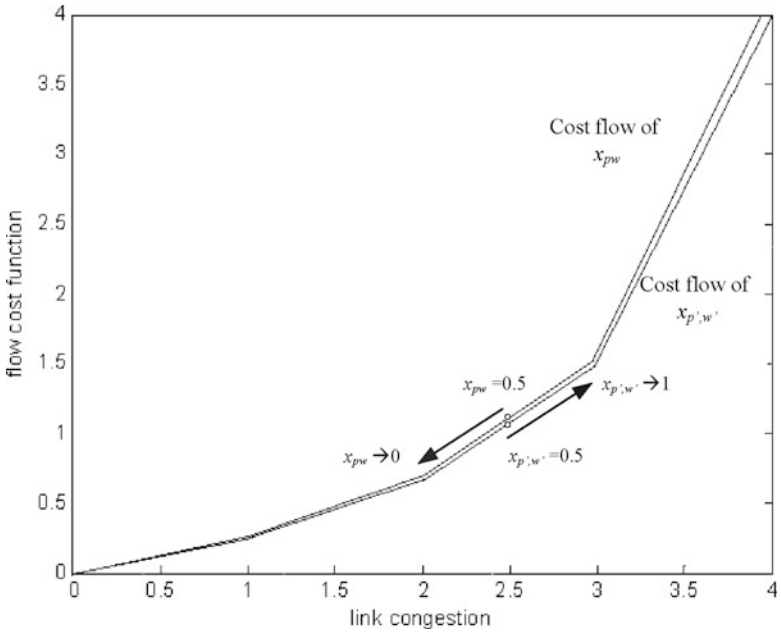


Fig. 5.5 Random perturbation mechanism. The first derivatives of the two variables x_{pw} and $x_{p',w'}$ are not equal on a specific link. Thus, the variable with the smaller derivative is selected (set to one), yielding an integer solution

4.4 Iterative Fixing and Rounding Techniques

If even with the piecewise linear cost function and the random perturbation technique presented above we do not obtain an integer solution, we continue by "fixing" and "rounding" the variables.

We start by fixing the variables; that is, we treat the variables that are integer as final and solve the reduced problem for the remaining ones. Fixing variables does not change the objective cost returned by the LP, so we move with each fixing from the previous solution to a solution with equal or more integers with the same cost. This is because when some variables are fixed and the RWA problem is reduced, Simplex uses different variables and constraints and starts from a different basic feasible solution (bfs) [34]. Thus, it ends with the same objective cost but with a solution that might consist of a different (and in particular greater or equal) number of integer variables. Since the objective cost does not change, if after successive fixings we reach an all-integer solution, we are sure that it is an optimal one. On the other hand, fixing variables is not guaranteed to return an integer optimal solution, if one exists, since the integer solution might consist of different integer values than the ones in the iterative fixing process. When we reach a point beyond which the process of fixing does not increase the integrality of the solution, we proceed to the rounding process. We round a single variable at a time, the one closest to 1, and continue solving the

reduced LP problem. While fixing variables helps us move to solutions with more integer variables and the same objective cost, rounding makes us move to higher objective values and search for an integer solution there. Rounding is inevitable when there is no integer solution with the same objective cost as the LP relaxation of the RWA instance. However, if after rounding the objective cost changes, we are not sure anymore that we will end up with an optimal solution. The maximum number of fixing and rounding iterations performed is the number of requests and is polynomial on the problem size.

We will refer to the algorithm outlined above as an "impairment-unaware" (IU)-RWA algorithm, in the sense that it ignores physical-layer impairments. Using this RWA algorithm and pruning the candidate paths under the worst-case interference assumption in the preprocessing phase (see discussion in Sect. 3), we obtain an impairment-aware (IA)-RWA algorithm that we will refer to as worst-case WC-IA-RWA. Note that the only difference between the RWA and the *WC-IA-RWA* algorithm is the set of candidate lightpaths that they use as input.

5 Impairment-Aware RWA Algorithm for Transparent Networks

In this section we extend the preceding RWA algorithm so as to make it impairment-aware.

5.1 Direct and Indirect IA-RWA Algorithms

To account for the PLIs in a cross-layer approach, an algorithm has to incorporate impairment-related constraints at some point in its formulation. An important distinction is whether these impairment constraints address *directly* or *indirectly* the effects of the impairments. To give an example, polarization mode dispersion (PMD) is proportional to the square root of the length of the path, and also other impairments, such as amplified spontaneous emission (ASE) noise, are affected by the path length. An algorithm that chooses paths with small lengths is bound to exhibit lower PMD and ASE effects; such an algorithm would be treating impairments *indirectly* by bounding the length of the paths. However, when the network is heterogeneous, the length of a path can be misleading even for PMD, since a path with small length is not ensured to have acceptable QoT, as the impairment effects are not verified. On the other hand, an algorithm that includes constraints that bound *directly* the effects of PMD and ASE in its formulation, using, for example, analytical formulas, can be sure (to a certain degree, depending on the other effects that are accounted for) that a lightpath satisfying these constraints has acceptable QoT. A key difference between these two classes

of algorithms is that indirect algorithms need a QoT evaluation tool in order to verify the feasibility of the chosen lightpaths, while direct algorithms do not require an evaluation module, assuming that all dominant impairments are accounted for.

5.2 *Direct IA-RWA Formulation*

We now describe a *direct* IA-RWA algorithm. For each candidate lightpath inserted in the RWA formulation, we calculate an upper bound on the interference noise variance it can tolerate after accounting for the impairments that do not depend on the utilization of the other lightpaths (impairments of class 1, Sect. 3). Then, we use this to bound the interfering noise variance caused by other lightpaths (impairments of class 2) by introducing appropriate constraints in the RWA formulation.

5.2.1 Calculating the Noise Variance Bound of a Lightpath

We start from the Q-factor definition presented in Sect. 3. In the approach adapted here [21, 36], $I_{`1`,p}(w)$ depends on the transmitter's power, the gains and losses over path p, and the "eye impairments": self-phase modulation (SPM), chromatic dispersion (CD), polarization mode dispersion (PMD), and filter concatenation (FC). The remaining impairments are considered as noise or noise-like. For the noise impairments and bits 1 and 0, we have

$$\sigma^2_{`1`}(p,w) = \sigma^2_{\mathrm{ASE},`1`}(p,w) + \sigma^2_{\mathrm{XT},`1`}(p,w) + \sigma^2_{\mathrm{XPM},`1`}(p,w) + \sigma^2_{\mathrm{FWM},`1`}(p,w),$$

$$\sigma^2_{`0`}(p,w) = \sigma^2_{\mathrm{ASE},`0`}(p,w) + \sigma^2_{\mathrm{XT},`0`}(p,w) + \sigma^2_{\mathrm{FWM},`0`}(p,w) \approx \sigma^2_{\mathrm{ASE},`0`}(p,w),$$

where $\sigma^2_{\mathrm{ASE}}, \sigma^2_{\mathrm{XT}}, \sigma^2_{\mathrm{XPM}}$, and σ^2_{FWM} are the electrical noise variances due to amplifier spontaneous emission (ASE), intra-channel crosstalk (XT), cross-phase modulation (XPM), and four-wave mixing (FWM), respectively. Note that the noise variances $\sigma^2_{\mathrm{XT},`0`}, \sigma^2_{\mathrm{FWM},`0`}$, of bit 0 are low and negligible compared to $\sigma^2_{\mathrm{ASE},`0`}$, and especially compared to the corresponding noise variances of bit 1. Also, note that XT, XPM, and FWM depend on the utilization of the other lightpaths (class 2 impairments). Let $Q_{\min}$ be the acceptable threshold for the QoT of a lightpath. Since we do not consider all the physical impairment and we also make various simplifying assumptions (such as not considering XT and FWM for bit 0), we will use a $Q_{\min}$ that is somewhat higher than the desired threshold. We have

$$\frac{I_{\text{`}1}(p,w)}{\sigma_{\text{`}1}(p,w)+\sigma_{\text{`}0}(p,w)} \geq Q_{\min} \Rightarrow \sigma^2_{\text{`}1}(p,w) \leq \left(\frac{I_{\text{`}1}(p,w)}{Q_{\min}} - \sigma_{\text{`}0}(p,w)\right)^2$$

$$\Rightarrow \sigma^2_{\text{XT,`}1}(p,w) + \sigma^2_{\text{XPM,`}1}(p,w) + c_{\text{FWM}} \leq \left(I_{\text{`}1}(p,w) Q_{\min} - \sigma_{\text{ASE,`}0}(p,w)\right)^2 - \sigma^2_{\text{ASE,`}1}(p,w)$$

$$\Rightarrow \sigma^2_{\text{XT,`}1}(p,w) + \sigma^2_{\text{XPM,`}1}(p,w) + c_{\text{FWM}} \leq \sigma^2_{\max}(p,w),$$

$$\text{where} \quad \sigma^2_{\max}(p,w) \stackrel{\text{def}}{=} \left(\frac{I_{\text{`}1}(p,w)}{Q_{\min}} - \sigma_{\text{ASE,`}0}(p,w)\right)^2 - \sigma^2_{\text{ASE,`}1}(p,w)$$

assuming that FWM contributes a constant c_{FWM} (c_{FWM} is small relative to the other effects and could be chosen as the worst-case FWM contribution, when all wavelengths are used). Thus, for a lightpath (p,w), this inequality gives a bound $\sigma^2_{\max}(p,w)$, calculated based on the impairments that do not depend on the interference among lightpaths (class 1 impairments), which constrains the total interference noise variances of the impairments that depend on the other lightpaths (class 2 impairments).

5.2.2 Defining the Noise Interference Constraints

We assume that for each link l and the optical cross connect (OXC) switch n it ends at, we know the following noise variance parameters:

- G_l (in dB): the power loss/gain of the link/OXC due to fiber attenuation, power leakage, and amplifiers' gains
- $s^2_{\text{1−XPM,`}1\text{`},l}$, $s^2_{\text{2−XPM,`}1\text{`},l}$: the noise variance due to XPM for bit 1 from an active adjacent and second channel, respectively
- $s^2_{\text{XT,`}1\text{`},n}$: the intra-channel crosstalk (intra-XT) noise variance for bit 1, contributed to a lightpath that also crosses switch n and uses the same wavelength

Note that we assume that $s^2_{\text{1−XPM,`}1\text{`},l}$, $s^2_{\text{2−XPM,`}1\text{`},l}$, $s^2_{\text{XT,`}1\text{`},n}$ are the same irrespectively of the wavelength w, but the algorithm can be extended to consider different parameters per wavelength. To obtain the above parameters, analytical models for the specific impairments can be used. For example, for NRZ modulation, we can use models described in [14] and [36].

For a path p that consists of links $l = 1,2,\ldots,m$, we have

$$\sigma^2_{\text{XT,`}1}(p,w) + \sigma^2_{\text{XPM,`}1}(p,w) = \sum_{\{l=1\,|\,n \text{ end of } l\}}^{m} \left(\left(s^2_{\text{XT,`}1\text{`},n} \cdot n_{\text{XT},n}(w) + s^2_{\text{1−XPM,`}1\text{`},l} \cdot n_{\text{adj},l}(w) + s^2_{\text{2−XPM,`}1\text{`},l} \cdot n_{\text{2−adj},l}(w) \right) \cdot \prod_{i=l+1}^{m} 10^{\frac{2 \cdot G_l}{10}} \right).$$

where $n_{\mathrm{XT},n}(w)$ is the number of intra-XT generating sources on switch n and wavelength w, $n_{\mathrm{adj},l}(w)$ and $n_{2-\mathrm{adj},l}(w) \in \{0,1,2\}$ is the number of utilized adjacent and second-adjacent channels on link l and wavelength w, respectively.

Based on the above, we constrain the interference accumulated over the lightpaths:

$$\sum_{\{l \in p | n \text{ end of } l\}} \left(\overbrace{s^2_{\mathrm{XT},n} \cdot \sum_{\{p' | n \in p'\}} x_{p',w}}^{\text{intra-XT}} + \overbrace{s^2_{\mathrm{XPM},l} \cdot \left(\sum_{\{p' | l \in p'\}} x_{p',w-1} + x_{p',w+1} \right)}^{\text{adjacent channel XPM}} \right.$$

$$\left. + \overbrace{s^2_{2-\mathrm{XPM},l} \cdot \left(\sum_{\{p' | l \in p'\}} x_{p',w-2} + x_{p',w+2} \right)}^{\text{second adjacent channel XPM}} \right) + c_{\mathrm{FWM}} + M \cdot x_{pw} - S_p \leq \sigma^2_{\max}(p,w) + M,$$

for all $p \in P$ and all $w \in C$.

We use constant M (large integer) to activate/deactivate the constraints and carry the surplus variables S_p in the objective cost. Lightpaths satisfying these constraints have by definition (assuming that the models used for calculating the noise variance parameters and the Q-factor are accurate) acceptable quality of transmission. Although in the above constraint we assumed that the signal power is totally compensated at the end of each link and each OXC, this assumption is not restrictive, and the above constraints can be modified for nonzero power gains or losses. We will refer to the direct IA-RWA algorithm described above as *Sigma-Bound* or *SB-IA-RWA* algorithm.

6 Impairment-Aware RWA Algorithms for Translucent Networks

We now turn our attention to translucent WDM networks. We assume a network of known physical topology that supports a given number of wavelengths, and wavelength conversion is possible only through 3R regeneration. We assume static traffic, given in the form of a traffic matrix specifying the number of connection requests between any pair of nodes. Each connection requires transmission rate equal to that of one wavelength, but there may be multiple connection requests for a given source–destination pair. The proposed algorithm is given a specific RWA instance, described as above, and returns the solution, in the form of paths and wavelengths used, including decisions on whether a connection will be served transparently or using regenerators. For the latter connections, the algorithm also

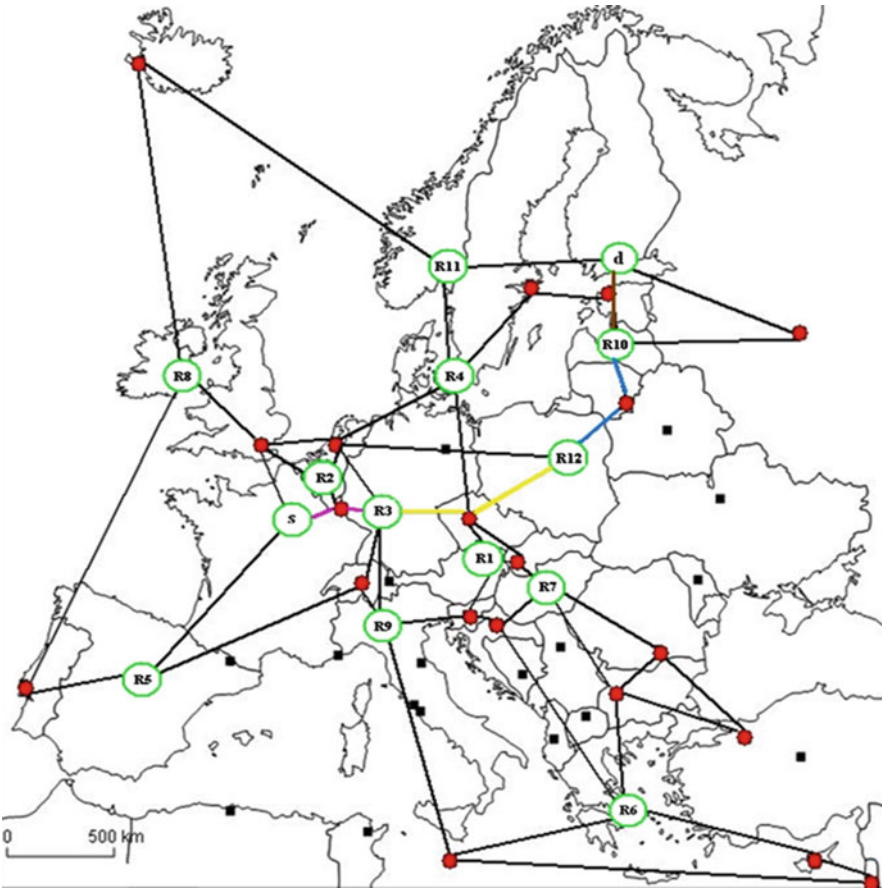


Fig. 5.6 The nontransparent connection request between source–destination pair (*s*,*d*) can be broken into four transparent sub-connection requests: *s-R3*, *R3-R12, R12-R10*, *and R10-d*. Each of the three sub-connection requests can be served using a different wavelength

decides the sequence of regenerators to be used. In case a solution serving all connections cannot be found with the given number of wavelengths, the algorithm returns the corresponding blocking ratio.

The IA-RWA algorithms for translucent WDM networks consist of three phases. In phase 1, described in Sect. 6.1, the algorithm decides which connections cannot be served transparently. For these nontransparent connections, we formulate and solve a virtual topology design problem to decide the regeneration sites they should use. The virtual topology consists of the 3R sites with (virtual) links between any pair of transparently connected 3R sites. Each virtual link of the paths chosen in the virtual topology to serve a connection corresponds to a sub-connection (lightpath) in the physical topology (Fig. 5.6). Thus, a nontransparent connection uses multiple consecutive segments (lightpaths), defined by the intermediate 3R sites it is going to

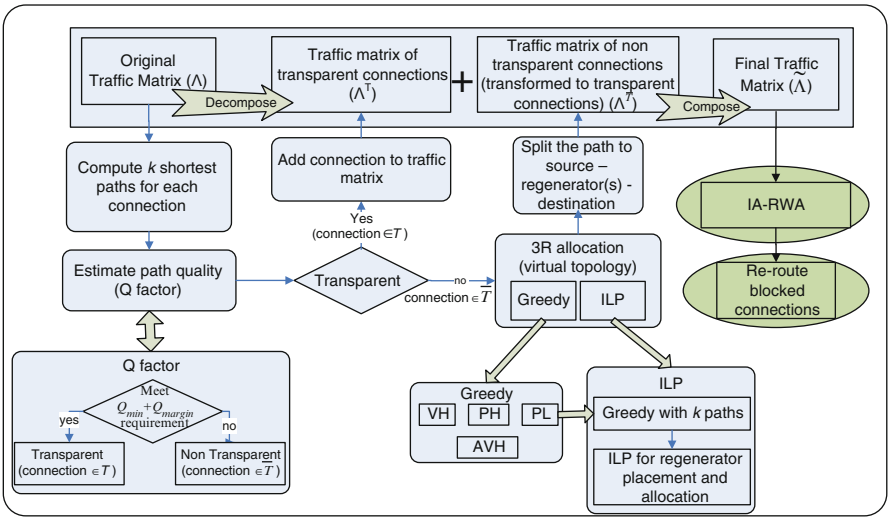


Fig. 5.7 Flow chart of the proposed IA-RWA algorithm for translucent networks

utilize. Based on these decisions, we transform the *initial traffic matrix* into a *transparent traffic matrix,* where nontransparent connection requests are replaced by a sequence of transparent sub-connection requests. In phase 2 of the algorithm, described in Sect. 6.2, we apply an impairment-aware (IA) RWA algorithm using as input the transformed transparent traffic matrix. The IA-RWA algorithm we used is the one proposed in Sect. 5, but other algorithms are also applicable. If some connections cannot be served while there are still regenerators not used in phase 1, in phase 3, described in Sect. 6.3, these physically blocked connections are reattempted through the remaining regenerators. Figure 5.7 provides an overview of the IA-RWA algorithm for heterogeneous translucent networks.

6.1 *Obtaining the Transparent Traffic Matrix (Phase 1)*

Phase 1 of the algorithm aims at transforming the initial traffic matrix Λ into a *transparent traffic matrix* $\tilde{\Lambda}$ that consists of connection requests that can be served without regenerators. To do so, we formulate in Sect. 6.1.1 a virtual topology problem that considers only nontransparent connection requests. Section 6.1.2 presents algorithms to solve this to obtain the intermediate regeneration sites. The traffic matrix is then transformed into a transparent traffic matrix, as described in Sect. 6.1.3.

We assume that 3R regenerators will be sparsely placed in the network, forming *pools* of regenerators at some nodes. We let $R \subseteq V$ be the set of nodes equipped with at least one 3R regenerator and r_n be the number of regenerators at node $n \in R$. In the

version of the problem where the algorithm is free to select the regeneration sites and the number of regenerators to deploy (regenerator placement problem), we assume that $R = V$ and that r_n is unbounded. In the version of the problem where the placement of regenerators is given (regenerator assignment problem), the set R and r_n are part of the input to the algorithm. The same problem can be given in the slightly different setting where, instead of the sparse regenerator placement, we are given the number of available transceivers at each node. Given the number of transceivers at a node, and subtracting the ones used by originating or terminating traffic (described in matrix Λ), we can find the number of spare transceivers at each node. These spare transceivers can be connected back-to-back so as to function as 3R regenerators, and thus, we can transform this problem to the typical regeneration assignment problem.

6.1.1 Constructing the Virtual Topology

In order to formulate the virtual topology problem, we first distinguish between transparent and nontransparent connection requests in the given traffic matrix. Specifically, for each source–destination pair (s,d), we examine if the QoT of at least one of its k-shortest length paths has acceptable performance, as indicated by its Q-factor, in an otherwise empty network. More formally, we let $P_k(s,d)$ be the set of k-shortest length paths between nodes s and d. We will say that a source–destination pair (s,d) is transparently connected and will denote that by $(s,d) \in T$, when the following holds:

$$(s,d) \in T \text{ iff } Q_p > Q_{\min} + Q_{\text{margin}} \quad \text{for some path} \quad p \in P_k(s,d),$$

where $Q_{\min}$ is the Q-factor value that corresponds to the minimum transmission quality requirement and Q_{margin} is a safety margin. We use k-shortest paths instead of a single path in the above definition, since in a heterogeneous network, the shortest length path does not always yield the best QoT. The set of nontransparent connections that do not satisfy the above constraint are denoted by $\overline{T}$.

For a source–destination pair $(s,d) \in T$, at least one of its k-shortest paths has acceptable Q performance, and it is possible *in principle* to serve a connection between them without using regenerators. However, this possibility depends on the IA-RWA solution since the Q-factor values used in distinguishing between transparent and nontransparent connections are calculated for an empty network. When all connections are present, some of the assumed transparent lightpaths may be infeasible in reality due to interference. The IA-RWA algorithm applied in phase 2 aims at avoiding such problems by finding lightpaths that are feasible even when all impairments are taken into account.

Given the set T of transparently connected (s,d) pairs and the set R of nodes with regeneration capabilities, we define the virtual topology as the graph $\widetilde{G} = (R, \widetilde{E})$, where $\widetilde{E} = \{(u,v) | \, u \in R, v \in R, (u,v) \in T\}$, denotes the set of virtual links

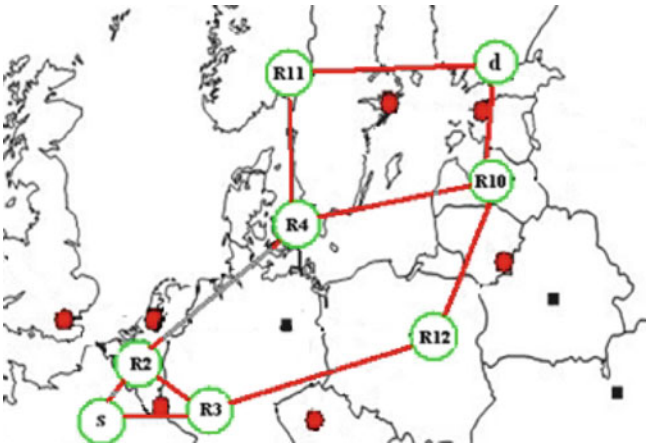


Fig. 5.8 The virtual topology $\tilde{G}(s,d)$, assuming regeneration sites R2, R3, R4, R10, R11, and R12. Source s and destination d are transparently connected to sites R2 and R3 and R10 and R11, respectively

between regeneration nodes that are transparently connected. Each virtual link corresponds to a transparent lightpath between two regeneration nodes, spanning one or several physical links. We also denote by R^s and R^d the set of regeneration sites the source s and the destination d are transparently connected to, respectively.

6.1.2 Choosing the Regenerators for the Nontransparent Connections

In this step of phase 1 we consider only the set $\overline{T}$ of source–destination pairs that are *not* transparently connected and that have to be routed through regenerators. Therefore, for a source–destination pair $(s,d) \in \overline{T}$, we have to choose the sequence of regeneration site(s) they are going to use. This is equivalent to finding a path from s to d in the virtual topology $\tilde{G}(s,d)$, obtained by adding to graph $\tilde{G}$ the source node s and the virtual links connecting it to the elements of R^s and, also, the destination node d and the virtual links connecting it to the elements of R^d. In the example of Fig. 5.8, we focus on six regenerator sites (assuming 12 regeneration sites as presented in Fig. 5.6) and present the virtual topology $\tilde{G}(s,d)$ for the source–destination pair (s,d). In order to select the virtual path from s to d (equivalently, the sequence of regenerators to be used), we investigated two classes of algorithms in the following two subsections:

(a) Greedy first-fit algorithms for routing in the virtual topology

We now present several greedy algorithms for routing nontransparent connections in the virtual topology. We denote by $\Lambda^{\overline{T}}$ the part of the traffic matrix Λ that corresponds to source–destination pairs in $\overline{T}$, that is,

$$\Lambda_{sd}^{\overline{T}} = \Lambda_{sd} \text{ if } (s,d) \in \overline{T}; \text{ and } \Lambda_{sd}^{\overline{T}} = 0, \text{otherwise.}$$

Nontransparent connection requests are treated by the greedy routing algorithms one by one. In particular, for each pair (*s*,*d*) for which $\Lambda_{sd}^{\overline{T}} \neq 0,$, we run Dijkstra's shortest path algorithm on graph $\widetilde{G}(s,d)$ with appropriate link costs. The way the costs of the virtual links in $\widetilde{G}(s,d)$ are defined is important for the performance of the algorithms and is described next.

A nontransparent connection may be blocked due to the unavailability of free regenerators, or of free wavelengths, or due to significant physical impairments (including inter-lightpath interference). The latter factor is important for paths that are long or use many hops. Given the above considerations, we studied the following definitions for the cost of the virtual links in the virtual topology:

1. *Virtual-hop (VH) shortest path algorithm*: All the links of the virtual graph have cost equal to 1. The optimal virtual path is the one consisting of the fewest regenerators (virtual hops).
2. *Physical-hop (PH) shortest path algorithm*: The cost of a virtual link is set equal to the number of physical links (physical hops) it consists of. The optimal virtual path is the one that traverses the minimum number of physical nodes.
3. *Physical-length (PL) shortest path algorithm*: The cost of a virtual link is set equal to the sum of the lengths of the physical links that comprise it. Thus, the PL algorithm selects the virtual path that has the shortest physical length.
4. *Adjusted-virtual-hop (AVH) shortest path algorithm*. Since in the three algorithms described above, Dijkstra's algorithm is executed for each nontransparent (*s*,*d*) pair separately and successively, many connections may try to use certain regeneration sites. To avoid this and also distribute the load more equitably among the regeneration sites, the AVH algorithms adjust the cost of using the regenerators. In our simulations, the cost of a virtual link l between virtual nodes i and j equipped with r_i and r_j regenerators, respectively, from which u_i and u_j regenerators have already been used by previous connections, is taken to be

$$w_l = \left\lceil \frac{u_i}{r_i} + \frac{u_j}{r_j} \right\rceil \cdot \max_{n \in R} (r_n).$$

(b) ILP routing algorithm for routing in the virtual topology

We also developed an ILP formulation to solve the virtual topology problem presented above. The ILP algorithm selects the routes of the nontransparent connections by minimizing one of the following: (1) the maximum number of regenerators used among all nodes or (2) the total number of regenerators used in the network or (3) the number of regeneration sites. Due to space limitations, we do not present this algorithm here and refer the reader to [8] for details. The ILP developed can be solved for large-sized networks since the number of variables and

constrains is small. When its solution is intractable (which may be the case for very large networks), the LP relaxation of the problem combined with piecewise linear link costs and iterative fixing and rounding, as presented in Sect. 4.4, can be used and also yield good solutions. In the simulation experiments we performed with realistic network and traffic input, we were always able to solve the corresponding ILP problems in a few seconds time.

6.1.3 Transforming the Traffic Matrix

Having found the sequence of regenerators to be used by nontransparent connections, the IA-RWA problem in the translucent network is transformed into a corresponding IA-RWA problem in a transparent network that has a different traffic matrix. In particular, after obtaining the solution to the virtual topology problem, which effectively breaks each nontransparent connection into multiple transparent segments, we transform appropriately the original traffic matrix, as follows. From the part of the initial traffic matrix that corresponds to the nontransparent connections, $\Lambda^{\overline{T}}$, we use the virtual paths chosen for the nontransparent connections to obtain the traffic matrix $\tilde{\Lambda}^{\overline{T}}$ corresponding to these virtual paths. By adding this to the matrix $\Lambda^T = \Lambda - \Lambda^{\overline{T}}$ that represents the initial transparent connections, we obtain the transparent traffic matrix $\tilde{\Lambda} = \Lambda^T + \tilde{\Lambda}^{\overline{T}}$ that if routed, will serve all requested (transparent and nontransparent) connections.

6.2 IA-RWA Using the Transformed Traffic Matrix (Phase 2)

Phase 2 of the algorithm takes as input the transparent traffic matrix $\tilde{\Lambda}$, calculated in the previous phase, which only contains connections that can be served transparently, and applies an IA-RWA algorithm designed for transparent networks to route and assign wavelengths to these connections. Given the traffic matrix $\tilde{\Lambda}$, the physical network topology G, the number of supported wavelengths W, and impairment-related parameters, the IA-RWA algorithm returns the RWA solution to the specific instance in the form of paths and wavelengths and the corresponding blocking probability. Blocking in phase 2 may occur for two reasons: (1) if the transformed traffic matrix $\tilde{\Lambda}$ cannot be served with the available wavelengths, some connections will have to be dropped, or (2) the interference among lightpaths may make the QoT of some lightpaths (which are feasible in an empty network) unacceptable. Blockings of category (1) are avoided by appropriately placing the regenerators in the network. Blockings of category (2) are avoided by a well-designed IA-RWA algorithm that accounts for the interference among lightpaths.

In any case, connections blocked in phase 2 are reattempted in phase 3 of the algorithm, as described in Sect. 6.3.

The IA-RWA algorithm presented in Sect. 5 was used in this phase, but other IA-RWA algorithms can also be used.

6.3 Rerouting the Blocked Connections (Phase 3)

In the final phase 3 of the algorithm, we try to reroute any connections blocked in phase 2. Given the solution of the IA-RWA algorithm of phase 2 and the set B of blocked (s,d) pairs, we want to serve the connections in B without altering previously accepted lightpaths. The set B of blocked connections refers to the initial traffic matrix Λ and not to the transformed traffic matrix $\tilde{\Lambda}$. We formulate a new virtual topology problem similar to that of phase 1 (Sect. 6.1). The input is the set of remaining regenerators in the network and the set B of blocked connections. We re-execute the IA-RWA algorithm of phase 2 (Sect. 6.2), assuming that the lightpaths calculated in the previous solution are established (i.e., we set equal to 1 the corresponding variables). The output of this reduced IA-RWA problem will indicate if we can route the connections in B with acceptable transmission quality without affecting the other connections. The proposed algorithm is terminated at this point and outputs the routed lightpaths from phase 2 in addition to any new lightpaths that were rerouted in phase 3. Connections that have unacceptable Q performance, after phase 3, are blocked.

7 Simulation Results

To evaluate the performance of the proposed IA-RWA algorithms, we carried out a number of simulation experiments. We implemented all the algorithms in Matlab and used LINDO [37] to solve the corresponding LP and ILP problems. We start, in Sect. 7.1, by presenting results for the impairment-unaware IU-RWA algorithm (Sect. 4). We evaluate the integrality and optimality performance of the proposed LP-relaxation algorithm and the random perturbation technique by comparing it to a typical min–max RWA formulation. We then, in Sect. 7.2, turn our attention to the RWA problem in the presence of physical impairments. We initially consider a transparent WDM network and evaluate the performance of the proposed Sigma-Bound IA-RWA algorithm of Sect. 5. We then turn our attention to a translucent WDM network and evaluate the performance of the algorithms presented in Sect. 6.

The network topology used in our simulations for the impairment-unaware and the impairment-aware transparent RWA algorithms (Sects. 7.1 and 7.2.1) was the

Deutsche Telekom network (DTnet), shown in Fig. 5.3, which is a candidate transparent network as identified by the DICONET project [33]. For the translucent network experiments (Sect. 7.2.2), we have used the GEANT-2 network topology, shown in Fig. 5.6. In both cases, the capacity of a wavelength was assumed equal to 10 Gbps.

7.1 Impairment-Unaware RWA Performance Results

In this section we evaluate the performance of the IU-RWA algorithm that is based on the LP-relaxation formulation of Sect. 4. To have a reference point, we also executed the same experiments using a typical min–max formulation (a formulation whose objective is to minimize the maximum number of wavelengths used), which was solved using the ILP branch-and-bound algorithm of [37]. Note that the maximum number of wavelengths used is the actual objective that we want to minimize. The piecewise linear cost function used in the proposed LP RWA algorithm (Sect. 4.2) tries to approximate the min–max objective, so as to exhibit a good integrality performance when the Simplex algorithm is used. Thus, the ILP-min–max algorithm sets the criterion in terms of optimality. We also used the same min–max formulation and solved its LP-relaxed version followed by iterative fixing and roundings. This LP-min–max algorithm sets a comparison criterion in terms of integrality and execution time since its difference to our proposed LP algorithm lies on the piecewise linear cost function that we utilize and the random perturbation technique. For all algorithms in this section, we have used $k = 3$.

In this set of experiments, we used the DTnet of Fig. 5.3. The results were averaged over 100 experiments obtained with different random traffic matrices of a given traffic load, defined as the ratio of the total number of connection requests over all possible source-destination pairs for loads ranging from 0.5 to 2 with 0.5 step. Figure 5.9a, b shows the number of RWA instances for which we are sure to have obtained an optimal solution and the average execution time of the algorithms.

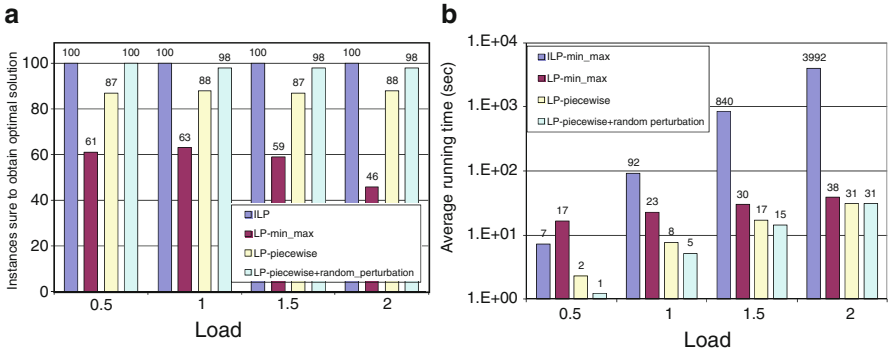


Fig. 5.9 (**a**) Number of RWA instances that we are sure to obtain optimal solution and (**b**) corresponding average running times, as a function of load ρ

We observe that the LP-piecewise algorithm has superior overall performance than the other algorithms examined. It finds with high probability an optimal solution in low execution times. The random perturbation technique further improves the LP-piecewise algorithm by increasing its integrality performance and reducing its running time. The good optimality performance of the proposed algorithm is maintained irrespectively of the load.

7.2 IA-RWA Performance Results

We now turn our attention to the case where physical-layer impairments are present. To evaluate the feasibility of the lightpaths in terms of QoT, we used a Q-factor estimator (Q-Tool) that relies on analytical models to account for the most important impairments (see section 3). The link model of the reference network is presented in Fig. 5.10. We assumed NRZ-OOK modulation format, 10 Gbps transmission rates, and 50 GHz channel spacing. The span length on each link was set to 100 km. Each link was assumed to consist exclusively of single-mode fibers (SMF) with dispersion parameter $D = 17$ ps/nm/km and attenuation parameter $a = 0.25$ dB/km. For the dispersion-compensating fibers (DCF), we used parameters $a = 0.5$ dB/km and $D = -80$ ps/nm/km. The launch power was 3 dBm/ch for every SMF span and -4 dBm/ch for the DCF modules. The Erbium Doped Fiber Amplifier's (EDFA) noise figure was 6 dB with small variations (± 0.5 dB), and each EDFA exactly compensates for the losses of the preceding fiber span. We assumed a switch architecture similar to [14] and a switch–crosstalk ratio $X_{sw} = 32$ dBs with small variations per node (± 1 dB). Regarding dispersion management, a pre-compensation module was used to achieve higher reach: initially the dispersion was set to -400 ps/nm/km, every span was under-compensated by a value of 30 ps/nm/km to alleviate nonlinear effects, and the accumulated dispersion at each switch input was fully compensated to zero using an appropriate post-compensation module at the end of the link. The acceptable Q-factor value was set to $Q_{\min} = 15.5$ dB.

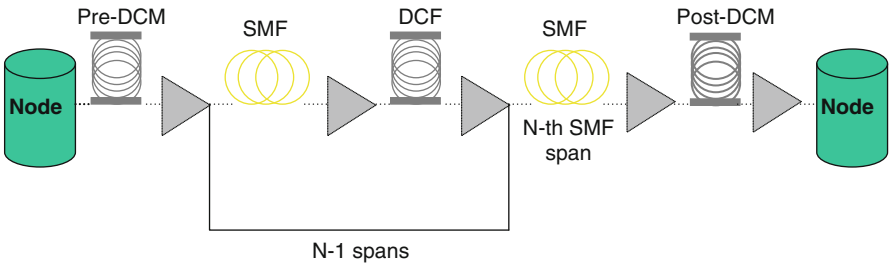


Fig. 5.10 Link model

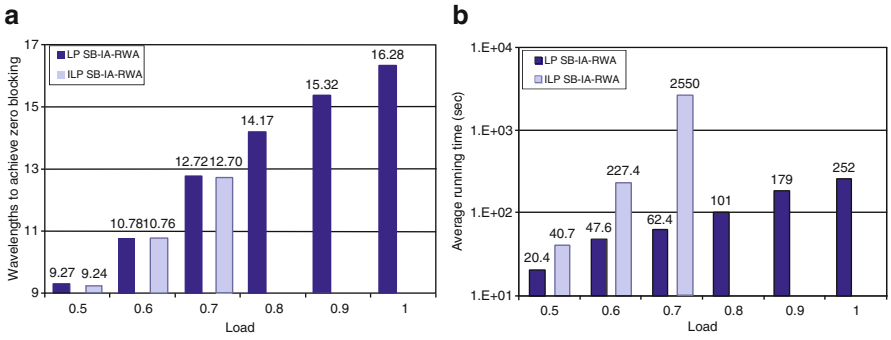


Fig. 5.11 (**a**) Average number of wavelengths required to obtain zero blocking and (**b**) average execution time for the SB-IA-RWA algorithm, solved with the proposed LP-relaxation technique and ILP

7.2.1 Transparent Network Simulation Experiments

In this section we present performance results for the Sigma-Bound SB-IA-RWA algorithm of Sect. 5, obtained for the DTnet topology shown in Fig. 5.3.

When planning a WDM network (offline problem), we are usually looking for a zero-blocking solution. The direct SB-IA-RWA algorithm can find a zero-blocking solution given enough wavelengths, indicating that DTnet is in principle a transparent network, in the sense that it can serve all traffic without regenerators. Thus, in these experiments, we focus on zero-blocking solutions and examine the SB-IA-RWA performance by solving either (1) its LP-relaxation version using iterative fixings and roundings to obtain an integer solution or (2) the ILP version with a branch-and-bound method.

In Fig. 5.11a we report the number of wavelengths required for zero blocking averaged over 100 traffic matrices and loads between 0.5 and 1 with a step of 0.1. For the ILP version of the algorithm, we were able to track solutions for loads up to $\rho = 0.7$ in a time limit of 5 h per instance (for certain instances of larger load, we were not able to find an optimal solution within 5 h). We see that the number of wavelengths required by the LP-relaxation algorithm is quite close to that required by the optimum ILP. The optimality was lost only in two or three instances (from the 100 examined) for the traffic loads for which we were able to find optimal solutions. Regarding execution times (Fig. 5.11b), we see that the LP-relaxation version has superior performance and maintains the average running time within a few hundreds of seconds, while the ILP version cannot solve certain hard instances even at medium load. Note that as the load increases, more lightpaths are activated and the interfering sources among them increase, making the problem more complicated and difficult to solve.

We also performed experiments assuming the actual traffic matrix of DTnet, consisting of 381 connection requests (corresponding to load $\rho \approx 2.05$) [33].

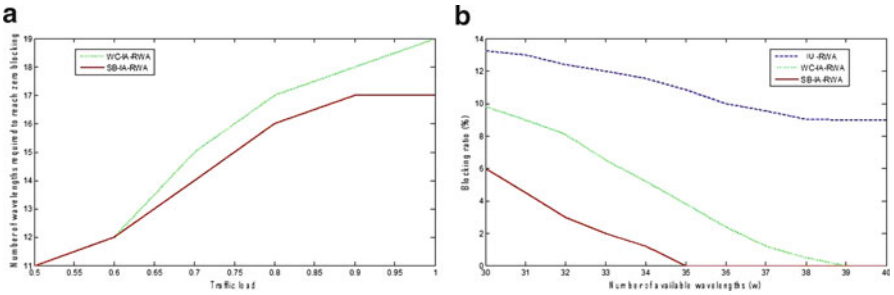


Fig. 5.12 (**a**) Number of wavelengths required to reach zero blocking for various loads obtained using the random traffic generator and (**b**) blocking probability versus number of available wavelengths W per link for realistic traffic load

We compared the performance of three algorithms: (1) the IU-RWA algorithm that does not consider physical impairments at all, (2) the WC-IA-RWA algorithm that prunes the candidate paths based on the impairments of class 1 and of class 2 under the worst interference scenario (Sect. 4.4), and (3) the cross-layer SB-IA-RWA algorithm that takes into account the interference among lightpaths in its RWA formulation (Sect. 5). A key difference of these algorithms is the set of candidate paths they take as input. In particular, referring to Table 5.1, the IU-RWA algorithm takes as input the set of paths that correspond to column (a), the WC-IA-RWA algorithm the set of paths that correspond to column (c), and the proposed cross-layer SB-IA-RWA algorithm the set of paths that correspond to column (b). To make the comparison more fair, we used $k = 3$ for the IU-RWA and the proposed SB-IA-RWA algorithm and $k = 4$ for the WC-IA-RWA.

Figure 5.12a shows the number of wavelengths required to achieve zero (physical and network) blocking for the WC-IA-RWA and the SB-IA-RWA algorithms using the randomly generated traffic matrices. We do not graph the performance of IU-RWA since this algorithm was unable to obtain zero-blocking solutions due to physical blocking. We observe that SB-IA-RWA achieves zero blocking using fewer wavelengths than WC-IA-RWA. WC-IA-RWA algorithm is only affected by network-layer blocking, since all candidate paths that are fed to the algorithm have acceptable QoT performance. However, the set of candidate paths that WC-IA-RWA uses turns out to be inferior, resulting in a solution that consumes more wavelengths than SB-IA-RWA. Note that path population is not the only parameter affecting performance. Even more important is the distribution of the number of different candidate paths that are available for each source–destination pair. Since, in the case of WC-IA-RWA, some paths are unnecessarily discarded, the choices on the routes that can be used for certain connections are very restricted. In contrast, the cross-layer SB-IA-RWA has more options in selecting lightpaths of acceptable quality to serve the traffic.

Figure 5.12b shows the average blocking probability versus the number of available wavelengths per link for all examined algorithms, assuming the realistic DTnet traffic matrix. The impairment-unaware IU-RWA algorithm has the highest

blocking rate since it cannot avoid physical-layer blocking. The WC-IA-RWA algorithm exhibits only network-layer blocking, while SB-IA-RWA exhibits both physical and network blocking. The SB-IA-RWA algorithm outperforms the WC-IA-RWA algorithm in terms of rejected connections and is able to reach zero blocking using fewer wavelengths. In particular, the difference between these two algorithms is 4 wavelengths for this realistic traffic load, a difference which is substantial. The running time of IU-RWA and WC-IA-RWA for $W = 36$ was around 30 s. The corresponding running time for the SB-IA-RWA algorithm was about 15 min, which is still quite low, considering that this is a realistic experiment and we are considering a planning (offline) problem. Comparing the results of Fig. 5.12a to those of Fig. 5.12b, we can deduce that the improvements obtained by SB-IA-RWA are more pronounced for heavy traffic; since then, the number of interfering sources is higher and cross-layer optimization is more important.

Concluding, SB-IA-RWA was shown to exhibit superior performance. It is able to serve all traffic (zero blocking) given enough wavelengths, which is particularly important in practice for the planning problem (telecom operators do not like to reject connections). The number of variables and constraints it requires is rather small, and its execution time is acceptable, using the LP-relaxation and iterative fixing and rounding methods. We also performed a small number of experiments with networks that are even larger than DTnet (see the following section) and found that the SB-IA-RWA algorithm scales very well with network size.

7.2.2 Translucent Network Simulation Experiments

We now focus on the algorithms described for translucent WDM networks. The topology used in this set of simulations was the GEANT-2 network, shown in Fig. 5.6 [33]. All single-hop connections in GEANT-2 were able to be served transparently, but some multi-hop connections were not, making the use of regenerators necessary for some connections. We assume that all nodes can host regenerators; that is, the number of regeneration sites is not restricted. The algorithms were asked to solve the regeneration placement problem in order to decide the regeneration sites and the number of regenerators to deploy at each site. The traffic matrix used in our simulations corresponds to a realistic traffic load with 826 connection demands. The acceptable Q-factor limit was taken equal to $Q_{\min} = 15.5$ dB and the transparency margin equal to $Q_{\text{margin}} = 0.5$ dB (Sect. 6.1.1). For the given topology, the described link models (Fig. 5.10), the given traffic matrix, and using $k = 3$ candidate paths for each source–destination pair, the set of nontransparent connections consists of 373 connections, which have to be routed through regenerators.

We evaluated the performance of the ILP virtual topology algorithms (Sect. 6.1.2) combined with VH and PH heuristics to pre-calculate their paths. Thus, the algorithms evaluated are VH-ILPsum, VH-ILPmax, VH-ILPsites, PH-ILPsum, PH-ILPmax, and PH-ILPsites. It is worth noting that we were able to obtain results for all these ILP algorithms with running times in the order of a few seconds.

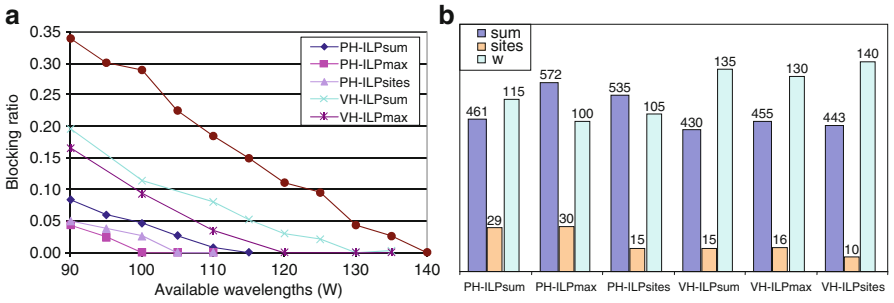


Fig. 5.13 (**a**) Blocking ratio for realistic traffic load versus number of available wavelengths and (**b**) total number of regenerators, total number of regeneration sites, and total number of wavelengths in the network required to obtain blocking equal to zero

In Fig. 5.13a we graph the blocking ratio as a function of the number of available wavelengths. From Fig. 5.13a, it is obvious that PH-based algorithms have better performance, since they have lower blocking probability and require fewer wavelengths to reach zero blocking. Among these algorithms, PH-ILPmax outperforms all the other examined algorithms, requiring the fewest number of wavelengths, in particular $W = 100$, to serve all the connections with zero blocking. This was expected, since the PH-ILPmax algorithm distributes the regenerators in the network so as to minimize the number of connections crossing a specific regeneration node, preventing regeneration nodes from becoming "bottlenecks." In this way, PH-ILPmax distributes the load and ends up using fewer different wavelengths to achieve zero blocking.

In Fig. 5.13b we graph (1) the total number of regenerators used in the network, (2) the number of regeneration sites, and (3) the number of wavelengths required to reach zero blocking. We can observe the difference between the VH- and PH-based algorithms. PH-based algorithms need a smaller number of wavelengths to reach zero blocking but utilize more regenerators and regeneration sites. On the other hand, VH-based algorithms need more wavelengths to reach zero blocking but have better regenerator utilization performance. For example, focusing on PH-ILPsum and VH-ILPsum, they both minimize the total number of regenerators used in the network. PH-ILPsum needs 115 wavelengths while VH-ILPsum needs 135 to reach zero blocking. To achieve this performance, PH-ILPsum needs 461 regenerators while VH-ILPsum needs 430. This difference can be explained by the fact that PH-based algorithms are fed with good candidate paths that have been calculated in the physical topology, while VH-based algorithms are fed with good candidate paths that have been calculated in the virtual topology. PH algorithm uses paths that minimize the number of physical hops utilized, which tends to give good wavelength utilization performance, since shorter physical-hop paths use less wavelengths than longer physical-hop paths. This is the reason shorter physical-hop paths are widely used in impairment-unaware RWA problems. On the other

hand, these paths are not directly related to the virtual topology, and thus, the ILP algorithm that runs over the virtual topology cannot find a good solution that minimizes its objective. The VH-based algorithms use as input good virtual paths that are not related to the physical topology. So, although these paths yield good solutions to the regenerator placement problem that is performed over the virtual topology, they may be long physical-hop paths that waste wavelength resources. From Fig. 5.13b, we see that there is a trade-off between the number of wavelengths and the number of regenerators required to reach zero blocking. All examined algorithms provide good solutions depending on the objective they are set to minimize, and there is no single algorithm that outperforms the others with respect to all criteria. For example, if the cost of the regenerators is the dominant expense, one has to use VH-base algorithms and probably VH-ILPsum; if the objective is to minimize the number of wavelengths, PH-ILPmax seems to perform better.

8 Conclusions

We presented algorithms for planning transparent and translucent WDM networks in the presence of physical-layer impairments. We initially presented an algorithm for solving the (impairment-unaware) RWA problem, where physical impairments are not considered, based on an LP-relaxation formulation that provides integer optimal solutions for a large subset of RWA input instances, despite the absence of integrality constraints. We then extended the RWA formulation and proposed an impairment-aware (IA)-RWA algorithm for transparent WDM networks that models the physical-layer impairments as additional constraints in its formulation, performing a cross-layer optimization between the network and the physical layers. The proposed Sigma-Bound IA-RWA is a direct algorithm that constrains the interference among the lightpaths so as to obtain acceptable transmission quality as defined by the Q-factor. We also addressed the planning problem of translucent WDM networks, in which case the algorithm has to select the 3R regeneration sites and the number of regenerators that need to be deployed on these sites. We formulated regenerator placement as a virtual topology design problem and addressed it using various algorithms, ranging from ILPs to simple greedy heuristics. After determining the sequence of regenerators to be used by the nontransparent connections, we transform the initial traffic matrix to a transparent traffic matrix and then apply a transparent IA-RWA algorithm to serve it. Using realistic traffic scenarios, our results quantified the blocking improvements obtained by considering physical-layer impairments according to the actual utilization of the network as opposed to worst-case utilization based approaches.

References

1. Stern T, Bala K (1999) Multiwavelength optical networks: a layered approach. Prentice Hall
2. Ramaswami R, Sivarajan KN (2001) Optical networks: a practical perspective, 2nd edn. Morgan Kaufmann, San Francisco
3. Agawal GP (2002) Fiber-optic communication systems, 3rd edn. Wiley-Interscience, New York
4. Chlamtac I, Ganz A, Karmi G (1992) Lightpath communications: an approach to high-bandwidth optical WANs. IEEE Trans Commun 40(7):1171–1182
5. Ramamurthy B, Datta D, Feng H, Heritage JP, Mukherjee B (1999) Transparent vs. opaque vs. translucent wavelength-routed optical networks. Optical Fiber Communication Conference (OFC) 1999
6. Christodoulopoulos K, Manousakis K, Varvarigos E (2008) Comparison of routing and wavelength assignment algorithms in WDM networks. Globecom 2008
7. Christodoulopoulos K, Manousakis K, Varvarigos E (2010) Offline routing and wavelength assignment in transparent WDM networks. IEEE/ACM Trans Netw 18(5):1557–1570
8. Manousakis K, Christodoulopoulos K, Kamitsas E, Tomkos I, Varvarigos E (2009) Offline impairment-aware routing and wavelength assignment algorithms in translucent WDM optical networks. IEEE/OSA J Lightwave Technol 27(12):1866–1877
9. Zang H, Jue JP, Mukherjee B (2000) A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks. Optical Networks Magazine, vol 1
10. Birman A, Kershenbaum A (1995) Routing and wavelength assignment methods in single-hop all-optical networks with blocking. IEEE Infocom 2:431–438
11. Banerjee D, Mukherjee B (1996) A practical approach for routing and wavelength assignment in large wavelength-routed optical networks. IEEE J Sel Areas Commun 14(5):903–908
12. Ozdaglar A, Bertsekas D (2003) Routing and wavelength assignment in optical networks. IEEE/ACM Trans Netw 11(2):259–272
13. Saad M, Luo Z (2004) On the routing and wavelength assignment in multifiber WDM networks. IEEE J Sel Areas Commun 22(9):1708–1717
14. Ramamurthy B, Datta D, Feng H, Heritage JP, Mukherjee B (1999) Impact of transmission impairments on the teletraffic performance of wavelength-routed networks. IEEE/OSA J Lightwave Technol 17(10):1713–1723
15. Huang Y, Heritage J, Mukherjee B (2005) Connection provisioning with transmission impairment consideration in optical WDM networks with high-speed channels. IEEE/OSA J Lightwave Technol 23(3):982–993
16. Cardillo R, Curri V, Mellia M (2005) Considering transmission impairments in wavelength routed optical networks. Cong. on Optical Network Design and Modeling (ONDM), 2005
17. Deng T, Subramaniam S, Xu J (2004) Crosstalk-aware wavelength assignment in Dynamic Wavelength-Routed Optical Networks. Broadnets, 2004
18. Anagnostopoulos V, Politi C, Matrakidis C, Stavdas A (2007) Physical layer impairment aware wavelength routing algorithms based on analytically calculated constraints. Opt Commun 270(2):247–254
19. He J, Brandt-Pearce M, Pointurier Y, Subramaniam S (2007) QoT aware routing in impairment-constrained optical networks. IEEE GLOBECOM, pp 2269–2274
20. Tomkos I, Vogiatzis D, Mas C, Zacharopoulos I, Tzanakaki A, Varvarigos E (2004) Performance engineering of metropolitan area optical networks through impairment constraint routing. IEEE Commun Mag, special issue on metro optical networks, 2004
21. Markidis G, Sygletos S, Tzanakaki A, Tomkos I (2007) Impairment aware based routing and wavelength assignment in transparent long haul networks. Conf. on Optical Network Design and Modeling, 2007
22. Hamad AM, Kamal AE (2005) Routing and wavelength assignment with power aware multicasting in WDM networks. Broadnets 2005, vol 1, pp 31–40

23. Saleh AM (2000) Transparent optical networking in backbone networks. Optical Fiber Communication Conference (OFC) 2000
24. Shen G, Grover W, Cheng T, Bose S (2002) Sparse placement of electronic switching nodes for low-blocking in translucent optical networks. OSA J Opt Netw 1:424–441
25. Ye Y, Chai TY, Cheng TH, Lu C (2003) Novel algorithm for upgrading of translucent optical networks. Opt Express 11(23):3022–3033
26. Yang X, Ramamurthy B (2005) Dynamic routing in translucent WDM optical networks: the intra-domain case. IEEE/OSA J Lightwave Technol 23(3):955–971
27. Ye Y, Chai TY, Lu C (2004) Routing and wavelength assignment algorithms for translucent optical networks. Opt Commun 233–239
28. Pachnicke S, Paschenda T, Krummrich P (2008) Assessment of a constraint-based routing algorithm for translucent 10 Gbits/s DWDM networks considering fiber nonlinearities. OSA J Opt Netw 7(4):365–377
29. Ye Y, Chai TY, Cheng TH, Lu C (2003) Algorithms for the design of WDM translucent optical networks. Opt Express 11(22):2917–2926
30. Shen G, Grover W, Cheng T, Bose S (2002) Sparse placement of electronic switching nodes for low blocking in translucent optical networks. J Opt Netw 1(12):424–441
31. Yang X, Ramamurthy B (2005) Sparse regeneration in translucent wavelength-routed optical networks: architecture, network design and wavelength routing. Photonic Netw Commun 10 (1):39–53
32. Ezzahdi MA, Zahr SA, Koubaa M, Puech N, Gagnaire M (2006) LERP: a quality of transmission dependent heuristic for routing and wavelength assignment in hybrid WDM networks. International Conference on Computer Communications and Networks (ICCCN), 2006
33. Dynamic Impairment Constraint Network for Transparent Mesh Optical Networks (DICONET) project, FP7 – GA 216338. http://www.diconet.eu
34. Papadimitriou C, Steiglitz K (1998) Combinatorial optimization: algorithms and complexity. Dover publications, Mineola, NY
35. Bertsekas D, Gallager R (1992) Data networks, 2nd edn. Prentice Hall, Englewood Cliffs
36. Pachnicke S, Reichert J, Spälter S, Voges E (2006) Fast analytical assessment of the signal quality in transparent optical networks. IEEE/OSA J Lightwave Technol 24:815–824
37. LINDO: http://www.lindo.com/

Chapter 6
Cross-Layer Control of Semitransparent Optical Networks Under Physical Parameter Uncertainty

Guido Maier, Eva Marìn, Marco Quagliotti, Walter Erangoli, Giovanni Tamiri, Marcelo Yannuzzi, Xavier Masip, and René Serral-Gracià

1 Introduction

Optical transport networks (OTNs) based on wavelength-division multiplexing (WDM) have achieved today a high degree of maturity. With the extremely rapid growth of Internet traffic generated by new applications (e.g., cloud computing, data center interconnection) and subscriber needs, a new challenge awaits OTN: that is, the capability of scaling up in capacity while keeping cost and required energy per transported bit at the same level or even lowering them.

It is well recognized that one of the major limitations in this process is represented by the optical signal regeneration. This function is needed to overcome signal degradation due to several impairments which accumulate as light propagates through a network. Effects that should be compensated include: loss, amplified spontaneous emission (ASE) noise, chromatic and polarization-mode dispersion, filter cascading, and nonlinear effects such as four-wave mixing, self- and cross-phase modulation, and Brillouin and Raman scattering. They are particularly severe in high-bit-rate and long-haul dense WDM (DWDM) systems.

G. Maier (✉)
Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Milano, Italy
e-mail: guido.maier@polimi.it

E. Marìn • M. Yannuzzi • X. Masip • R. Serral-Gracià
Advanced Network Architectures Lab (CRAAX), UPC, Vilanova y la Geltrù, Spain

M. Quagliotti
Telecom Italia, Torino, Italy

W. Erangoli
Formerly with Politecnico di Milano, DEI; now with ICT Consulting, Milano, Italy

G. Tamiri
Formerly with Politecnico di Milano, DEI; now free lance consultant

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*, Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_6,

Since all-optical regeneration is still mainly a research topic, the opto-electronic regenerator (also known as *transponder*) is the most commonly adopted solution: this device converts signals from optical to electronic form, regenerates, and converts back to optical. The double conversion from optical to electronic and back to optical (OEO conversion) is highly power consuming. Moreover, although transponders are less expensive than all-optical regenerators, they still are among the most expensive network elements of an OTN.

In order to save on CAPEX, regeneration capability can be sparsely provided in the network, installing transponders where needed, possibly trying to minimize their number. Networks in which this strategy is adopted are called *semitransparent* (alias "translucent"). Off-line planning of a semitransparent network requires deciding where to deploy transponders in an optimal way, on the basis of a given forecast traffic demand. Once the network is in operation, the on-line routing and wavelength assignment (RWA) process of a new connection has to take into account the location of the available regeneration points.

With semitransparent OTNs, both off-line design and on-line RWA become cross-layer problems coupling the physical layer, where constraints due to transmission impairments are generated, to the logical layer, which manages the optical end-to-end connections (lightpaths). Cross-layer design and routing techniques are generally quite complicated: in our case, they are even more complex especially due to the difficulty in modeling some optical transmission impairments, as, for instance, those caused by nonlinear effects in light propagation.

The first part of this chapter is preliminary. We present a design method to perform off-line planning and regenerator placement in a semitransparent OTN. Network resources are dimensioned, and transponders are properly installed in order to support a given set of static connections while guaranteeing that the signal quality at the receiver is above a fixed threshold for each connection. The method relies upon a semiempirical model of the physical layer, which requires a few input parameters describing the optical DWDM transmission systems to estimate the propagation impairments.

The second part of this chapter contains the main idea. We propose two different RWA algorithms for on-line connection control. The first (named "deterministic") bases its routing decisions upon the knowledge of the input parameters of the physical-layer model. The second (called "predictive") selects the routes for the lightpaths on a topological basis, "agnostically" of the physical layer, but self-learns from successes and failures when signal quality of the routed connections is compared a posteriori to the fixed threshold.

The main focus of the study we are presenting in this chapter is to investigate the behavior of the two RWA algorithms when the input parameters of the physical-layer model are affected by *uncertainty*, that is, they do not faithfully represent the real state of the network. In fact methods to provide the input values to the parameters (e.g., inference from data sheets) and to keep such values updated (e.g., infield measurements) may lead to deviations between the actual state and the model, which can potentially jeopardize cross-layer design and control. This is an important issue which has been up to now only covered by a few studies (see,

e.g., [1], which focuses on infield measurement of physical-layer parameters). Our novel contribution to this topic is the identification of two possible ways in which uncertainty about the physical-layer state can manifest:

- *Imperfect matching*: model input parameter values used in the planning phase are different from the actual values. This may occur, for example, because the network design relied upon a nominal set of values declared in data-sheet components which revealed to be too optimistic, the design was done long time ago and aging has worsened equipment behavior, and the model adopted for the physical layer does not represent all relevant impairments or is not accurate enough.
- *Imperfect knowledge*: model input parameter values used by the control system, and in particular for routing connections, deviate from the actual values. This may happen, for example, because the values are provided to the control plane by periodic measurements that are either affected by errors or too infrequent to correctly sample parameter variations and the physical model adopted by the control plane does not represent all relevant impairments or is not accurate enough. Our physical model does not take into account interchannel effects such as the crosstalk.

We have simulated on a case-study network combinations of imperfect matching and knowledge in order to compare the behavior of the two RWA algorithms. As will be apparent from the results of our dynamic traffic simulations presented at the end, the deterministic algorithm achieves the best results in ideal conditions, but the predictive is more robust to uncertainty, providing remarkable advantages, especially when imperfect matching and knowledge jointly occur. Thus, in a realistic scenario, a network operator should choose to rely for RWA computation on detailed and accurate physical-layer information or rather to rely on prediction according to the confidence the operator has on physical-layer parameters.

The outline of this chapter is as follows. Section 2 is dedicated to network design, illustrating the physical-layer model we have adopted based on the Personick Q factor. Section 3 deals with network control, presenting the two algorithms for semitransparent cross-layer routing we are proposing and comparing. This section also defines the different uncertainty scenarios we have considered. Section 4 shows and comments on the results of the case study by which we have compared the behavior of the routing algorithms under the uncertainty conditions presented in Sect. 3.

2 Semitransparent Optical Network Design

The problem of network design for a semitransparent OTN can be defined as follows. The topology of the network in terms of switching nodes and links is given. Requirements are set by providing the features of the traffic that has to be supported by the network. The final purpose is to dimension network resources

(transmission resources and regenerators), so that traffic requirements are met under the constraints (propagation impairments) imposed by the physical layer.

Recent studies have been dedicated to modeling transmission impairments, both linear and nonlinear; the reader is referred to [1–7]. A physical-layer impairment model must be simple enough to be usable. Namely, a limited number of input parameters should be sufficient to characterize each optical link; then, a single output parameter should be used to represent the effects of all the considered impairments on each optical circuit, see [8–10]. Many works adopt the *Personick Q factor*[1] (see Sect. 2.2) as single output parameter [8]. OTN design under physical impairments—intended as network dimensioning and regenerator placement for a given static traffic demand—has been treated by different works, like [8–10].

However, all these previous proposals consider that the physical information is completely accurate. Few recent works in the literature deal with the modeling/dimensioning and routing taking into account the possible inaccuracy in the physical information. In [11], a method is proposed, based on Q factor and optical signal-to-noise ratio (OSNR) measurements, to compute the error (or inaccuracy) associated with each link on the network. Authors in [12] propose a physical model that interpolates the bit error rate (BER) from experimental measurements. This interpolation introduces uncertainties which have to be considered when evaluating the feasibility of a lightpath. In [13], these uncertainties are considered by means of an extra fixed margin. A lightpath is only considered feasible if its Q factor is higher than a threshold plus this extra fixed margin. This fixed value is computed from the standard deviation of the difference between the real BER and the interpolated BER values. Finally, in [14] this extra margin is proposed to be variable, and it is based on the amount of residual chromatic dispersion and nonlinear phase experienced by the signal.

The following two sections explain how we model the *lightpath* and *physical layers*.

2.1 Optical Circuit-Layer Model

In optical circuit-switched networks, the basic managed entity is the *lightpath*, an end-to-end optical connection between a source and a destination node, requiring one wavelength channel per network link crossed. Traffic is defined by a matrix of demands between the node pairs of the network. In our design problem, we have assumed that traffic is static (all connections are permanent) and lightpaths are unprotected (the protected case will be considered in Chap. 9).

Each network link is equipped with a certain number of DWDM *transmission systems* (the details of the DWDM transmission systems will be specified in Sect. 2.2). We have assumed that all the DWDM systems in the network are

[1] From this point on, we will indicate the Personick Q factor with the abbreviation *Q factor*.

equal and provide the same preassigned maximum number of WDM channels W. The network design procedure determines how many DWDM systems have to be installed on each link to support a given traffic.

The type of approach to regeneration adopted in an OTN has a deep impact on the complexity of the planning problem.

In most OTNs (so-called *opaque*), each optical link of the network is terminated at both ends by optoelectronic interfaces hosted in the nodes. This approach simplifies network management, design, and control, as it implies a full independence of the logical layer from the physical layer. Each optical transmission system can be engineered separately, regardless of traffic conditions. Also the network at the optical circuit layer can be optimized regardless of the physical-layer transmission impairments. As OEO converters jointly perform regeneration and wavelength conversion, the two problems of wavelength assignment and of routing are also decoupled. Actually, the particular channel assignment of a connection on a specific link becomes a link local problem; hence, only routing has to be planned network-wide for each connection. The design problem is not substantially different from a classical transport network (e.g., SDH).

At the opposite extreme of the opaque approach, we have the *fully transparent* OTN. With optical transparent switching, OTN design and operation become cross-layer problems coupling the physical to the logical layer: transmission impairments and wavelength assignment have to be taken into account when the lightpath is set up and routed across the network, as well as transmission systems have to be planned on the basis of a given traffic demand. All-optical wavelength converters, as the all-optical regenerators, are not yet commercially available products. Thus, in a *fully transparent* OTN, we can assume that no wavelength conversion is allowed, and thus, routing and wavelength assignment (RWA) has to be found for each lightpath during the network design procedure. Transmission impairments limit the maximum distance reachable from the source node and thus the geographical extension of the network.

Eliminating OEO conversions completely from the network is possible only for limited-size plants. In most wide area networks, the only viable option is the *semitransparent* approach. The regenerators break up the optical continuity of those lightpaths that would be impossible to set up in transparency. Semitransparency is likely to be preferable also in smaller networks for control and monitoring reasons, at least until such functions be all optically available. A regenerator completely restores the quality of the signal along a lightpath, as if it were back-to-back to the transmitter. Regenerators are available only at the network nodes. It should be noted that an optoelectronic regenerator is also able to operate as a wavelength converter.

In the example of Fig. 6.1a, while lightpath $\mathbf{P_2}$ can be set up transparently, lightpath $\mathbf{P_1}$ needs at least a regenerator point. A regenerator is thus installed in node C (enlarged in Fig. 6.1b), splitting $\mathbf{P_1}$ into two sub-paths. The regenerator is also used as a wavelength converter. In general when a regenerator is necessary on a path, there may be several options for its placement. For instance, in the example of Fig. 6.1a, the regenerator could be installed in node B instead of node C.

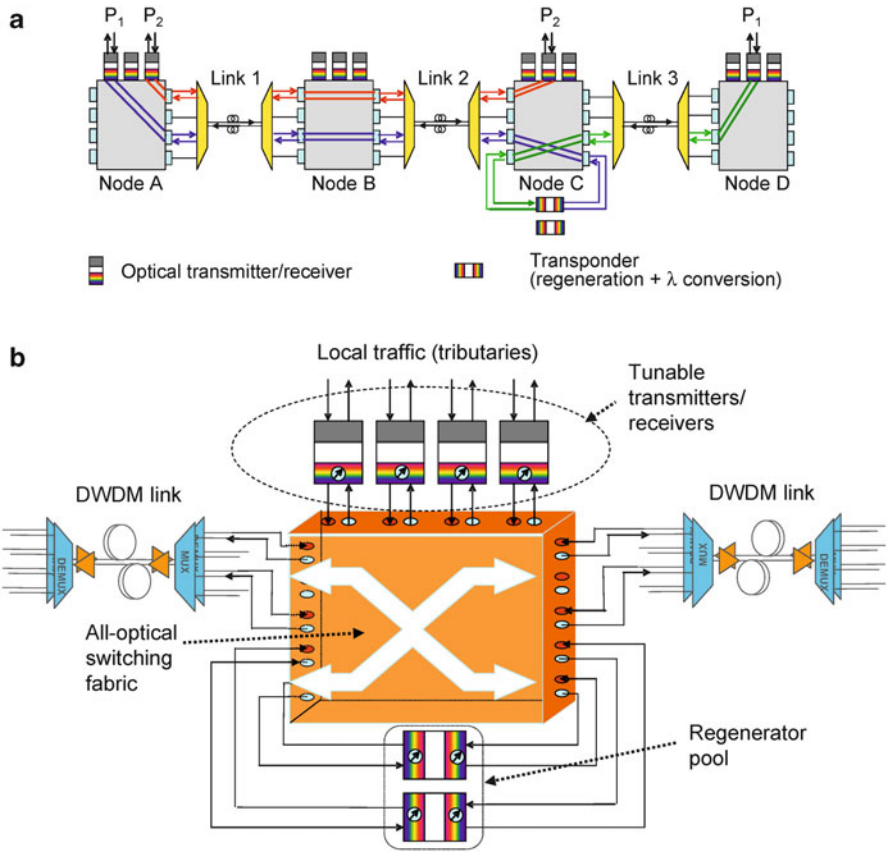


Fig. 6.1 (**a**) Semitransparent network: routing a connection in wavelength continuity (*red*) or with regeneration and wavelength conversion (*blue*/*green*). (**b**) Node model for a semitransparent network

In conclusion, designing a semitransparent network implies three problems to be solved under constraints from the transmission layer: routing, wavelength assignment, and regenerator placement (RWARP). Section 3.3 presents our design algorithm which jointly solves the three problems.

2.2 Physical-Layer Model

An optical signal is subject to impairments (linear and nonlinear) which degrade its quality as it transparently propagates through the network: the physical-layer model allows us to relate such degradation to the physical parameters of the network elements crossed along the path. Our model is based on two network elements:

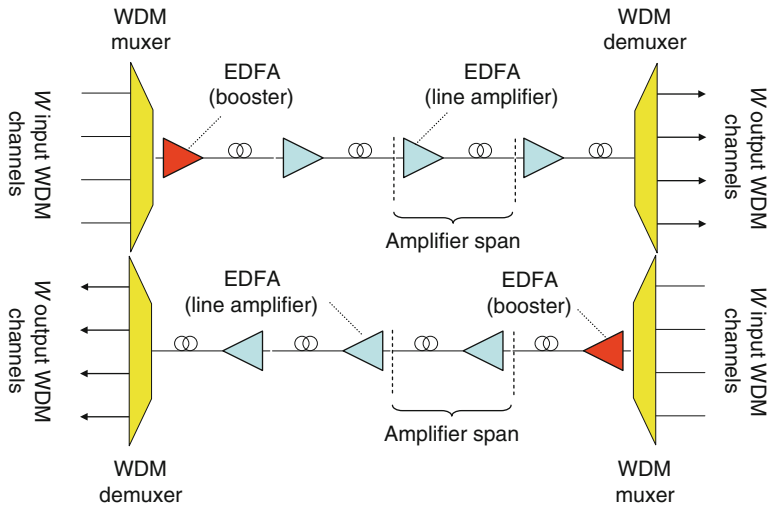


Fig. 6.2 DWDM transmission system

the optical switching node and the optical transmission system. In a semitransparent OTN, the *optical switching nodes* are the sites where transmitters, receivers, and regenerators are located (see Fig. 6.1b). The core of the node is a non-blocking all-optical switching fabric, assumed to introduce only attenuation (crosstalk, i.e., interference between WDM channels within the node is neglected in this work).

Each *optical transmission system* (Fig. 6.2) is bidirectional and composed of (1) a couple of counter-propagating fibers and (2) the set of optical erbium-doped fiber amplifiers (EDFAs) necessary to completely recover loss due to fiber propagation. An optical amplifier span is the optical fiber segment connecting an EDFA to the next one or to the end node of the link. We assume that each EDFA is placed at the beginning of its span. The first EDFA of a link in one direction is located at the output of the source node of the link (i.e., it is used as a booster); all the other EDFAs of a link are used as line amplifiers.

If a connection is unfeasible end-to-end in transparency, a regenerator is added in a transit node and dedicated to the connection: the lightpath is split by each regenerator into two contiguous *transparent sub-paths*. If R_P is the number of regenerators dedicated to an H-hop lightpath $\mathbf{P}$ ($0 \leq R_P \leq H - 1$), then the number of transparent sub-paths composing the lightpath is $R_P + 1$.[2] At the end of each sub-path (except the last one), a regenerator (OEO transponder) renews the signal. In this work we are considering 3R (re-amplification, reshaping, and retiming) regenerators, which fully restore signal quality as if at the transmitter. Thus, from a signal impairment point of view, the regeneration operation implies a complete

[2] If $R_P = 0$, the connection is end-to-end transparently feasible; hence, the single sub-path is coincident with the lightpath.

loss of memory of the history of the signal along the path followed to reach the regenerator. The regeneration of a bidirectional WDM channel crossing a node requires four unidirectional (or two bidirectional) node ports and one (bidirectional) regenerator (see Fig. 6.1b).

The key parameter to measure the signal quality is the BER, which directly contributes to the quality of service perceived by the user of a circuit-switched network: an optical circuit can be set up if the BER at the receiver is above a threshold. A BER threshold translates by well-known relations into a threshold value of the Q factor [8], which in turn can be evaluated as a function of the transmission system parameters and the transmission impairments.

The computation of the Q factor can be carried out by different methods, from analysis in the simplest cases to physical-layer simulations in the most complicated ones. In particular, we have adopted in our model a semiempirical method proposed in the European Project NOBEL 1 [15, 16]. The computation of the Q factor takes the following impairments into account: ASE, loss (linear), and self- and cross-phase modulation (nonlinear). The other effects (and, in particular, polarization mode and chromatic dispersions) are not considered.

Let us consider an h-hop sub-path $\mathbf{p}$, crossing h links having lengths $L_{(1,p)}, \ldots, L_{(h,p)}$ from the source to the end node. The Q factor at the end of the sub-path is given by the equation

$$Q_p\ [\mathrm{dB}] = a_0 + a_1 \mathrm{OSNR}_p + a_2 N_p + a_3 \left(P_0 \cdot N_p\right)^B \tag{6.1}$$

Coefficients a_0, a_1, a_2, a_3 and B depend only on the type of DWDM transmission system deployed in the network links. They are evaluated by parametric identification starting from a set of experimental measurements. The third and fourth terms of the equation take nonlinear effects into account. P_0 [dBm] is the power level at the sub-path channel signal launch. N_p is the total number of EDFA amplifier spans crossed by the sub-path, which is given by $N_p = \Sigma_{j=1}^{h} \left\lceil L_{(j,p)}/s \right\rceil$, where s is the maximum span length and all the spans of a link j have equal length $L_{(j,p)}/\left\lceil L_{(j,p)}/s \right\rceil$ ($\approx s$ for $L_{(j,p)} > $). It is assumed that each span contributes to the nonlinear effects.

OSNR_p is the optical signal-to-noise ratio over a fixed optical bandwidth (dependent on the bit rate and the modulation format of the transmitters):

$$\mathrm{OSNR}_p\ [\mathrm{dB}] = P_0 - \mathrm{QN} - 10\mathrm{Log}_{10} \left(\sum_{i=1}^{N_p} (1 - x_i) \cdot \alpha\ [\mathrm{lin}] \cdot L_{(i,p)} + x_i \cdot \mathrm{TN}\ [\mathrm{lin}] \right) - \mathrm{NF} \tag{6.2}$$

where QN is the quantum noise; $\alpha\,[\mathrm{lin}]$ is the fiber loss per km (in linear units); TN [lin] is the loss of the transparent optical switching matrix of a node (in linear units); i indicates the EDFA spans; $x_i = 1$ when i is pointing to the last span of each link; $x_i = 0$ otherwise; and NF [dB] is the noise figure of the optical amplifiers

(assumed equal for all the EDFAs, regardless if boosters or line amplifiers). A detailed list of the values used for all the parameters is reported in Sect. 4.

A minimum threshold value of Q, called $Q_{\min}$, is required at the end of each sub-path: $Q_p \geq Q_{\min} \; \forall p$. If the minimum threshold on the Q factor is not satisfied by a sub-path, this has to be further divided into sub-paths by adding other regenerators. $Q_{\min}$ is chosen according to the maximum acceptable BER (see Sect. 4 for numerical values).

2.3 Planning Procedure

RWARP is a nontrivial optimization problem (even when performed for a single connection): in fact, it contains the restricted shortest path (RSP) as a subproblem. RSP consists of finding a minimum-cost route satisfying a constraint on some parameter (e.g., delay, noise, Q factor), which monotonically increases (or decreases) with the distance from the source. RSP is known to be NP complete, and thus, RWARP is at least NP complete. Compared to RSP, RWARP is surely more complex because the Q factor is not monotonic due to the presence of the regenerators along the path. Moreover, there is an additional objective function to minimize, that is, the number of regenerators. Since exact solution methods are likely to be too complex to be useful in realistic dimension scenarios, we have adopted a heuristic approach, briefly described in what follows.

Our network planning procedure makes use of the RWARP approach described in [16, 17]. At the beginning, the network is equipped with one regenerator per node and one DWDM transmission system per link (we recall that all DWDM systems have the same maximum number of WDM channels W). The set of permanent connection requests provided by the traffic matrix is randomly sorted.[3] Then, requests are processed one by one by the RWARP algorithm, which computes the RWA under the Q factor constraint, trying to minimize the number of regenerators used by the connections and the amount of allocated resources. After serving each request, the nodes with no more free regenerators are provided with one extra regenerator and each link with no more free wavelengths on the already installed DWDM systems is provided with one additional DWDM system. After all connections have been set up, unused regenerators and DWDM systems, if any, are removed.

The RWARP algorithm is based on the definition of a metric that leads to small-length paths using the lowest number of regenerators: the cost of an H-hop path $\mathbf{P}$, routed on links having lengths $L_{(1,\,P)}, \ldots, L_{(H,\,P)}$, is $C_P = \sum_i^H L_{(i,P)} + R_P \cdot C_{\text{reg}}$, where R_P is the number of regenerators used by $\mathbf{P}$ and C_{reg} is the cost of a regenerator (normalized to the cost of a unit length of optical link). The metric is

[3] Random sorting is the simplest approach. Other more efficient sorting rules will be studied in future work.

additive and can be used in a minimum-cost path-searching algorithm. To do this, each optical cross-connect of the network is represented in an auxiliary graph as three graph nodes: *ingress*, *egress,* and *regenerator*. A direct connection (*ingress* → *egress*) corresponds to a transparent bypass, while an *ingress* → *regenerator* → *egress* connection means that the lightpath **P** has made use of a regenerator available at the node (splitting **P** into two adjacent sub-paths). C_{reg} is assigned to the graph arc *ingress* → *regenerator*, while the arc *ingress* → *egress* has null cost. Thanks to the metric and to this auxiliary graph construction, a path is sent through a regenerator (increasing C_P) only if this is necessary, that is, if otherwise the Q factor constraint would be violated. Moreover, in most cases sub-paths are routed in such a way that their Q_p is maximized, since an important component of Q-factor degradation is proportional to the total sub-path length (see Eq. 6.2) and the first term of C_P tends to minimize the length of each sub-path of a lightpath.

Due to the non-monotonic behavior of the Q factor (because of regenerator crossings), the minimum-cost path-searching algorithm cannot be as simple as Djikstra. The classical version of this algorithm finds the minimum-cost path in a sequence of steps in which at each step a comparison is carried out between two or more candidate paths. By these comparisons, candidate paths are progressively discarded, until only one shortest path remains. In our case, candidate paths cannot be immediately discarded simply on the basis of their costs: a more expensive sub-path may become the best solution if the Q factor of the cheapest sub-path falls below the threshold Q_{min}. However, not discarding candidate paths implies that the computational complexity tends to grow exponentially as the network is explored starting from the source node. Luckily, the known concept of *domination* comes into aid: given an H-hop path $\mathbf{P_1}$ and a K-hop path $\mathbf{P_2}$, both from the source node S to the same intermediate node N, $\mathbf{P_1}$ *dominates* $\mathbf{P_2}$ if $[C_{P_1} \leq C_{P_2}$ *and* $H \leq K$ *and* $Q_{P_1} \geq Q_{P_2}]$. In our algorithm, the set of candidate paths from S to N is restricted to the dominating paths, that is, each time a path of the set is dominated it is discarded.[4] The already-mentioned property that a vast component of Q-factor degradation is proportional to the sub-path length contributes to keep the number of candidate paths relatively low. In conclusion, our RWARP algorithm is a Q-constrained breadth-first search over non-dominated paths.

3 Semitransparent Optical Network Control

Let us now focus on the second cross-layer problem in semitransparent OTNs that is network control. Numerous publications concern the control of transparent and semitransparent networks in dynamic traffic conditions and propose various routing

[4] It should be remarked that discarding dominated paths may occur at each intermediate node N, that is, we do not need to know the entire paths from source to destination.

and wavelength assignment algorithms under transmission impairment constraints [18, 19] and also different extensions to the control plane for encompassing the physical impairments [20, 21].

Our approach to the problem is the following. A semitransparent network formerly designed for a given static traffic, with RWARP performed for each connection, is now considered under dynamic traffic, with lightpaths set up and torn-down on demand. Events of request and holding times are randomly generated (in our case, we have assumed a Poisson traffic model, with exponential distribution of both inter-arrival and holding times), but dynamic traffic is not entirely uncorrelated to the original static traffic matrix used as input for the design phase: in fact, we set the parameters of traffic generators so that on average the number of connections requested between each node pair is the same as in the static case.

As regenerators are already deployed according to the design phase, we assume they can only be used or not used by each dynamic connection, but not added or moved from a node to another. Thus, the control plane performs RWA for each new optical connection. If this operation is successful, the lightpath can be set up and resources for it (one WDM channel for each crossed link and all the *regenerators* it uses) are held for the whole duration of the connection. If RWA is unsuccessful, the connection request is blocked. It should be noted that blocking can occur for two reasons: lack of free resources and lack of free regenerators.

We have proposed two different RWA algorithms that will be presented in the next sections. Then we will introduce the issue of physical-layer parameter uncertainty and how it can affect RWA.

3.1 Deterministic RWA Algorithm

The first RWA algorithm we are presenting is *deterministic* in the sense that it makes use of the knowledge of the state of both lightpath and physical layers to take its RWA decisions.

The algorithm is run by the control plane at each occurrence of a connection request event. In practice, it is exactly the same as the heuristic used for the dimensioning phase (RWARP) and presented in Sect. 2.3, but with the simplification that it is used for one connection at the time, rather than on a set of connections as during planning. The medium complexity makes it more suitable for a centralized control implementation rather than a distributed one.

There are differences in its behavior which can be summarized as follows:

- Access to optical regeneration functionality is forbidden in nodes having no free regenerators: regenerators cannot be added.
- Connections have a finite duration: when a connection ends, its resources in terms of used wavelengths/regenerators are freed.

If no path can be found guaranteeing the Q-factor quality, then the connection is blocked.

3.2 Predictive RWA Algorithm

The *predictive* RWA algorithm, *prediction route according to the Q factor* (PR-Q) [16], takes into account the Q factor, and it is also based on the *prediction-based routing* (PBR) [22] mechanism. This PR-Q algorithm does not route the lightpaths minimizing the number of regenerators or maximizing Q (routes are pre-calculated). On the other hand, it is designed to counteract the negative effects of signal quality information inaccuracy by a self-learning capability. PR-Q utilizes K routes previously computed by means of the *minimum coincidence and distance* (MINCOD) routing algorithm [19].

- MINCOD algorithm exploits the concept of minimum coincidence between paths (a purely topological property) to balance the traffic load, hence reducing the network congestion. MINCOD finds the K paths from a source to a destination node having minimum distance and fewest shared links. Firstly, it chooses the shortest path (in distance); secondly, it associates a metric to the rest of routes. This metric is named minimum shared link (MSL) and is computed according to Eq. 6.3, where DP is the end-to-end distance of the particular path and SL is the number of links shared between this path and the paths selected in previous steps. The MINCOD algorithm selects the next $K - 1$ paths with minimum MSL value.

$$\mathrm{MSL} = \mathrm{DP} \cdot (1 + \mathrm{SL}) \tag{6.3}$$

- PR-Q algorithm: To implement the predictive mechanism (PBR) [22], w two-bit counters are introduced for every sub-path, where w is the number of possible wavelengths that can be assigned to the sub-path, that is, $w = W \cdot \min\{S_{(j,p)}\}$, where $S_{(j,p)}$ is the number of transmission systems installed on link j of the path. The algorithm selects the first route (and wavelength) fulfilling that the Q factor of every sub-path (and wavelength) is higher than $Q_{\min,}$ the two-bit counters of the sub-paths (and wavelengths) are lower than 2, and there is wavelength availability on all the sub-paths. In the PR-Q algorithm, a connection is set up allocating an available regenerator on all the intermediate nodes equipped with regenerators forming the path. When a lightpath is selected and it is blocked, the corresponding two-bit counters of the sub-paths causing the blocking (i.e., not accomplishing the minimum Q-factor threshold) are increased; otherwise, if the connection can be established, the two-bit counters are decreased.

3.3 Network Control in Case of Uncertainty on the Physical-Layer Parameters

The main focus of this work is to study the behavior of the control algorithms when the network is operated under uncertainty on the physical-layer parameters.

In order to simulate uncertainty, we developed an ad hoc dynamic traffic network simulator. In particular, the process of connection setup has been implemented in the following way. Upon a connection request, first, the network control system computes the RWA for the lightpath using one of the two algorithms presented in Sects. 3.1 and 3.2. If the lightpath allocation is feasible, then an attempt is made to set it up according to the computed RWA. Since the values of the physical-layer parameters that the control system has used in the RWA computation may now deviate from the real values, there is no guarantee that the lightpath can be actually set up according to the computed RWA. We have assumed that if the computed RWA is unfeasible because the signal degradation is more than foreseen, then the connection is blocked forever (there is no RWA recalculation by the control system).

In the above context, the following definitions are given:

- Q_{p_RWA} is the Q factor of a lightpath or a sub-path **p** computed by the control plane at the end of the RWA procedure using the set of values of physical-layer parameters known by the control plane itself. By definition, $Q_{p\_\mathrm{RWA}} \geq Q_{\min}$ for the sub-paths.
- Q_p_act is the real Q factor, computed with the actual set of physical-layer parameters, which may not match the values used by the control plane to compute Q_{p_RWA}. If Q_p_act $< Q_{\min}$ for any sub-path of the connection, then the connection is blocked.

In ideal conditions (Perfect Knowledge and Perfect Matching—PKPM), there is no uncertainty, hence if Q_p is the value of Q that path **p** would have in the designed network: $Q_p = Q_{p_\mathrm{RWA}} = Q_p$_act. In such situation, the physical layer behaves as predicted in the design phase and the control plane is able to exactly measure impairments.

When uncertainty on the physical-layer parameters is assumed, then the Q factor evaluated in the design phase (Q_p) and/or Q_{p_RWA} evaluated by the control plane starts deviating from Q_p_act. Obviously, the interesting cases are those in which the design procedure and/or the control plane *overestimate* Q.

We had to face the problem of how to reproduce such deviations in our simulations, as there are multiple ways in which estimations can be inaccurate. For example, one could create random deviations evenly distributed in the whole network, or concentrated only in few transmission systems, or else could generate deviations concerning all or just a subset of impairments. We have chosen a very simple approach: we have assumed that all misestimations (no matter in which link they are generated and to which effect they are due) combine together producing a fixed overestimation X on the calculated values of Q. More precisely, we have reproduced the following two uncertainty scenarios:

- *Perfect Knowledge and Imperfect Matching (PKIM)*: $Q_{p_\mathrm{RWA}} = Q$_act for all the sub-paths set up for the dynamic connection requests, but Q_p_act $= Q_p - X$ [dB]. In this case, the design procedure has been too optimistic: all the transmission equipment behaves worse than expected. The combined effects of

the worsened performance result in the X degradation factor that measures the mismatching between design and reality. Still uncertainty does not affect routing directly: the control plane automatically adapts to the unpredicted situation because impairments are perfectly measured during RWA calculation for the connections. However, in PKIM, uncertainty indirectly affects the network performance (in terms of blocking probability) because less resources and regenerators have been deployed compared to the PKPM case.

- *Imperfect Knowledge and Imperfect Matching (IKIM)*: $Q_{p\_\mathrm{RWA}} = Q_{p\_}\mathrm{act} - X$ [dB] for all the sub-paths set up for the dynamic connection requests; furthermore, $Q_{p\_}\mathrm{act} = Q_p - X$ [dB]. To the overestimations in the design phase, in IKIM, we add control-plane overestimations. There is the possibility that $Q_{p_}$act falls below the threshold value for some sub-path, despite $Q_{p\_\mathrm{RWA}} \geq Q_{\min}$ for all sub-paths. Thus, some connections that were thought to be feasible are actually blocked when their setup is attempted, after their feasibility is checked with the real physical-layer parameter.

This approach of creating mismatching by a fixed amount of decibels of degradation on each sub-path is simple but effective in producing a degradation of network blocking performance. It should be noted that, although not considered in this study, also sub-estimation of the Q factor is interesting, at least in the case of imperfect matching. Q-factor sub-estimation would not lead to unexpected performance degradation, but rather to unnecessary CAPEX. Extra expenditure due to sub-estimations will be compared to the loss of revenues (higher number of blocked connections) due to overestimations in a future development of this work.

4 Simulation Results

In this case study, transparent network planning is followed by a set of simulation sessions performed with both *predictive* and *deterministic* algorithms. The aim of the whole set of dimensioning and simulation sessions is to evaluate the behavior of the two algorithms and in particular to evaluate the advantage of the *predictive* algorithms to face the uncertainties in the network parameters or imperfect matching of the design and operational conditions.

The network used in this study is the Pan-European Network with 28 nodes and 41 links (see Fig. 6.3) reported in Deliverable D2.1 of the European Project NOBEL 2 [17]. All links are equipped with systems of $W = 40$ wavelengths each, each WDM channel modulated at 10 Gbit/s (assuming an optical bandwidth per channel of 0.1 nm). The number of systems installed in parallel on each link is calculated by the design procedure.

Static traffic is defined by a uniform matrix of demands including one bidirectional request for a 10 Gbit/s connection between each pair of nodes (i.e., 378 bidirectional connections in total).

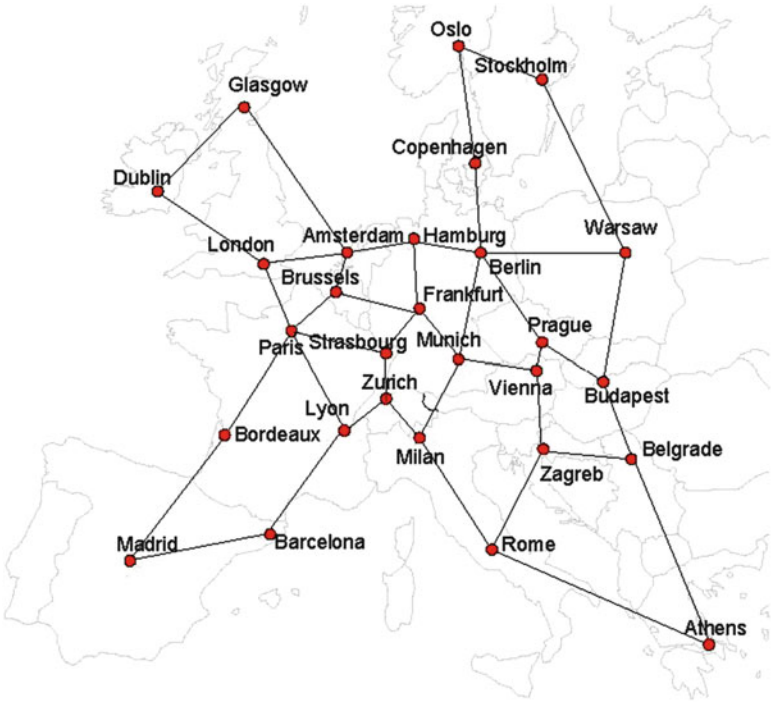


Fig. 6.3 Pan-European Network used in the study

Table 6.1 Physical model parameters

s—max span length [km]	85
α—cable attenuation [dB/km]	0.23
QN—quantum noise	58
NF—EDFA noise figure (booster and line amplifiers) [dB]	5
P_0 – [dBm]—source power	3
TN [dB]—node attenuation in dB	13.0
a_0—first coefficient, Eq. 6.1	0.4
a_1—second coefficient, Eq. 6.1	0.96
a_2—third coefficient, Eq. 6.1	−0.041
a_3—fourth coefficient, Eq. 6.1	0.02
B—power, Eq. 6.1	0.2

The set of values of the input parameters used to model the physical layer (according to Sect. 2.2) is reported in Table 6.1. The table refers to the PKPM case. The values appearing in the table have been proposed and adopted in studies developed within the European Project NOBEL 1 [15]. They have been obtained by experimental measurements carried out on commercially available DWDM transmission systems.

Table 6.2 Regenerators and system required for different values of $Q_{\min}$

$Q_{\min}$	Installed regenerators	Required full line systems (40 λ each)
17 dB	219	56
16 dB	165	56
15 dB	129	56

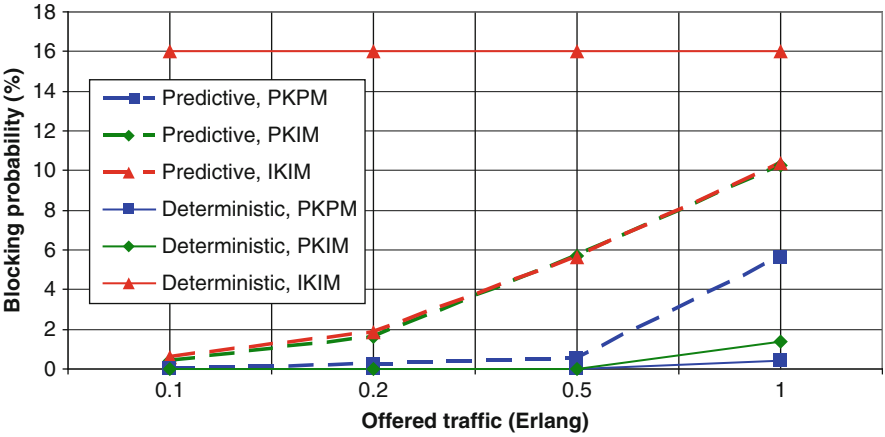


Fig. 6.4 Blocking probability for the two algorithms

Results of the dimensioning phase in terms of total number of installed regenerators and installed 40 lambda systems are reported in Table 6.2, for three different values of $Q_{\min}$. $Q_{\min} = 17$ dB roughly correspond to a BER of 10^{-12} (assuming no forward error correction is performed). The table shows that the number of systems does not depend on $Q_{\min}$, while the number of regenerators needed decreases with $Q_{\min}$.

Let us now present the dynamic traffic simulation results. We recall that the network has been previously dimensioned on the basis of a uniform static matrix. In order to have a good topological matching between dynamic traffic and network capacity, a uniform dynamic traffic has been generated as well. Different average load values between each node pair in the network have been tested, from 0.1 to 1 Erlang. All dynamic simulations are performed with $Q_{\min} = 15$ dB, that is, with the lowest number of regenerators installed.

An overestimation factor $X = 2$ dB has been introduced both for the imperfect matching and for imperfect knowledge. The objective of the simulations is the evaluation of the blocking probability in different conditions for the two RWA algorithms presented in Sect. 3. Figure 6.4 shows the blocking probability as a function of load obtained by simulating the two RWA algorithms in the three uncertainty conditions PKPM, PKIM, and IKIM. We observe that the *deterministic*

algorithm performs better in case of perfect knowledge, regardless of the matching condition (only under high traffic load a little worsening in performance is observed in case of imperfect matching). This is expected: the *deterministic* algorithm can route each connection on the base of available resources without mistakes in Q-factor computation. Under IK conditions, the *deterministic* algorithm suffers of systematic Q-factor overestimation over a set of routes. This is the reason of the flat profile of blocking probability as a function of load for the IKIM case: independently of the load the same routes are blocked due to the inaccuracy in Q; the percentage of blocking equals the percentage of routes on which Q is systematically overestimated. The *predictive* algorithm performs better only in case of imperfect knowledge and demonstrates its robustness to cope with the uncertainty on physical-layer parameters. The reason for this is that, thanks to its ability to learn from past mistakes and successes, the *predictive* algorithm can avoid systematic blocking events.

5 Conclusions

In this chapter we have compared *deterministic* and *predictive* RWA algorithms on semitransparent networks in order to evaluate their robustness under uncertainty on the values of the parameters of the physical layer. After a planning/dimensioning phase, we have simulated a semitransparent network in dynamic traffic conditions. In the dimensioning phase, network resources and regenerators can be dimensioned correctly or by overestimating signal quality when *imperfect matching* occurs between assumed and real physical-layer parameter values. In the dynamic traffic phase, the control plane can compute RWA according to reality or again overestimating signal quality when *imperfect knowledge* of the physical layer occurs.

The main outcome of this study is the evidence that, at least in the cases tested, a *deterministic* RWA approach is able to benefit from the availability of detailed physical-layer information as long as this is accurate. In such conditions, the *predictive* approach is not useful. When inaccuracy increases, however, the *deterministic* algorithm starts losing its effectiveness, while *predictive* routing reveals its robustness. For the highest uncertainty, when imperfect matching combines with imperfect knowledge, prediction routing becomes significantly better than the *deterministic* one. In the scenario of simultaneous imperfect matching and imperfect knowledge, realistic if we consider a degradation factor X of 2 dB, the gain in blocking probability we get from prediction ranges from 40% for 1 Erlang traffic to over 90% for low traffic loads.

A practical guideline message could be drawn with potential interest for optical network operators, control plane developers, and researchers working on new standard preparation. The best approach to adopt when routing connections in a semitransparent network depends on the level of accuracy of the characterization of the physical layer and of the transmission impairments. If information is guaranteed

to be very accurate and there is an actual correspondence between nominal and real parameters, then a *deterministic* routing approach is the best choice. Otherwise, it is better to rely upon a routing algorithm which ignores the exact physical layer at the beginning, allowing it to be discovered during network operation, learning from the success or failure of the connection setup attempts.

References

1. Lavigne B et al (2007) Method for the determination of a quality-of-transmission estimator along lightpaths of partially transparent networks. In: Proceedings of ECOC 2007, VDE Verlag, vol 3, Berlin, Sep 2007, pp 287–288
2. Pointurier Y et al (2007) Analysis of blocking probability in noise and crosstalk impaired all-optical networks. In: Proceedings of IEEE INFOCOM 2007, Anchorage, USA
3. Pachnicke S et al (2006) Physically constrained routing in 10Gb/s DWDM networks including fiber non linearities and polarization effects. IEEE/OSA J Lightwave Tech 24(9): 3418–3426
4. Gagnaire M, Al Zahr S (2009) Impairment-aware routing and wavelength assignment in translucent networks: state of the art. IEEE Commun Mag 47(5):55–61
5. Azodolmolky S et al (2009) A survey on physical layer impairments aware routing and wavelength assignment algorithms in optical networks. Comput Netw 53(7), Elsevier, ISSN: 1389–1286
6. Vijaya Saradhi C, Subramaniam S (2009) Physical layer impairment aware routing (PLIAR) in WDM optical networks: issues and challenges. IEEE Commun Surv Tutorials 11(4):109–130
7. Yang X, Ramamurthy B (2005) Dynamic routing in translucent WDM optical networks the intradomain case. IEEE/OSA J Lightwave Technol 23(3):955–971
8. Personick SD (1973) Receiver design for digital fiber optic communication systems, I. Bell Syst Tech J 52(6):843–874
9. Kulkarni P et al (2005) Benefits of Q-factor based routing in WDM metro networks. In: Proceedings of ECOC 2005, Glasgow, UK
10. Ezzahdi AM et al (2006) LERP a quality of transmission dependent heuristic for routing and wavelength assignment in hybrid WDM networks. In: Proceedings of ICCCN 2006, Arlington, VA, USA
11. Friskney R et al (2002) Link-based photonic path performance prediction and control. In: Proceedings of ECOC 2002, Copenhagen, DK
12. Penninckx D et al (2003) New physical analysis of 10 Gb/s transparent optical networks. IEEE Photonics Technol Lett 15(5):778–780
13. Zami T et al (2008) The relevant impact of the physical parameters uncertainties when dimensioning an optical core transparent network. In: Proceedings of ECOC 2008, Brussels, BE, September 2008
14. Leplingard F et al (2009) Interest of an adaptive margin for the quality of transmission estimation for lightpath establishment. In: Proceedings of OFC 2009, San Diego, California, Mar 2009
15. Politi C et al (2007) Integrated design and operation of a transparent optical network: a systematic approach to include physical layer awareness and cost function. IEEE Commun Mag 45(2):40–47
16. Yannuzzi M, Quagliotti M, Maier G, Marin-Tordera E, Masip-Bruin X, Sanchez-Lopez S, Sole-Pareta J, Erangoli W, Tamiri G (2009) Performance of translucent optical networks under dynamic traffic and uncertain physical-layer information. In: Proceedings of the ONDM 2009, Braunschweig, Germany, Feb 2009
17. Deliverable D2.1 of Nobel Project, Phase 2, “Preliminary Report on Multilayer Traffic Engineering and Resilience Mechanism” (A2.1 part)

18. Yang X et al (2005) Dynamic routing in translucent WDM optical networks the intradomain case. IEEE/OSA J Lightwave Tech 23(3):955–971, Mar 2005
19. Marín-Tordera E et al (2007) MINCOD-MTD: a RWA algorithm in semi-transparent optical networks. In: Proceedings of ECOC 2007, Berlin, GE
20. Strand J et al (2005) Impairments and other constraints on optical layer routing – RFC4054, May 2005
21. Castoldi P et al (2007) Centralized vs. distributed approaches for encompassing physical impairments in transparent optical networks. In: Proceedings of ONDM 2007, Athens, GR
22. Marín-Tordera E et al (2006) The prediction-based routing in optical transport networks. Comp Commun Journal, Elsevier, 19(7):865–878

Chapter 7
Analytical Models for QoT-Aware RWA Performance

Yvan Pointurier and Jun He

1 Introduction

As can be seen in Chap. 3, a large number of QoT-Aware or Impairment-Aware Routing and Wavelength assignment algorithms (IA-RWA) were designed to minimize blocking in dynamic transparent optical networks. The very vast majority of those algorithms were evaluated through extensive simulations. With time, proposed IA-RWA grew in complexity, actually making their accurate evaluation possible only using full-scale simulations. Full-scale simulations, however, tend to be lengthy for the following three reasons: (a) the growing complexity of the proposed IA-RWA techniques; (b) the increasing complexity of the networks that must be modeled (spurred for instance by the increase in the number of wavelengths that can be routed in the network—note that this could be somewhat offset by the deployment of networks with fewer, higher-capacity channels); and (c) the inclusion of more complex QoT models in the simulations—more accurate QoT models are typically more simulation intensive. In addition, establishing a new lightpath may disrupt lightpaths that are already established through the addition of cross-channel effects, such as node crosstalk or non-linear effects. Such disruption is not desirable in a transparent network, and hence in simulations of IA-RWA the QoT of any lightpath that may be disrupted by the arrival of a new demand should be evaluated. Hence, for each new demand, the QoT of many lightpaths may have to be evaluated. If a blocking rate of 10^{-5} or less is desired, then the simulation of the arrival of (many times more than) 10^5 lightpaths is required; if a network operator wants to test an IA-RWA in less than 10 min, then a decision must be reached for

Y. Pointurier (✉)
Alcatel-Lucent, Bell Labs, Nozay, France
e-mail: yvan@ieee.org

J. He
The University of Arizona, Tucson, AZ, USA
e-mail: jhe@optics.arizona.edu

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*, Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_7,

each demand in (much) less than 60 ms. This can prove difficult to achieve if QoT is to be estimated accurately.

An alternate solution to this dimensioning problem is the utilization of fast analytical models to estimate the blocking rate in a QoT-aware network. Similar to analytical QoT estimators, the accuracy of analytical blocking probability estimators may be somewhat degraded, yet sufficient for a designer to make a decision on the dimensioning of a network. Network designs that are shown to be clearly unacceptable (e.g. too high a blocking rate) when evaluated by an analytical technique can be ruled out and more promising designs kept for more thorough investigations. Analytical techniques for blocking probability evaluation can therefore be seen as useful tools to help the planning of transparent optical networks.

As it turns out, the field of blocking rate analysis in QoT-aware networks is still in its infancy, in spite of the development of several analytical techniques to analyze blocking rates in QoT-unaware optical networks, as we will see in Sect. 7.2. Essentially, only two analytical techniques were developed for QoT-aware networks. The first technique applies only to random-pick wavelength assignment (Sect. 7.3), while the second applies to networks with first-fit wavelength assignment (Sect. 7.4).

2 Blocking Probability Analysis in QoT-Unaware Optical Networks

Before presenting analytical techniques to evaluate the blocking probability of a call or traffic demand in a transparent optical network when physical layer impairments are included, it is important to understand what was done at the time when physical layer impairments were considered to be a non-issue. Indeed, both techniques that include those impairments are rooted on techniques where the impairments are not accounted for. The key features of the techniques that compute blocking probabilities analytically in transparent optical networks are summarized in Table 7.1.

A seminal attempt to analyze wavelength blocking in transparent optical networks dates back to 1995 [1]. Two key assumptions are made: (a) the blocking process on a link is independent of blocking on other links (a characteristic of the "reduced load approximation" method); and (b) the utilization of a wavelength is link-independent. These assumptions lead to large inaccuracies especially in sparse networks, where the average node degree is small. In addition the model is computationally expensive: its complexity scales exponentially with the number of nodes of the considered network.

The wavelength independence assumption is also made in [2], where wavelength conversion is introduced for the first time. In further works, the wavelength utilization independence assumption is relaxed, making the algorithms more accurate and applicable to a wider variety of networking scenarios. For instance, the

Table 7.1 Analytical estimation of blocking probability

Reference	Routing	Wavelength assignment	Wavelength conversion	Topology	Other
[1]	Fixed, least loaded	Random	No	Arbitrary	Up to 3 links per path Up to 2 links per path
[2]	N/A	N/A	Yes	Arbitrary	
[3]	Fixed	Random	Yes	Arbitrary	
[4]	Fixed, alternate	First fit	Yes	Arbitrary	
[5]	Fixed, alternate	Random	Yes	Arbitrary	
[6]	Fixed, alternate	First fit	No	Arbitrary	
[7]	Fixed	Random, first fit	No	Arbitrary	
[8]	Fixed, alternate, least-loaded	Random	No	Arbitrary	
[9]	Fixed	Random	No	Arbitrary	Includes blocking due to outdated information
[10]	Fixed	Random, first fit	No	Arbitrary	

technique outlined in [3] includes (partial) wavelength conversion and was successfully tested on regular as well as arbitrary topologies.

The vast majority of analytical works focus on fixed routing, where a single route of a demand between 2 nodes is pre-computed and fixed over the life of the network, and wavelength assignment is typically assumed to be random. With alternate routing, several routes are precomputed between each source and destination, such that a demand may use any of those routes if the others are not available. In order to analyze networks with alternate routing, a "layered approach" is often used: the network is modeled as a layered graph where each layer corresponds to a wavelength and arriving demands are modeled as flowing from one layer to the other until a path and a wavelength are jointly found. In particular, in [4] the traffic that flows from one layer to the next ("overflow traffic") is modeled as a Bernoulli–Poisson–Pascal (BPP) process with an assumption of link independence, that is, the probability of a link being available at any given wavelength is independent of the states of the other links. In [5], wavelength conversion is additionally included.

In [6], the Erlang fixed-point approximation, originally used for generic circuit switched networks [11], is extended by developing a new link blocking model, based on a continuous time Markov chain. The authors analyze the performance of the single hop traffic grooming heuristic where demands with same source and destination can be groomed into the same lightpath, with the alternate routing and sequential wavelength assignment in the optical networks. Link independence

is a general assumption of the link decomposition technique adopted by the Erlang fixed-point approximation, therefore it overestimates blocking rates, especially in bus networks [7, 10]. In [8], a new algorithm was proposed, that proves to be very accurate despite a relatively simple formulation, which relaxes the link independence assumption into a two-link correlation assumption, i.e., the wavelength utilization on a link of a given route depends only on that of the one previous or next link of the route, but not of other links' utilizations. In [7], the traditional link independence assumption, which is known to yield inaccuracies in the estimation of blocking probabilities, is relaxed into an "object independence assumption", where an object can be, for the first time, a complete path, or, more traditionally, a link.

In [10], the author uses the "layered" approach to divide the network into layers. Wavelength independence is assumed and an iterative model to calculate the blocking probability for fixed routing in networks with any topology is derived. Moment matching and equivalent path methods are used to characterize the overflow traffic. This method underestimates the wavelength blocking probability because of the approximation for the rate of the overflow traffic [12].

Furthermore, blocking may occur not only because resources are lacking or QoT is unacceptable, but also because propagating state information in a network takes some time (essentially due to the required signalling), such that information maintained locally in the network may become outdated. Wrong decisions may be made due to the lack of up-to-date information: such scenario was analyzed in [9].

Note that none of these techniques accounts for physical blocking due to insufficient QoT, that is, blocking that is due to physical layer impairments instead of the lack of available resources. In the next sections, we describe two algorithms that include physical layer blocking.

3 Blocking Probability Analysis with Random-Pick Wavelength Assignment

The first analytical technique where the QoT blocking probability in a transparent optical network is estimated makes a number of relatively strong assumptions [13]. First, routing is fixed and wavelength assignment is random. Other routing schemes, such as alternate routing, and other wavelength assignment schemes, such as first-fit, or other combined IA-RWA schemes, are not supported by this technique. The "random-pick" wavelength assignment assumption stems from a more fundamental assumption: the strict equivalence of all wavelengths in the network, which has an impact on the physical layer modeling of the network. Indeed, considering all channels equivalent makes it impossible to account for per-channel effects; this includes Optical Signal-to-Noise Ratio (OSNR), which may vary with the considered channel for instance due to amplifiers' tilt, but also the inter-channel nonlinear effects (cross phase modulation, XPM and four-wave

mixing, FWM) that could be limiting impairments in certain next-generation transparent networks. This assumption, however, is valid for metro or regional networks where routes are not more than a few hundreds of kilometers, and where node crosstalk could be an important impairment.

The proposed technique makes further assumptions, however, they are less important in the sense that the algorithm could be modified to weaken or remove those assumptions. In particular, full transparency, that is, the absence of wavelength regeneration, is assumed, and the only physical effects that were included in the study were: Amplified Spontaneous Emission (ASE) noise, chromatic dispersion/Self Phase Modulation (SPM).

Despite a limited applicability, this first analytical technique is computationally efficient and enables a quick dimensioning of a transparent optical network including some physical impairments. In addition, the technique not only enables the inclusion of static impairments (impairments that do not change with the network state, i.e. that do not depend on the establishment of other lightpaths) in the physical layer modeling, but it also permits the inclusion of dynamic impairments—in this case, node crosstalk caused by poor port isolation. Here node crosstalk is assumed to be of the "adjacent wavelength self-crosstalk" type [14], that is, the in-band crosstalk produced when a leaked copy of the signal follows a different path through an optical cross-connect (OXC) from the input to the output port of the main signal. This happens when a lightpath on an adjacent wavelength enters and exits the same OXC ports as the channel of interest. Other types of crosstalk, such as switching crosstalk and neighbor-crosstalk are ignored here; for networks utilizing optical switches based on MicroElectroMechanical Systems (MEMS) technology, these two sources of crosstalk are much weaker.

3.1 Network and QoT Models

Consider a network with bidirectional links, each supporting exactly W channels or wavelengths. A traffic demand is a lightpath request between a source and a destination node. Assume that traffic demands arrive according to a Poisson process with mean rate $\Lambda_{R(s,\,d)}$ for route $R(s,\,d)$ from node s to d, and that their durations are distributed exponentially with mean $M_{R(s,\,d)} = 1$, without loss of generality, such that the offered load in Erlang on a route R is Λ_R. We further assume that (A1) wavelength occupancies on disjoint routes are independent of each other; and (A2) the establishments of lightpaths on two different routes are independent events. These assumptions make analysis tractable and fast, yet approximate well the behavior of transparent optical networks. Assumption (A1) is made in other papers focusing on wavelength blocking, most notably, in [3, 8]. Assumption (A2) is used

further in this chapter; numerical results show that the resulting approximations do not affect much the accuracy of the method.

Now consider QoT estimation. The QoT of a lightpath is modeled through its Q factor. Since ASE noise, intersymbol interference (ISI) caused by chromatic dispersion and SPM, and node crosstalk are accounted for, the Q factor for an intensity-modulated lightpath routed over R can be expressed as

$$Q_R = \frac{\mu_{1,R} - \mu_{0,R}}{\sigma_{0,R} + \sigma_{1,R}} = \frac{\mu_{1,R} - \mu_{0,R}}{\sigma_{0,R} + \sqrt{\sigma_{i,R}^2 + \sigma_{n,R}^2 + \sigma_{X,R}^2}} \tag{7.1}$$

where $\sigma_{1,R} = \sqrt{\sigma_{i,R}^2 + \sigma_{n,R}^2 + \sigma_{X,R}^2}$, and $\sigma_{i,R}^2$, $\sigma_{n,R}^2$, and $\sigma_{X,R}^2$ are the variance contributions due to intersymbol interference, amplifier noise, and node crosstalk, respectively; while $\mu_{1,R}$ and $\mu_{0,R}$ are the "1" and "0" symbol means after photodetection and $\sigma_{0,R}$ is the standard deviation of the "0" symbols after photodetection. Here we make the (usual) assumption of a high transmitter extinction ratio, such that ASE noise and crosstalk impairments can be ignored for the "0" bits. In addition, we assume that all signals are in the same polarization state, a worst-case scenario typically used to design networks (see for instance [15] in the context of crosstalk modeling).

Node crosstalk is a dynamic effect that depends on the network state, that is, on what lightpaths are established in the network. The crosstalk $\sigma_{X,R}$ consists of the accumulation of crosstalk at each node along route R. Therefore, recalling that node crosstalk corresponds to optical leaks within the nodes, and calling $\sigma_{x,R}^2$ the variance contribution of each of these leaks, then

$$\sigma_{X,R}^2 = n\sigma_{x,R}^2, \tag{7.2}$$

if n of those leaks have accumulated over route R. In the following, we do not differentiate between leaks coming from adjacent channels (or ports) and non-adjacent channels. If non-adjacent crosstalk is actually weaker than adjacent crosstalk, then the technique would actually result in over-dimensioning, as the weaker non-adjacent channel crosstalk components would then be on an equal footing with the stronger adjacent channel crosstalk. Nevertheless, because the technique takes *dynamic* effects into account, all in all it reduces the amount of over-dimensioning required at network design time.

Before we describe how the blocking probability can be estimated analytically, we note that the exact nature of the QoT estimator described by (7.1) is not essential in the algorithm presented next: other QoT estimators could be used. In particular, the QoT estimator presented here applies only to intensity modulation or on–off keying (OOK), while modern long-haul networks increasingly use coherent modulation. Any QoT estimator where static and dynamic impairments are separable, where static impairments can be pre-estimated, and where dynamic impairments

can be counted (e.g., as in (7.2)), are candidates for inclusion in the following algorithm.

Algorithm 7.1 Blocking probability computation for fixed routing and random pick wavelength assignment.

1: Initialize $B_R^{(\lambda)}$ for all R (as in Section 7.3.2).
2: Initialize $B_R^{(q)} = 0$ and $B_R = B_R^{(\lambda)}$ for every route R.
3: **repeat**
4: Let $\tilde{B}_R = B_R$.
5: Compute $B_R^{(q)}$ for all R: using Equations (7.4), (7.5), (7.6), (7.7), (7.8), (7.9).
6: Compute B_R for all R: using (7.3).
7: **until** $(B_R - \tilde{B}_R)/B_R < \varepsilon$ for every route R

3.2 Performance Analysis Under QoT Constraints

3.2.1 Overview

We now assume that wavelength blocking probabilities are known, for instance using one of the techniques described in Sect. 7.2. Notice that, for a route R, the blocking probabilities due to wavelength continuity $B_R{}^{(\lambda)}$ and due to QoT $B_R{}^{(q)}$ are related by:

$$B_R = B_R^{(\lambda)} + (1 - B_R^{(\lambda)})B_R^{(q)}, \tag{7.3}$$

where B_R is the blocking probability for a traffic demand on route R. Indeed, a traffic demand can be blocked due to QoT only if a wavelength is available on the route assigned to the traffic demand, that is, when the traffic demand is not blocked due to the wavelength continuity constraint. Assuming $B_R^{(\lambda)}$ is known for all R, we need to determine the QoT blockings $B_R^{(q)}$ in order to compute B_R. The steps to do so are outlined next, and the general algorithm to compute the blocking probability in a transparent network with fixed routing and random-pick wavelength assignment accounting for physical layer impairments is outlined in Algorithm 1. The QoT blocking computation algorithm is iterative and stops when a convergence criterion (e.g., blocking rate difference for each route between two consecutive iterations lower than a preset threshold ε) is met.

We emphasize that Algorithm 1 is largely independent of the algorithm used to compute wavelength blockings $B_R^{(\lambda)}$ for every route R. However, some algorithms, e.g. the algorithm proposed in [8], rely on the computations of *conditional blocking probabilities*, for instance $B_{R|X_j}^{(\lambda)}$ where X_j is the number of free wavelengths on link j. How to handle the interaction between such algorithms and Algorithm 1 is out of the scope of this chapter, however, the interested reader can refer to [13] for further details on this topic.

In order to compute the blocking probability due to QoT, we first assume that the wavelength blocking probabilities $B_R^{(\lambda)}$ are known, and we determine the distribution of the number of crosstalk components XT_R that impair each route R in Sect. 7.3.2.2. Then, we relate the crosstalk distributions to the blocking probability due to QoT in Sect. 7.3.2.3.

3.2.2 Distribution of the Number of Crosstalk Terms

We seek to determine the distribution of the number of crosstalk components XT_R that impair each route R.

The probability that a lightpath is established on route R is given by:

$$p_R = \frac{\Lambda_R}{M_R}\frac{1-B_R}{W} = \Lambda_R\frac{1-B_R}{W}. \tag{7.4}$$

Let $U_R(k)$ be the probability that $0 \leq k \leq W$ traffic demands use route R. Using assumption (A2), U_R can be approximated as a binomial random variable. Considering that establishing a lightpath on route R is a Bernoulli trial with success probability p_R, we can infer that the probability that exactly k traffic demands are established on route R is:

$$U_R(k) \approx \frac{W}{k}p_R^k(1-p_R)^{W-k}, k = 0, 1, \ldots, W. \tag{7.5}$$

Let $U'_{R,\,R'}(m)$ be the probability that some route R' injects m crosstalk components on route R. Route R' can inject crosstalk on R only in the nodes that are at the same time on R and R'. Let $n_{xt}(R, R')$ be this number of common nodes. If k lightpaths are routed on R then the number of crosstalk components injected by R' on R is $kn_{xt}(R, R')$, therefore:

$$U'_{R,R'}(kn_{xt}(R,R')) = U_{R'}(k). \tag{7.6}$$

Finally, let $I_R^{xt} = \{R_1, \ldots, R_p\}$ be the set of the routes that are potential sources of crosstalk for lightpaths established on route R, and $XT_R(k)$ be the probability that route R is subject to exactly k crosstalk components (from any other route $R' \in I_R^{xt}$). The total number of crosstalk components k seen by route R is the sum of all crosstalk components injected at each node of R by all routes that intersect R. Using assumption (A2), the probabilities for establishing lightpaths on different routes are independent, such that the random variables $U'_{R,R'}$ are independent. Therefore, $*$ denoting the convolution operator, the distribution of XT_R is:

$$XT_R = U'_{R,R_1} * \ldots * U'_{R,R_p}. \tag{7.7}$$

This distribution is used next to compute the QoT blocking probability.

3.2.3 Blocking Probability Due to QoT

We now make the link between the physical and the network layer of a transparent optical network.

Since we assumed that all impairments—node crosstalk excepted—are static, $\mu_{1,R}$, $\mu_{0,R}$, $\sigma_{0,R}$, $\sigma_{i,R}$ and $\sigma_{x,R}$ can be assumed to be known (through offline computations) for all routes in the network. Hence, we can compute the maximum number of crosstalk components $N_{\max}(R)$ that a route R can accommodate to maintain a Q factor above a predetermined threshold Q_{th} using (7.1) and (7.2):

$$N_{\max}(R) = \left\lfloor \frac{\left(\frac{\mu_1-\mu_0}{Q_{th}} - \sigma_{0,R}\right)^2 - \sigma_{i,R}^2 - \sigma_{n,R}^2}{\sigma_{x,R}^2} \right\rfloor. \tag{7.8}$$

Therefore, the probability that a lightpath is blocked because it does not meet the QoT constraint is the probability that this lightpath is subject to $N_{\max}(R)$ crosstalk components or more; that is:

$$B_R^{(q)} = \sum_{k>N_{\max}(R)} XT_R(k). \tag{7.9}$$

Finally the blocking probabilities due to wavelength continuity and due to QoT are related through (7.3).

3.3 Validation with Simulations

To validate this analysis, we use the following sample scenario. We consider 2 different networks where all links are made of 70 km of SMF and full post-dispersion compensation is assumed at every node. The port isolation is 30 dB. The first network is a simple bidirectional ring of 6 nodes, where there is a single span of standard single-mode fiber (SSMF) between two adjacent nodes. The second network (see Fig. 7.1) based on the NSFNET topology, modified as follows: each link was scaled down by a factor of 10, and split into fiber spans of 70 km (roundings are used so that each link consists of an integer number of spans). Both topologies therefore emulate a metro or regional network. All lightpaths, OOK-modulated at 10 Gb/s with a peak power of 2 mW per channel, maintain a Q factor greater than $Q_{th} = 6$, corresponding to a BER of 10^{-9} when no FEC is used. FEC can easily be included through a simple modification of the threshold Q_{th}.

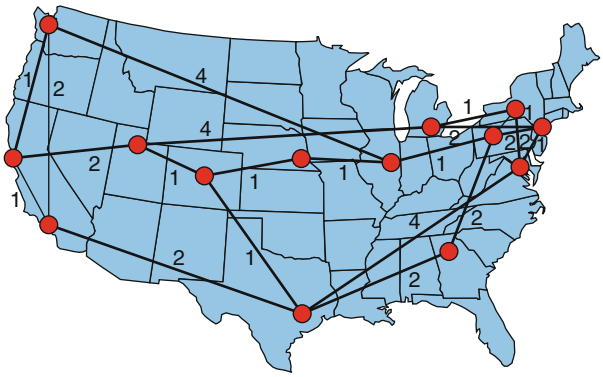


Fig. 7.1 Down-scaled version of the NSFNET topology (scaling factor: 1/10). In the figure, the weights represent the number of 70-km long spans for the links

Let B be the (total) mean traffic demand blocking probability over all pairs of nodes for a network. Similarly, denote by $B^{(\lambda)}$ the mean traffic demand blocking probability incurred by breaking the wavelength continuity constraint and by $B^{(q)}$ the mean traffic demand blocking probability due to the impossibility of satisfying the QoT constraint.

We simulated the resource allocation for 10^5 traffic demands and compared the blocking rates with the proposed analytical method. In order to obtain the values for $B^{(\lambda)}$, which are needed by Algorithm 1, we used the technique described in [8] for QoT-unaware networks. Algorithm 1 is iterative and was assumed to have converged whenever the difference between the blocking rates for each route differed by less than 1% between two consecutive iterations.

First consider the 6-node ring topology, with 32 wavelengths per link ($W = 32$). Blocking probabilities for this topology are shown in Fig. 7.2. In this case, QoT blockings dominate and $B^{(q)} \approx B$. The analytical technique outlined in this section estimates accurately blocking probability in a wide operation range (blocking probabilities varying over 4 orders of magnitude).

In Fig. 7.3, we show the blocking probabilities for the NSFNET topology for a total number of wavelengths of only $W = 8$. Although impractical, such a small number of wavelengths forces the wavelength blocking probability to be non-negligible, and permits to assess the accuracy of the proposed technique in a more complex context. The analytical technique is less accurate than with the more simple ring topology with QoT blocking only. However, analytical values for the overall blocking probability are well within an order of magnitude compared with the values obtained through simulation. For instance, QoT blocking dominates wavelength blocking for loads lower than 30 Erlangs, a result that is well predicted with the analytical technique.

On standard computing hardware, blocking probabilities estimation with the analytical method takes at most a few minutes, which is compatible with the rough estimation of the behavior of a network by the network designer.

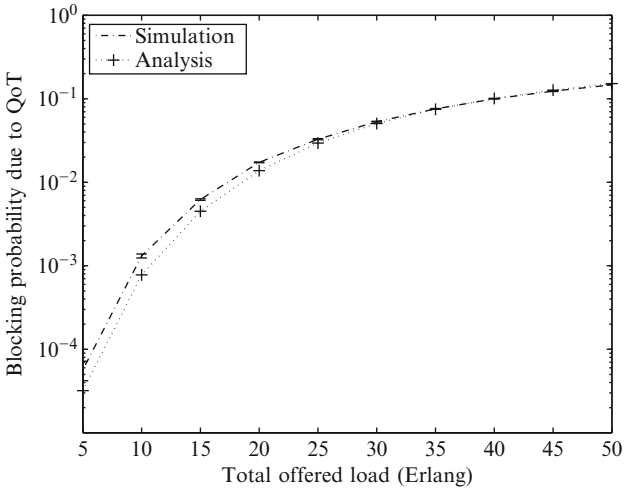


Fig. 7.2 Blocking probability for the ring of 6 nodes, 32 wavelengths, −30 dB crosstalk; 95% confidence intervals are given for the simulation curve

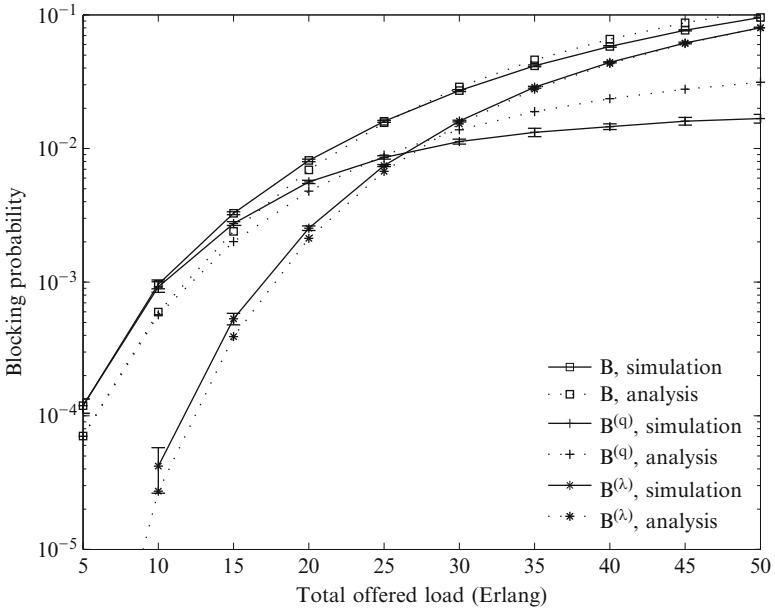


Fig. 7.3 Blocking probability for the NSFNET topology, 8 wavelengths, −30 dB crosstalk; 95% confidence intervals are given for the simulation curve

4 Blocking Probability Analysis with First-Fit Wavelength Assignment

In this section, we present the first analytical model to compute the blocking probabilities for first-fit (FF) wavelength assignment for networks with QoT constraints [12].

The FF wavelength assignment algorithm has been shown to be more powerful at reducing wavelength blocking probability in networks not susceptible to physical impairments, and has therefore become one of the preferred wavelength assignment techniques [16, 17]. However, when node crosstalk or nonlinear effects (XPM, FWM) are one of the primary causes of QoT constraint violations, the FF approach is more vulnerable to physical layer degradations than RP because of FF's successive wavelength usage. Node crosstalk is due to imperfect WDM demultiplexing, and is significant when the channel spacing is tight compared to the data rates (e.g., 10 Gbps or higher, using a 50 GHz spacing or less). Even though a state-of-the-art system can be designed to suppress some of the node crosstalk, as the throughput increases, node crosstalk level can be as high as -20 dB, becoming an important source of physical layer impairments [14]. Hence, it is of great interest to analytically and accurately predict how vulnerable these FF systems are to this type of degradation.

Both wavelength blocking and QoT blocking are included in this analytical model, where each path is treated individually and the total blocking probability is calculated by averaging all path blocking probabilities. Here, no wavelength conversion is assumed.

The analysis includes QoT blocking due to node crosstalk, ASE noise, thermal noise and shot noise. In particular, we consider as above "adjacent wavelength self-crosstalk" as the type of node crosstalk. This analytical model can be extended to analyze the impact of other physical impairments on the network performance, such as Polarization Mode Dispersion (PMD), XPM, and FWM. It is also capable to evaluate other variants of FF algorithms, such as *first-fit with ordering* proposed in [18].

4.1 Analytical Framework

The transparent optical network with physical impairments is decomposed by wavelengths as a layered system, as shown in Fig. 7.4. A network with W wavelengths in each link is decomposed into W layered networks, each of which has the same topology but a capacity corresponding to a single wavelength in each link. The offered network traffic enters into the first layer ("Wavelength 1" in Fig. 7.4), then the traffic that is blocked on this layer (also called "overflow traffic")

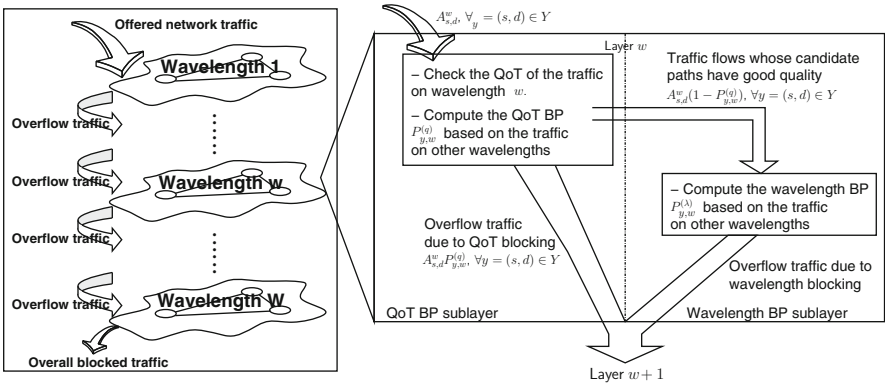


Fig. 7.4 Layered network model for transparent optical networks, showing the analytical technique for the first layer for a QoT-aware FF wavelength assignment algorithm

flows down to the second layer ("Wavelength 2"), etc. The overflow traffic is modeled as an offered load to the next layer, adjusted for its bursty nature. At layer W the overflow traffic is the overall network blocked traffic, and the blocking probability can be computed. The wavelength continuity constraint, which forces a traffic demand to remain on the same wavelength along the path, is automatically enforced in this approach. Conditional independence of the probability of each wavelength being used is assumed, given the arrival rates. Since this model deals with each wavelength separately and since the traffic flows from one wavelength to the next, it is natural to use this method to analyze FF wavelength assignment.

First, the assumptions used in the model are given. We assume that routing is fixed (e.g., shortest path first) and that wavelength assignment is first-fit. We list the important notations as follows. Unless otherwise stated, other notations and assumptions are the same as those already used in this chapter.

- W: the total number of accessible wavelengths on the fiber; w is used as the wavelength (layer) index.
- Y: the set of all source–destination pairs in the network, i.e., $Y = \{y_1, \ldots y_{|Y|}\}$[1]; y without a subscript is used to indicate any source–destination pair (s, d).
- $R(s, d)$ or $R(y)$: a route from source node s to destination node d, i.e., originating from node s and ending at node d.
- $\Lambda_{R(y)}$: the Poisson arrival rate for source–destination pair $y \in Y$.

[1] We use the notation $|X|$ to denote the cardinality of the set X.

- $A_{s,d}^{w}$: the equivalent Poisson arrival rate (offered load) to wavelength (layer) w for $R(s, d)$. Then clearly $A_y^{1} = \Lambda_y$.
- $\bar{A}_y^{w+1}$: the mean value of the overflow traffic from layer w to layer $w + 1$. Note that the overflow traffic is not Poisson distributed.
- $N_{\max}(y)$: the maximum number of crosstalk terms allowed in $R(y)$ before the QoT constraint is violated.
- $\underline{I}^w$: a $1 \times |Y|$ random vector, $\underline{I}^w = \left[I_{y_1}^w, \ldots, I_{y_{|Y|}}^w\right]$ for wavelength w, where $I_y^w = 1$ if $R(y)$ is active on wavelength w and $I_y^w = 0$ otherwise.
- $\xi_k^w[y]$: kth QoT blocking event.
- $P_{y,w}^{(\lambda)}$: the wavelength blocking probability for source–destination pair $y \in Y$ due to lack of resource on wavelength w.
- $P_{y,w}^{(q)}$: the QoT blocking probability (BP) for $R(y)$ on wavelength w, due to unsatisfactory QoT.
- P_y: the overall blocking probability for path $R(y)$, including both wavelength blocking and QoT blocking.
- P: the overall network blocking probability.

In Fig. 7.4 we show the first layer of a system if a QoT-aware RWA algorithm is applied. Expressions for the blocking probabilities are derived in Sect. 7.4.2. Each layer is divided into two sublayers, a QoT blocking probability sublayer and a wavelength blocking probability sublayer. After baseline values for the wavelength blocking probability and flow rates for a QoT-unaware system have been computed using one of the models described in Table 7.1, the offered traffic is checked for QoT compliance in the QoT blocking probability sublayer for each path. Then the offered traffic enters the wavelength blocking probability sublayer to update the flows by rechecking whether the lightpath is available. Then the total blocked traffic from the two sublayers overflows to the next layer. The QoT blocking probability depends on the usage of other layers and is computed by using the stationary state probabilities in other layers. Thus we have to iterate this process until a steady-state is reached. Figure 7.4 shows one iteration of the algorithm.

4.2 RWA Analytical Model with QoT Constraints

We assume that the wavelength blocking probabilities $P_y^{(\lambda)}$ are known, for instance using one of the techniques described in Table 7.1. In this section, we use the technique in [10] to compute wavelength blocking probability. Thus, for each wavelength layer w, we obtain its non-Poisson traffic arrival rate $\bar{A}_y^{w}$ and use it to estimate the equivalent Poisson arrival rate $A_{s,d}^{w}$.

4.2.1 Counting QoT Blocking Events

In order to calculate the QoT blocking probability, we have to count QoT blocking events, i.e. to enumerate all possible ways that a QoT blocking event can occur when a traffic demand for source–destination pair y reaches wavelength w. We use the physical model in [18] to estimate the bit-error-rate (BER) of the lightpath. Given the network topology, the QoT blocking probability depends on the level of noise, the strength of self-crosstalk, and the number of terms of self-crosstalk, which are all determined by the path $R(y)$ for all $y \in Y$. Among them, the number of crosstalk terms for y is the only variable influenced by the instantaneous state of the network. As listed above, let notation $N_{\max}(y)$ represent the maximum number of crosstalk terms allowed in $R(y)$ before the QoT constraint is violated. $N_{\max}(y)$ can be computed for each routing path $R(y)$, for all $y \in Y$, based on the model for counting the number of crosstalk terms in [19].

List all source–destination pairs in Y as $y_1, \ldots, y_{|Y|}$. Random vector $\underline{I}^w \left(\underline{I}^w = \left[I^w_{y_1}, \ldots, I^w_{y_{|Y|}}\right]\right)$ represents the utilizations of paths at layer w, i.e., the state of the network layer. We use the notation $\underline{I}^w(y') = \left[I^w_{y_1}, \ldots, I^w_{y'} = 1, \ldots, I^w_{y_{|Y|}}\right]$ to represent a state such that path $R(y')$ is known to be used at wavelength w. All used paths at wavelength w form a subset of Y, called $Y^w = \{y : I^w_y = 1\}$. Because of the wavelength continuity constraint, two paths cannot be used in the same layer simultaneously if they share a common link.

Based on [14], self-crosstalk occurs if and only if two paths share two or more consecutive hops on adjacent wavelengths. Then, one unit of self-crosstalk is inserted to both traffic demands for each set of two consecutive hops. A $1 \times |Y|$ deterministic vector $\underline{X}_{y'} = \left[X^{y_1}_{y'}, \ldots, X^{y_{|Y|}}_{y'}\right]$ is defined to describe the possible crosstalk interference at a chosen y' from other source–destination pairs. Given the routing table and self-crosstalk definition, it is straightforward to count how many units of self-crosstalk, say $X^y_{y'}$, are induced by $y \in Y$ on a selected y' when $R(y)$ is lit on an adjacent wavelength.

In this work, we only consider crosstalk from the nearest adjacent wavelengths. Thus, only lightpaths on wavelength $w - 1$ and wavelength $w + 1$ can introduce crosstalk to $R(y)$ in layer w. Then when a traffic demand arrives on y, a QoT blocking event happens at wavelength w if the current network state at $w - 1$ and $w + 1$ satisfies the following condition:

$$(\underline{I}^{w-1} + \underline{I}^{w+1}) \cdot \underline{X}_y^{\mathrm{T}} > N_{\max}(y), \tag{7.10}$$

where the superscript T indicates a matrix transpose.

Moreover, a new assigned lightpath at wavelength w cannot cause the QoT of other existing lightpaths to violate the QoT requirement. Thus, wavelengths $w + i$ for $i = -2, -1, 0, 1, 2$ need to be considered as well. Note that for $w = 1, 2, W - 1, W$, we do not count QoT blocking events that refer to wavelengths outside the

spectrum band. To preserve the transmission quality of existing traffic, a traffic demand arriving on y has to be blocked if either (7.10) holds, or if

$$\exists y' \in Y^{w+1} \; s.t. \; (\underline{I}^{w}(y) + \underline{I}^{w+2}) \cdot \underline{X}_{y'}^{\mathrm{T}} > N_{\max}(y'); \tag{7.11}$$

or if

$$\exists y' \in Y^{w-1} \; s.t. \; (\underline{I}^{w}(y) + \underline{I}^{w-2}) \cdot \underline{X}_{y'}^{\mathrm{T}} > N_{\max}(y'). \tag{7.12}$$

When a traffic demand arrives on y, we can enumerate all associated possible QoT blocking events, which satisfy either (7.10), (7.11), or (7.12). Each QoT blocking event is named ${\xi_k}^{w}[y]$, $k = 1, \ldots, K$, where K is the total number of QoT blocking events.

4.2.2 Computing QoT Blocking Probability

Because of the wavelength independence assumption, the probability of each event can be directly computed. Define $\pi^{w'}$ the stationary probability of the state of layer w' (as computed in [12]). Any QoT blocking event probability can be computed as

$$\begin{aligned} Pr({\xi_i}^{w}[y]) = & \prod_{j=-2}^{-1} \frac{1}{1 - \pi^{w+j}(R(y)\ idle)} \\ & \times \prod_{j=-2}^{2} \left[\prod_{y':[{\xi_i}^{w}[y]]_{(w+j,y')}=1} (1 - P_{y',w+j}^{(q)}) \pi^{w+j} \left([{\xi_i}^{w}[y]]_{(w+j)} \right) \right] \end{aligned} \tag{7.13}$$

for $2 < w < W - 1$, where ${\xi_i}^{w}[y]_{(w',\, y')}$ is the (w', y') element of the matrix ${\xi_i}^{w}[y]$ and $[{\xi_i}^{w}[y]]_{(w\,+\,j)}$ is the $w + j$ row of the matrix ${\xi_i}^{w}[y]$. Note that for layers greater than or equal to w ($j = 0$, 1, and 2), the network can be in any state, but, for layers below w, $(j = -1 \; and \; -2)$, $R(y)$ cannot be idle because y has to have flowed down to w from the upper layer. This is represented by the factor of $\frac{1}{1-\pi^{w+j}(R(y)\ idle)}$ in (7.13).

The QoT blocking probability is the sum of the probabilities of the K disjoint blocking events, given by

$$P_{y,w}^{(q)} = \sum_{i=1}^{K} Pr({\xi_i}^{w}[y]). \tag{7.14}$$

Generally K is a large number. We can decrease the computational complexity of the algorithm by lumping the states together, as more general blocking events. The QoT blocking probability can then be calculated as the probability of the union of

these (non-disjoint) events. For example, a blocking event will occur if either $w - 1$ or $w + 1$ have non-zero crosstalk; the probability of these events are more easily computed, yet more complex to combine since they are not disjoint.

4.2.3 Total Blocking Probability for QoT-Aware FF Wavelength Assignment

The overall blocking probability for path y, including both wavelength blocking and QoT blocking, is computed as:

$$P_y = \frac{A_y^W \cdot P_y^W}{\Lambda_y} = \frac{\bar{A}_y^{W+1}}{\Lambda_y}. \tag{7.15}$$

The overall network blocking probability is given by:

$$P = \frac{\sum_{y \in Y} \bar{A}_y^{W+1}}{\sum_{y \in Y} \Lambda_y}, \tag{7.16}$$

Algorithm 7.2 Algorithm to compute the total blocking for QoT-aware FF RWA from wavelength $w = 1$ to W given $\bar{A}_y^1 = \Lambda_y$ for all $y \in Y$.

1: Set $w = 1$.
2: Wavelength blocking sublayer: compute $P_{y,w'}^{(\lambda)}$ and $A_y^{w'}$ for all $y \in Y$ and $w' = w$ to W using [14].
3: QoT blocking sublayer: compute $P_{y,w}^{(q)}$ for all $y \in Y$ using Equation (7.13) and (7.14).
4: Wavelength blocking sublayer: using the wavelength blocking probability model in [14], recalculate arrival rate $\bar{A}_y^w$ for wavelength blocking sublayer for all $y \in Y$; recalculate $P_{y,w}^{(\lambda)}$ and obtain equivalent Poisson traffic rate (at wavelength blocking sublayer) $\hat{A}_y^w$; then calculate $\bar{A}_y^{w+1}$.
5: Set $w = w + 1$.
6: **if** $w \leq W$ **then**
7: Go to Step 2.
8: **else**
9: Calculate P_y using Equation (7.15).
10: **if** $\exists y \in Y$ such that $\frac{|P_y - P_y'|}{P_y} > \varepsilon$ **then**
11: $P_y' = P_y$. Go to Step 1.
12: **else**
13: Calculate P using Equation (7.16).
14: **return** P and P_y for all $y \in Y$.
15: **end if**
16: **end if**

where $\bar{A}_y^{w+1}$ now includes the QoT blocking traffic. The detailed steps are shown in Algorithm 2.

We first obtain the arrival rates and state probabilities, which are used as the initial values in the QoT estimation, in Step 2 for the network without physical impairments based on the algorithm in [10]. Then, the flow proceeds to the QoT blocking probability sublayer, after which part of the flow with rate

$$\bar{A}_y^w = A_y^w \cdot (1 - P_{y,w}^{(q)}) \tag{7.17}$$

flows to the wavelength blocking probability sublayer. We obtain the blocking probability and equivalent Poisson arrival rate for this sublayer in Step 4. Then the overflow traffic for layer $w + 1$ is equal to the sum of the flow due to QoT blocking and wavelength blocking, which is computed as

$$\bar{A}_y^{w+1} = A_y^w \cdot (1 - P_{y,w}^{(q)}) + \hat{A}_y^w P_{y,w}^{(\lambda)}, \tag{7.18}$$

where $\hat{A}_y^w$ is the equivalent Poisson traffic for the wavelength sublayer derived from the wavelength blocking model [10] in Step 4 of Algorithm 2. Then, we update the probabilities and arrival rates for layers 2 to W in the same way. After the last layer is examined, if the overall blocking has converged, the results are returned; otherwise, the algorithm returns to Step 1 and iterates again from the first layer.

4.2.4 Approximation to Compute QoT Blocking

The complexity of computing all QoT events is only reasonable for small networks and is prohibitive in large networks. We introduce an approximation to estimate the QoT blocking probability for large networks with any topology [12]. In this method, instead of looking at the whole network, we isolate the path considered, y', to a tandem (bus) network with its related 5 (including y' itself) or fewer wavelengths (if $w' = 1, 2, W - 1$, or W), which are independent one from another. Sources of crosstalk are modeled as originating from segments of y', instead of considering all source–destination pairs that intersect y'. Note that the impact of other traffic flows is still included. Since the major part of the crosstalk comes from the links shared with the path being considered, this approximation can include most of the impact of crosstalk, yet with lower complexity.

4.3 Validation with Simulations

Here, we validate the analytical approximate solution via simulation for the 14-node NSFNET topology depicted in Fig. 7.1. We run simulation and analysis for a 16 wavelengths case ($W = 16$) for a data rate of 10 Gb/s. The QoT for the lightpaths being considered for (and affected by) each traffic demand is checked by counting the crosstalk terms occurring under the simulated network state and comparing the resulting BER to a threshold. BER threshold is 10^{-12}, assuming uncoded transmission; the crosstalk ratio is -20 dB and the wavelength spacing is 50 GHz.

Because of the high level of noise, we limit the length of the transparent lightpaths to three hops in the NSF network; longer paths require optical-electro-optical (OEO) or optical 3-R regeneration to restore the quality of transmission and

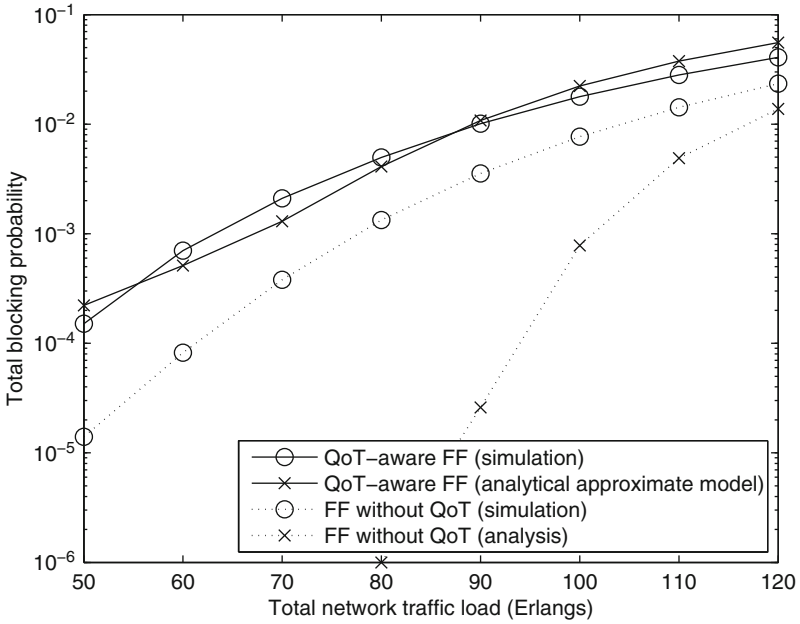


Fig. 7.5 FF wavelength assignment blocking probability computed using the analytical approximate method and simulation for 14-node NSF network, $W = 16$. In this plot, the bottom curve (FF without QoT) is based on the model in [10]

are not considered here. A uniform load is assumed for both analysis and simulation. Simulation results are obtained by running 10^6–10^7 traffic demands.

In the plot, circles indicate simulation data, while crosses indicate analytical model data. We also plot the blocking probabilities without considering any physical impairments, where only wavelength blocking is counted and marked with a dotted line in order to show the accuracy of the wavelength blocking probability model that we used.

As shown in Fig. 7.5 the presented analytical model is accurate and correctly estimates the network blocking probability as the simulation does for all traffic loads. The analytical model also correctly predicts the behavior when the wavelength blocking probability model is less accurate, e.g. at low traffic cases. The wavelength blocking probability model in [10] underestimates the wavelength blocking probability because of the wavelength independence assumption and of the approximation in the overflow traffic estimation. With $W = 16$, QoT blocking probability dominates the total network blocking probability. The analytical model can accurately compute the QoT blocking probability and thus the overall network blocking probability.

For all tested cases, computing the analytical results using the approximate method required computer running times orders of magnitude shorter than for simulation at this level of accuracy. Moreover, the analytical model presented

here is flexible. Better analytical predictions for the wavelength blocking probability in each layer may be found: the proposed analytical approach can be directly interfaced with them so that more accurate results can be obtained.

5 Conclusion

Analytical models for blocking probability in QoT-unaware transparent optical networks are plentiful and can now apply to a wide range of network scenarios: presence of wavelength converters or not, fixed routing, alternate routing, least loaded routing, arbitrary mesh networks, impact of signaling delay. The inclusion of physical impairments in those models is relatively recent, and much work is still needed to remove the many assumptions that are needed to obtain tractable models. Nevertheless, the two models presented in this chapter cover both random-pick and first-fit wavelength assignment schemes and were shown to capture the behavior of mesh networks that are subject to at least some physical impairments with sufficient accuracy for a network planner to make basic planning decisions, such as rejecting designs that will lead to unacceptable blocking rates due to QoT blocking or the combination of wavelength blocking and QoT blocking. In many cases the limitations of the proposed algorithms are really limitations of the underlying physical models, and better or additional physical models could simply be plugged in the analyses to make them more comprehensive (e.g.: inclusion of PMD in the models; definition of an analytical Q factor for coherent systems, etc.). In other cases, limitations are more fundamental (e.g.: ability to include some wavelength-dependent effects such as XPM and FWM, or even wavelength-dependent ASE noise), such that simulations are still unavoidable when it comes to fine-tune the planning of an optical network.

References

1. Birman A (1996) Computing approximate blocking probabilities for a class of all-optical networks. IEEE J Sel Areas Comm 14(5):852–857
2. Barry R, Humblet P (1996) Models of blocking probability in all-optical networks with and without wavelength changers. IEEE J Sel Area Comm 14(5):858–867
3. Subramaniam S, Azizoğlu M, Somani A (1996) All-optical networks with sparse wavelength conversion. IEEE/ACM Trans Netw 4(4):544–557
4. Harai H, Murata M, Miyahara H (1998) Performance analysis of wavelength assignment policies in all-optical networks with limited-range wavelength conversion. IEEE J Sel Area Comm 16(7):1051–1060
5. Zhu Y, Rouskas G, Perros H (2000) A path decomposition approach for computing blocking probabilities in wavelength-routing networks. IEEE/ACM Trans Netw 8(6):747–762
6. Xin C, Qiao C, Dixit S (2003) Analysis of single-hop traffic grooming in mesh WDM optical networks. Proc SPIE 5285:91–101
7. Waldman H, Campelo DR, Almeida RC Jr (2003) A new analytical approach for the estimation of blocking probabilities in wavelength-routing networks. Proc SPIE 5285:324–335

8. Sridharan A, Sivarajan K (2004) Blocking in all-optical networks. IEEE/ACM Trans Netw 12(2):384–397
9. Lu K, Xiao G, Chlamtac I (2005) Analysis of blocking probability for distributed lightpath establishment in WDM optical networks. IEEE/ACM Trans Netw 13(1):187–197
10. Alyatama A (2005) Wavelength decomposition approach for computing blocking probabilities in WDM optical networks without wavelength conversions. Elsevier Comput Netw 49(6):727–742
11. Kelly F (1991) Loss networks. Ann Appl Probab 1:319–378
12. He J, Brandt-Pearce M, Subramaniam S (2011) Analysis of blocking probability for first-fit wavelength assignment in transmission-impaired optical networks. IEEE/OSA J Opt Comm Netw 3(5):411–425
13. Pointurier Y, Brandt-Pearce M, Subramaniam S (2009) Analysis of blocking probability in noise and crosstalk-impaired all-optical networks. IEEE/OSA J Opt Comm Netw 1(6):543–554
14. Deng T, Subramaniam S, Xu J (2004) Crosstalk-aware wavelength assignment in dynamic wavelength-routed optical networks. In: Proceedings of IEEE Broadnets, pp. 140–149
15. Goldstein E, Eskildsen L (1995) Scaling limitations in transparent optical networks due to low-level crosstalk. IEEE Photon Technol Lett 7(1):93–94
16. Mukherjee B, Huang Y, Heritage J (2004) Impairment-aware routing in wavelength-routed optical networks. In: IEEE LEOS 2004, vol. 1, pp. 428–429
17. Azodolmolky S, Klinkowski M, Marin E, Careglio D, Solé-Pareta J, Tomkos I (2009) A survey on physical layer impairments aware routing and wavelength assignment algorithms in optical networks. Elsevier Comput Netw 53(7):926–944
18. He J, Brandt-Pearce M, Subramaniam S (2009) QoS-aware wavelength assignment with BER and latency constraints for all-optical networks. IEEE/OSA J Lightwave Tech 27(5):462–474
19. He J, Brandt-Pearce M, Subramaniam S (2008) Optimal RWA for static traffic in transmission-impaired wavelength-routed networks. IEEE Comm Lett 12(9):694–695

Chapter 8
Impairment-Aware Control Plane Architectures to Handle Optical Networks of the Future

E. Salvadori, A. Zanardi, and Chava Vijaya Saradhi

1 Introduction

In WDM optical networks, the physical layer impairments (PLIs) and their significance depend on network type (opaque, translucent, or transparent), the optical reach (longer lightpaths without regeneration are subject to impairment accumulation), the number and type of network elements (NEs) (fiber, wavelengths, amplifiers, switching elements, etc.), and the type of applications (real time, non-real time, mission critical, etc.). In transparent optical networks, PLIs incurred by nonideal optical transmission media accumulate along an optical path, and the overall effect determines the feasibility of the lightpaths [1–6]. If the received signal quality is not within the receiver sensitivity threshold, the receiver may not be able to correctly detect the optical signal, and this may result in high bit-error rates. Hence, it is important to understand various PLIs and their effect on optical feasibility, analytical models, and monitoring and mitigation techniques. Introducing optical transparency in the physical layer on one hand leads to a dynamic, flexible optical layer with the possibility of adding intelligence such as optical performance monitoring and fault management. On the other hand, transparency reduces the possibility of client layer interaction with the optical layer at intermediate nodes along the path. This has an impact on network design, planning, control, and management of an optical network. The purpose of this chapter is to discuss and explore the issues related to control and management plane protocols with specific emphasis on the interaction between physical layer and control and management plane protocols.

It is important to understand the techniques that provide PLI information to the control plane protocols and that use this information efficiently to compute feasible routes and wavelengths. A recent approach to network control and management using the GMPLS framework developed by the Internet Engineering Task Force

E. Salvadori • A. Zanardi • C.V. Saradhi (✉)
CREATE-NET, via alla Cascata 56D, 38100 Povo (Trento), Italy
e-mail: saradhi@ntu.edu.sg

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*,
Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_8,

(IETF) seems to be emerging as the winning control plane solution for the next-generation optical network. One of the main applications of GMPLS [4, 7, 8] in the context of optical networks is the dynamic establishment of lightpaths. However, it suffers from a lack of physical layer details such as PLI, transponder, and other physical layer component characteristics. The availability of this information would make a GMPLS-capable node capable of evaluating the effects of PLIs and to decide whether a proposed path is feasible in the optical domain. In addition, GMPLS also suffers from the lack of good techniques to disseminate and utilize physical layer details. Hence, there is a strong need for the development of efficient techniques to address these issues, without which it would be impossible to automatically initiate a lightpath establishment from client layers, for example, by a switch or IP router. In addition, control plane protocols need several extensions to make them aware of PLIs. Various GMPLS protocols can be extended to carry PLI information. This information is used by RWA protocols to compute impairment-aware routes and makes sure that the established path is feasible in the optical domain. The various approaches considered in this work are (1) signaling-based approach, (2) routing-based approach, (3) hybrid approach, and (4) PCE-based approach, which are all described in detail in the following sections. This document provides a summary of the extensions required in various architectures with block diagrams.

2 PLI-Aware Control Plane Architectures

Two main approaches to introduce impairment awareness in GMPLS-based transparent optical networks are centralized and distributed [1, 5]. Centralized approaches assume the availability of a centralized server that is reachable by all network elements (NEs) and that is aware of the complete network topology, resource availability, and physical layer parameters through the traffic engineering database (TED), which are used during routing and wavelength assignment (RWA). Hence, centralized approaches are able to guarantee and satisfy a set of lightpath specific requirements (e.g., bandwidth, diversity, QoT, and latency). Two different mechanisms can be used to implement the *centralized approach*:

- Network management system (NMS): For a connection request, the NMS computes the route and selects a wavelength considering both the current TED and the requirements of the connection request. Then, it configures various optical components (e.g., optical cross-connects) along the route in parallel using the network management interface to set up an optically feasible lightpath.
- Path computation element (PCE): For a connection request, the source node sends a request to the central PCE which computes the route, taking into account the TED and the connection requirements. PCE returns the computed route to the source node which establishes the lightpath using distributed signaling protocol such as RSVP-TE.

Centralized approaches have advantages due to the awareness of the complete and a detailed view of the whole network through the central TED and can find the optimal route. However, both NMS- and PCE-based approaches have scalability problems. NMS has additional interoperability problems, and PCE lacks in flexibility and cannot realize rapid restoration in case of multiple failures. However, PCE-based approaches can be extended to provide rapid restoration by computing protection paths for all failed lightpaths in a single computation step and by using a table lookup.

Distributed approach (distributed optical control plane) is performed via all NEs in the network using a common distributed GMPLS control plane which manages the required procedures for establishing lightpaths, that is, routing and signaling. There are two different mechanisms under the distributed approach:

- Signaling-based optical control plane (S-OCP) extends the RSVP-TE protocol to include PLI information. Each node computes a route using standard OSPF-TE protocol without the knowledge of PLIs, and then RSVP-TE carries the PLI information along the route until the destination node, which evaluates the optical feasibility and establishes the lightpath.
- Routing-based optical control plane (R-OCP) introduces PLI information into the routing protocol such as OSPF-TE, as suggested in IETF RFC4054 [8]. By flooding link-state advertisements (LSAs) enhanced with PLI information, all nodes populate their TED, thereby allowing each node to compute a feasible route while standard RSVP-TE is used for lightpath establishment. In R-OCP each node has its own PCE, which means that it is functionally equal to the centralized PCE approach, though the coordination among different PCEs is missing.
- Hybrid approach extends both RSVP-TE and OSPF-TE, the latter in order to carry and disseminate PLI information. Several variants of hybrid approaches are possible to provide a trade-off between the information carried in OSPF-TE and RSVP-TE protocols, the complexity of protocol extensions, control overhead, and network performance. In general it is sufficient to extend OSPF-TE to disseminate PLI information that is useful for intelligent RWA computation, while RSVP-TE carries full information to validate the optical feasibility during lightpath setup to accommodate the network status changes and inconsistent PLI information.

When dealing with the evaluation of the optical feasibility of a lightpath in a transparent network, several approaches can be found in literature according to the assumptions and the mathematical models used to evaluate the effect of the physical impairments themselves. For a more detailed analysis of these approaches, the reader is referred to [4], where they have been defined as impairment-aware service level agreements. The choice of the IA-SLA to use generally lies with the optical node vendor; however, there are no studies showing one approach is more reliable than others. In the remaining of this chapter, we will consider the assumptions of leveraging an overall Q-factor which takes into account the effect of both linear and nonlinear impairments when evaluating the optical feasibility of a

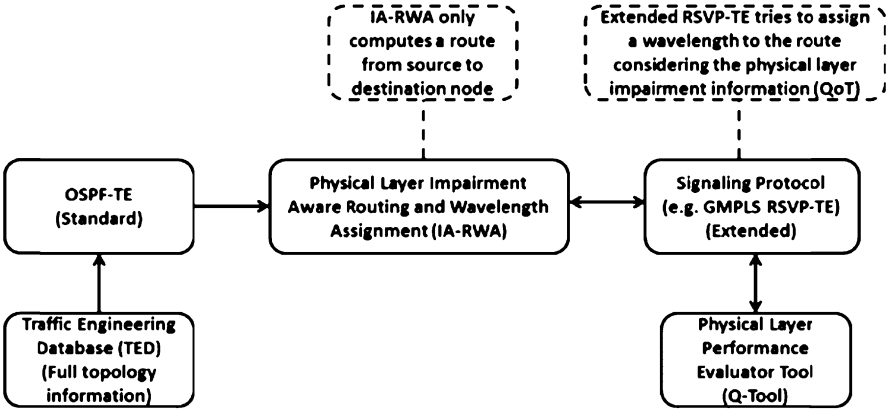


Fig. 8.1 Signaling-based (RSVP-TE) architecture

lightpath. This is generally performed through some kind of "Q-Tool" including the mathematical formulas to evaluate the quality of transmission (QoT). However, the presented architectures are only partially influenced by the decision of the IA-SLA used.

In this section, we present four architectural options in brief with schematic diagrams with interfaces between various modules.

2.1 Signaling-Based (RSVP-TE) Architecture

In the signaling-based approach, the routing component of the control plane (i.e., OSPF-TE) is not extended. However, the signaling component (i.e., RSVP-TE) is extended to consider the PLIs and QoT metrics. The metrics are those that the Q-Tool takes as inputs. In this approach the lightpath establishment procedure takes place as follows (see Fig. 8.1):

1. The OSPF-TE component of the control plane provides the IA-RWA with the full network topology information. Note that unlike in routing-based approach, the topology information and active connection information are not sent to the IA-RWA component. Hence, the path calculated is not aware of PLI information in signaling-based approach.
2. The IA-RWA component in this approach only computes a route (or K routes) from source to destination without PLI information. The computed route is used in an extended PATH message. When the extended PATH message reaches the destination node, it has all available wavelengths. The first available wavelength is checked for Q-factor value using the Q-Tool interface and local information.
3. If a wavelength is found for the given route, the QoT of the lightpath is checked at the destination node using the interfaces provided by Q-Tool, and the lightpath

will be established. Otherwise, the extended signaling protocol requests next possible route from the IA-RWA component. If none of the candidate routes satisfy the QoT requirement, the demand is blocked.

In RSVP-TE-based architecture, the TED is built using the usual TE extensions to OSPF without PLI information and is available at all nodes in the network. As the lightpath feasibility is evaluated at the destination nodes, each node needs to run an instance of Q-Tool. IA-RWA computes the route at source without using any impairment information, and hence, it needs to be implemented on all nodes in the network.

This approach has been proposed and studied by different authors, among them [3, 9]. Castoldi et al. [9] were the first to propose leveraging this mechanism to evaluate the feasibility of a new lightpath, but the model used to take into account was oversimplified. In [10] these authors have extended that mechanism to take into account the presence of regenerators, in the case of translucent networks. Salvadori et al. [3] generalize this mechanism by proposing four distributed architectures based on extensions of RSVP-TE and a more accurate modeling for impairments. Further extensions to the architectures presented in those works have been considering the potential disruption of active connections when a new lightpath introduces excessive nonlinear crosstalk (XPM, FWM) on them [11, 12]. Some standardization efforts are also happening at the IETF level on this specific architecture [13], even though the progress is highly affected by the current standardization effort for the more general WSON-based networks [14].

2.2 Routing-Based (OSPF-TE) Architecture

The building blocks of this approach are depicted in Fig. 8.2. Only the routing component (i.e., OSPF-TE) of the control plane (i.e., GMPLS) is extended. These building blocks have to be implemented in all nodes in the network. In this integration approach, the lightpath establishment process takes place as follows:

1. Since the OSPF-TE is extended, the full network topology, information about all active (already established) lightpaths (connections), and PLI information are at the disposal of physical layer impairment-aware routing and wavelength assignment (IA-RWA) component. Note that IA-RWA cannot directly access the physical layer database (PLD)/physical parameters database (PPD) or the TED. The proper interface to these two databases is provided through the extended OSPF-TE component.
2. The IA-RWA component utilizes its interface to the Q-Tool and computes a route and available wavelength (if any is available) for the requested demand.

The IA-RWA interfaces to the signaling part of the control plane (i.e., RSVP-TE for GMPLS) and establishes the lightpath. Note that the signaling protocol (i.e., RSVP-TE) is not extended in this approach. In this architecture, IA-RWA computes

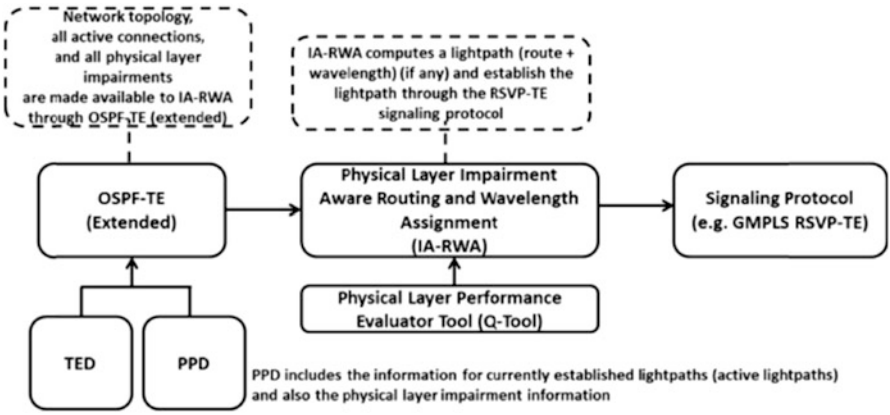


Fig. 8.2 Routing-based (OSPF-TE) architecture

the explicit routes with the knowledge of impairments provided by PPD and TED through proper OSPF-TE interfaces; hence, an instance of IA-RWA needs to be run on all nodes in the network. Also as the Q-Tool verifies the feasibility of lightpath at the source nodes in conjunction with IA-RWA, it also needs to be run on all the nodes in the network.

Even though proposed by a very pioneering RFC [8], a "pure" routing-based approach has generally not been considered in literature due to the fact that in very dynamic networks, where the inter-arrival time of the connections is small or arrival rate is high, it can easily lead to inconsistent databases which result in optically infeasible routes. Some results showing this behavior have been shown in [15], where a comparison between this approach and an approach based on signaling has been presented.

2.3 Hybrid Architecture

Both stand-alone signaling and routing approaches suffer from some drawbacks which may affect their performance. Therefore, it is a matter of fact that by combining the approaches, in a so-called hybrid approach, the aforementioned drawbacks could be overcome. In a hybrid approach, both the signaling and routing component of the control plane are extended in order to encompass and disseminate the PLI information. As explained earlier, several variations of hybrid approaches are possible to provide a trade-off between the information carried in OSPF-TE and RSVP-TE protocols, the complexity of protocol extensions, control overhead, and network performance. In general it is sufficient to extend OSPF-TE to disseminate PLI information that is useful for intelligent RWA computation, while RSVP-TE carries full information to validate the optical feasibility during lightpath setup to accommodate the network status changes and inconsistent PLI information.

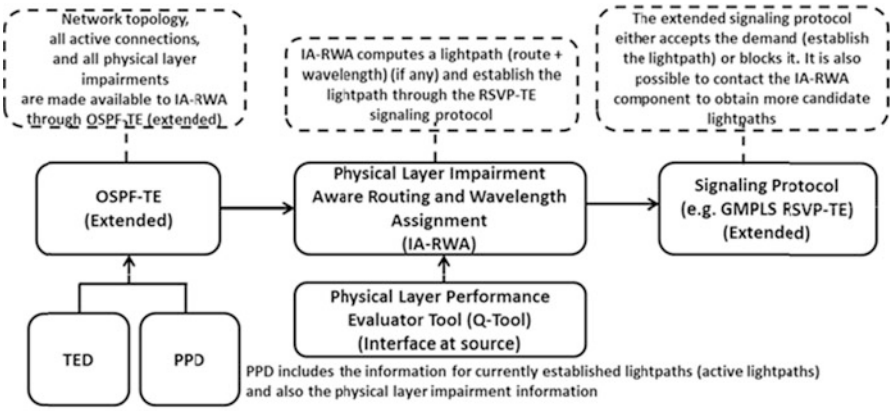


Fig. 8.3 Hybrid integration approach

In this section, we explain one of the possible examples (an extreme case) where both protocols are extended to carry all the required information. The input to the IA-RWA component includes the full network topology, all active lightpaths (currently established lightpaths), and PLIs. This information is kept in the PPD (however, we can think about separate databases for each of them, which is a matter of design and implementation). The lightpath establishment process takes place as follows (Fig. 8.3):

1. The IA-RWA component computes a route and wavelengths (i.e., lightpath) from source to the destination node, considering the PLIs.
2. The candidate lightpath is forwarded to the extended signaling protocol (i.e., RSVP-TE) to establish the lightpath while double-checking the QoT using proper Q-Tool interfaces. Due to nonzero convergence time, the PPD and TED used to evaluate the feasibility of lightpath at the source node may not contain up-to-date information on PLIs. Also there can be network status changes, because of either a new lightpath establishment or active lightpath teardown between the feasibility computation at source node and the actual reservation in RESV phase of RSVP-TE. Hence, the RSVP-TE protocol again checks the feasibility during RESV phase to accommodate these changes and to make sure that no active lightpath is disrupted due to inconsistent database information.

If the QoT of the candidate path is above a given threshold, then the lightpath will be established. Otherwise, a PATH_ERROR message will be sent back to the source node. In this architecture as well, IA-RWA and Q-Tool need to be implemented on all the nodes in the network as explained earlier.

Recent works have proposed such an approach in order to leverage the intrinsic advantages of each of the two previous mechanisms. In [16], the authors have extended OSPF-TE by introducing a few time–length–value (TLV) fields into the link-state advertisement (LSA), mainly referring to more static physical parameters

of the network, plus the information about the availability of the wavelength per link. Instead, RSVP-TE has been extended to carry the more dynamic parameters in order to assure the evaluation of the optical feasibility at the destination node with the more updated potential information capturing the status of the fiber spans. Results show a definitive improvement against purely routing- or signaling-based mechanisms and a good trade-off between control overhead and network performance. In [17], the authors have extended previous contributions [9, 11] toward a translucent scenario where regenerators are sparsely distributed in the network. Compared to those, the authors have adopted a hybrid mechanism based on a three-phase approach—reachability graph (RG) construction, route computation on RG, and signaling, for routing lightpaths with multiple transparent segments. As reachability graph construction is based on approximate PLI models, the authors have focused on analyzing the effect of PLI modeling errors on the network performance.

2.4 PCE-Based Architecture

In this approach the whole process of PLI-aware RWA is performed in a single computing element, called PCE. In the PCE framework, all nodes run a particular protocol to disseminate their status to the central PCE. This software module at each node is usually called PCE agent. The counterpart software component in the centralized PCE server is named PCE manager. This protocol can be considered as an SNMP-like management protocol. Therefore, it provides the central PCE with the topology information, already established connections, and also the PLI information. Note that monitors can communicate with each node using a local protocol, and then the node disseminates this information to the central PCE server. The lightpath establishment process in this approach takes place as follows (Fig. 8.4):

1. The source node contacts the PCE server asking for a lightpath establishment to a particular node in the network. We can also think about an operator, which can utilize the PCE interface for establishing a lightpath from a central server (i.e., PCE server–user interface).
2. The IA-RWA, Q-Tool, and all the required information (i.e., network topology, currently established connections, and PLI information) are already collected in the PCE central database either in PPD or TED.
3. PCE computes a route and wavelength (i.e., lightpath) for the requested demand using IA-RWA and communicates it with the standard signaling component of the control plane (i.e., GMPLS/RSVP-TE).
4. The signaling component at the source node uses the standard signaling for establishing the lightpath.

This approach has been proposed by many authors, among them [18–20]. The main differences among them refer to the parameters exchanged between the control unit on the optical nodes and the PCE, which are strictly connected to

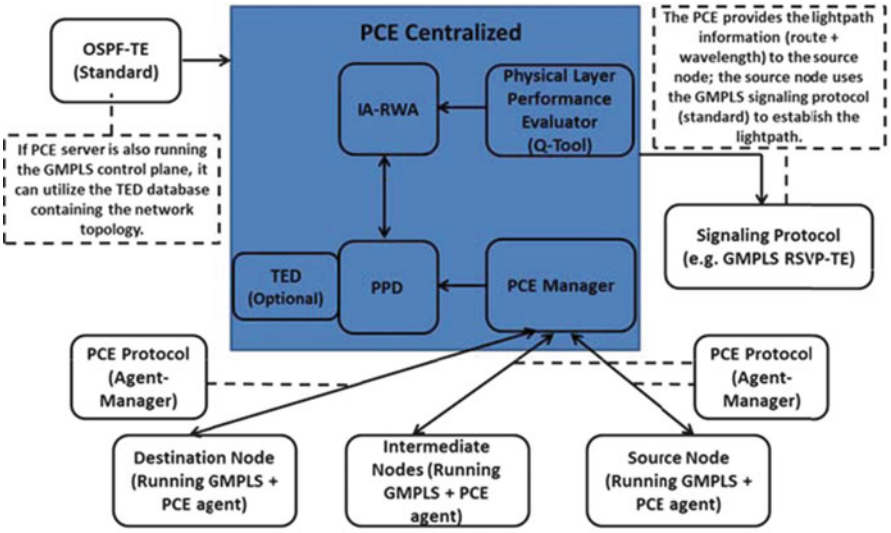


Fig. 8.4 PCE-based approach

the Q-Tool used to compute the feasibility of the lightpath, and the potential involvement of OSPF-TE protocol to exchange some of them. Cugini et al. [18] have validated their PCE implementation experimentally in an opaque network composed of commercially available MPLS routers and via simulation in a transparent network affected by excessive physical impairments. The same authors in [20] have extended their work to compare the major strengths and weaknesses of both a centralized and a distributed approach. However, both works are characterized by a simplified physical impairment model, where nonlinear impairments have been linearized. Azodolmolky et al. [20] have presented the experimental results for both a centralized and a distributed control plane scheme, leveraging on an accurate Q-Tool (results of a European project [12]). It is important to highlight that none of the current works available in the literature have proposed specific impairment-related extensions to the so-called PCEP protocol, most likely due to the lack of maturity of this protocol within IETF.

3 Performance Evaluation for IA Control Plane Solutions

The research on impairment-aware control plane architectures is still a hot topic due to the lack of clear indicators coming from the scientific works currently available in the literature. Each approach discussed in the previous sections has its own advantages and drawbacks. Furthermore, the quite heterogeneous adoption of different modeling approaches to physical impairments does not help in finding out which of these potential strategies could be the one which should be adopted

universally in transparent and translucent optical networks of the future. Actually even within IETF, the current attitude toward this issue is to let vendors have the freedom to propose their own solutions at control plane level, instead of trying to identify a common definition and characterization of optical fiber, subsystems, and network elements, of course leveraging on existing ITU-T recommendations [21].

Due to these reasons, we decided to show a comparison among two fully distributed approaches, outcome of the results of a recently concluded European project (DICONET) [12] whose focus was on considering the impact of PLIs in the planning and operation of all-optical (and translucent) networks.

For this purpose, a specific simulation environment has been developed as an extension of the Java-based GLASS (GMPLS Lightwave Agile Switching Simulator) [22], and then a simulation campaign has been conducted. GLASS is developed by the US National Institute of Standards and Technology (NIST) using the scalable simulation framework (SSF). It provides network components such as hosts, routers, links, and a number of network protocols. GLASS extends these components with an implementation of MPLS, optical components such as optical cross connects (OXC), edge routers, optical links, fibers, and lambdas. However, the protocols implemented in GLASS do not have support for interacting with the optical layer. Hence, the protocols included in GLASS suffer from lack of PLI information and corresponding protocol extensions to carry and evaluate the feasibility of lightpaths in the optical domain. In particular, the GLASS framework has been extended to interact with the optical layer to get the required information and corresponding protocols to carry this information. In addition several other mechanisms as discussed earlier in this chapter are implemented. These mechanisms include (a) PLI modeling, (b) evaluation of Q-factor, (c) proper interfaces to Q-Tool, and (d) mechanisms to deal with potential active lightpath disruption. The objective is to check the implementation feasibility of proposed mechanisms and protocol extensions. In addition, it is also important to study the performance of different approaches to check the performance for various real-world scenarios to see which protocol extensions work well in those scenarios. A statistical processing module ANCLES [23] has been integrated to collect simulation results with a target accuracy of 0.05 and confidence level of 95%.

3.1 Performance Metrics

The parameters used to test the performance of the proposed mechanisms are listed below:

- Blocking probability: the ratio of the number of rejected connection requests to all requested connections.
  - Blocking due to wavelength unavailability called lambda blocking.
  - Blocking due to optical infeasibility (i.e., blocking due to infeasibility of the lightpath under consideration).

  - Blocking due to affected lightpath disruption (i.e., blocking due to the potential existing lightpath disruption if a new lightpath is accepted into the network).

- *LSP setup time*: elapsed time between the first *PATH* message sent and the *RESV* message received at the source node. This metric reflects how fast a connection request can be established.
- *Average number of setup attempts*: average number of attempts required for a successfully established lightpath over all connection requests.
- *Average number of hops*: average number of hops for a successfully established lightpath over all connection requests.

3.2 Network Topology and Characteristics

The network scenarios considered so far are only transparent networks where neither wavelength converters nor wavelength regenerators are available at intermediate nodes. Therefore, an end-to-end lightpath must respect the so-called *wavelength continuity constraint*.

The simulation results reported in the next section are for the DTAG/T-Systems National Core Network [24] whose topology is depicted in Fig. 8.5 and characteristics are summarized in Table 8.1.

3.3 Traffic Models

Simulations have been conducted by employing a dynamic traffic model where lightpath setup requests are generated according to a Poisson process. Following this model, lightpath requests arrive at an average rate of $1/\alpha$, where α is the average request inter-arrival time expressed in seconds. The traffic requests considered in simulations are uniformly distributed among all the nodes.

An installed lightpath has an average exponential duration of μ seconds. The traffic load ρ is defined as the average network resource (link wavelength) usage computed in percentage as:

$$\rho = \frac{\overline{N}_c \times \overline{L}_c}{M \times W} \times 100\%,$$

where $\overline{N}_c$ is the average number of active connections and equals μ/α, $\overline{L}_c$ is the average number of hops in the network (considering only the shortest paths between nodes and a uniformly distributed traffic matrix), M is the number of network links, and W is the number of wavelengths inside a fiber.

Fig. 8.5 DTAG/T-Systems National Core Network

Table 8.1 Parameters of DTAG/T-Systems National Core Network

Parameter	Value
Number of nodes	14
Number of links	23
Node degree	3.29 (min = 2, max = 6)
Link length (km)	186 km (min = 37, max = 353)
Span length (km)	25 km (long aggr.), (min = 0.21, max = 65) 97 km (short aggr.), (min = 37, max = 149)
Number of spans	7.5 (long aggr.), (min = 1, max = 17) 1.87 (short aggr.), (min = 1, max = 3)
Path length (km)	410 km (min = 37, max = 874)
Hop count	2.35 (min = 1, max = 5)

3.4 Performance Study

The two distributed approaches considered in our simulation experiments are the signaling-based approach and the hybrid approach. For these two scenarios, the following parameters have been used: when a node receives a connection request, the *K-CSPF* algorithm computes *K* routes. The Constrained Shortest Path First

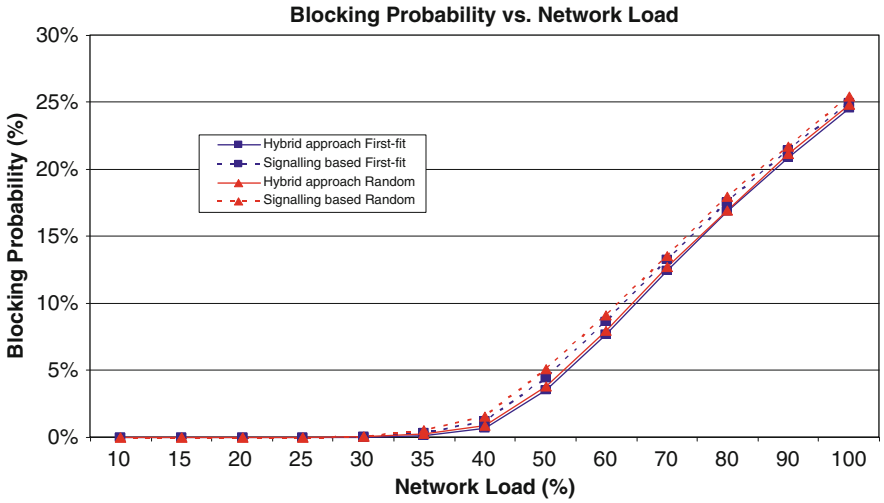


Fig. 8.6 Blocking probability vs. network load for RSVP-TE and hybrid approach

(CSPF) metric used for route computation is the number of hops with $K = 3$. Two kinds of wavelength selection policies are implemented: first-fit and random-fit. In general, each node takes some time to process various RSVP-TE messages and to set up/clear OXC configuration. To simulate real-world scenarios, we have considered the following: (1) *PATH* message processing time = 5 ms; (2) egress node (where time-consuming Q computation is carried out) *PATH* message processing time = 80 ms; (3) *RESV*, *PATH_ERROR*, and *RESV_ERROR* message processing time = 5 ms; (4) *Q_CHK_REQ* processing time = 5 ms; and (5) OXC switch configuration time = 25 ms. The inter-arrival (I) time is set to 200 s.

Figure 8.6 shows blocking probability (BP) vs. load for first-fit and random-fit for both signaling-based approach and hybrid approach. Note that BP has the following main contributions: (1) blocking due to wavelength continuity constraint called wavelength blocking, that is, no free common wavelength along the route; (2) optical failures due to the infeasibility of connection request under consideration (as shown in Fig. 8.7); and (3) optical failures due to the active connections, that is, setup of a new lightpath may introduce excessive crosstalk on active connections potentially disrupting them; hence, the new connection needs to blocked (as shown in Fig. 8.8).

BP is lower in the case of hybrid approach due to exact information regarding wavelength availability which reduces number of attempts below $K = 3$ before a successful feasible path is found as shown in Fig. 8.7. However, in case of signaling-based approach, as there is no wavelength availability information during route computation time, the route is computed only based on the shortest path. When the path message reaches the destination nodes, these realize that there is no common free wavelength along the path, leading to more number of attempts ultimately reaching the maximum number of allowed attempts (which in this case is $K = 3$) and hence more blocking due to wavelength blocking as shown in Fig. 8.7.

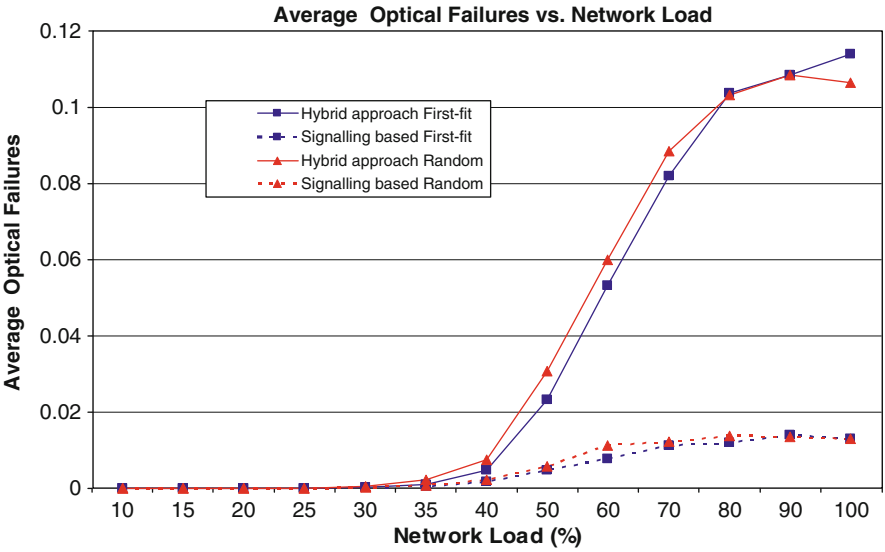


Fig. 8.7 Percentage of optical failures (due to infeasibility of lightpath itself) vs. network load for RSVP-TE and hybrid approach

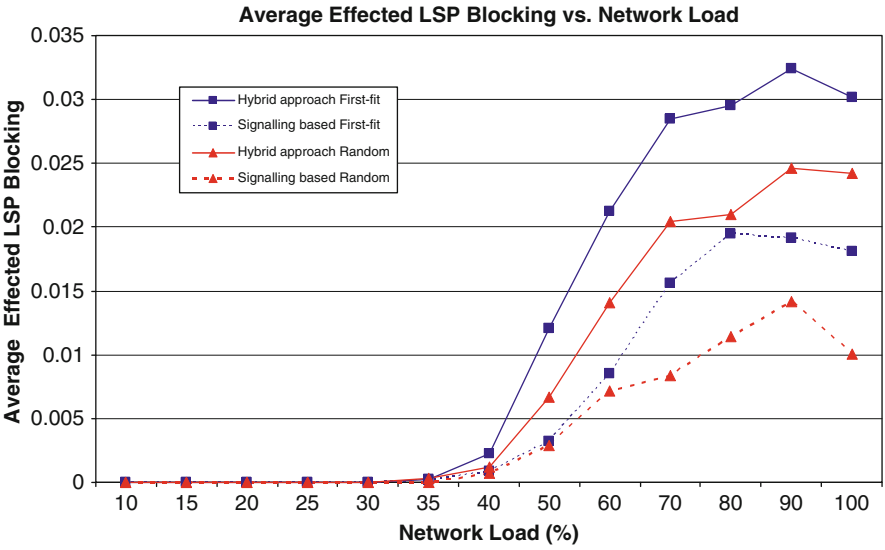


Fig. 8.8 Percentage of blocking due to active (potentially affected) LSPs vs. network load for RSVP-TE and hybrid approach

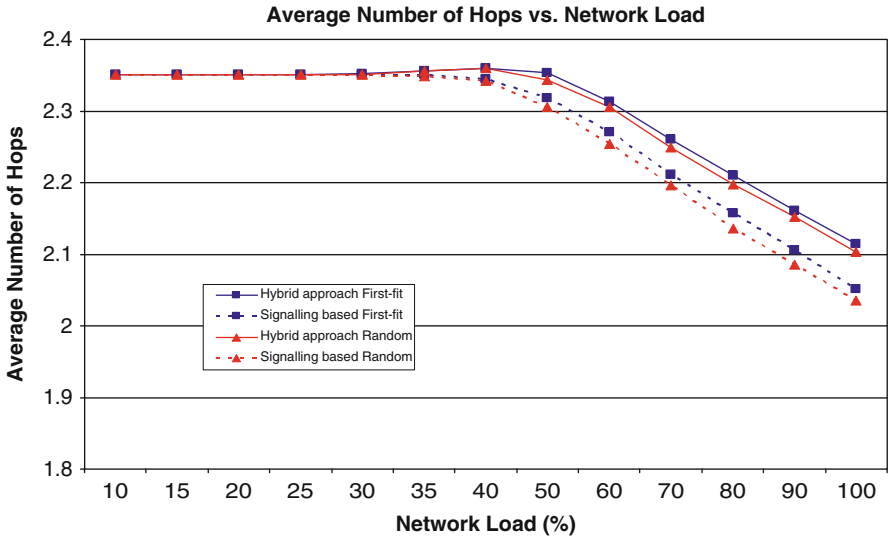


Fig. 8.9 Average number of hops vs. network load for RSVP-TE and hybrid approach

As the load increases, the BP for both architectures increases because of the following reasons: (1) the number of active connections in the network increases, and hence, the chances of finding a free common wavelength along the route decrease, leading to higher wavelength blocking as shown in Fig. 8.7; (2) as the number of active connections increases in the network, the effect of these active lightpaths on the new lightpath requests increases (due to multichannel effects), leading to higher blocking due to the optical failures as shown in Fig. 8.7; (3) as the number of active connections in the network increases, the number of lightpaths that may be potentially disrupted due to the setup of new connection increases, leading to higher blocking due to affected lightpaths as shown in Fig. 8.8. Note that the blocking due to optical failures is higher for the hybrid approach as the average path length/number of hops in this case is higher compared to signaling-based approach due the computation of route using CSPF with wavelength availability information. However, the wavelength blocking for the signaling-based approach is higher compared to the hybrid approach, as signaling-based approach computes the shortest path without wavelength availability information, and when the PATH message reaches the destination, it realizes that there is no common free wavelength, leading to more attempts and higher wavelength blocking.

The average setup time depends on the average number of attempts and average number of hops in the route. As the average number of hops in the route is higher in case of hybrid approach (as shown in Fig. 8.9), the average setup time is high. The wavelength availability information in hybrid approach reduces the wavelength continuity constraint failures at the intermediate and destination nodes due to

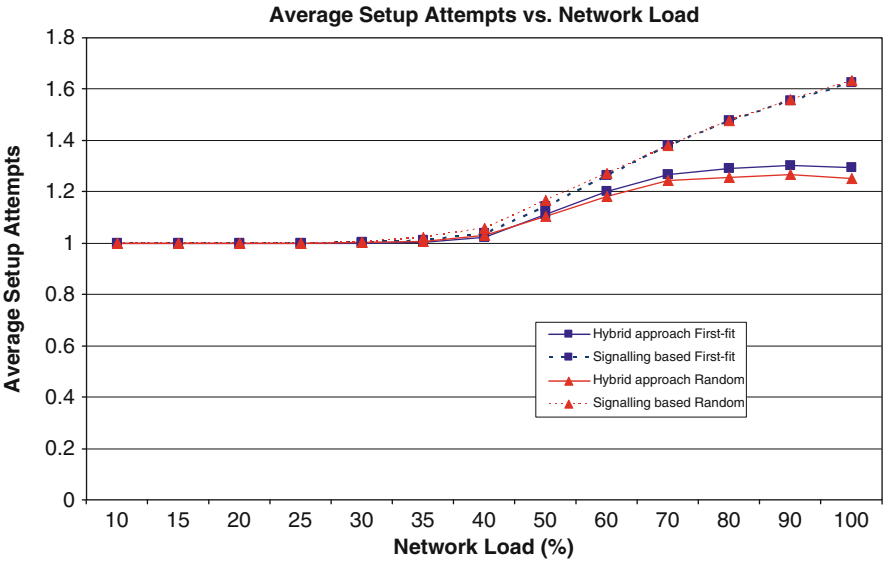


Fig. 8.10 Average number of setup attempts vs. network load for RSVP-TE and hybrid approach

wavelength unavailability, leading to less average number of attempts, as shown in Fig. 8.10. As the load increases, the average number of hops decreases, as the chances of finding a free common wavelength on longer hop paths decrease, leading to decrease in lower setup time. The average number of attempts in case of signaling-based approach is higher due to lack of wavelength availability information, and it increases with the load due to increase in both lambda failures and both kinds of optical failures.

The average number of PATH messages sent by each node in the network depends on the average number of hops and average number of attempts. As the average number of hops in hybrid approach is higher (due to the fact that the longer paths are more affected by wavelength unavailability failures in the signaling-based approach), the average number of PATH messages is high as shown in Fig. 8.11.

4 Conclusions

In this chapter, various impairment-aware control plane approaches have been described and compared: signaling-based approach, routing-based approach, hybrid approach, and PCE-based approach, and their properties qualitatively studied. Each approach discussed in the previous sections has its own advantages and drawbacks, and no clear indicators are coming from the scientific works currently available in

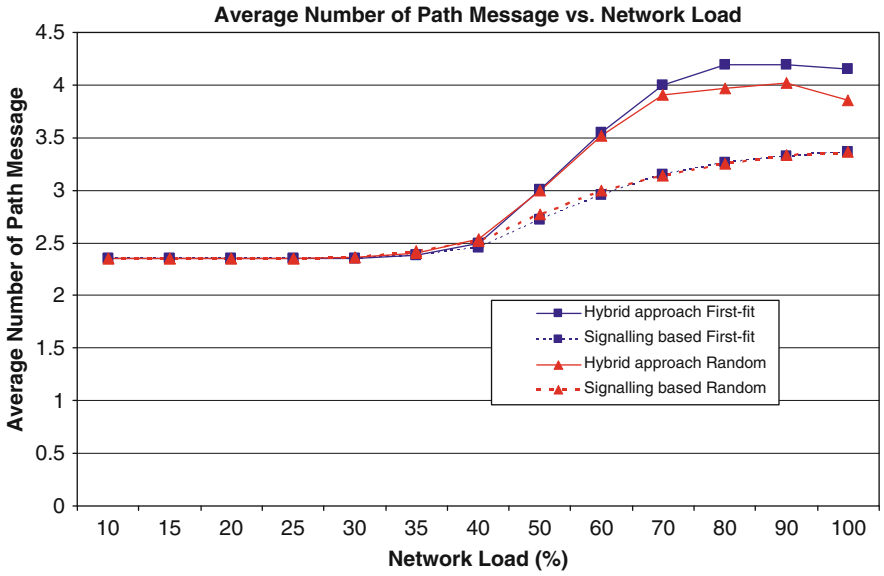


Fig. 8.11 Average number of PATH messages vs. network load for RSVP-TE and hybrid approach

the literature to suggest which among these architectures is the most effective. The performance study of two impairment-aware control plane approaches is discussed through results coming from simulation experiments, showing that a hybrid approach seems to perform better compared to other approaches at the expense of slightly increased control plane overhead.

References

1. Salvadori E et al (2008) Towards deployment of signalling based approaches for impairment aware lightpath setup in transparent WDM optical networks. In: Proceedings of the ECOC, Brussels, 21–25 Sept 2008
2. Martínez R et al (2006) Challenges and requirements for introducing impairment-awareness into the management and control planes of ASON/GMPLS WDM networks. IEEE Commun Mag 44(12):76–85
3. Salvadori E et al (2009) Distributed optical control plane architectures for handling transmission impairments in transparent optical networks. J Lightwave Technol 27(13):2224–2239
4. Saradhi CV, Subramaniam S (2009) Physical layer impairment aware routing (PLIAR) in WDM optical networks: issues and challenges. IEEE Commun Surv Tuts 11(4):109–130
5. Farrel A, Bryskin I (2005) GMPLS: architecture and applications. Morgan Kaufmann Publishers Inc. San Francisco, CA
6. Castoldi P et al (2007) Centralized vs. distributed approaches for encompassing physical impairments in transparent optical networks. In: Proceedings of the ONDM, May 2007, Athens, Greece, pp 68–77
7. Berger L (2003) Generalized multi-protocol label switching (GMPLS) signaling resource reservation protocol-traffic engineering (RSVP-TE) extensions. RFC 3473, Jan 2003

8. Strand J, Chiu A (2005) Impairments and other constraints on optical layer routing. IETF RFC 4054, May 2005
9. Cugini F et al (2004) A novel signaling approach to encompass physical impairments in GMPLS networks. In: Proceedings of the IEEE Globecom 2004, Dallas, TX, pp 369–373
10. Sambo N, Giorgetti A, Cugini F, Andriolli N, Valcarenghi L, Castoldi P (2009) Accounting for shared regenerators in GMPLS-controlled translucent optical networks. J Lightwave Technol 27(19):4338–4347
11. Saradhi CV, Salvadori E, Zanardi A, Galimberti G, Martinelli G, Pastorelli R, Vercelli ES, Tanzi A, La Fauci D (2009) Novel signalling based approach for handling linear and non-linear impairments in transparent optical networks. In: Proceedings of the Broadnets, Invited Paper. Madrid, Sept 2009
12. DICONET FP7 EU project (2009) (http://www.diconet.eu), Deliverable Report D2.3, Sept 2009
13. Martinelli G, Zanardi A (2009) GMPLS signaling extensions for optical impairment aware lightpath setup. IETF draft, draft-martinelli-ccamp-optical-imp-signaling-02, July 2009
14. Lee Y, Bernstein G, Imajuku W (2011) Framework for GMPLS and PCE control of wavelength switched optical networks (WSON). Draft-ietf-ccamp-rwa-wson-framework-12.txt, Feb 2011
15. Salvadori E, Ye Y, Zanardi A, Woesner H, Carcagnì M, Galimberti G, Martinelli G, Tanzi A, La Fauci D (2007) A study of connection management approaches for an impairment-aware optical control plane. In: 11th conference on optical network design and modeling (ONDM07), May 2007, Athens, Greece
16. Saradhi CV, Salvadori E, Zanardi A, Dalsass S, Piesiewicz R (2009) Hybrid control plane architecture for dynamic impairment-aware routing in transparent optical networks. In: Proceedings of a meeting held 15–19 September 2009, Pisa, Italy.
17. Saradhi CV, Salvadori E, Zanardi A, Dalsass S, Galimberti G, Tanzi A, Martinelli G, Gerstel O (2011) Effect of impairment modeling errors on reachability graph based lightpath setup in translucent optical networks. In: 15th conference on optical network design and modeling (ONDM11), Bologna, Italy, on 8–10 February, 2011
18. Cugini F et al (2007) Implementing a path computation element (PCE) to encompass physical impairments in transparent networks. In: Proceedings of the OFC 2007, Anaheim, California
19. Castoldi P, Cugini F, Valcarenghi L, Sambo N, Le Rouzic E, Poirrier MJ, Andriolli N, Paolucci F, Giorgetti A (2007) Centralized vs. distributed approaches for encompassing physical impairments in transparent optical networks. In: Proceedings of the ONDM 2007, Athens, Greece
20. Agraz F, Azodolmolky S, Angelou M, Perelló J, Velasco L, Spadaro S, Francescon A, Saradhi C, Pointurier Y, Kokkinos P, Varvarigos E, Gunkel M, Tomkos I (2010) Experimental demonstration of centralized and distributed impairment-aware control place schemes for dynamic transparent optical networks. In: Proceedings of the IEEE/OSA OFC/NFOEC, PDP5, Mar 2010, San Diego, CA
21. Bernstein G, Lee Y, Li D, Martinelli G (2009) A framework for the control of wavelength switched optical networks (WSON) with impairments. Draft-bernstein-ccamp-wson-impairments-05.txt, May 2009
22. GMPLS Lightwave Agile Switching Simulator (GLASS) Version: Draft 1.0. Nationals Institute of Standards and Technology. http://www-x.antd.nist.gov/glass/Main.htm
23. ANCLES: ATM Networks Call Level Simulator (2002) Networks Group - Politecnico di Torino. http://www.telematica.polito.it/oldsite/ancles/
24. Deliverable Report D2.1. Definition of dynamic optical network architectures, DICONET FP7 project, June, 2008. http://www.diconet.eu/deliverables.asp
25. Karz D, Kompella K, Yeung D (2003) Traffic engineering (TE) extensions to OSPF Version 2. RFC 3630, Sept 2003

Chapter 9
QoT-Aware Survivable Network Design

Suresh Subramaniam and Maïté Brandt-Pearce

1 Introduction

Modern society relies on communications networks for every aspect of life, from economic transactions to personal connectivity. Yet the communication infrastructure consisting largely of optical networks remains vulnerable to failures that can occur with no warning, resulting in loss of connectivity over potentially large geographical regions. Disruptions, whether intentional or unintentional, can bring down single links or entire network nodes. Mechanisms must be implemented so that networks can either remain immune to or quickly recover from such failures. Techniques that allow networks to survive a failure are referred to as network protection, and typically include significant redundancy, and thus inefficiency. Algorithms for fast recovery from a failure are called restoration techniques. In this chapter we discuss the protection and restoration of optical networks affected by potential failures and also physical layer impairments.

Wavelength routed optical networks benefit from the flexibility of all-optical transmission, yet suffer from the inevitable lack of isolation in terms of physical impairments and network failures. In this chapter we address the problem of designing and operating transparent networks that suffer simultaneously from degradations caused by an imperfect physical layer and failures caused by unexpected faults. As discussed in Chap. 2, designing algorithms for routing and wavelength assignment (RWA) for dynamic traffic must be approached as a cross-layer problem as both the physical and the lambda layers are strongly affected by the decisions. Protocols for survivability and restoration are even more

S. Subramaniam (✉)
George Washington University, Washington, DC, USA
e-mail: suresh@gwu.edu

M. Brandt-Pearce
University of Virginia, Charlottesville, VA, USA
e-mail: mb-p@virginia.edu

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*,
Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_9,

vulnerable to the state of the physical network, and must therefore be approached with as much information about this state as possible.

We begin the chapter by discussing appropriate metrics that can be used to measure the success of a cross-layer survivable approach in Sect. 9.2, including the quality of transmission, i.e., the performance of the physical layer, and the performance of the network as a whole, both in terms of providing service and reliability. In Sect. 9.3, we propose several cross-layer RWA techniques that enhance the network reliability. We then consider path and link protection and restoration approaches for network survivability in Sects. 9.4 and 9.5. Section 9.6 concludes the chapter.

2 Network Access, Quality of Transmission, and Survivability Metrics

In our model a centralized network management system controls call admissions, including routing and wavelength assignment, call establishment and termination, and restoration. The network is assumed to have N nodes and L bidirectional links with C equally spaced wavelengths in each direction. We consider a fully transparent network, or an all-optical subnetwork within a larger translucent network. In this work we assume that wavelength conversion is not available. Calls arrive at the source node and, if accepted, travel all-optically to the destination. A central call-admission algorithm assigns each call a wavelength and a route, jointly called a lightpath. Calls are assumed to arrive at the nodes in the network according to a Poisson process with average arrival rate ℓ and the call durations are assumed to follow an exponential distribution with unit mean, without loss of generality, such that ℓ is the offered load at each node of the network, in Erlang. Call sources and destinations are uniformly distributed over the set of network nodes.

2.1 Blocking Probability

Arguably the most critical measure of network performance is the call blocking probability (BP). In optical networks, a connection request can be denied, or blocked, if an end-to-end lightpath is not available due to a shortage of wavelength channels, referred to as wavelength blocking, or because the lightpaths that are considered have insufficient Quality of Transmission (QoT; as described below), resulting in what is referred to as QoT blocking. The call can also be blocked if a backup protection path satisfying specified conditions cannot be found. The total blocking probability is the ratio of calls blocked (for whatever reason) to the number requested. This metric is the most relevant to network operators as it is tied to utilization and revenue.

The time required to set up a call becomes critical to both call acceptance and restoration algorithms, being part of the quality of service (QoS) provided. We refer to as latency the elapsed time from the point that a call set-up is requested to the point that all RWA decisions and QoT checks have been made so that the call can be accepted. The RWA should have a time-out constraint so that if a call cannot be found within a time window, it is not accepted. The latency is affected by both the complexity of the RWA algorithm and the QoT computation requirements. Algorithms that must find a solution before a time-out period expires, or block the call, are called *latency guaranteed*.

2.2 Q Factor

A signal traversing an all-optical network segment experiences degradations in fidelity that must be tracked to ensure the connection has sufficient QoT. In this chapter the QoT of interest is taken to be the bit error rate (BER) of the end-to-end lightpath, which is to be kept below some threshold value. The QoT must be estimated during the call admission process, and this algorithm must therefore be simple. Note that monitoring instead of estimation can also be used to establish QoT, but cannot be used exclusively to predict how accepting a new connection might impact existing calls.

We assume on-off keying signaling and, denoting by μ_1 and μ_0 the expected signal for "1" and "0" samples, and by σ_1 and σ_0 their respective standard deviations, the *Q factor* of a signal is defined as:

$$Q = \frac{\mu_1 - \mu_0}{\sigma_0 + \sigma_1}. \tag{9.1}$$

Using a Gaussian assumption [1], the BER and the Q factor of a signal are related by $BER = 0.5\ erfc\left(Q/\sqrt{2}\right)$ for uncoded signals. For instance, a BER of 10^{-9} corresponds to a Q factor of $Q = 6$.

Four main physical impairments are known to affect lightpaths in all-optical networks [2]: intersymbol interference (ISI), amplifier spontaneous emission (ASE) noise, nonlinear crosstalk, and node crosstalk, as illustrated in Fig. 9.1. In this work we account for these effects by adding a noise variance in the Q factor computation for a lightpath, which becomes

$$Q = \frac{\mu_1 - \mu_0}{\sigma_0 + \sqrt{\sigma_i^2 + \sigma_n^2 + \sigma_{nlx}^2 + \sigma_{nx}^2}} \tag{9.2}$$

where σ_i^2, σ_n^2, σ_{nlx}^2 and σ_{nx}^2 are the variances due to intersymbol interference, amplifier noise, nonlinear crosstalk and node crosstalk, respectively. σ_0 is dominated by the ISI term because there is no power transmitted for a "0" bit.

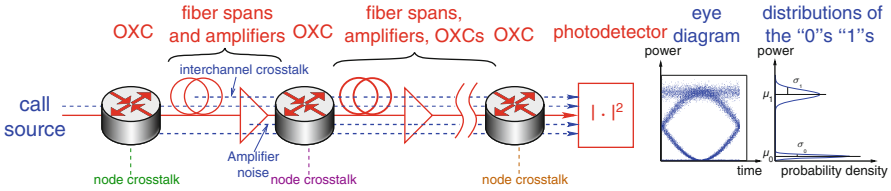


Fig. 9.1 Model of a transmission path used to compute the Q factor. Amplifiers inject ASE noise, interplay between channels in fiber spans cause nonlinear crosstalk, while leaks in the optical cross-connects (OXCs) cause node crosstalk [3]

We ignore polarization mode dispersion (PMD) effects, rejecting any candidate lightpath with substantial PMD as not viable. Higher-order modulation techniques, such as Differential Quadrature Phase Shift Keying (DQPSK), have similar Q factor formulations and are also primarily affected by the same four impairments.

ISI, self-phase modulation, concatenated filtering, and amplifier noise are single-channel effects independent of the network state, and can therefore be predetermined given a network topology and physical layer parameters. Nonlinear crosstalk, occurring when signals co-propagate in the fiber, is a network state-dependent impairment that cannot be precomputed. In this chapter the variance due to nonlinearity is calculated as in [4]. Node crosstalk originates from signal leaks in the optical switches and from the imperfect wavelength demultiplexing, and is therefore also network state dependent. We use the formulation in [5, 6] to estimate the added variance due to node crosstalk.

2.3 *Fairness*

Optical networks often have physical topologies, traffic requirements, and performance needs that are diverse. Lightpaths can range from a few hundred kilometers to several thousand. Some may cross many nodes while others are single-hop. Lightpaths between physically distant nodes or crossing congested nodes habitually experience worse performance, as measured by either the BER or the blocking probability. One important metric to consider is therefore *fairness*, which is a measure of the similarity between the performance of all services rendered. We consider fairness in blocking probability, so that long-distance links and paths are not disadvantaged, and so that survivability remains uniform throughout the network (one link or node down does not kill the entire network).

In this work, we use a *fairness index*, introduced by Jain in [7] for circuit-switched networks. The fairness index $0 \leq f_{BP} \leq 1$ over N nodes measures how fairly the resources are shared:

$$f_{BP} = \frac{\left(\sum_{n=1}^{N} BP_n\right)^2}{\sum_{n=1}^{N} BP_n^2} \tag{9.3}$$

where BP_n is the blocking probability experienced by node n. A fairness index near 1 indicates a high level of fairness, while a small value of f_{BP} shows that some connections suffer from more blocking than others.

2.4 Vulnerability Ratio

The usefulness of a restoration algorithm can be measured by its ability to recover from failures. In this work we use the *Vulnerability Ratio* (VR) proposed in [8] as a metric to describe the performance of our algorithms. Considering only link failures, the VR is defined as the probability that a randomly picked connection cannot be restored because of unacceptable QoT, if a single random link fails at a random point of time during the operation of the network. Since the vulnerability of a connection can only vary as the network state changes, we calculate the VR by averaging over all network states S. For a failure on link j in network state i, the probability that a random ongoing connection fails is

$$P_i^j = \frac{D_i^j}{T_i} \tag{9.4}$$

where D_i^j is the number of the connections that are dropped due to unacceptable QoT, and T_i is the total number of ongoing connections in state i. For each network state period, any of the links can fail with equal probability (other failure probability distributions can easily be incorporated if needed), so the average failures over all links is

$$P_i = \frac{1}{L}\sum_{j=1}^{L} P_i^j = \frac{1}{L}\sum_{j=1}^{L} \frac{D_i^j}{T_i}. \tag{9.5}$$

Then, the VR, $\mathcal{V}$, is computed by averaging P_i over all the states:

$$\mathcal{V} = E[P_i] = \frac{1}{\sum_{i=1}^{S}\tau_i}\sum_{i=1}^{S} P_i\tau_i = \frac{1}{L}\frac{1}{\sum_{i=1}^{S}\tau_i}\sum_{i=1}^{S}\sum_{j=1}^{L}\frac{D_i^j\tau_i}{T_i} \tag{9.6}$$

where τ_i is the duration of state i.

3 Robust Routing and Wavelength Assignment

One of the most important functions of the network control plane is assigning a lightpath to each incoming connection request, the so-called dynamic routing and wavelength assignment (RWA) problem. In this section we describe various approaches to RWA that consider the physical layer impairments; a more complete survey of applicable methods can be found in [9]. In physically impaired networks, the survivability of a network is closely tied to the RWA used. The delay in call set-up (the latency) is also crucial because restoration must be accomplished as quickly as possible so that service is not severely disrupted.

We consider four non-QoT aware RWA algorithms, named First Fit (FF), Random Pick (RP), Best Fit (BF), and FF with ordering (FFwO). In FF, the routing is performed using a shortest path (SP) and the connection is established using the lowest-index available-wavelength on that path (if any). In RP, the routing is also shortest path but the wavelength used is randomly selected. In BF the shortest path is computed for each wavelength (based on its availability), and the shortest among those is selected. FFwO uses FF on an ordered list of wavelengths so that adjacent wavelengths are less often used, thus minimizing adjacent channel crosstalk [10]. These four algorithms can be improved by allowing the call admission procedure to consider multiple wavelengths successively for QoT and latency compliance. A call is then blocked if, once all wavelengths of the route have been tested, none satisfies the QoT constraint, or a timeout constraint is exceeded. These techniques are called QoS-FF, QoS-RP, and QoS-FFwO, depending on the order in which the wavelengths are considered.

Simulations are conducted using the parameters listed in Table 9.1, typical for current regional networks. The network load is the total arrival rate of calls to the network, assuming the source destination pairs are uniformly selected.

The average blocking probability (BP) for the NSFNET, illustrated in Fig. 7.1 in chapter 7, is shown for various network loads and several RWA algorithms in Fig. 9.2. A call can be blocked by an insufficient number of wavelengths, too high a latency, or insufficient QoT, as defined above. QoS-FFwO can be seen to perform the best out of the six RWAs, yielding the lowest average BP. QoS-FF is second-best in our simulation scenario. Note that in low traffic cases, the wavelength blocking is negligible and BP is dominated by the timeout and BER blocking. Further insight on the effects on the blocking probability of imposing a latency constraint can be found in [12].

We also test survivability methods on QoT-aware RWA algorithms that belong to the class of adaptive RWA algorithms as defined in [13] and provide various degrees of fairness and QoT enhancements. Our QoT-optimized adaptive RWA algorithms enforce the QoT constraint by computing the Q factors of the tentative lightpath and of all lightpaths that share at least a link or node with the tentative lightpath, and rejecting the tentative lightpath from the list of candidate lightpaths if any of these Q factors is below the preset threshold. We propose the following five policies to select the route on which an incoming call is to be accommodated [11]:

Table 9.1 Physical parameters for the simulated network [3]

Description	Value
Span length	70 km
Signal peak power	2 mW
Bit duration	100 ps (10 Gbps)
Pulse shape	NRZ
Fabric crosstalk	− 40 dB
Adj. port crosstalk	− 30 dB
Non adj. port crosstalk	− 60 dB
Adj. wavelength crosstalk	− 25 dB
Fiber loss	0. 2 dB/km
Nonlinear coefficient	2. 2 $(\text{W km})^{-1}$
Linear dispersion	17 ps/nm/km
Dispersion compensation	100% post-DC
ASE noise factor	2
Receiver electrical bandwidth	7 GHz
Number of wavelengths	8
Minimum Q factor	6

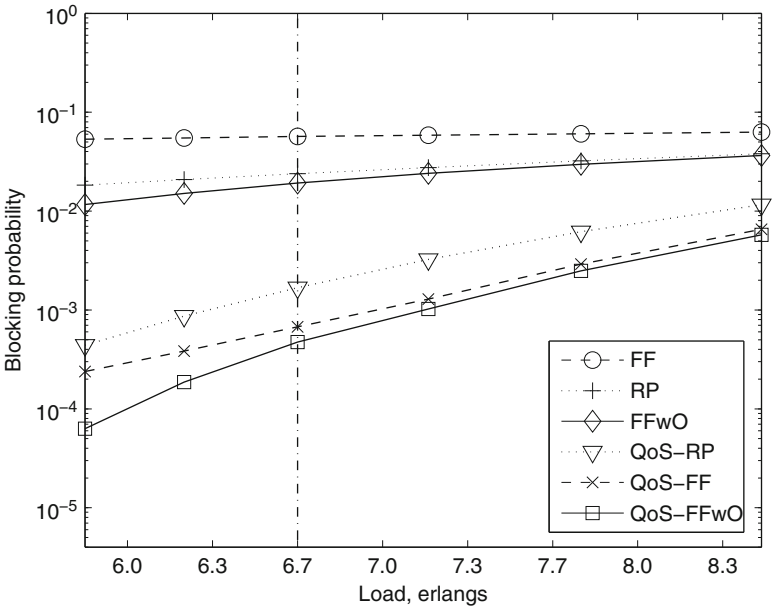


Fig. 9.2 Call blocking probability for the six RWA algorithms with BER and latency guarantees [14]

- *SP2* is the shortest path algorithm but with some longer paths reserved for long calls, as proposed in [15].
- *HQ* (highest Q factor) selects the lightpath with the highest Q factor among the set of candidates.

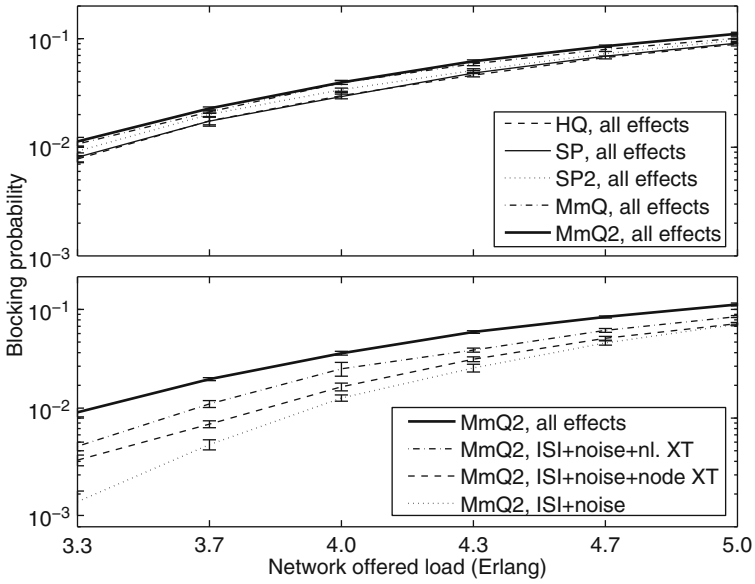


Fig. 9.3 Call blocking probability for the various QoT-aware RWA algorithms (*top panel*) and for the MmQ2 algorithm with various physical impairments (*bottom panel*) [11]

- *MmQ* (max–min *Q* factor) maximizes the margin of QoT operation in the network by selecting the lightpath that maximizes (over the candidates) the minimum *Q* factor. The idea is to push all lightpaths to lower BER together in a fairer manner.
- *MmQ2* is MmQ augmented with the same path reservation technique as in SP2.

Fairness-enhanced algorithms (SP2, MmQ2) may result in higher blocking probability than the reference algorithm (SP). Similarly, although HQ, MmQ and MmQ2, which are designed with QoT in mind, attempt to lower blocking probability due to the QoT constraint, the path they select is not necessarily the shortest path. Therefore these three algorithms may waste network resources, making wavelength continuity constraint harder to fulfill for future calls, and making link failure more likely on those paths.

In the top panel of Fig. 9.3, we plot the blocking probability when all physical impairments are accounted for as the total offered network loads increases. Among the four QoT-aware RWA algorithms listed above, the HQ algorithm tends to perform best and the MmQ2 worst; however, the differences in average blocking probabilities for all algorithms are small and overall all algorithms perform very similarly. Further investigating the origin of blocking probabilities for one of the algorithms, MmQ2, consider the bottom panel in Fig. 9.3. The difference between the topology-related impairments only (ISI and noise) case and the case with all impairments (all-effects: ISI, noise, nonlinear and node crosstalk) is close to one order of magnitude in average call blocking probability. Furthermore, we observe that the blocking probabilities for

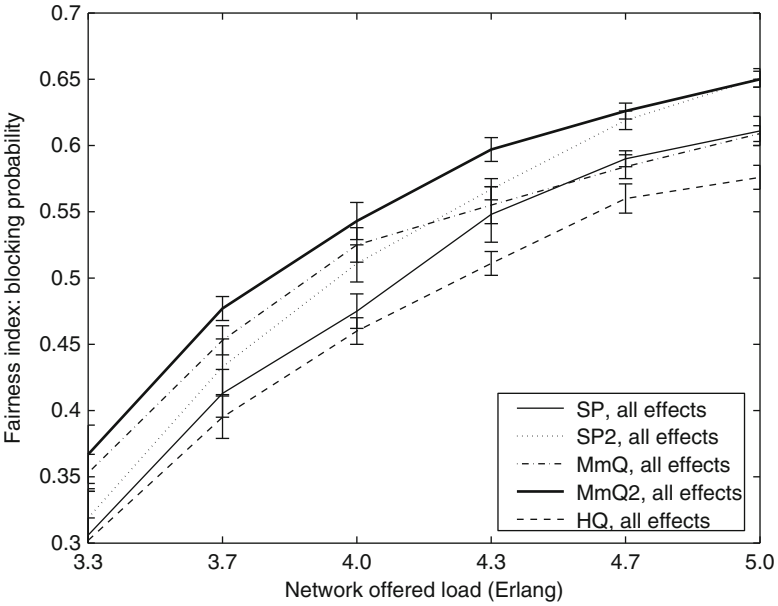


Fig. 9.4 Call blocking probability fairness for the various QoT-aware RWA algorithms [11]

cases including only nonlinear crosstalk is only slightly higher than the case considering only node crosstalk. Therefore, inclusion of both kinds of crosstalk is important when evaluating RWA algorithms for blocking probability.

We show how our algorithms perform in terms of blocking probability fairness in Fig. 9.4. For the NSFNET topology, the MmQ2 algorithm exhibits the highest fairness even compared to SP2 which was designed to improve fairness. The HQ algorithm performs significantly lower than other algorithms as it is only designed to improve QoT.

4 Protection and Restoration in Physically Impaired Networks

In this section, we use these robust RWA algorithms within networks designed to survive or recover quickly from link failures. Survivability in DWDM networks can be achieved by protection or restoration, which in turn can be path-wise or link-wise; furthermore, protection algorithms can use either a shared protection path or a dedicated one [16]. We first address path protection and restoration, followed by a description of link protection and restoration techniques, and provide simulation results showing the performance of these techniques in a network suffering from physical impairments.

4.1 Path Protection and Restoration

We consider dedicated (1 + 1) path protection, where every connection has two link-disjoint lightpaths to handle single-link failures, a *primary* path and a *backup* path. In networks with regeneration (such as SONET), both the primary and backup paths are simultaneously used, and the strongest SNR path at the receiver is chosen. This ensures quick traffic restoration in case one of the paths fails. However, in physical impairment-sensitive optical networks, the QoT can be significantly worsened by keeping the backup path lit, increasing the blocking probability of lightpaths. On the other hand, keeping the backup path dark (until it is needed) can lead to increased traffic restoration times. We study the impact of dark and lit backup paths on the blocking probability and the vulnerability ratio.

In the path protection approach with a lit backup path, the chosen RWA algorithm is run twice in order to compute two link-disjoint paths, and the call is blocked if they cannot be found. Both paths are assumed to be lit and the call is accepted if at least one of the two paths meets the BER threshold requirement. We also check that no connection in the network sees both of its lightpaths disrupted (due to an unmet QoT constraint) at the same time by the crosstalk added due to the introduction of the two newly lit paths; otherwise the new call is blocked.

In the path protection scheme with dark backup path, the RWA approach is the same as above except that only one of the two established paths is lit. The QoT for all primary paths of every ongoing connection must be sufficient. If the QoT constraint is violated with the one path, then the same procedure is repeated with the second path found for the incoming connection. If both fail, the call is blocked. We evaluate three RWA algorithms for dedicated path protection, SP, BF (Best Fit), and HQ. In SP, the shortest path and the first fit wavelength is chosen for each of the two paths. HQ, as mentioned earlier, finds the shortest path on each wavelength and picks the one with the highest Q factor.

When a link fails, data on each affected primary lightpath must be switched over to its corresponding backup. In the lit backup scheme, the backup path must be checked for sufficient QoT. If the QoT is good enough, then the restoration is successful; otherwise it is not. In the dark backup scheme, the backup path must be lit, and the QoT of the dark backup as well as the primary lightpath of every ongoing connection must be checked for sufficient QoT. If any of these QoTs is not satisfactory, then the restoration is deemed to be not successful.

Path restoration is achieved by using one RWA scheme to set up a path for a requested connection, and using possibly another RWA scheme to find the restoration path between the connection source and destination in case of link failure. In our nomenclature, the first acronym refers to the RWA to the primary route and the second refers to the restoration route; for example a *FF–FF* path restoration algorithm uses FF for both the primary lightpath and the restoration lightpath. We consider three simple schemes for comparison. In the first, called FF–FF, shortest routing with first-fit wavelength assignment (FF) is used for selecting the

primary path as well as the restoration path. Another scheme is HQ–HQ, wherein HQ is used for choosing both the primary path and the restoration path. The final algorithm is BF–BF, in which the shortest path on each wavelength is found and the shortest of these is selected for both primary and restoration paths. Results for other combinations can be found in [17].

4.2 Link Protection and Restoration

In link protection, a backup path between the end nodes of every link is determined off-line and is activated if the link fails. There are many ways of determining these backup paths; we consider two ways.

First, we select the backup paths for links using the method presented in [18], wherein an algorithm is presented to find a 2-connected directed subgraph of the network graph. Let us call this directed subgraph *blue* digraph (directed graph). At the same time another subgraph, which we refer to as *red* digraph, is generated that is the same as the blue digraph except that the edge directions are reversed. Half of the available wavelengths, say set Λ_1, would be assigned to be used by primary paths on the blue graph and by protection paths on the red digraph. The remaining set of wavelengths, which we refer to as set Λ_2, would be used to carry primary data on the red digraph and protection data on the blue digraph. Upon arrival, a connection can be routed on either of the two digraphs, using the wavelength set assigned for the primary data on that digraph. We use the Shortest Path routing algorithm with First-Fit Wavelength assignment *(FF)* to find the primary path. Note that the shortest path may be either on the blue digraph or the red digraph. In case of a link failure, those connections that were using that link on the blue digraph direction, which are using a wavelength in set Λ_1, would be routed on the backup path on the red digraph around that link. Since set $\Lambda_1$is reserved for protection in the red digraph, it would be available for all those connections. The same approach would be followed for data on the other digraph.

The backup path for each link is static and can be found off-line, hence in case of failure the backup path is already known around each link. In our approach, we assign the shortest path around each link (on the other digraph) for protecting a connection passing that link in each direction. Notice that all the connections on the failed link would follow the same backup path, therefore there is no need for demultiplexing and re-multiplexing these connections at either end of the failed link.

In the link restoration scheme, arriving calls are routed according to FF. In case of link failure, for each affected connection we find the shortest available path around that link on the same wavelength the connection is already using. It is clear that different connections may need to use different restoration paths.

The second way of providing link protection will be presented in Sect. 9.5.

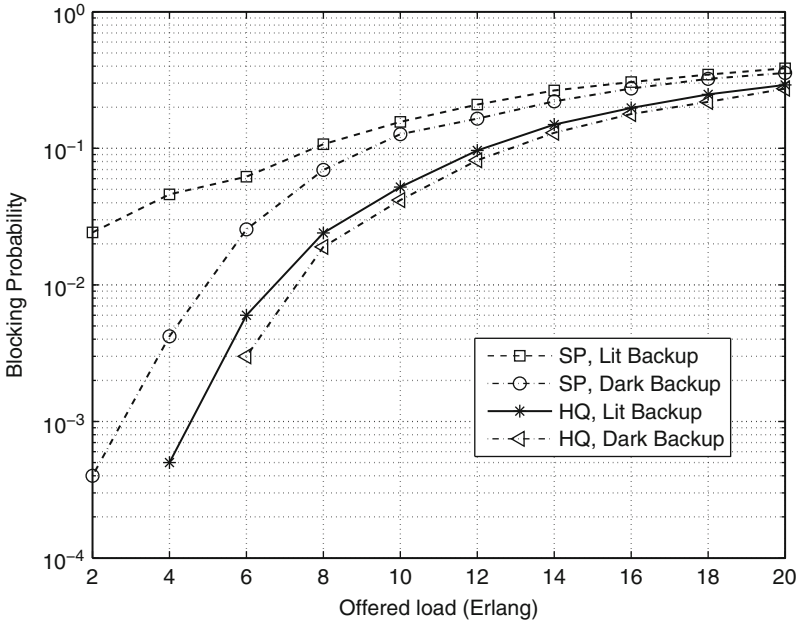


Fig. 9.5 Blocking probability vs. traffic load for dedicated path protection [8]

4.3 Performance of Protection and Restoration Algorithms

We now present numerical results for the performance of the various algorithms for path and link protection and restoration. The performance of the path protection algorithms for both lit and dark backup paths is shown in Figs. 9.5 and 9.6. These results show that the lit backup protection schemes have much worse blocking performance than the dark backup scheme, especially at lower loads. This is due to the increased interference caused by lighting up the backup paths, and the resulting lower QoT. The price paid for this lower blocking probability is a slower traffic restoration time for the dark backup scheme compared to the lit backup scenario. The slower restoration time is because not only must the dark backup path be activated, but also because the effect of lighting up the backup on the QoT of *all* ongoing connections must be checked before activating it. In this figure we also consider the QoT-aware HQ RWA technique, which significantly improves the performance of both the lit and dark backup schemes. At higher loads, all algorithms become wavelength-blocking limited, and there is little that QoT-aware algorithms can do to improve blocking performance.

When the network experiences a link failure, some backup paths that become primary paths may not have sufficiently high QoT to be used. The vulnerability ratio of protection schemes, shown in Fig. 9.6, is lowest for systems using the HQ RWA algorithm for low offered loads, especially for the lit scheme. In the dark

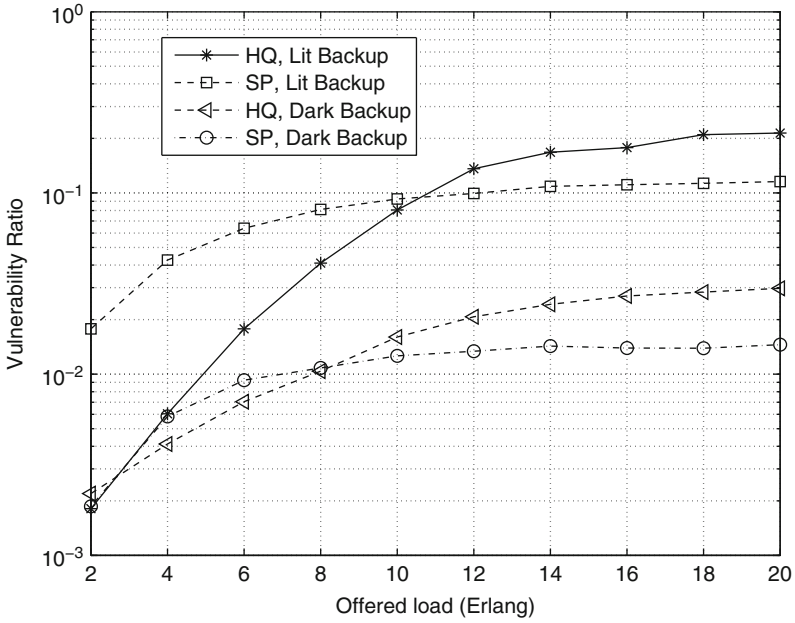


Fig. 9.6 Vulnerability ratio vs. traffic load for dedicated path protection [8]

backup scheme, the HQ algorithm cannot predict the performance of the backup path. At higher loads (where networks avoid operating), HQ has a negative impact, which is an artifact of our definition of the vulnerability ratio: because HQ blocks fewer calls, there are more calls on the network that can be blocked; this does not mean that more calls are serviced properly in an absolute sense. The QoT-unaware techniques do not perform well, justifying our emphasis on cross-layer algorithms.

A comparison of the performance of the link protection and restoration algorithms based on [18] and path protection and restoration is given in Figs. 9.7 and 9.8. As we can see from Fig. 9.7, link protection has high blocking probability due to severe degradation of QoT. This is due to the fact that in link protection the route of a connection is restricted to one of the two digraphs; therefore, it may not be the shortest path. Increasing the path length leads to more noise and crosstalk in more intermediate nodes and finally lower quality of the signal at the receiver. As for the restoration algorithms, we see from Fig. 9.7 that the BF scheme is much less advantageous than the HQ alternative. The proposed HQ–HQ path restoration algorithm, which is the most complex, strongly outperforms other algorithms in the presence of physical layer impairments, which supports the idea of using cross layer approaches in all-optical network design.

The high vulnerability ratio of link protection algorithms (Fig. 9.8) is also due to the long paths used by connections; when the protection path of a link is lit up, it can deteriorate the quality of a large number of ongoing connections to a degree that would not be acceptable. Note that Fig. 9.8 shows the vulnerability of these

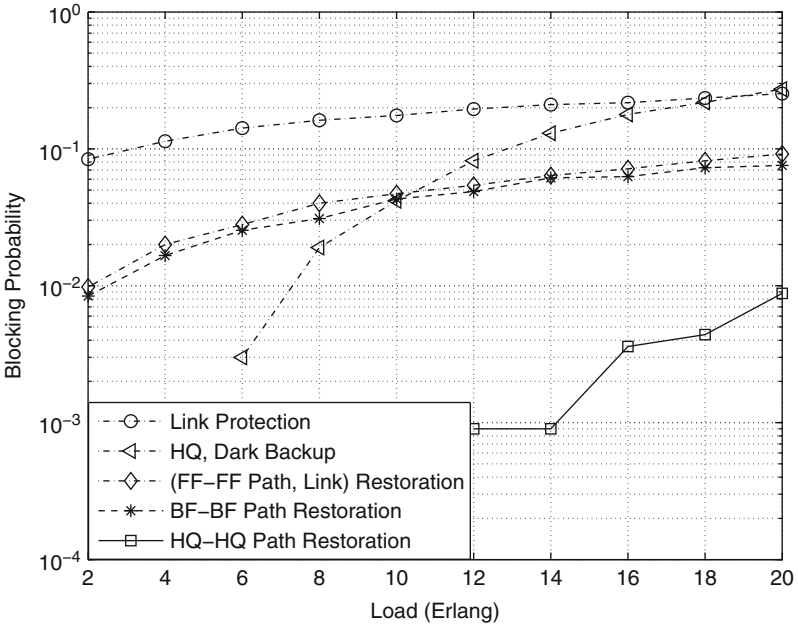


Fig. 9.7 Blocking probability vs. traffic load for link and path protection and restoration [8]

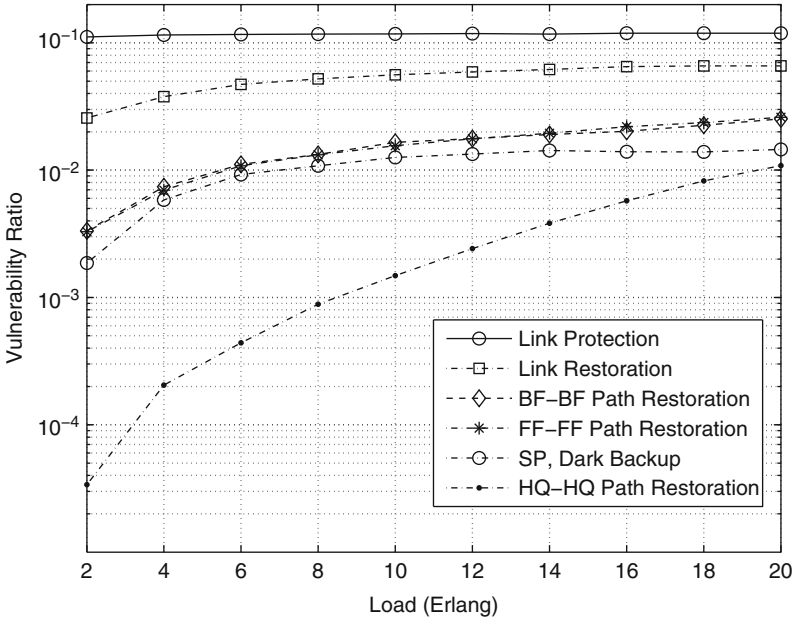


Fig. 9.8 Vulnerability ratio vs. traffic load for link and path protection and restoration [8]

algorithms taking the physical layer effects into account. If we had a perfect physical layer, this metric would be zero except for restoration schemes, which do not reserve resources to handle failure effects.

5 Link Protection Using P-Cycles

In this section, we present two impairment-aware p-cycle approaches to link protection, as they promise to provide high recovery speed and high efficiency in wavelength usage. In the p-cycle approach to protection, the network is decomposed into ring-like structures (called p-cycles) with pre-configured switches, which allows the switching of traffic from the primary path to the protection path to be very fast. The first approach we propose assumes a single protection class while the second results in better performance by defining two service classes.

5.1 Single Protection Class

P-cycles (pre-planned, or pre-configured cycles) have been used as a protection method for many years, but most of the literature has ignored physical impairments when studying p-cycles. In this approach, cycles on the network are chosen such that each link is either on a cycle or is a straddling link [19], i.e., p-cycles form a set of overlay cycles that cover the network such that each link is either on a cycle or has its two ends on a single cycle.

The approach is illustrated using a simple example. Figure 9.9 shows a simple grid network with a choice of two p-cycles, namely A-B-F-G-H-A and B-C-D-E-F-B. With these two p-cycles, links B-H, G-H, C-F, and D-F are straddling links, while the other links are on-cycle links. In each cycle half the available wavelength set, say set Λ_1, is assigned to carry primary traffic on the blue (clock-wise) directed

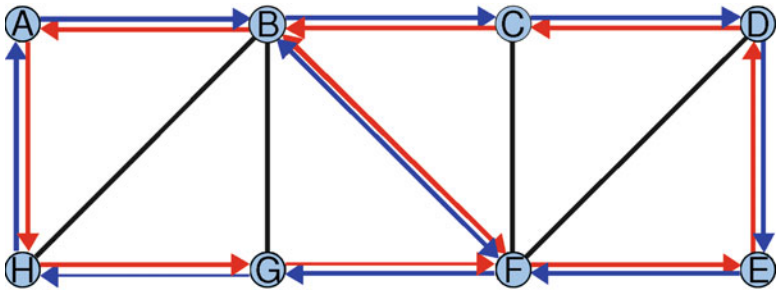


Fig. 9.9 Grid topology with multiple (two) p-cycles. Each link consists of two spans, 70 km each [20]

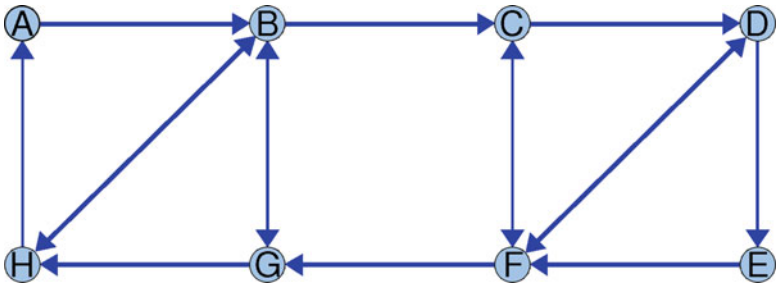


Fig. 9.10 Blue graph for call set up on wavelength set Λ_1 [20]

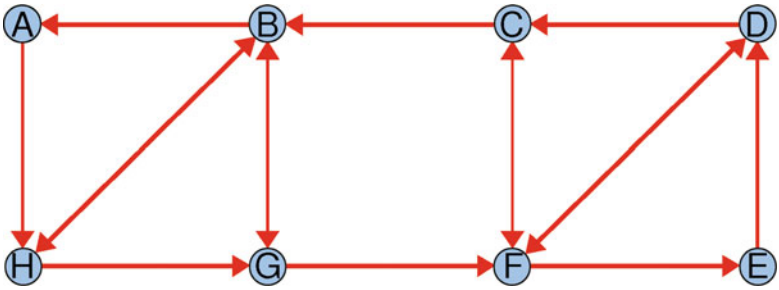


Fig. 9.11 Red graph for call set up on wavelength set Λ_2 [20]

cycles and the same set is reserved for protection on the red (counter-clock-wise) cycles. The other half of the wavelength set (Λ_2) is assigned to carry primary traffic on the red cycles and is reserved as protection on the blue directed cycles. All the wavelengths on the straddling links are available to the primary routes.

The direction of either of the cycles (red or blue) should be the same on both p-cycles. This is the only way that pre-configuration is possible for those switches that are shared between the two p-cycles. Consequently, all the wavelengths on link B-F should be assigned for protection, because both cycles contain that link (in both directions). As can be seen in the figure, both blue and red cycles pass this link in both directions. Figure 9.10 shows the links and directions available for primary path setup on wavelength set Λ_1 and Fig. 9.11 shows the same for set Λ_2.

A new connection is routed by searching for a path on one of the two digraphs, using a prescribed algorithm for finding a path on each digraph, and using the "better" of the two. In case of a link failure, connections that were using a wavelength from set Λ_1 will be rerouted on the red cycle of the corresponding p-cycle and those using set Λ_2 will be rerouted on the blue cycle. Obviously, if the failed link is a straddling link, the traffic on that is split on the two cycles of the corresponding p-cycle.

Generally, if a Hamiltonian cycle exists in a network, then that is considered to be the most efficient p-cycle because it uses the fewest links and therefore reserves

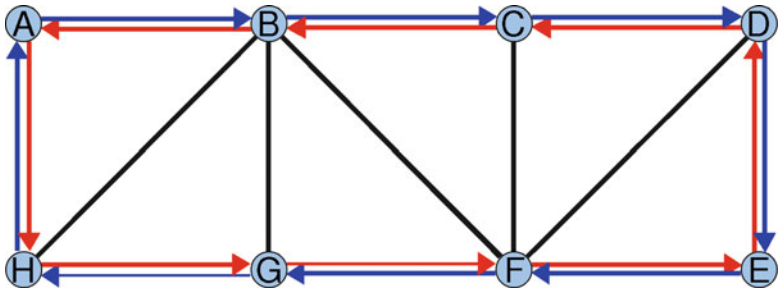


Fig. 9.12 Grid topology with Hamiltonian p-cycle [20]

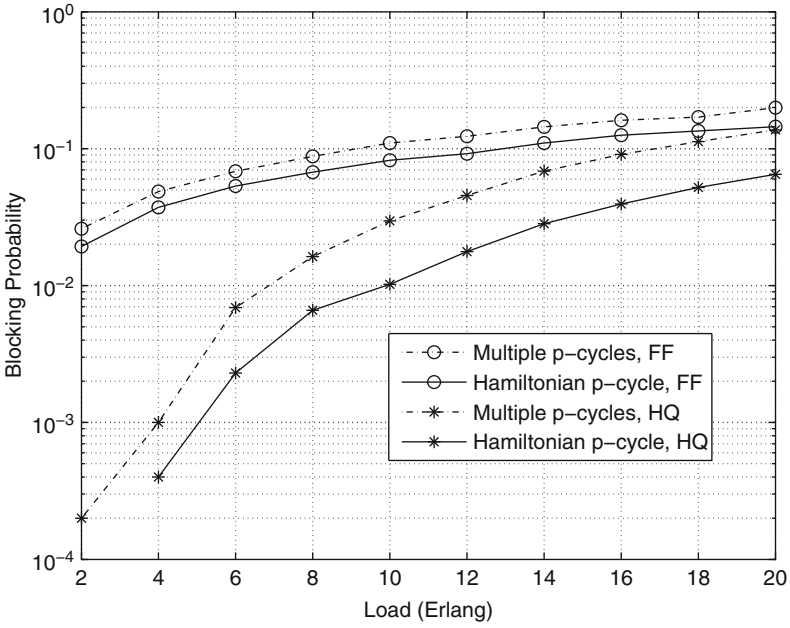


Fig. 9.13 Blocking probability vs. traffic load for two p-cycle designs [20]

the fewest wavelengths for protection [21]. As such, this is expected to provide a better blocking performance than a p-cycle design using more than one p-cycle. A Hamiltonian cycle for the same grid network is shown in Fig. 9.12.

Figures 9.13 and 9.14 show the blocking probability and vulnerability ratio for the above two p-cycle options (the Hamiltonian vs. two smaller cycles) applied to the 14-node NSFNET. Additional simulation results are given in [20]. As might be expected, the Hamiltonian cycle shows better blocking performance compared the two-cycle case. However, it also has a higher vulnerability ratio, because the

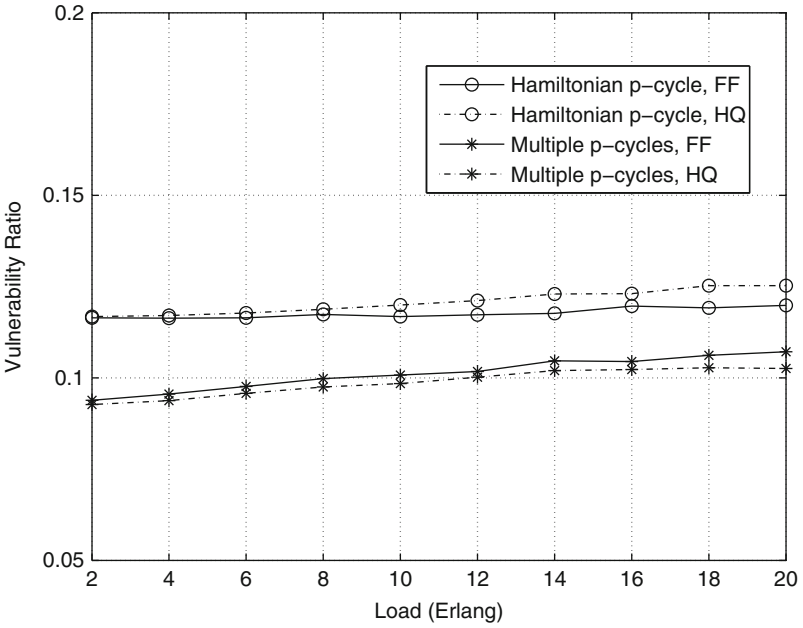


Fig. 9.14 Vulnerability ratio vs. traffic load for two p-cycle designs [20]

backup path for a failed link is generally longer in this case and may not have a good enough QoT. Once again, HQ significantly outperforms FF, validating the need for cross-layer RWA.

5.2 Providing Multiple Protection Classes Using Partial Preconfiguration

In the previous section, we saw that if a single p-cycle can be found in a network, it may not necessarily provide the best performance because of the long paths. For example, consider the network shown in Fig. 9.15.

All the links of this network can be protected from failure by the single p-cycle 1-2-5-3-6-5-4-1 in the above network. All the crossconnects on the cycle can be pre-configured. Note that although node 5 appears two times in the cycle, it is traversed on different input/output link pairs each time, thus allowing its crossconnect to be pre-configured. This cycle is a long cycle and even if the primary paths are short, the backup paths that are activated upon failure may be too long and may not have sufficient QoT. However, if we relax the requirement that *every crossconnect* has to be pre-configured, then it is possible to choose multiple cycles as follows. Suppose two cycles are chosen: 1-2-5-4-1, and 3-6-5-3. Then, the crossconnects at all of the nodes on these cycles can be pre-configured to protect

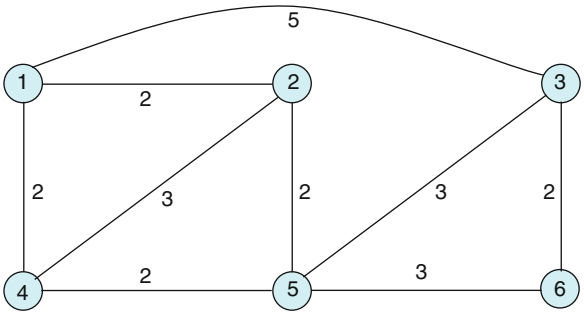


Fig. 9.15 A 6-node network. The number next to each link corresponds to the number of spans, 70 km each

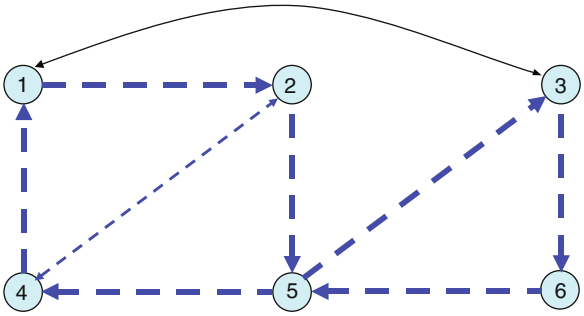


Fig. 9.16 The $Blue_{\text{setup}}$ graph

the failure of any on-cycle link or a link straddling one of the two cycles. Link (1,3) however, straddles the two cycles, and even though there is a backup path, namely 1-2-5-3, node 5 is not pre-configured for this path, and must therefore be configured *after* the link (1,3) fails. This leads to slower traffic recovery for connections using link (1,3). Using this idea, we define two types of links, namely, *fast-recovery* and *slow-recovery* links. Fast-recovery links are those that have all the crossconnects in their protection path preconfigured and slow-recovery links are those that have their protection path precomputed but the crossconnects are not preconfigured for their backup paths.

A connection is called a *fast-recovery* connection when all the links in its path are fast-recovery links, and as a *slow-recovery* connection if it passes through at least one slow-recovery link. As before, four digraphs, $Blue_{\text{setup}}$, $Blue_{\text{protection}}$, Red_{setup}, and $Red_{\text{protection}}$, are formed from the two cycles. Figures 9.16 and 9.17 show the $Blue_{\text{setup}}$ and $Blue_{\text{protection}}$ graphs, respectively. The red digraphs are exactly the same as the blue digraphs, but with on-cycle links oriented in the opposite direction. If the whole wavelength set is divided into two sets of equal cardinality, $\Lambda = \Lambda_1 \cup \Lambda_2$, then Λ_1 is used for setup and protection on the blue

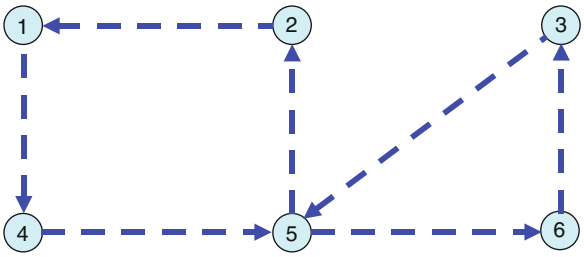


Fig. 9.17 The $Blue_{\text{protection}}$ graph

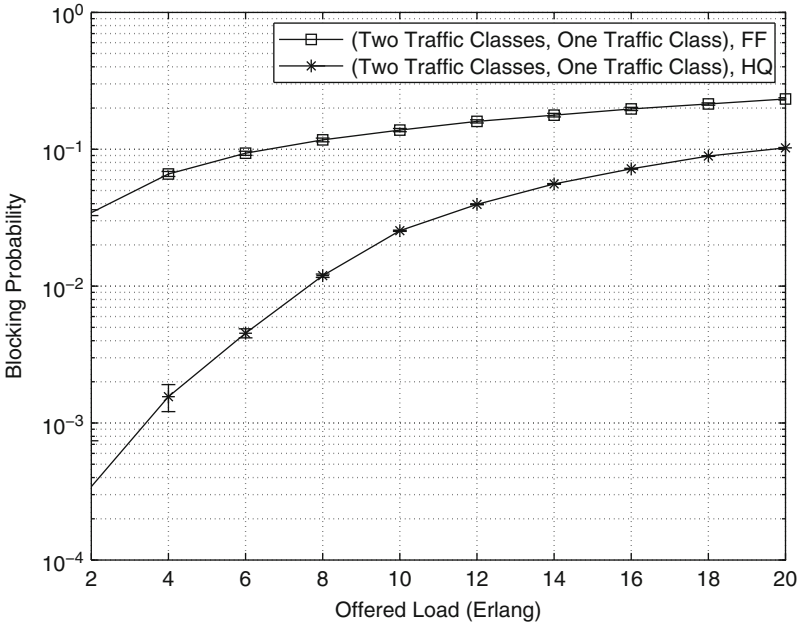


Fig. 9.18 Blocking probability vs. traffic load for the 6-node grid network

digraphs and Λ_2 is used on the red digraphs. When a new connection request arrives, the RWA algorithm chosen is run on both digraphs using the corresponding wavelength sets and selects the lightpath according to the RWA criterion. A directed link on the $Blue_{\text{setup}}(Red_{\text{setup}})$ digraph is protected by the shortest path on the $Blue_{\text{protection}}(Red_{\text{protection}})$ path.

An general integer-linear programming formulation for selecting cycles that balance wavelength usage and length of cycles is given in [22]. We present in Figs. 9.18 and 9.19 the blocking probability and vulnerability ratio for the single-cycle (one class) case and two-cycle (two classes) case illustrated in the example above.

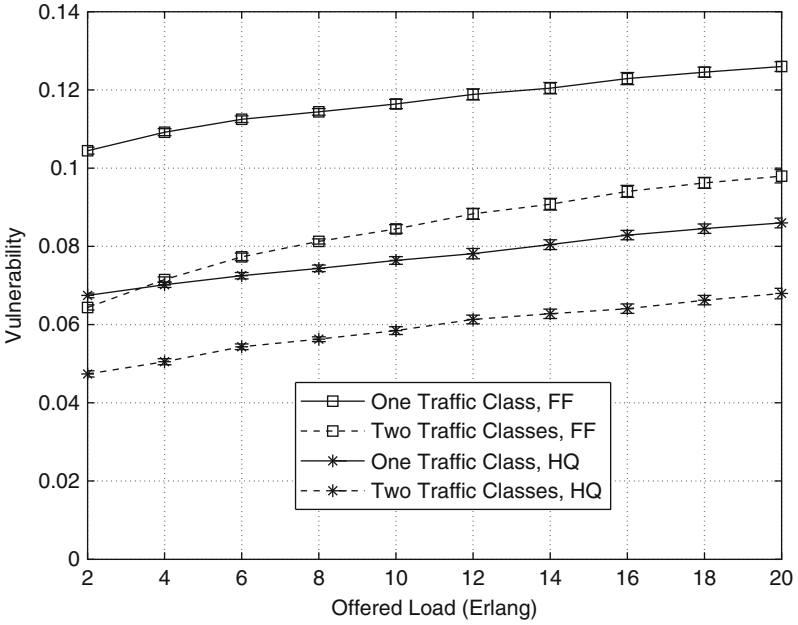


Fig. 9.19 Vulnerability ratio vs. traffic load for 6-node grid network

Table 9.2 Percentage of fast-recovery connections in the two-cycle case

Load (Erlang)	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**	**20**
Percentage (%)	78	79	80	81	81	82	82	83	83	84

First, we see that the blocking probabilities for the two cases are exactly the same. This is because the set of wavelengths that are reserved on the network are identical for the two cases. We see once again that HQ performs much better than FF. The difference in performance between the two cases can clearly be seen in the vulnerability ratios. The vulnerability ratio for single-class is almost double that for the slow-recovery case, for both FF and HQ. This is due to the fact that the backup paths are much shorter when there are two cycles. It must be noted that this improvement comes at the expense of online reconfiguration of the crossconnect at node 5 when link (1,3) fails.

Table 9.2 shows the percentage of connections that are fast-recovery in the two-cycle case. As can be seen, the significant improvement in vulnerability can be obtained with only a modest increase in the number of slow-recovery connections.

6 Chapter Summary

Transparent optical networks are particularly susceptible to link failures, as the effects of these failures can result in long restoration paths that may have excessive physical-layer impairments. In this chapter we addressed the design of protection and restoration techniques suitable for wavelength switched optical networks suffering from link failures and physical-layer impairments. We considered multiple routing and wavelength assignment algorithms, ranging from completely QoT unaware to the most complex QoT-optimizing schemes. Metrics used to evaluate the performance of the proposed techniques included the blocking probability and the vulnerability ratio, both important measures for network operators.

Through our results we conclude that for path protection schemes, it is far better to leave the backup paths dark if the network is able to accommodate the resulting longer path restoration time. For both path protection and restoration, the QoT-aware HQ RWA algorithm significantly outperforms the simple QoT-unaware alternatives, providing both lower blocking probability and lower vulnerability ratio.

Link protection techniques tend to be more susceptible to impairments, and require careful design. We examine the use of p-cycles for link protection, and present algorithms and performance comparisons for both a single protection class and the more powerful two protection class case. The results show that when some connection are allowed to have slightly longer restoration times, the overall system performance, both in blocking probability and vulnerability ratio, is much improved.

References

1. Agrawal G (2002) Fiber-optic communications systems.Wiley, New York
2. Mukherjee B (2000) WDM optical communication networks: Progress and challenges. IEEE J Sel Area Comm 18(10):1810–1824
3. Askarian A, Subramaniam S, Zhai Y, Brandt-Pearce M (2009) A cross-layer ILP formulation for finding p-cycles in all-optical networks. In: Proceedings of IEEE GLOBECOM, Honolulu, Hawaii
4. Xu B, Brandt-Pearce M (2003) Comparison of FWM- and XPM-induced crosstalk using the Volterra Series Transfer Function method. J Lightwave Tech 21(1):40–53
5. Pointurier Y, Brandt-Pearce M (2005) Analytical study of crosstalk propagation in all-optical nNetworks using perturbation theory. J Lightwave Tech 23(12):4074–4083
6. Deng T, Subramaniam S, Xu J (2004) Crosstalk-aware wavelength assignment in dynamic wavelength-routed optical networks. In: Proceedings of Broadnets, San Jose, CA, USA
7. Jain R, Chiu D, Hawe W (1984) A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. DEC Research Report TR-301
8. Askarian A, Zhai Y, Subramaniam S, Pointurier Y, Brandt-Pearce M (2008) Protection and restoration from link failure in DWDM networks: A cross-layer study. In: IEEE ICC, Beijing, China

9. Azodolmolky S, Klinkowski M, Marin E, Careglio D, Pareta J, Tomkos I (2009) A survey on physical layer impairments aware routing and wavelength assignment algorithms in optical networks. Comput Netw 53(7):926–944
10. He J, Brandt-Pearce M (2006) RWA using wavelength ordering for crosstalk limited networks. In: Proceedings of the IEEE/OSA optical fiber conference (OFC), Anaheim, CA, USA
11. Pointurier Y, Brandt-Pearce M, Subramaniam S, Xu B (2008) Cross-layer adaptive routing and wavelength assignment in all-optical networks. IEEE J Sel Area Comm 26(6):32–44
12. He J, Brandt-Pearce M, Subramaniam S (2009) QoS-aware wavelength assignment with BER and latency constraints for all-optical networks. IEEE/OSA J Lightwave Tech
13. Mokhtar A, Azizoglu M (1998) Adaptive wavelength routing in all-optical networks. IEEE/ACM Trans Netw 6(2):197–206
14. He J, Brandt-Pearce M, Subramaniam S (2007) Qos-aware wavelength assignment with BER and latency guarantees for crosstalk limited networks. In: Proceedings of the IEEE international conference on communications (ICC), Glasgow, UK
15. Birman A, Kershenbaum A (1995) Routing and wavelength assignment methods in single-hop all-optical networks with blocking. In: Proceedings of the IEEE conference on computer communications (INFOCOM), vol. 2, pp. 431–438
16. Zhou D, Subramaniam S (2000) Survivability in optical networks. IEEE Netw Mag 14(6): 16–23
17. Askarian A, Zhai Y, Subramaniam S, Pointurier Y, Brandt-Pearce M (2010) Cross-layer approach to survivable DWDM network design. IEEE/OSA J Opt Comm Netw 2:319–331
18. Medard M, Barry RA, Finn SG, He W, Lumetta S (2002) Generalized loop-back recovery in optical mesh networks. IEEE/ACM Trans Netw 10(1):153–164
19. Grover W (2004) Mesh-based survivable networks. Prentice Hall, Englewood Cliffs, NJ
20. Askarian A, Subramaniam S, Brandt-Pearce M (2009) Evaluation of link protection schemes in physically impaired optical networks. In: IEEE ICC, Dresden, Germany
21. Schupke D (2005) On Hamiltonian cycles as optimal p-Cycles. IEEE Comm Lett 9:360–362
22. Askarian A, Subramaniam S, Brandt-Pearce M (2010) Implementing protection classes through p-cycles in impairment-constrained optical networks. In: Proc. IEEE ICC

Chapter 10
Energy-Efficient Traffic Engineering

Bart Puype, Ward Van Heddeghem, Didier Colle, Mario Pickavet, and Piet Demeester

1 Introduction

The interest in the energy footprint of information and communication technology (ICT) has only recently been sparked. One of the first papers on this topic is probably "Greening of the Internet," written by Gupta and Singh in 2003 [1], and discusses the power consumption of the Internet and a number of approaches to increase its energy efficiency. Since then, the number of related publications has been rising steadily, and national and international projects have sprung up.

There have been three main drivers behind this increased interest from research communities, companies, and governmental bodies alike. Rising energy prices start to affect both end users and companies and provide a financial stimulus to choose energy-efficient devices. Technically, increased operating times for mobile devices, and energy supply and heat disposal issues for large systems, press to provide solutions featuring reduced power consumption. Finally, the looming global climate change scenario has increased funding for research to reduce electrical energy consumption—as a major contributor to greenhouse gas emissions—across a wide range of technologies, including ICT equipment.

If we look at the actual figures, the energy consumed by ICT equipment was estimated to be about 4% of the global primary energy consumption[1] in 2008 and projected to double in 2020 [2]. This estimate includes energy both consumed for manufacturing and during use of the equipment. Subdividing ICT equipment further into data centers, personal computers, network equipment (data and telecom only), TV sets, and a general "other" category (containing devices such as telephone handsets,

[1] Primary energy includes all energy found in nature, not only electrical energy.

B. Puype (✉) • W. Van Heddeghem • D. Colle • M. Pickavet • P. Demeester
Department of Information Technology (INTEC), Ghent University – iMinds,
Gaston Crommenlaan 8, 9050 Gent, Belgium
e-mail: bart.puype@intec.ugent.be

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*,
Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_10,

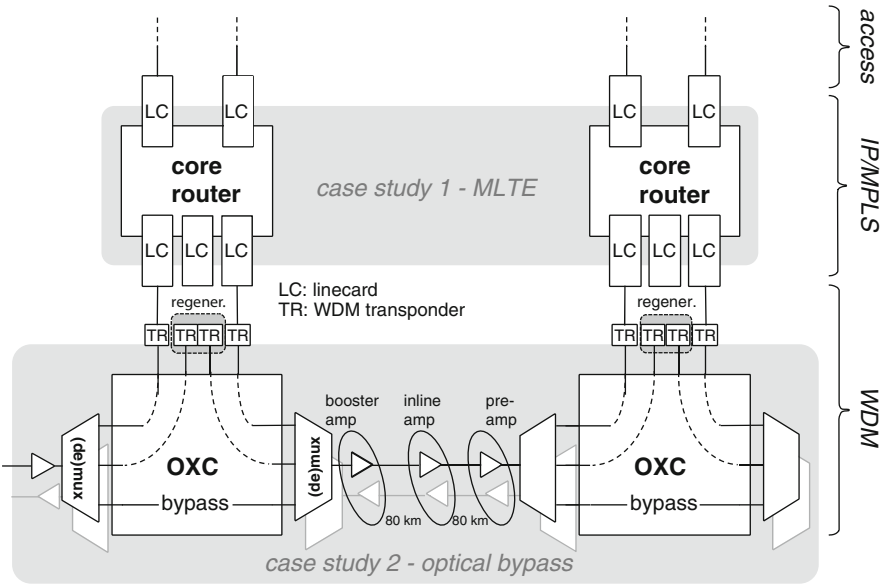


Fig. 10.1 Core network model and case study scope

printers, fax machines, gaming consoles), it appears that each of these five categories roughly consumes an equal share of electrical energy during use. As such, network equipment constitutes about 15% of the ICT electrical energy consumption.

Generalized, telecommunication networks are typically split up into a large number of access networks and a single or small number of core networks. The access network allows end users to connect to the core network. The core network provides a high-speed intermediate connection system that links the access networks of the engaging end users. The access networks currently take up by far the highest share of the network power consumption. However, with rising traffic demands, consumption is expected to shift to the core network [3, 4]. It is in this context that approaches to limit or reduce the energy consumption in core networks have to be seen.

While a detailed survey of existing techniques to increase energy efficiency in ICT networks is outside the scope of this chapter, a number of high-level optimization approaches can be considered. One general approach consists of switching off components during low traffic load situations, for example, by employing dynamic topology optimization through multilayer traffic engineering (MLTE) (see further). With the maximum number of idle components switched off, a next step consists of reducing the load on the remaining components, in effect avoiding overdimensioning the capacity. The optical bypass technique discussed further in this chapter is an example of such an approach. Finally, the power consumption of the individual components itself should be reduced. A more extensive discussion can be found in [5].

In the remainder of this chapter, two energy optimization studies are presented. Consider the model as shown in Fig. 10.1. A first study looks at core networks from

the view point of a layer 3 operator, that is, managing an IP/multiprotocol label switching (MPLS) layer networks on top of circuit-switched infrastructure typically based on wavelength-division multiplexing (WDM). Often, information about lower layers will be shielded, leading to an overlay operating model—this issue will extend to knowledge about lower layer energy consumption as well. The study presents energy-efficient MLTE building on an IP/MPLS layer power model. As can be expected, some similarities to typical capacity and quality of service (QoS) optimizing traffic engineering can be seen. Some results for short-term daily traffic variations are included.

On the other hand, a second study looks at core networks below the IP layer. In optical transmission systems such as WDM, the main contributing power consumers are transponders, fiber amplifiers, and regeneration. Two WDM power models are presented, and in this case, IP layer energy consumption is abstracted instead. Optical bypass is evaluated as a technique to achieve considerable power savings over per-hop optical–electronic–optical regeneration as the core network experiences longer term traffic volume increases.

2 Energy-Efficient Multilayer Traffic Engineering

Automatically switched multilayer IP-over-optical networks offer extensive flexibility in adapting the network to offered IP/MPLS traffic. Multilayer traffic engineering takes advantage of this through online IP logical topology reconfiguration in addition to the more traditional rerouting. The main functionality of MLTE is to optimize towards resource usage, bandwidth throughput and QoS performance. This functionality can be extended toward energy efficiency of IP/MPLS routing and logical topology. Moreover, these can be adapted dynamically to achieve optimal energy efficiency even for scenarios with short-term traffic changes, using the flexibility of MLTE strategies.

When including energy efficiency as an optimization goal, we need some kind of energy-based objective function or energy "cost." In the next subsections, we present a power model to abstract energy requirement allowing an IP/MPLS MLTE strategy with limited to no information about lower layers to optimize for this. Next we point out the similarities of such a model with more traditional MLTE objective functions such as capacity usage cost. Finally we adapt an MLTE strategy to include this power model and show some results comparing full mesh, slow (static) MLTE, and fast MLTE under a scenario with daily traffic variations.

2.1 Power Model for Multilayer Traffic Engineering

In a general case, assuming no suboptimal configuration, we may say that the total volume of packets processed (routed, forwarded) in a router node is dictated by the total volume of traffic offered at the interfaces. These interfaces consist of both tributary interfaces, which connect to access nodes, and line cards (LCs),

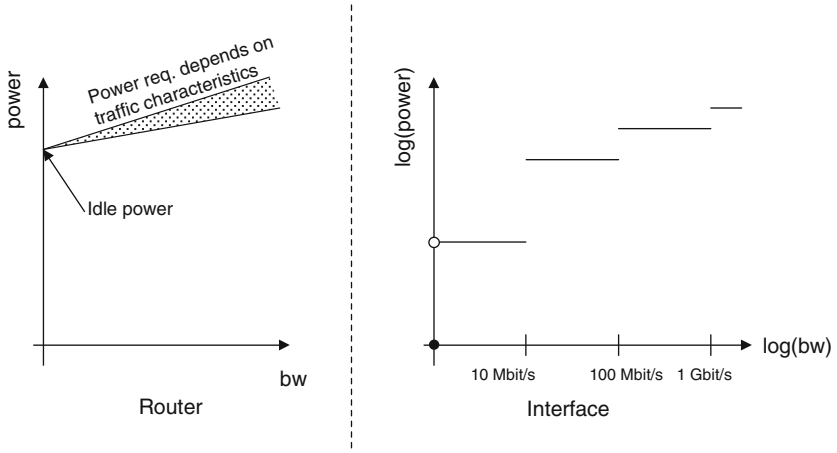


Fig. 10.2 Bandwidth dependency of energy consumption for router chassis and interfaces

which connect to the optical cross connect (OXC)/optical add/drop multiplexer (OADM) associated with the router.

Looking at interface traffic volume, we need to make a distinction between traffic volume sent and received over tributary interfaces to the client layer, and volume carried over LCs towards the optical layer. Client layer traffic is well defined through traffic matrices applied during simulation case studies or, in real-life situations, assumed not to be subject to operator-imposed energy limitations (which would amount to client traffic throttling). Therefore, we can say that the total amount of traffic processing in the IP/MPLS node depends on traffic volumes exchanged with its OXC/OADM, and it is these volumes that are subject to energy efficiency optimization.

Assuming all interfaces are bidirectional (i.e., connect to a RX and TX fiber), moving transit traffic from full optical switching into the electronic layer instead will increase LC bandwidth as well as router packet processing (and therefore power requirements). Of course, transit traffic in a node originates and terminates as client traffic in other nodes eventually.

Even if this transit traffic is the main area of optimization, the client traffic terminated at the node cannot simply be ignored in the bandwidth and power requirement of the core interfaces of that node. This is because energy consumption is not linear with traffic volume. This is the case both for the complete IP/MPLS router chassis and for single LCs, as shown in Fig. 10.2.

Firstly, the layer 3 router hardware itself will have a certain idle power specified. This is the power requirement when no (external) traffic is processed. Energy consumption increases with bandwidth, but there may not be a one-to-one relation, since, in addition to bulk bandwidth, traffic characteristics such as packet size or type (IP or MPLS) may also influence power requirements [6].

Secondly, ignoring traffic processing, layer 2 interfaces may show similar behavior, as shown on the right side of the figure (using logarithmic axes).

For example, for an Ethernet interface, several line rates may be supported. For each line rate regime, power requirement may be unrelated to transported traffic. However, higher line rates are generally more efficient than lower ones (in watt per Mbit/s). For example, a 1 Gbit/s line rate consumes less power than ten times the 100 Mbit/s power figure. For (electronic–optical) core network interfaces, we will typically see only 1 Gbit/s vs. 10 Gbit/s line rate, but one can assume similar effects for SONET/SDH interfaces supporting several framing rates (e.g., OC-48 vs. OC-192). Functionality such as the ability to power down parts of interface silicon in-between frames sent/received or throttling the interface to a lower line rate during lower bandwidth usage periods may lead to a smoother, more continuous curve for the interface power vs. bandwidth characteristics as well.

With the assumptions that router idle power is constant and interfaces carrying no traffic can be shut down, traffic processing is dictated by interface traffic volume. We adopt the LC power vs. bandwidth characteristic as a combined model of the bandwidth dependency of both chassis and interfaces and assume it to be sufficient in describing power requirement optimization. In a generalized case, such a power characteristic will take the power law form $p = p_0 + (1 - p_0)\ b^a$ (p normalized power, b normalized bandwidth, and parameters p_0 idle power, $a > 0$). Here p_0 is an abstraction of combined chassis idle power, energy dissipation in unloaded LC, etc. Simulations have shown in fact that a linear power characteristic will suffice in attaining near-optimal energy-efficient MLTE. Linear characteristics will correspond with $a = 1$; p_0 can be modified slightly to take the linear approximation into account.

2.2 Comparison with Traditional Capacity Usage Optimization

The introduction of the power vs. bandwidth characteristic suggests similarities with more traditional MLTE optimization objectives. Logical topology design and routing in traffic grooming and MLTE are generally centered on optimizing network resource usage for a certain traffic scenario, reducing optical layer capacity requirements, or maximizing IP layer throughput. For resource usage, we aim at minimizing number of utilized IP/MPLS links, wavelength channels, etc. For example, a resource usage performance metric could be total number of set up IP/MPLS links (as opposed to total power requirements of the IP/MPLS links).

Reducing the number of logical topology lightpaths (LPs) as per traditional resource usage studies will, of course, also lower energy requirements, but the relation between logical topology and energy consumption is in fact more complex. For example, sparser logical topology meshes require longer multi-hop IP paths, which lead to increased transit traffic processing energy consumption, as explained in the previous section.

Clearly, energy-efficient logical topology design is neither about maximum optical capacity saving (sparse mesh) nor maximum IP/MPLS router power reduction (short paths; full mesh). Optimal energy efficiency should be reached

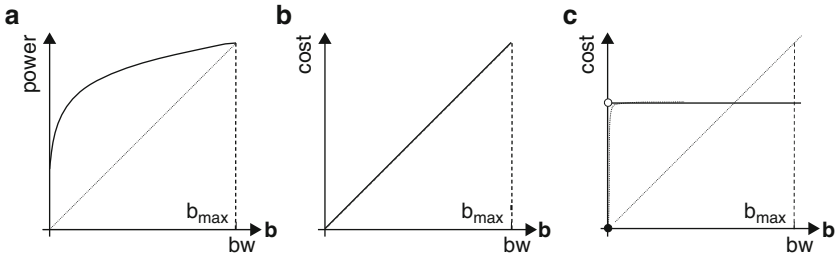


Fig. 10.3 Bandwidth-dependent cost functions used in optimization. (**a**) Power vs. bandwidth characteristic; (**b**) transit bandwidth optimization cost function; (**c**) lower layer capacity optimization cost function

somewhere in-between end-to-end and point-to-point grooming. However, link power consumption p vs. traffic load b is a nonlinear function; the shape of this function will determine the optimal parameters for the MLTE algorithm.

A sample power vs. used link bandwidth function is shown in Fig. 10.3a. When it is adopted as an MLTE cost function, the objective achieved is the minimization of total power consumption, which is the sum of all logical link *power* (b) values. Used as a cost function, it relates to other basic optimization scenarios with their own distinct "cost" shape:

- *Minimization of IP/MPLS router traffic volume, Fig.* 10.3b—router traffic volume is directly related to link traffic volume; therefore, such optimization corresponds with a proportional cost (vs. link bandwidth). Lightpaths in this case are "free" according to the cost function (meaning there is no optimization towards optical capacity requirements); therefore, an unloaded link has zero cost. Such optimization will lead to a (almost) full mesh logical topology, with minimal bandwidth utilization in all IP/MPLS links.
- *Minimization of optical layer capacity, Fig.* 10.3c—in this case, IP/MPLS router transit processing capacity and very long IP/MPLS paths are assumed to be "free." This suggests a constant cost regardless of bandwidth utilization, except for unloaded links, which can be assigned a zero cost (they can be removed from the logical topology for an optical layer capacity reduction). Using this cost function, the objective is to minimize the total number of IP links, that is, LPs.

Of course, a typical MLTE capacity optimization scenario combines the two fringe cases. Additionally, one can assume that for multi-hop grooming (as is typical with MLTE) to appear in the network, we will require a concave shape in this link cost vs. bandwidth utilization function, or at least a cost function which lies higher than the proportional one for $b < b_{\max}$. Note that the function on the right is not constant over the interval $[0, b_{\max}]$ as one may think at first glance but in fact (discontinuously) concave because of the zero cost for unloaded ($b = 0$) LPs. The middle function is neither concave nor convex; indeed, there is no reason to groom traffic on multi-hop paths with such a function, and the resulting routing will consequently be a shortest-path one on an IP/MPLS full mesh logical topology.

2.3 Multilayer Traffic Engineering Strategy

The MLTE algorithm [7] adopted in this section uses a routing cost function (not to be confused with the power characteristic which is used to calculate power requirement after an MLTE solution is found) as shown in Fig. 10.4.

The routing cost function is dependent on IP/MPLS link load and is such that an optimal cost is seen for links between a certain range LLT and HLT (low and high load threshold, respectively). High costs beyond HLT avoid overloading IP/MPLS links; a cost "bump" is added in order to raise costs for links with load below LLT. This makes sure that no lightly loaded links are present in the network. In fact, when routing separate traffic streams, any IP/MPLS links that have their load drop below LLT will eventually become more and more costly, leading to all traffic being deviated from those links. When applying such an algorithm on a full mesh, this leads to some of the links not carrying any traffic, with many streams being routed along multi-hop paths on the remaining links. The algorithm can route over a virtual full mesh first; the subset of links that do carry traffic after the MLTE finishes is then matched on the actual multilayer IP-over-optical network, using flexible optical switching and LP setup. Links not carrying traffic do not need setup, which is how the strategy optimizes logical topology (as well as routing).

Eventually, when looking at how IP/MPLS are then loaded in the network, we can plot a load histogram. An example histogram is also shown in Fig. 10.4, showing number of IP/MPLS links with a certain load for a sample traffic pattern, using this routing cost function. Obviously, the link load distribution is important as the power characteristic depends on core interface bandwidth utilization, that is,

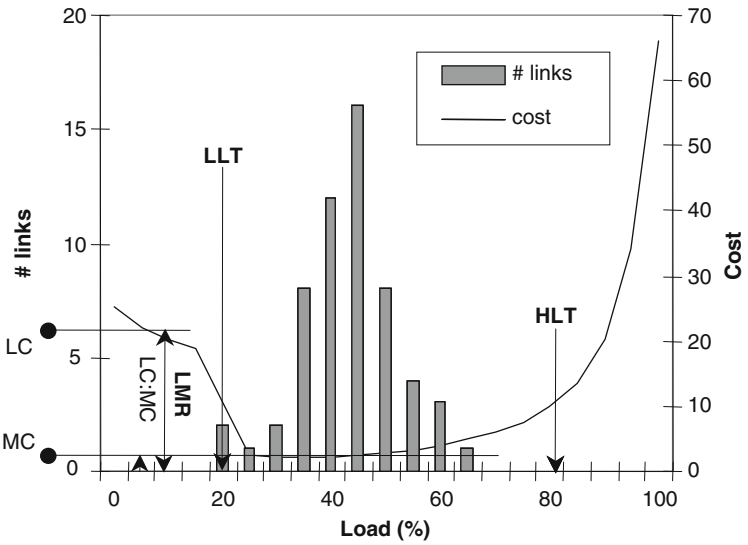


Fig. 10.4 IP/MPLS MLTE routing cost function

IP/MPLS link load. This distribution is influenced not only by MLTE parameters such as the LLT but also by the total traffic offered to the network, or rather how total traffic relates to link capacity $b_{\max}$. For example, for low traffic volumes, a very sparse (tree-like) logical topology may suffice, but this will not be the case for higher volumes.

2.4 Energy-Efficient MLTE for Daily Traffic Variations

In this section, we look at typical operator networks and note that often such networks see daily traffic variations. Although MLTE is suited for very large traffic variations, there is still merit in adapting the network to the typical daily alternation between peak and off-peak volumes seen in access as well as core networks. In [8], the rate of off-peak vs. peak volume is 50%. Depending on the type of network and devices accessing it, this rate may be 20% for WiFi networks, or even 10% for a network servicing smart phones [9]. Obviously, a large part of the background off-peak volume is always-on traffic such as P2P services. However, even on wired access networks where such services are prevalent, daily variations are increasing in recent years [8]. Peak (day) traffic largely consists of more interactive services, which need increasingly larger bandwidths (e.g., consider gaining popularity of on-demand streaming video), making for a stronger diurnal traffic oscillation.

As MLTE provides an optimized topology adapted to offered traffic demand, we assume that power requirement can be optimized accordingly for the special case of daily variations, switching links on and off along with traffic volume changes. There clearly is room for specific energy efficiency-oriented optimization by somehow modifying the MLTE strategy. For the case of optical layer resource usage, we used a so-called optical metric in [10]. This optical metric provides a cost figure for each possible LP, each corresponding to an IP/MPLS link, so the optical metric leads to a cost assignment in the IP/MPLS layer as well. It is incorporated into the MLTE strategy by multiplying it with the original MLTE routing cost function and using this resulting function instead for routing. The optical metric in [10] is a linear function of the number of optical hops.

In the study presented here, we repeat this approach for a power-based metric: we take the interface power characteristic (as shown in Fig. 10.5) and multiply it with the original routing cost function (Fig. 10.4) to reach a new cost function yielding automatic energy efficiency optimization. This new cost function incorporates a power metric (similar to the optical metric from [10]) that allows to take energy efficiency of IP/MPLS links into account. Specifically, it allows preferring those energy-efficient links over the less efficient links in routing the client traffic over the network and constructing the logical topology. The distinction between normal and energy-efficient links is determined by the idle power requirement of the link interfaces, as shown in Fig. 10.5 for two types of links: normal ones, which have an idle power P_0 of 90% of $P_{\max}$, and energy-efficient ones at 25% $P_{\max}$.

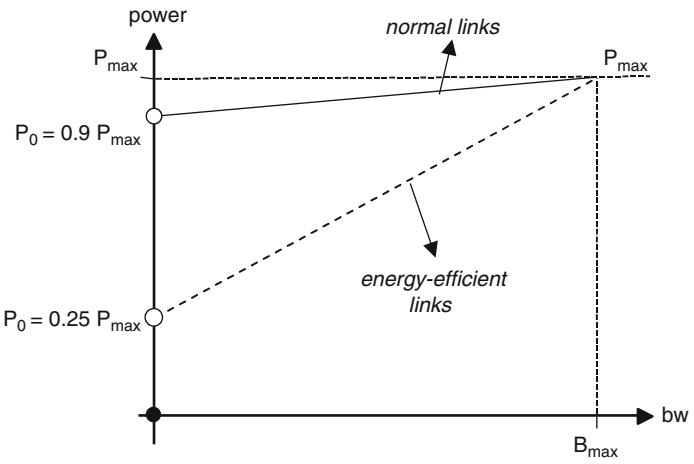


Fig. 10.5 Power characteristic for normal and energy-efficient links

Table 10.1 Peak vs. off-peak power requirements (P_{tot}/P_{max}) for three TE scenarios

TE scenario	Peak (100% traffic volume)	Off-peak (25% traffic volume)
Full mesh	56.4	47.7
"Slow" MLTE	40.0	28.0
"Fast" MLTE	40.0	13.2

To indicate the impact of MLTE optimization with a power-based metric and how it interacts with the presence of links of different energy efficiency in a network, we consider an example scenario. The sample network consists of 14 nodes, of which four have power-efficient LCs ($P_0 = 25\%$ of P_{max} as in Fig. 10.5). For the peak volume traffic pattern, source-destination traffic is uniformly distributed between 0% and 50% maximum link capacity B_{max}. To evaluate savings under daily traffic variations, we set the off-peak (i.e., night) traffic volume at 25% of the peak demand.

In a first case, we look at the impact of the MLTE optimization itself, by comparing it to full mesh routing. Furthermore, we run the MLTE optimization in any of two regimes, which have different reaction times. In Table 10.1, total network power requirement P_{tot} after routing (normalized as P_{tot}/P_{max}) is shown for both peak and off-peak demand, and this for full mesh routing and the two MLTE regimes.

The full mesh routing requires an IP/MPLS link to be set up between any two of the 14 IP/MPLS nodes; we notice some small reduction in power requirement for the off-peak volume, but since the idle power is quite substantial for most LCs, the savings are minimal (15%). It does, however, set a benchmark figure for the difference between peak and off-peak power requirements under MLTE optimization techniques.

Both MLTE regimes use the same power-based metric but have different reaction times. The first one, "slow," has logical topology updates slower than the daily variations, that is, a reaction time in terms of days. This may in fact even be some kind of dynamic grooming which acts on a large time scale (months). In any case, the general idea of such optimization is to cope with slowly increasing traffic demands over longer time periods, as seen in most networks.

For the "fast" variant, we allow MLTE actions in-between peak and off-hour periods, but this requires protocol support (e.g., GMPLS) for fast resource reservation and LP setup/teardown. The "slow" variant necessarily optimizes for peak volume (so there is always enough bandwidth available for carrying traffic), which is why power requirements are equal for "slow" and "fast" during peak hours. Still, just the ability to optimize logical topology and switch off some of the links that are required for a full mesh routing leads to a power reduction of 30% compared to the full mesh peak case.

For off-peak hours, power requirements are 45% lower (vs. full mesh off-peak) for "slow" MLTE simply because of lower traffic volume, similar to the full mesh case. More importantly, for "fast" MLTE, the MTLE optimization yields additional energy savings: drop of 70% vs. full mesh off-peak, since the "fast" reoptimization switches off unneeded links and reroutes traffic accordingly as traffic volume drops to off-peak numbers.

Next, we will ignore the full mesh scenario and instead look again at the "slow" and "fast" MLTE cases, but in this case, we will incorporate the power-based metric in the MLTE strategy. The power metric allows energy consumption optimization by taking into account the energy efficiency of links and configures optimal traffic loads for each type of LC.

In Fig. 10.6, for the same network and traffic scenario, we show peak, off-peak ("slow" MLTE), and again off-peak ("fast" MLTE) power requirements, and this for both regular MLTE (top pie charts) and MLTE with a power-based metric (bottom). For each case, total network power requirement is indicated relatively through pie-chart area; the darker/lighter part of each pie chart shows total power required by all LCs having high/low P_0, respectively.

In the "slow" MLTE scenario, power requirement of high P_0 link stays mostly the same during off-peak hours, as high P_0 suggests link power requirement to be mostly insensitive to a traffic volume reduction. In energy-efficient links, however, power requirement will lower with traffic volume; this explains why the relative portion of high P_0 (vs. low P_0) power requirement increases for off-peak traffic volume.

The power metric enabled scheme tends to avoid and shut down inefficient LCs (links), especially at off-peak ("low") volumes. Although the slow regime offers no reconfiguration at off-peak volumes, avoiding inefficient LCs leads to better (compared to regular MLTE) scaling of total power requirement with traffic volume.

Comparing "fast" regime for regular vs. power metric, a reduction of 12% in power is seen for off-peak hours (13% for the regular vs. power metric in the "slow" regime). The difference for peak traffic is minimal as no interfaces are run near their idle operating point, so differences in power efficiency are minimal also.

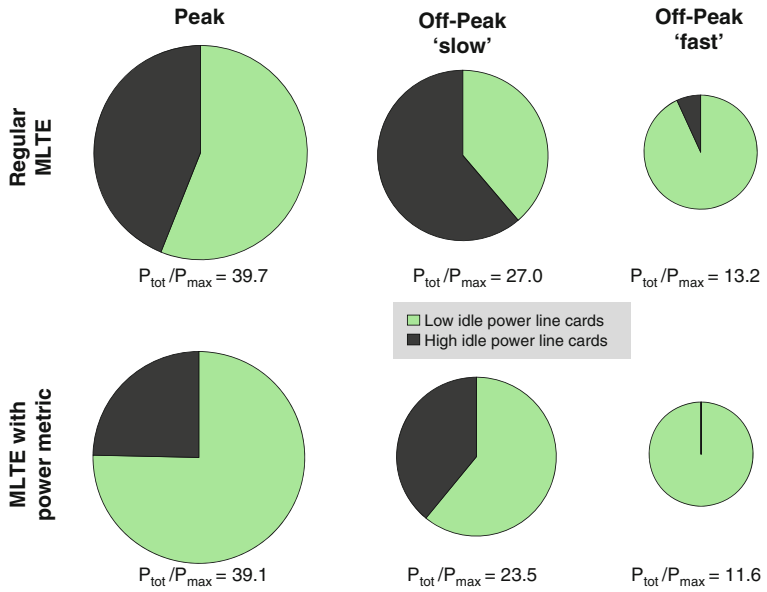


Fig. 10.6 Comparison of regular MLTE and MLTE with power metric

Energy efficiency aware MLTE serves to traffic engineer around, and shut down power inefficient parts of the network, leading to lower total energy consumption. Even for slow response regimes, per-interface power considerations lead to logical topologies scaling better with offered traffic volume.

3 Optical Bypass

3.1 Core Network Architecture

A core network consists of a relatively small number of core routers that are interconnected through WDM optical fiber links, usually in a mesh or ring topology. These core routers (also referred to as layer 3 switches) serve as an access point for the access network and route the traffic using the high-capacity WDM links to other, distant routers of the core network. A WDM fiber link carries multiple optical signals over one single fiber by employing wavelength-division multiplexing. Each wavelength, or channel, is capable of carrying, for example, 10 Gbit/s or 40 Gbit/s, with 40, 80, or more wavelengths multiplexed per fiber.

Current core network architectures are typically a mix of several layers of technologies on top of each other, such as IP-over-ATM-over-SDH, as illustrated in Fig. 10.7a. However, there is a trend to move to more homogeneous architectures where IP is routed directly over WDM links, as shown in Fig. 10.7b.

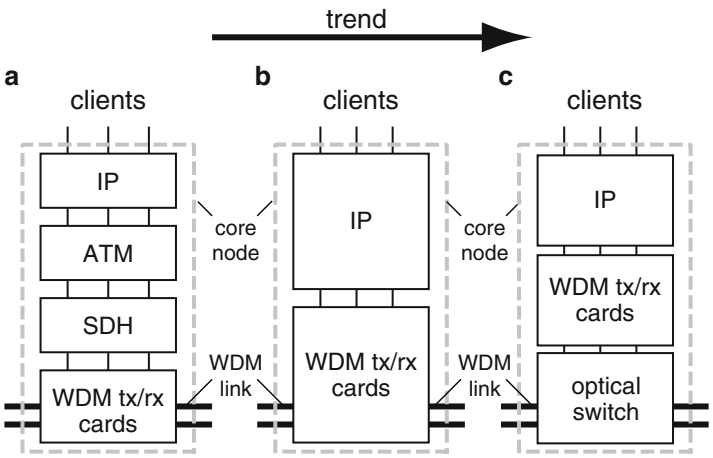


Fig. 10.7 Evolution of the core node architecture. (**a**) Example of current architecture; (**b**) IP-focused architecture; (**c**) optical bypass architecture

A technique already in use for cost reduction and router capacity offloading is optical bypass [11], as shown in Fig. 10.7c. Traffic not intended for the intermediate node remains in the optical domain and is not processed by the router. The lightpath is switched, using OADMs or OXCs from an incoming fiber link directly on the appropriate outgoing fiber link. This allows reducing the capacity of the router and the corresponding power consumption. Optical bypass is possible on single-wavelength granularity or on waveband granularity (requiring fewer ports in the OXC or OADM since multiple wavelengths are switched at the same time).

As stated, while optical bypass is already in use for cost reduction and router capacity offloading, in this study, we investigate the potential of single-wavelength optical bypass for power consumption reduction.

The node and link architecture supporting optical bypass, and that we consider for the rest of this paper, is shown in Fig. 10.1. The core node consists of an IP router equipped with a number of LCs having a certain capacity, for example, 1 Gbit/s at the client side and 10 Gbit/s at the network side. The optical WDM signals are generated using long-haul transponders (TR).

3.2 Case Study

For evaluating the power consumption and potential savings by employing optical bypass, we will consider the fictional but realistic pan-European core network shown in Fig. 10.8. It is based on the European Géant research network that has

Fig. 10.8 The logical topology of the exemplary core network

Table 10.2 Network study parameters

Parameter	Value
Network—number of nodes	34
Network—number of links	54
Network—mean node degree	3.09
Traffic matrix—mesh degree (%)	65
Protection	1 + 1
Node client-side capacity interface granularity	1 Gbit/s
Node network-side capacity interface granularity	10 Gbit/s

been modified to represent a commercial transport network; for example, each node is connected to at least two other nodes to protect against link failures (Table 10.2).

A 34-by-34 traffic demand matrix specifies the IP traffic between the nodes, with a smallest granularity of 1 Gbit/s. It represents a realistic demand scenario with randomized traffic considering higher demands for nodes located geographically nearer to each other. All traffic is assumed to be bidirectional. The demand matrix is approximately 65% meshed, that is, 65% of all potential node-to-node demands are nonzero.

Furthermore, the network will provide 1 + 1 protection, which means that each demand is continuously transmitted over two different paths at the same time; if one path fails, the traffic will still be available without interruption over the other path.

The granularity for the interfaces available in a core node differs: the access or client-side traffic connects to a core node using 1 Gbit/s interfaces, and the core network-side channel interfaces are all 10 Gbit/s interfaces.

Apart from this we will consider the following two different scenarios for calculating the power consumption:

- A *non-bypass scenario*, where all the traffic in a node—both the traffic that starts or ends in the node and the bypass traffic—will be processed by the IP router. This provides the opportunity for the IP router to groom—that is, bundle—traffic from different sources destined for the same outgoing link. This assures that optical channels can be optimally filled.
- An *optical bypass scenario*, where a dedicated optical path (channel) is set up from end node to end node. This way, the bypass traffic destined for another node will not have to be handled by the IP router and consequently not have to be converted from the optical to the electronic domain and back to the optical domain. On the other hand, if an end-to-end traffic demand is smaller than the available channel capacity, the channel will not be optimally used, resulting in a higher number of channels and equipment required.

3.3 Power Consumption Model

We will apply two different methods to calculate the power consumption of the total network and then compare the resulting values:

- The first method will be based on dimensioning the network, that is, calculating for each traffic demand the path that will be followed across all nodes, and subsequently determine the equipment required. By multiplying the equipment count with its respective power consumption, the total power will be known.
- The second alternative method is based on a simplified model or formula that takes a number of parameters and outputs the resulting total power consumption. This will be less accurate than the first method but has the advantage of being less computationally intensive.

In both cases, we will take into account an extra amount of power consumed, mainly for cooling the locations from the heat generated by the equipment. This overhead is commonly characterized by the power usage effectiveness (PUE) [12]. The PUE is the ratio of the total amount of power consumed over the useful power consumed.

We assume a typical PUE value of 2, indicating that for each watt consumed by the actual useful equipment (such as the routers and transponder line cards), an additional watt is consumed in overhead equipment. In highly optimized and efficient cooled data centers, lower PUE values are possible, but this is not yet commonplace.

3.3.1 Dimensioning-Based Power Model

The power consumption in the core network will be the sum of the various power consuming components in the different logical layers. Descending from the IP to the optical layer, we will consider the power consumption of the *IP routers*, the *WDM transponders*, the *optical amplifiers*, and for very long optical connections the optical signal *regenerators*.

To calculate the corresponding number of each type of the above equipment, every node and link in the network has to be dimensioned. Dimensioning entails the calculation of the required number of client and network interfaces at the core nodes, and the required number of channels and fibers in the links. Following from this, we can calculate the router capacity, number of transponders, amplifiers, and regenerations required. To do so, the traffic demands are routed through the network using a shortest cycle path algorithm. The shortest cycle path algorithm provides 1 + 1 protection. Taking all this into account, for each node-to-node traffic demand, a data path will be set up, likely traversing multiple intermediate links and nodes. As a result, the required traffic over each link and in each node will be known.

An overview of the relevant parameters and values that will be discussed in more detail in the remainder of this section is given in Table 10.3.

Router Power Consumption—the router power consumption is largely independent from the router load, as we have already mentioned in Section 2 above. Therefore, the total installed router capacity is a good metric to calculate the router power consumption. The number of commercially available core routers is limited. Notable manufacturers include Cisco (CRS series), Juniper Networks (T-series), and Huawei (NetEngine 5000E).

Plotting the maximum power consumption of various router configurations of the first two vendors shows an almost linear relationship between capacity and power consumption approximated by $P_R = 14$ W/Gbit/s for bidirectional communication. This implies that a 1 Gbit/s interface card will consume 14 W, and a 10 Gbit/s interface card 140 W.

Table 10.3 Power consumption model values

Element	Value (unit)
Router efficiency, P_R/C_R	14 (W/Gbit/s)
Transponder (10G, bidirectional), P_{TR}	50 (W)
Optical amplifier (bidirectional, per fiber pair), P_A	100 (W)
Optical amplifier span	80 (km)
Regenerator (bidirectional, per channel), P_{RE}	20 (W)
Regenerator span	1,500 (km)
Channels per fiber	40
Channel capacity	10 (Gbit/s)
Protection	1 + 1
Power usage effectiveness (PUE)	2

Transponder Power Consumption—a WDM transponder is required both at the sending and the receiving node. It converts the signal from a short-range, site-specific wavelength to an ITU standardized wavelength for long-haul DWDM communication and vice versa.

The power consumption of transponders is mainly dependent on the supported data rate. We do not differentiate by the modulation scheme or type of the transponders (e.g., DPSK or coherent), as it is often an implicit requirement for achieving the required data rate.

Based on a number of vendor data sheets, power consumption values for 10 Gbit/s bidirectional transponders under typical loads are around 50 W. This value includes overhead power consumption, mainly for management of the transponders. As WDM transponders are typically only required at the network side, we thus get for the power consumption per connection:

$$P^{\text{conn}}_{\text{transponders}} = 2 \cdot P_{\text{TR}}$$

Amplifier Power Consumption—a number of optical amplifiers are required for each fiber pair, as can be seen in Fig. 10.1. A booster amplifier and preamplifier amplify all signals in a fiber pair respectively upon leaving and entering a node. Furthermore, an optical inline amplifier (also referred to as optical line amplifier or OLA) inserted typically every 80 km corrects for intermediate signal attenuation. Commercially available amplifiers usually can be used for all of the above three functions.

The power consumption values found in vendor data sheets vary wildly, ranging from as low as 6 W to over 600 W. However, based on a detailed analysis, a representative value for common EDFA long span amplifiers (as opposed to more powerful Raman amplifiers) is around 100 W. This value includes a 20% overhead attributed to management cards and monitoring support systems at the amplifier sites.

The total amplifier power consumption for a physical link is thus a function of the individual amplifier power consumption P_{A}, the number of required fiber pairs N_{fp}, and the link length L_{link}:

$$P^{\text{link}}_{\text{amps}} = P_{\text{A}} \cdot N_{\text{fp}} \cdot \left(\left\lfloor \frac{L_{\text{link}}}{80\text{km}} \right\rfloor + 2 \right)$$

The number of required fiber pairs N_{fp} for a link can be calculated—assuming full wavelength conversion—based on the number of bidirectional channels required N_{c} and the number of bidirectional channel pairs N_{cfp} available per fiber pair:

$$N_{\text{fp}} = \left\lceil \frac{N_{\text{c}}}{N_{\text{cfp}}} \right\rceil$$

Regeneration Power Consumption—even with intermediate optical amplification, the length of an optical path is limited due to attenuation and increasing signal-to-noise ratio. As a result, optical paths exceeding this maximum transmission distance will require optical 3R regeneration (reamplification, reshaping, and retiming of the data pulse). The maximum transmission distance for this study is taken to be 1,500 km [13].

Regeneration has to be performed per channel (and not per fiber) and is functionally similar to two back-to-back transponders. As such, its power consumption is also dependent on the data rate. Based on a number of vendor data sheets, a typical value for bidirectional 10 Gbit/s regenerators is around 20 W. Again, this includes a (roughly) 20% management overhead.

With the number of regenerations per link depending on the number of required bidirectional channels per link N_c and the length of the bidirectional connection lightpath L_{conn}, the regeneration power consumption per bidirectional connection then becomes

$$P_{\text{regen}}^{\text{conn}} = P_{\text{RE}} \cdot N_c \cdot \left(\left\lfloor \frac{L_{\text{conn}}}{1,500\text{km}} \right\rfloor \right)$$

3.3.2 Alternative Power Model

It is also possible to estimate the network power consumption without calculating the individual equipment count. Similar to what is described in [14], this alternative model is based on a number of high-level parameters and properties of the network and provides a way to calculate the power consumption of the same components as considered above.

The total power consumption of the core network is again simply the sum of the power consumption of the constituting components:

$$P_{\text{core}} = P_{\text{routing}} + P_{\text{transponders}} + P_{\text{amps}} + P_{\text{regens}}$$

The individual power consumption is directly given by the following equations:

$$P_{\text{routing}} = \eta_{\text{PUE}} \cdot \frac{P_R}{C_R} \cdot N_C \cdot 2 \cdot D_C \cdot \left(1 + \eta_{\text{pr}} \cdot H\right)$$

$$P_{\text{transponders}} = \eta_{\text{PUE}} \cdot \frac{P_{\text{TR}}}{C_{\text{TR}}} \cdot N_C \cdot 2 \cdot D_C \cdot \eta_{\text{pr}} \cdot H$$

$$P_{\text{amps}} = \eta_{\text{PUE}} \cdot \frac{1}{f} \cdot \frac{P_A}{C_A} \cdot N_C \cdot D_C \cdot \eta_{\text{pr}} \cdot H \cdot \left(\frac{\alpha}{80\text{km}} + 2 \right)$$

$$P_{\text{regens}} = \eta_{\text{PUE}} \cdot \frac{P_{\text{RE}}}{C_{\text{RE}}} \cdot N_C \cdot D_C \cdot \eta_{\text{pr}} \cdot H \cdot \left\lfloor \frac{\alpha}{1,500\text{km}} \right\rfloor$$

Table 10.4 Symbols and values

Quantity	Symbol	Value (non-bypass)	Value (bypass)
Core router efficiency	P_R/C_R	14 W/Gbit/s	
Transponder efficiency	P_{TR}/C_{TR}	5 W/Gbit/s	
Amplifier efficiency	P_A/C_A	0.25 W/Gbit/s	
Regenerator efficiency	P_{RE}/C_{RE}	2 W/Gbit/s	
Hop count	H	3.83	1
Provisioning factor for protection	η_{pr}	2	
Provisioning factor for cooling (=PUE)	η_{PUE}	2	
Number of connections	N_C	367	
Average demand per connection	D_C	See further	
Fiber filling (% of used channels in fiber)	f	100	
Average link length	α	753 km	3.83 × 753 km

Table 10.4 gives an explanation of the symbols and parameter values.

These equations can be deduced naturally from multiplying the component efficiency with the number of connections and the amount of traffic over that connection. Only the last equation expressing the regeneration power consumption will be very crude since the average link length is not a very good indicator for the number of regenerations, especially since its value is of the same order of the generation span of 1,500 km.

The parameter values were determined as follows:

- The core router efficiency value P_R/C_R is as given in the previous power model, the transponder efficiency P_{TR}/C_{TR} has been derived by dividing the transponder consumption by the channel capacity, and the amplifier efficiency P_A/C_A by dividing the power consumption by the total fiber capacity (40 × 10 Gbit/s).
- Following the network global expectation model proposed in [15], the hop count H in a uniform network can be approximated by the following equation, with N the total number of nodes in the network and L the number of bidirectional links in the network:

$$H \cong \sqrt{\frac{N-2}{\frac{2L}{N}-1}}$$

For the non-bypass scenario, filling in $N = 34$, $L = 54$ gives $H = 3.83$. For comparison, the actual average hop count is around 4. For the non-bypass scenario, we will assume the hop count to be 1, as direct end-to-end paths will be set up.

- The number of connections is directly available from the traffic demand matrix, as well as the average demand per connection. The latter will be increasing however.
- The fiber filling is optimistically estimated to be 100%, which will be a good approximation for large demands. The average connection length is calculated

directly from the network topology. Again, it could also be estimated; the network global expectation model [15] provides an approximation based on the geographic area A covered by the network $\alpha = \sqrt{A}/(\sqrt{N} - 1)$. For an estimated area of 3,000 × 3,300 km^2, this would give $\alpha = 653$ km. As for the bypass scenario we have assumed a hop count equal to one, the average link length will have to be multiplied by the non-bypass hop count to get a reasonable representation.

3.4 Results

3.4.1 Comparison of Both Models

Figure 10.9 shows the results of applying the above two models to our exemplary European core network for an increasing traffic demand. The solid lines show the total power consumption by dimensioning the network; the dashed lines indicate the result from the alternative model.

For the non-bypass scenario, the approximation of the alternative model is reasonably accurate; for average demands higher than 10 Gbit/s (the interface), the estimate is over 93% of the actual value. A more detailed analysis shows that the deviation results from a slight underestimate of the router power consumption; the alternative model does not take into account suboptimal used interfaces.

For the bypass scenario, the approximation is much less accurate, certainly for low demands. There are three main reasons why our alternative model does not hold well for the bypass scenario:

- At traffic demands below the network interface capacity (i.e., below 10 Gbit/s), the model does not take into account the suboptimal used interfaces, thereby overestimating the router efficiency; the underestimate is worse than for the non-bypass scenario because in the latter the grooming dampens the suboptimality.
- The amplifier power consumption is estimated much lower because a higher number of fibers will be required per link (to accommodate all the extra channels) for the same amount of demand traffic; this results in an overestimate of the amplifier efficiency.
- The regeneration estimate is too low as our simplified equation is too crude.

3.4.2 Total Power Consumption

As can be seen in Fig. 10.9, the total power consumption in the network is in the order of a few megawatts, scaling linearly with increasing traffic demands. Taking the non-bypass 10 Gbit/s average demand as a convenient reference, 3 MW is roughly equal to the electrical power consumption of a nearly 6,000

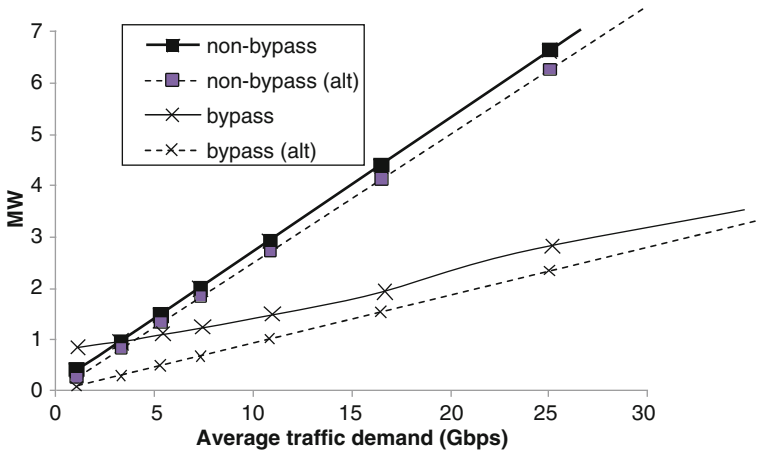


Fig. 10.9 Total power consumption of the pan-European core network as a function of the bidirectional average node traffic demand

households.[2] From a power generation perspective, this would require about four medium to large wind turbines.[3]

When comparing the power consumption of both the non-bypass and bypass scenario, we can make the following observations.

For very low demands (<3 Gbit/s), the power consumed in the bypass scenario is higher than in the non-bypass scenario. In the optical bypass scenario, a dedicated optical channel is required for each end-to-end communication. As at the network side we only have 10 Gbit/s interfaces available, for demands below 10 Gbit/s, the channels will not be optimally filled. This is less the case for the non-bypass scenario where all the traffic is "pulled" up into the IP routing layer: the router can groom all demands for the same outgoing links, thereby optimally filling the channels, saving on the number of 10 Gbit/s interfaces and subsequently power consumption.

With rising traffic demands (from around 4 Gbit/s), the bypass strategy starts to pay off, consuming less energy. The power consumption of the bypass strategy initially rises more slowly than for the non-bypass strategy. This is because the underutilized 10 Gbit/s channels can carry the additional traffic demands at almost no energy increase, whereas for the non-bypass, this is not the case. For example, consider a link which has to carry two demand streams of 4 Gbit/s. In the bypass scenario, this will require two 10 Gbit/s interfaces and two channels, for a total of 20 Gbit/s. In the non-bypass scenario, the two demands will be groomed to

[2] Assuming an average power consumption of 4,500 kWh/year per household, it takes about 5,800 households to consume 3 MW.

[3] For good measure, the average power produced by a modern wind turbine at a good site is about 30% of the peak output [16]. The peak output of a medium to large wind turbine is around 3 MW (or more), which means that in practice at least three or four wind turbines would be required.

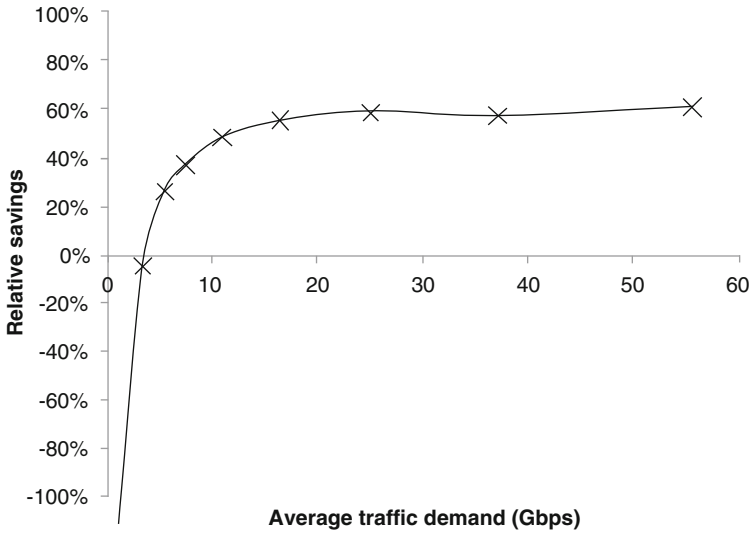


Fig. 10.10 Relative savings of the bypass scenario power consumption over the non-bypass scenario

8 Gbit/s and will require only one 10 Gbit/s router interface. Now, if both demands double to 8 Gbit/s, the bypass scenario requirements will remain unchanged (two 10 Gbit/s, and two channels), whereas in the non-bypass scenario, the groomed demands will equal 16 Gbit/s, now requiring two 10 Gbit/s interfaces and channels instead of one.

3.4.3 Relative Savings

The actual relative savings of the bypass scenario over the non-bypass scenario are shown in Fig. 10.10. As can be expected roughly when the average demand equals the channel capacity, the power savings stabilize, leveling off in this case to about 60%. Thus, by employing optical bypass, we are able to consume only 40% of the power that would be used when no optical bypass would be employed.

It is important to point out that the 60% value is no magic number. As shown in [17], the maximum energy savings achieved depend on the size of the network (in terms of nodes). For a network with similar connectivity, gains will be lower for smaller networks and higher for larger networks. This is because for larger networks, the chance of establishing longer LPs increases, bypassing more intermediate nodes, and thus saving on router interfaces.

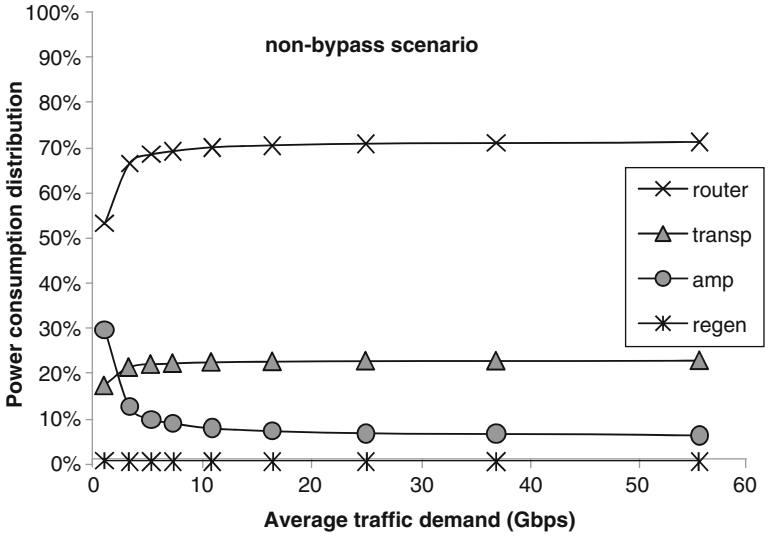


Fig. 10.11 Distribution of the power consumption

3.4.4 Distribution

Figure 10.11 shows the distribution of the power consumption among the four contributing components:

- The total power consumption is clearly dominated by the IP router, accounting for 60–70%.
- The transponder power consumption accounts for about 20%, with a slightly lower value for the bypass scenario: the longer end-to-end links only require a transponder at the end points, thereby skipping on the intermediate transponders in the bypass nodes.
- Amplifier power consumption is higher for the bypass scenario because the higher number of wavelengths results in more fiber pairs and thus more amplifiers.
- The regenerator power consumption is completely negligible in the non-bypass scenario where it accounts for only 0.3%. Obviously with longer optical paths in the bypass scenario, regeneration is more prominent, but still only <5% of the total power consumption.

Optical bypass allows to save up to 50% of the electrical energy over a non-bypass scenario in the core network. A key requirement is sufficient filling of the channel capacity. The main share of the power is consumed by the IP routers (50–70% of the total power), followed by the transponders (around 20%).

References

1. Gupta M, Singh S (2003) Greening of the internet. ACM, New York. doi:10.1145/863955.863959
2. Pickavet M, Vereecken W, Demeyer S, Audenaert P, Vermeulen B, Develder C, Colle D, Dhoedt B, Demeester P (2008) Worldwide energy needs for ICT: the rise of power-aware networking. In: Second international symposium on advanced networks and telecommunication systems, IEEE, Bombay, pp 1–3
3. Baliga J, Ayre R, Hinton K, Sorin WV, Tucker RS (2009) Energy consumption in optical IP networks. J Lightwave Technol 27:2391–2403. doi:10.1109/JLT.2008.2010142
4. Lange C, Kosiankowski D, Gerlach C, Westphal FJ, Gladisch A (2009) Energy consumption of telecommunication networks. In: 35th European conference on ECOC, optical communication, IEEE, pp 1–2
5. Vereecken W, Van Heddeghem W, Deruyck M, Puype B, Lannoo B, Joseph W, Colle D, Martens L, Demeester P (2011) Power consumption in telecommunication networks: overview and reduction strategies. Commun Mag 49(6):62–69
6. Chabarek J, Sommers J, Barford P, Estan C, Tsiang D, Wright S (2008) Power awareness in network design and routing. In: Proceedings of the 27th conference on computer communications (IEEE INFOCOM 2008), Phoenix, April 2008, pp 457–465. doi:10.1109/INFOCOM.2008.93
7. Puype B, Yan Q, Colle D, De Maesschalck S, Lievens I, Pickavet M, Demeester P (2003) Multi-layer traffic engineering in data-centric optical networks. In: Proceedings of the COST266/IST OPTIMIST workshop (ONDM 2003), Budapest, February 2003, pp 211–226
8. Cho K, Fukuda K, Esaki H, Kato A (2008) Observing slow crustal movement in residential user traffic. In: Proceedings of the 4th annual ACM international conference on emerging networking experiments and technologies (ACM CoNEXT 2008), Madrid, December 2008. doi:10.1145/1544012.1544024
9. Afanasyev M, Chen T, Voelker GM, Snoeren AC (2008) Analysis of a mixed-use urban wifi network: when metropolitan becomes Neapolitan. In: Proceedings of the 8th ACM SIGCOMM conference on internet measurement (IMC), Vouliagmeni, October 2008, pp 85–98. doi:10.1145/1452520.1452531
10. Puype B, Yan Q, De Maesschalck S, Colle D, Pickavet M, Demeester P (2004) Optical cost metrics in multi-layer traffic engineering for IP-over-optical networks. In: Proceedings of the 6th international conference on transparent optical networks (ICTON), vol 1, Wroclaw, pp 75–80. doi:10.1109/ICTON.2004.1360248
11. Simmons JM, Saleh AAM (1999) The value of optical bypass in reducing router size in gigabit networks. In: IEEE international conference on communications, Vancouver. doi:10.1109/ICC.1999.768007
12. Belady C, Rawson A, Pfleuger J, Cader T (2008) Green grid data center power efficiency metrics: PUE and DCIE (white paper)
13. Gunkel M, Leppla R, Wade M, Lord A, Schupke D, Lehmann G, Furst C, Bodamer S, Bollenz B, Haunstein H, Nakajima H, Martensson J (2006) A Cost Model for the WDM layer. In: 2006 international conference on photonics in switching, Heraklion, October 2006, pp 1–6. doi:10.1109/PS.2006.4350152
14. Kilper DC, Atkinson G, Korotky SK, Goyal S, Vetter P, Suvakovic D, Blume O (2011) Power trends in communication networks. IEEE J Sel Top Quantum Electron 17(2):275–284 doi:10.1109/JSTQE.2010.2074187. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5611565&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5611565

15. Korotky SK (2004) Network Global Expectation Model: a statistical formalism for quickly quantifying network needs and costs. J Lightwave Technol 22:703–722. doi:10.1109/JLT.2004.825756
16. Mackay D (2009) Sustainable energy—without the hot air. UIT Cambridge Ltd, Cambridge
17. Shen G, Tucker RS (2009) Energy-minimized design for IP over WDM networks. J Opt Commun Netw 1:176. doi:10.1364/JOCN.1.000176

Chapter 11
Multilayer Protection with Integrated Routing in MPLS-Over-WDM Optical Networks

Qin Zheng and Mohan Gurusamy

1 Introduction

In multiprotocol label switching (MPLS)-over-wavelength-division multiplexing (WDM) networks, the physical topology consists of optical cross-connects (OXCs) and fibers, and the logical topology consists of IP/MPLS routers and logical links (lightpaths). An example is shown in Fig. 11.1. Such multilayer networks can use either an *overlay model* or a *peer model*. In the overlay model, there are two separate control planes, and the IP/MPLS routing and signaling protocols are independent of the routing and signaling protocols of the optical layer. Lightpaths (LPs) are routed over fibers in the physical topology (using wavelengths as their resource) forming a logical topology, and label-switched paths (LSPs) are routed over LPs (logical links) in the logical topology (using bandwidth on LPs as their resource). Here, a routing node is comprised of an OXC and a label-switched router (LSR). A LP starts at an LSR and traverses one or more OXCs and terminates at an LSR. An LSP traverses one or more LPs.

In the peer model, a single instance of the control plane spans an administrative domain consisting of the optical and IP domains. Thus, OXCs are treated just like any other routers (IP/MPLS routers and OXCs act as peers), and there is only a single instance of routing and signaling protocols for the network. To obtain the topology and resource usage information, one possibility is to run an OSPF-like protocol on both routers and OXCs to distribute both link-state and resource usage information to all network elements. The topology perceived by the network nodes

Q. Zheng
Computing Science (CS) Department, Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore
e-mail: qinzheng.sg@gmail.com

M. Gurusamy (✉)
Electrical and Computer Engineering Department, National University of Singapore, Singapore 117576, Singapore
e-mail: elegm@nus.edu.sg

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*, Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_11,

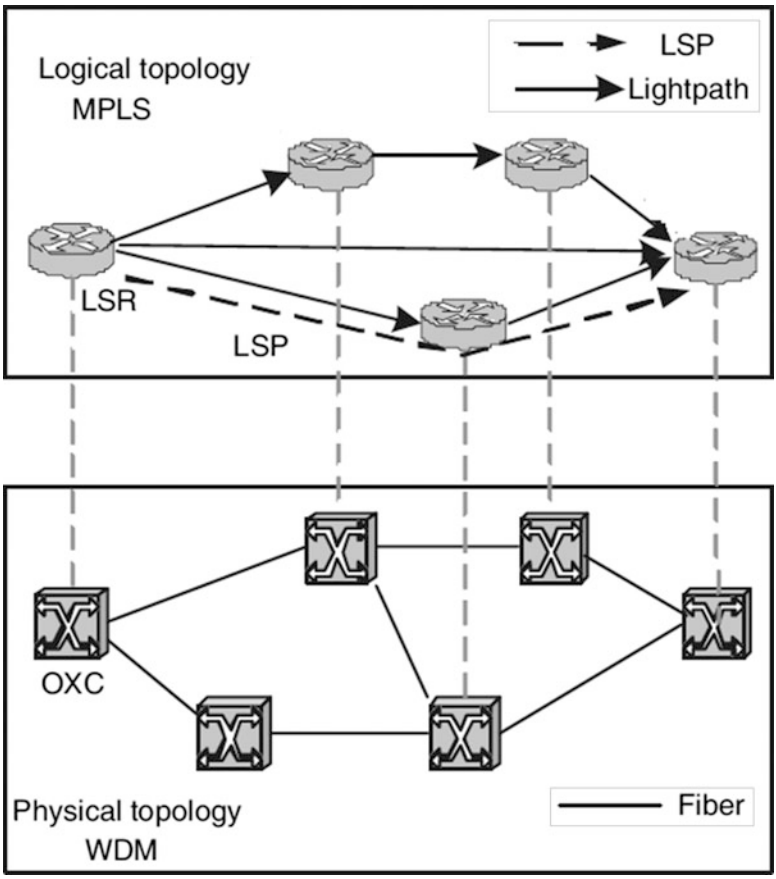


Fig. 11.1 An MPLS-over-WDM optical network with an LSP routed over lightpaths

is the integrated IP/WDM topology wherein wavelength channels (wavelengths on fibers) and LPs coexist. The topology contains complete information about the wavelength usage on fibers and bandwidth usage on logical links for efficient management and usage of the network resources. A standard interface can be defined permitting routers to exchange information and to dynamically request wavelength paths (LPs) from the optical network (http://www:sycamorenet:com=solutions=technology=frameodsi:html) [3].

Survivability is a critical problem in MPLS-over-WDM networks as a fiber cut can cause failure of a number of LPs, each carrying a number of LSPs. Single-fiber failure model is assumed in most of the survivability solutions. An LSP can be protected at the LP level where each LP traversed by it is protected by a fiber-disjoint LP. It is called as *lightpath protection*. An LSP can also be protected at the LSP level by a fiber-disjoint backup LSP wherein the LSPs traverse unprotected LPs. It is called as *LSP protection*. Figure 11.2 illustrates the two protection approaches.

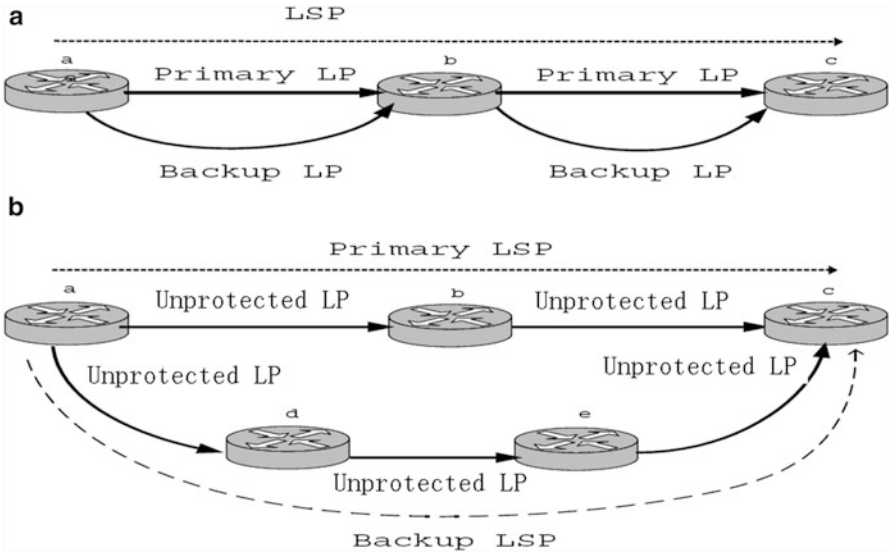


Fig. 11.2 Protection in an MPLS-over-WDM optical network: (**a**) lightpath protection and (**b**) LSP protection

We note that in the LP protection there is no explicit backup LSP since the LSP traverses protected LPs. The attractiveness of the LP protection is that it guarantees fast recovery within a few tens of milliseconds [1]. However, in this scheme, the primary and backup capacity are at a large granularity level and are isolated leading to poor resource usage. By "isolation," we mean that a primary LP carries only working traffic and a backup LP is designated to carry only protection traffic.

In the LSP protection, backup LSPs are established together with working LSPs when LSP requests are honored. This scheme has better resource efficiency as the LPs are not distinguished as primary and backup LPs, and primary LSPs (of some connections) as well as backup LSPs (of some other connections) could be multiplexed onto a LP, leading to better utilization of LP capacity. When a fault occurs, the MPLS layer can detect it using its own detection mechanism such as exchange of "Hello" messages [2]. The detection time is usually large, and it could be reduced by increasing the frequency of Hello messages at the expense of increased bandwidth overhead. Alternatively, the lower layer (optical layer) can detect the failure and propagate it to the MPLS layer through signaling messages [2]. The LSR, upon detecting the fault, needs to notify the source LSR to switch the affected traffic, for all the affected LSPs. Note that the traffic on all the failed LSPs needs to be rerouted and the number of failed LSPs is much larger when compared to the number of failed LPs. Therefore, the number of notification messages generated is quite high. All these result in longer recovery time for the LSP protection.

In the remainder of this chapter, we first present some key issues in multilayer MPLS-over-WDM networks. Then, we present two integrated algorithms [3] for the LSP protection. Next, we explain two multilayer protection schemes with inter-level resource sharing [4]. Finally, we present an LSP partial spatial protection (PSP) scheme [5].

2 Key Issues in Multilayer MPLS-Over-WDM Networks

2.1 Integrated Routing

We now describe how a dynamically arriving LSP request with the specified source, destination, and bandwidth can be routed. An LSP is routed on either existing LPs or fibers (leading to the creation of new LPs). An LSP can also be routed sequentially in the logical topology first before it is routed in the physical topology [6]. The sequential routing approach first tries to route requests over the residual capacity on existing logical links. Otherwise, it requires a new LP to be created between the ingress and egress routers. As a result, the path found traverses either existing logical links or a sequence of wavelength channels on fibers. An LSP can also be routed over LPs and wavelength channels in an integrated manner. In one such integrated routing algorithm [7], the objective is to choose the LSP that minimizes interference with future requests. This is achieved by identifying the critical links in the network, using the max-flow–min-cut principle. By choosing the shortest path with the least cost in terms of criticality, the route determined is least likely to interfere with future requests.

We use Fig. 11.3 to illustrate the network modeling for integrated routing [7]. Figure 11.3a shows a physical network with four nodes A, B, C, and D and two wavelengths w_1 and w_2 per fiber. Each layer in the graph shown in Fig. 11.3b corresponds to a wavelength. A node on a wavelength layer is referred to as a wavelength node. It is connected to its corresponding routing node (representing the LSR) through optical–electrical–optical (OEO) edges which represents OEO conversions. Initially, the topology at each layer resembles the physical network. Whenever a new LP is set up on some wavelength i, the corresponding wavelength channels on layer i are deleted. Lightpaths are modeled using cut-through arcs that replace the wavelength channels traversed. Suppose that the wavelength capacity is c units and a connection request with bandwidth b units requires a LP to be created on wavelength w_1 between nodes B and D. As a result, a cut-through arc with residual bandwidth c–b will be set up replacing wavelength w_1 on fibers between nodes B and D on wavelength layer w_1. Future requests may be routed on this arc. The wavelength channels on an arc will be restored when the arc (lightpath) is torn-down. The topology of the graph is dynamic which changes with each accepted or released request. Such a model enables direct application of the Dijkstra's algorithm on the network graph for

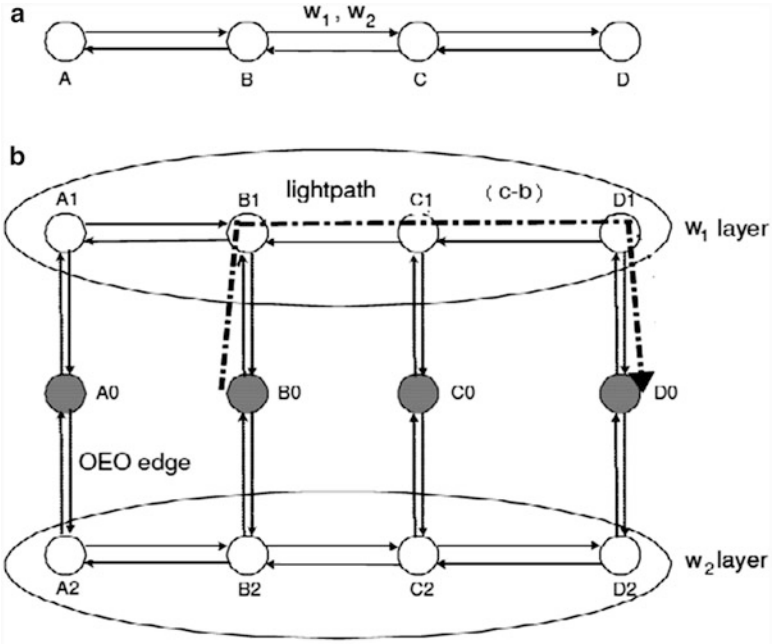


Fig. 11.3 (**a**) A physical network and (**b**) layered graph modeling for integrated routing

integrated routing where an arriving bandwidth request can either be routed over existing logical links or by setting up new LPs on fibers or use a mixture of them.

Integrated routing of LSP connections without survivability requirements has been studied in the literature. Routing algorithms providing service differentiation between classes of high- and normal-priority traffic can be found in [8]. The quality of service (QoS) delay requirements are assumed to be translated into bandwidth and OEO conversion requirements. A threshold-based routing algorithm was developed, which admits high-priority LSPs in preference over normal-priority LSPs and satisfies the bandwidth and OEO constraint requirements. It was proved in [9] that the constraint imposed by IP subnets transforms the problem of finding the shortest integrated IP hop path into an NP-hard problem. Two integrated routing algorithms were developed to select the shortest path in the presence of subnets. Different network models (overlay, augmented, and peer IP/WDM network models) were studied for dynamic LSP provisioning in [10]. In the augmented model, summarized capacity information from the WDM layer is used along with the IP layer information. A routing algorithm was developed considering both the number of wavelengths available in the WDM layer and ports in the IP layer. An integrated routing and grooming algorithm was developed in [11] where an enhanced blocking island graph network model was introduced.

The above works consider integrated routing of LSPs without taking into account the survivability requirements. Integrated routing of restorable connections

was considered in [3], and two integrated routing algorithms were developed to route the primary LSP and backup LSP. We will present the algorithms in Sect. 11.3.

2.2 Multilayer Protection

Failures can be handled at multiple layers in IP/MPLS-over-WDM networks. Failures can be handled through proactive protection where backup resources are allocated when connection requests are honored. They can also be handled through reactive restoration where backup paths are found dynamically upon failures. A few multilayer restoration approaches were studied in [12–14]. Resilience in a multilayer network with ATM and SONET layers was studied [12]. Guidelines are given as to which layer should be responsible for each failure and how to plan spare capacity among multiple layers. To account for both router and link failures, the logical IP topology is reconfigured to work around IP router failures, and optical link or node failures are recovered in the optical layer using an appropriate recovery mechanism [13]. Two schemes—intelligent optical networks (ION) local reconfiguration and ION global reconfiguration—were developed based on the number of logical link reconfigurations. A sequential two-layer recovery scheme was developed in [14] where the optical layer takes recovery action first, and subsequently the upper IP layer initiates its own recovery mechanism if the optical layer does not restore all affected traffic.

A few multilayer protection approaches were studied [2, 15, 16]. A number of issues were discussed such as how to handle the protection responsibility between the optical and client (MPLS) layers and how recovery actions at these layers can be coordinated. In particular, a mechanism was proposed which imposes a hold-off timer in the client layer whereby the client layer recovery mechanism is delayed for a certain period before it is invoked, giving the optical layer sufficient time to complete its recovery [15, 16]. There is also an approach which considers static traffic demands with the objective of capacity planning [2]. Here, the LP protection and the LSP protection are provided against link failure and router failure, respectively.

There have been approaches which provide protection either at the LP level or at the LSP level [3, 4]. Here, dynamic traffic with online routing has been considered, and a key objective is to optimize the blocking performance. We will present these approaches in Sect. 11.4.

2.3 Partial Protection

Providing several quality-of-protection (QoP) classes has been studied in the literature. Partial protection approaches provide a range of *protection grades*,

and they can be classified into three categories: (a) partial traffic protection, (b) partial temporal protection, and (c) PSP. In partial traffic protection, the percentage of working traffic to be protected depends on the specified protection grade. Such a partial traffic protection has been considered in [17, 18]. Partial temporal protection has been considered in [19] where protection bandwidth can be shared with some working paths, which allows the connection to be unprotected during some periods of time. In PSP, a connection is unprotected by its backup path against some fiber failures along its primary path. These fibers are called *unprotected fibers*, and when they fail backup paths are not guaranteed. Here, each fiber has a certain failure probability, and the accumulative failure probability of unprotected fibers must not exceed a value as required by the protection grade. The differentiated reliability (DiR) problem studied in [20] belongs to this category. It studied off-line LP routing of static traffic and considered LP-level partial protection wherein some fibers along the primary LP are not protected by the corresponding backup LP.

The LSP-level PSP has been considered in [5] where certain fibers along the working LSP are not protected by the corresponding backup LSP. As the primary LSP of a connection can traverse one or more LPs, providing partial protection at the LSP level makes it more efficient to meet the specified protection grade of an end-to-end connection. It differs from the shared LSP protection in that unprotected fibers can be selected before and/or after routing the backup LSP. The challenge is to find the backup LSP and unprotected fibers that minimize the total bandwidth required while guaranteeing the protection grade. We will discuss this approach in Sect. 11.5.

3 Integrated Routing Algorithms

In this section, we explain how integrated routing can be used to route restorable connections at LSP level in MPLS-over-WDM networks. Consider a new LSP request $\langle s, d, b \rangle$ between nodes s and d with a specific bandwidth requirement of b units. A primary LSP and a fiber-disjoint backup LSP need to be selected. The routing algorithms basically model the network as a graph, assign weights to different edges in the graph, and use a shortest-path selection algorithm such as Dijkstra's algorithm to choose the primary and backup LSPs. Based on the cost metric such as number of hops and amount of bandwidth, they determine whether to route a connection request on existing LPs, open new LPs, or use some existing LPs and create additional ones. The path is set up by updating the residual bandwidth of the existing logical links and by setting the residual bandwidth to $(c–b)$ units for the new logical links. If there is not enough bandwidth available for either the primary or the backup path, the request is blocked.

3.1 LSP Protection with Resource Sharing

Once a primary LSP is chosen for the current LSP request, a backup LSP needs to be selected. Two backup LSPs can share some backup bandwidth on a logical link if their primary LSPs do not fail simultaneously. This guarantees that all the failed working traffic will be restored upon any single-fiber failure in the network. Because of sharing, the additional amount of bandwidth needed on an existing logical link to accommodate the current backup LSP could be less than the demand b. We explain below how the *additional bandwidth* required can be computed.

For every logical link, a table is maintained to record the amount of (backup) bandwidth required for each fiber in the network. Since the single-fiber failure model is assumed, the amount of backup capacity needed on this logical link is the maximum among all the values in the table. Accommodating the current backup LSP would require an update on the table entries that correspond to the fibers traversed by the primary LSP. For each fiber traversed by the primary LSP, the corresponding entry in the table associated with the logical link is increased by b. The additional bandwidth needed on the logical link is the amount by which the maximum value is increased. The basic idea is to ensure that, in the event of any fiber failure, this logical link is guaranteed to restore all the traffic on the failed LSPs whose backup LSPs use this logical link.

3.2 Routing Algorithms

We present two integrated routing algorithms: (physical) hop-based integrated routing algorithm (HIRA) and bandwidth-based integrated routing algorithm (BIRA) [3]. The objective of HIRA is to minimize the number of physical hops used by the LSP. By doing so, it attempts to minimize the resource usage. Note that a fiber traverses exactly one physical hop and a logical link may traverse one or more physical hops. The objective of BIRA is to minimize the amount of bandwidth required by the LSP. For the backup LSP, it takes into account backup sharing and selects logical links based on the amount of additional bandwidth required on them.

Now, we explain how link weights are assigned in these two algorithms. For the primary LSP, in HIRA, a wavelength on a fiber and a logical link with residual capacity of at least b are assigned weights based on the physical hops. The other links are assigned infinite weight. In BIRA, a wavelength on a fiber and a logical link with residual capacity of at least b are assigned weights, which are b times the physical hops traversed by them. The other links are assigned infinite weight. The Dijkstra's algorithm is then used to choose a minimum cost path.

For the backup LSP, in HIRA, the weights are assigned in a similar way except that a fiber or a logical link will be assigned finite weights if they do not traverse any fiber traversed by the primary LSP. Further, a logical link must have residual capacity of at least the additional bandwidth required. In BIRA, a fiber or a

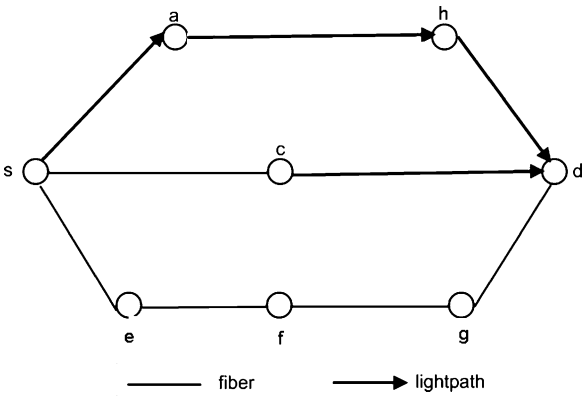


Fig. 11.4 An illustration of integrated routing algorithms

logical link will be assigned weights if they do not traverse any fiber traversed by the primary LSP. Further, a logical link must have residual capacity of at least the additional bandwidth required. While a wavelength on a fiber is still assigned weight *b*, a logical link is assigned weight that is the additional bandwidth times the physical hops traversed by it. The other links are assigned infinite weight.

We use an example to illustrate HIRA and BIRA. Figure 11.4 shows a network graph where LPs and fibers coexist. Recall that in traditional approaches, an LSP is routed on either existing LPs or fibers (leading to the creation of new LPs). For the sequential routing approach [6], it first tries to route requests over the residual capacity on existing logical links (s–a–h–d). Otherwise, it requires a new LP to be created between the ingress and egress routers (s–e–f–g–d). As a result, the path found traverses either existing logical links or a sequence of wavelength channels on fibers. With integrated routing, the LSP can be routed on both LPs and fibers (s–c–d). For the routing of the primary and backup LSPs in HIRA as well as the routing of the primary LSP in BIRA, the path traversing the minimum number of physical hops will be chosen. Suppose that all logical links have residual bandwidth of at least b and traverse two fibers. In this case, the path s–c–d will be chosen, among the three paths. For the routing of the backup LSP in BIRA, further suppose that b is 3 and the additional bandwidth required on logical links s–a, a–h, h–d is 0, 1, and 2, respectively. Also, the remaining two paths are fiber disjoint with the primary LSP chosen. Then, BIRA will select the path s–a–h–d as the backup LSP because it requires less bandwidth ($2 \times 0 + 2 \times 1 + 2 \times 2$) compared to the path s–e–f–g–d (3 + 3+ 3+ 3).

3.3 Performance Comparison

From simulation results, it has been observed in [3] that in terms of blocking probability, the shared LSP protection performs better than the shared LP protection and the integrated routing algorithms outperform the sequential routing algorithm.

Further, BIRA performs better than HIRA when the load increases as BIRA minimizes the additional bandwidth required along the backup LSP to accommodate each connection request.

4 Multilayer Protection and Inter-level Sharing

It is to be noted that IP traffic may have differentiated protection requirements on service recovery time. They can be classified into high-priority traffic and low-priority traffic based on their requirements. We recall that the LP-level protection ensures faster recovery when compared to the LSP-level protection. Providing protection at these two levels for differentiated traffic has been studied in [4]. It develops the multilayer protection schemes to protect high-priority traffic at the LP level and low-priority traffic at the LSP level. The objective is to satisfy the protection requirements of the LSPs and at the same time utilize network resources efficiently. Further, dividing the protection responsibility between the LP and LSP levels results in reduced number of recovery actions at each level when faults occur.

Consider a request specified as $\langle s, d, b, \text{pl}\rangle$ where s is the source node, d is the destination node, b is the amount of bandwidth, and pl is the specified protection level ($\text{pl} = 0$ corresponds to high-priority traffic which is provided LP-level protection, and $\text{pl} = 1$ corresponds to low-priority traffic which is provided LSP-level protection). To satisfy a high-priority connection request, a primary path traversing a sequence of LPs, each of which is protected by a fiber-disjoint backup LP, needs to be chosen. To satisfy a low-priority connection request, a physical fiber-disjoint pair of primary LSP and backup LSP needs to be chosen. The objective is to minimize blocking probability and satisfy the recovery time requirements of connections. There are three kinds of LPs existing in the network: primary LPs and backup LPs to route high-priority traffic and unprotected LPs to route low-priority primary and backup LSPs.

Two multilayer protection schemes have been developed in [4]. In both schemes, for the LSP protection, primary and backup LSPs are routed over unprotected LPs and backup resources (bandwidth on unprotected LPs) are shared among backup LSPs. For the LP protection, in one scheme, the shared LP protection is used where backup resources (wavelengths on fiber links) can be shared among backup LPs. This scheme is called multilayer protection with backup lightpath sharing (MLP-LS). In the other scheme, backup LPs are preconfigured, and thus resource sharing among them is not allowed among such dedicated protection paths. This scheme is called multilayer protection with no backup lightpath sharing (MLP-NLS).

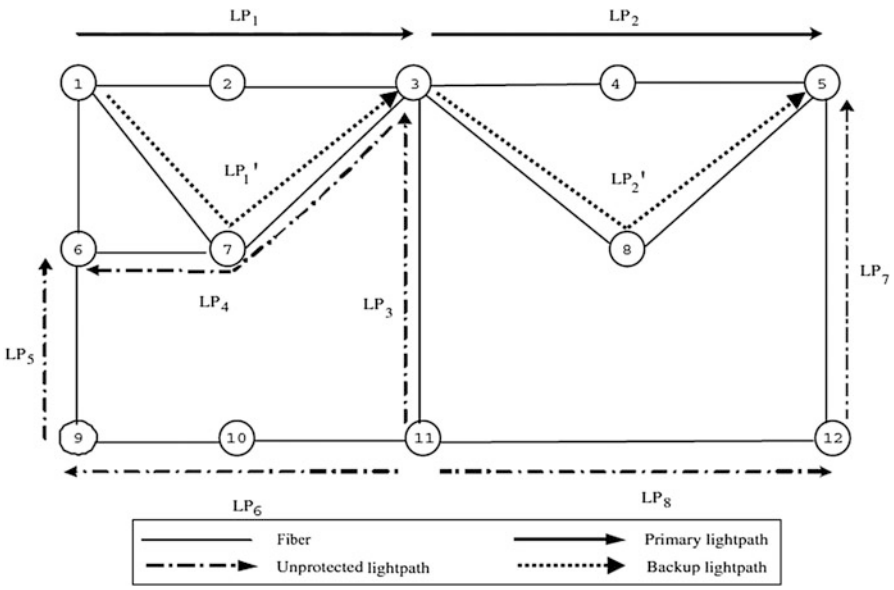


Fig. 11.5 An illustration of multilayer protection and inter-level sharing

4.1 Inter-level Sharing

In order to improve resource efficiency in MLP-NLS, *inter-level sharing* (ILS) method can be used. In ILS, a preconfigured backup LP can be used by a backup LSP if it and its corresponding primary LP are fiber disjoint with the selected primary LSP. This is to guarantee that recovery actions at the two levels will not interfere with each other upon a link failure, i.e., compete on the bandwidth resource on the backup LP. The objective of ILS is to improve the usage of dedicated backup LPs and hence save wavelength resources (by routing backup LSPs on backup LPs instead of creating new LPs).

Now, we illustrate the ILS method using an example [4]. In Fig. 11.5, there is a high-priority connection from node 1 to node 5. Its primary path (say LSP_1) traverses lightpaths LP_1 and LP_2 and is protected at the LP level. There is a low-priority connection from node 11 to node 6. Its primary LSP (say LSP_2) traverses LP_6 and LP_5 and is protected at the LSP level by a backup LSP (say LSP'_2) that traverses LP_3 and LP_4. Suppose that there is another low-priority connection which is from node 11 to node 5. Its primary LSP (say LSP_3) traverses LP_8 and LP_7. Note that a backup LSP for LSP_3 cannot be routed only through unprotected LPs and at least one additional LP needs to be created. However, using ILS, LSP_3 can be protected by routing a backup LSP (say LSP'_3) over LP_3 and a preconfigured backup

lightpath LP'_2. Here, ILS is permissible because LSP_3 (LPs traversed by LSP_3) and LP_2 traverse disjoint set of fibers implying that they do not fail simultaneously when a single-fiber model is assumed. We also note that the primary LSP and its backup LSP traverse disjoint sets of fibers. We can observe that ILS is able to increase the utilization of dedicated backup LPs and save wavelength resources.

4.2 Routing Algorithms

In the following, we present the routing algorithms for the LP protection and the LSP protection [4]. Integrated routing algorithms are used to select paths. For a high-priority connection, an LSP that traverses the minimum number of OEO conversions is chosen. Recall that a high-priority connection is protected at the LP level. The LSP can be routed on primary LPs and wavelength channels on fibers. Each OEO edge is assigned a sufficiently large weight. Wavelength channels are assigned very small weights. Primary LPs with at least b residual capacity are assigned very small weights. Otherwise, an infinite weight is assigned. An infinite weight is also assigned to unprotected LPs and reserved wavelength channels for backup LPs (as in MLP-LS) and preconfigured backup LPs (as in MLP-NLS). We note that the routing algorithm prefers to choose LPs and wavelength channels compared to OEO edges. A shortest-path algorithm can then be used to compute the minimum cost path as the primary LSP. When a new primary LP is required to be created (over wavelength channels), a physical hop-based routing algorithm is used to compute a fiber-disjoint backup LP with appropriate sharing.

For a low-priority connection, the primary LSP and backup LSP can be routed on unprotected LPs and wavelength channels on fibers. Recall that a low-priority connection is protected at the LSP level. In the selection of primary LSPs and backup LSPs, unprotected LPs are preferred than wavelength channels. The objective is to improve resource usage and sharing efficiency on LPs and save wavelength channels. This is done by assigning different weights on a wavelength channel and each physical hop of an unprotected LP. Specifically, each OEO edge is assigned a very small weight. Primary LPs are assigned infinite weights. Wavelength channels are assigned a sufficiently large weight. Unprotected LPs are assigned a small weight if the residual capacity is at least b in the primary LSP section and at least the additional bandwidth in the backup LSP selection (and infinity otherwise). Reserved wavelength channels (for backup LPs in the shared LP protection) are assigned infinite weight. Preconfigured backup LPs are assigned infinite weight in the primary LSP selection and a sufficiently large weight or infinite in the backup LSP selection depending on whether ILS is allowed. A shortest-path selection algorithm can then be used to choose the minimum cost path.

4.3 Performance Comparison

We compare the performance of the two multilayer protection schemes (MLP-LS and MLP-NLS) with the shared LP protection and the dedicated LP protection, respectively. It can be observed from the simulation experiments in [4] that the two multilayer protection schemes perform better than the LP protection in terms of blocking probability. Further, ILS can reduce blocking considerably by allowing backup LSPs to use dedicated backup LPs in MLP-NLS.

5 LSP Partial Spatial Protection

In the previous section, we considered traffic requests with differentiated protection requirements on service recovery time and discussed how they can be satisfied through protection at an appropriate level. In this section, we consider requests with differentiated protection grade requirements and protect them at the LSP level. Consider a connection request $\langle s, d, b, \text{pg}\rangle$ with bandwidth demand b from node s to node d and a protection grade pg ($0 \leq \text{pg} \leq 1$). Protection grade is the probability that a connection can be restored upon faults. The objective of the LSP PSP is to find a primary LSP, a backup LSP, and a set of unprotected fibers on the primary LSP satisfying the required protection grade, so that the total bandwidth required for the current connection is minimized. Recall that each fiber has a certain failure probability and the accumulative failure probability of unprotected fibers must not exceed a value as required by the protection grade.

5.1 An Illustrative Example

We use an example to illustrate the LSP PSP [5]. Figure 11.6a shows the logical topology of a network consisting of six nodes and seven bidirectional logical links (solid arrows). For simplicity, we assume that each logical link traverses only one fiber, which has the same index as the logical link that traverses it. All fibers are assumed to have the same failure probability, and thus the conditional failure probability (if a failure happens in the network) of each fiber is 1/7. The capacity of a logical link is assumed to be 10 units. Currently, there are two connections (LSP1 and LSP2) with 100% protection grades. We focus on logical link 7 with four units of reserved working bandwidth, three units of reserved backup bandwidth, and three units of residual bandwidth.

Now, we consider a new request (LSP3) which is from E to A with two units of bandwidth and 85% protection grade, as shown in Fig. 11.6b. Therefore, only one fiber can be selected as unprotected. The shared LSP protection can find a primary LSP and a backup LSP, as shown in the figure. As the primary LSP3 and primary

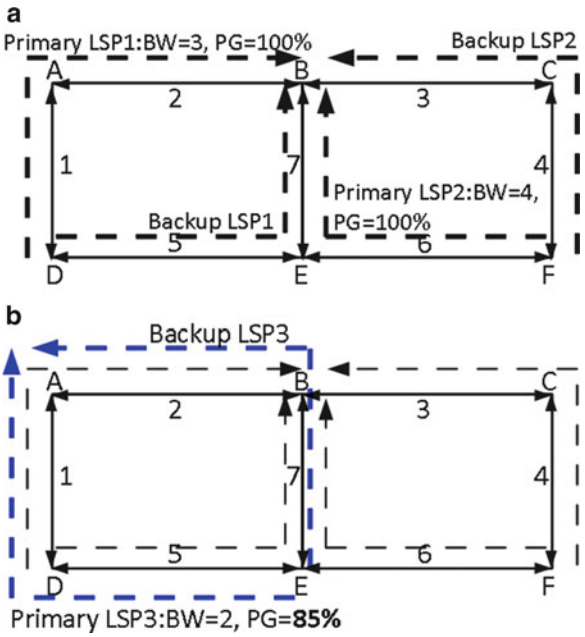


Fig. 11.6 An illustration of the LSP partial spatial protection: (**a**) two connections (LSP1 and LSP2) with 100% protection grades and (**b**) a new request (LSP3) with 85% protection grade

LSP1 traverse a common fiber 1 (contained by logical link 1), their backup LSPs cannot share backup bandwidth. As a result, the shared LSP protection requires $3 + 2 = 5$ units of backup bandwidth on logical link 7. However, with PSP, fiber 1 can be selected as unprotected, and thus three units of backup bandwidth on logical link 7 are sufficient. Next, we consider the case where the bandwidth requirement of the new request is six units instead of two units. In this case, logical link 7 is not eligible (no sufficient capacity available) for the backup LSP routing as it does not have sufficient bandwidth. Note that the additional bandwidth required is six units while its residual capacity is six units. As a result, the shared LSP protection is not able to use the backup LSP3. However, PSP can still use this backup LSP because selecting fiber 1 as unprotected would result in the case that logical link 7 becomes eligible (sufficient capacity available). Note that now the additional bandwidth required is three units (to protect fiber 5) and in the end six units of backup bandwidth in total are reserved on logical link 7.

5.2 Integer Linear Programming Formulation

In this section, we present an ILP formulation that can optimally solve the LSP–PSP problem [5]. In the following, we use (i, j) to denote a logical link and (u, v) to denote a fiber. Let E and E_p denote the set of logical links and fibers, respectively. The following binary decision variables are used in the ILP:

- $x_{i,j}$ 1 if logical link (i, j) is used in the primary LSP
- $y_{i,j}$ 1 if logical link (i, j) is used in the backup LSP
- $\alpha_{u,v}$ 1 if fiber (u, v) is traversed by the primary LSP
- $\gamma_{u,v}$ 1 if fiber (u, v) is traversed by the primary LSP and is selected as an unprotected fiber

The objective is to minimize:

$$b \sum_{(i,j)\in E} x_{i,j} + \sum_{(i,j)\in E} c_{i,j}.$$

The first term in the objective function indicates that the amount of bandwidth reserved for every logical link on the primary LSP is b. The second term indicates that the amount of bandwidth reserved for every logical link on the backup LSP is the additional bandwidth required, denoted as $c_{i,j}$.

The detailed discussion of the constraints can be found in [5]. They include the flow conservation for the primary LSP and the backup LSP of the current connection. The following two constraints imply that if fiber (u, v) is traversed by an LSP, then any logical link containing fiber (u, v) may be used in this LSP; otherwise, any logical link containing fiber (u, v) cannot be used. Clearly, a logical link can be used in an LSP if and only if all of its fibers are traversed by this LSP. We need to have $|E_P|$ as an LSP can traverse a number of logical links containing the same fiber:

$$h_{u,v}x \leq |E_P|\alpha_{u,v} \ \forall (u, v) \in E_P,$$

$$h_{u,v}y \leq |E_P|\beta_{u,v} \ \forall (u, v) \in E_P.$$

The constraint below specifies that unprotected fibers can only be selected from fibers traversed by the primary LSP and they must satisfy the protection grade pg. Here, $P^F_{u,v}$ denotes the conditional failure probability of (u, v) given a fiber failure in the network. This probability can be derived from normalized fiber downtime ratio and the length of fibers in the network [20]:

$$\gamma_{u,v} < \alpha_{u,v} \ \forall (u, v) \in E_P, \quad \sum_{(u,v)\in E_P} \gamma_{(u,v)} P^F_{u,v} < 1 - pg.$$

The constraint below specifies that the primary and backup LSP must not traverse common fibers, unless these common fibers are unprotected fibers. This is because although the connection cannot be restored in the event of the failure of any of these fibers, the failure is allowed to be unprotected as permitted by the protection grade:

$$\alpha_{u,v} + \beta_{u,v} - \gamma_{u,v} \leq 1 \forall (u, v) \in E_P.$$

In the following, $\theta_{i,j}^{u,v}$ is used to reflect backup bandwidth sharing on logical links. It is the cost of using logical link (i, j) in the backup LSP, given that fiber (u, v) is traversed by the primary LSP. The cost is infinity if either (i, j) traverses (u, v) (which corresponds to the case that the primary and backup LSPs are not fiber disjoint) or (i, j) does not have sufficient bandwidth to protect (u, v). The latter can be determined by observing the entry corresponding to (u, v) in the table associated with (i, j). These logical links are not *eligible* for the backup LSP routing, and we denote them as *noneligible* logical links. The value of $\theta_{i,j}^{u,v}$ is 0 if (u, v) can be protected without increasing the backup bandwidth on (i, j). Otherwise, $\theta_{i,j}^{u,v}$ is the amount by which the backup bandwidth on (i, j) needs to be increased to protect (u, v). Specifically, it is the difference between the value in the entry corresponding to $(u, v) + b$ and the backup bandwidth reserved on (i, j).

The term $\alpha_{u,v} + y_{i,j} - \gamma_{u,v} - 1$ equals to 1 only in the case that $\alpha_{u,v}$ and $y_{i,j}$ are set to 1, while $\gamma_{u,v}$ is set to 0. It corresponds to the case that logical link (i, j) is used by the backup LSP and fiber (u, v) is traversed by the primary LSP, but it is not an unprotected fiber. In all other cases, the quantity inside the bracket is 0 or negative, and thus the constraint is nonbinding. This equation indicates that the additional bandwidth $c_{i,j}$ must be at least the largest possible value of $\theta_{i,j}^{u,v}$ among all fibers traversed by the primary LSP that are not unprotected. To minimize the objective function, $c_{i,j}$ is set to be equal to the highest $\theta_{i,j}^{u,v}$ value:

$$c_{i,j} \geq 0 \ \forall (i,j) \in E,$$

$$c_{i,j} \geq \theta_{i,j}^{u,v}(\alpha_{u,v} + y_{i,j} - \gamma_{u,v} - 1), \quad \forall (i,j) \in E, \quad \forall (u,v) \in E_P.$$

5.3 Routing Algorithms

In this section, we describe the problem of routing backup LSPs. The objective is to find a backup LSP with the minimum amount of bandwidth requirement. This algorithm in [5] was developed for general LSP routing and may also be extended to integrated routing. The backup LSP is routed over all the logical links. If the chosen LSP uses only eligible logical links, it can proceed to select unprotected fibers. Otherwise, it needs to determine the set of fibers that must be unprotected. Specifically, a fiber traversed by the primary LSP must be unprotected if a noneligible logical link on the backup LSP does not have sufficient bandwidth to protect it.

Now, we explain how weights are assigned to logical links. A cost value k is assigned to noneligible logical links, where k is a large positive value much larger than the logical link capacity. The rationale behind doing this is to allow the use of these logical links but defer their usage. This is because it would result in more unprotected fibers, which is likely to increase the cumulative failure probability.

Eligible logical links are assigned weights that are the additional backup bandwidth required on each of them. A shortest-path selection algorithm is used to choose a minimum cost path as the backup LSP.

5.4 Restorable Probability

In the LSP protection, when a fiber fails, the traffic carried by the affected primary LSPs can be rerouted to their corresponding backup LSPs. In LSP PSP, if the failed fiber is not an unprotected fiber with respect to an affected primary LSP, the traffic carried by this primary LSP can be rerouted to its backup LSP. We note that the traffic is always protected at a certain protection grade. Otherwise, if the failed fiber is an unprotected fiber, whether the connection can be restored or not depends on whether all the logical links along the backup LSP have sufficient bandwidth at that time. This connection restorable probability was discussed in [5]. In practice, before a failure occurs on an unprotected fiber, the backup bandwidth reserved on one such logical link may have been increased by newly honored connections that use it in their backup LSPs. On the other hand, the amount of bandwidth needed on the logical link to protect the failure of the fiber may have been decreased if connections that traverse that fiber in their primary LSPs and use the logical link in their backup LSPs terminate. As a result, although bandwidth was not reserved for unprotected fibers at the time of honoring the request, there may be sufficient bandwidth available to restore connections at the time when the unprotected fiber fails. Thus, the restorable probability of a connection could be more than the specified protection grade.

5.5 Performance Comparison

We compare the performance of the LSP PSP with the LSP protection. In both approaches, primary LSPs are routed over a minimum number of logical links, and sharing among backup LSPs is allowed. It is observed in [5] through simulation results that LSP PSP performs better than LSP protection in terms of blocking probability and resource overbuild, which is defined as the ratio between the amount of bandwidth reserved for the backup LSP over that for the primary LSP. When a fiber fails, all connections can be restored with probabilities that are at least their required protection grades. Further, most of them have 100% restorable probability and can be recovered even when the failed fiber is one of their unprotected fibers.

The LSP PSP [5] can be used by service providers to allocate bandwidth with respect to users' protection grade requirements. Even for connections with 100% protection requirements, if the shared LSP protection cannot find a backup LSP, the LSP PSP can be invoked to avoid blocking by providing a backup LSP with a

certain protection grade. Given that most of connections have 100% restorable probability and can be recovered even when one of their unprotected fibers fails, users may be willing to gauge the grade of protection that they require to keep the cost low.

6 Conclusion

In this chapter, we first presented two integrated routing algorithms (HIRA and BIRA), for providing LSP protection in MPLS-over-WDM networks. Then, we explained two multilayer protection schemes for traffic requests with various recovery time requirements where high-priority traffic are protected at the LP level, while low-priority traffic are protected at the LSP level. An ILS method was described, which improves resource utilization by allowing backup LSPs to use preconfigured backup LPs under certain conditions. Finally, we presented a PSP approach at the LSP level for connections with various protection grades. An algorithm was presented which selects backup LSPs and unprotected fibers with the objective to minimize the bandwidth requirement. The LSP PSP can be used in multilayer protection schemes to further improve resource utilization by satisfying the specified protection grades.

References

1. Sahasrabuddhe L, Ramamurthy S, Mukherjee B (2002) Fault management in IP-over-WDM networks: WDM protection versus IP restoration. IEEE J Selec Areas Commun 20:21–33
2. Colle D et al (2002) Data-centric optical networks and their survivability. IEEE J Selec Areas Commun 20:6–20
3. Zheng Q, Mohan G (2003) Protection approaches for dynamic traffic in IP/MPLS-over-WDM networks. IEEE Commun Mag 41:S24–S29
4. Zheng Q, Mohan G (2006) Multi-layer protection in IP-over-WDM networks with and with no backup lightpath sharing. Comput Netw 50:301–316
5. Zheng Q, Mohan G (2009) LSP partial spatial-protection in MPLS over WDM optical networks. IEEE Trans Commun 57:1109–1118
6. Ye Y, Assi C, Dixit S, Ali MA (2001) A simple dynamic integrated provisioning/protection scheme in IP over WDM networks. IEEE Commun Mag 174–182
7. Kodialam M, Lakshman TV (2001) Integrated dynamic IP and wavelength routing in IP over WDM networks. In: IEEE INFOCOM, Anchorage, Alaska, USA, pp 358–366
8. Cheng Tien E, Mohan G (2002) Differentiated QoS routing in GMPLS-based IP/WDM networks. In: IEEE international conference on Globecom, Taipei, Taiwan, pp 2757–2761
9. Acharya S, Gupta B, Risbood P, Srivastava A (2003) IP-subnet aware routing in WDM mesh networks. In: IEEE INFOCOM, San Francisco, California, USA, pp 1333–1343
10. Koo S, Sahin G, Subramaniam S (2004) Dynamic LSP provisioning in overlay, augmented, and peer architectures for IP/MPLS over WDM networks. In: IEEE INFOCOM, Hong Kong, pp 514–523

11. Zheming D, Hamdi M, Lee JYB, Li VOK (2004) Integrated routing and grooming in GMPLS-based optical networks. In: IEEE international conference on Globecom, Dallas, Texas, USA, pp 1584–1588
12. Demeester P et al (1999) Resilience in multi-layer networks. IEEE Commun Mag 37:70–76
13. Maesschalck SD et al (2002) Intelligent optical networking for multilayer survivability. IEEE Commun Mag 40:42–49
14. Qin Y, Mason L, Jia K (2003) Study on a joint multiple layer restoration scheme for IP over WDM networks. IEEE Netw 17:43–48
15. Gerstel O, Ramaswami R (2000) Optical layer survivability-an implementation perspective. IEEE J Selec Areas Commun 18:1885–1899
16. Lai WS et al (2002) Network hierarchy and multilayer survivability. Internet-Draft, work in progress, draft-ietf-tewg-restore-hierarchy-01.txt
17. Gerstel O, Sasaki G (2002) Quality of protection (QoP): a quantitative unifying paradigm to protection service grades. Opt Netw Mag 3:40–49
18. Fang J, Sivakumar M, Somani AK, Sivalingam KM (2005) On partial protection in groomed optical WDM mesh networks. In: International conference on dependable systems networks (DSN)
19. Mohan G, Murthy CSR, Somani AK (2001) Efficient algorithms for routing dependable connections in WDM optical networks. IEEE/ACM Trans Netw 9:553–566
20. Fumagalli A et al (2003) Differentiated reliability in optical networks: theoretical and practical results. J Lightwave Technol 21:2576–2586

Chapter 12
Cross-Layer Survivability*

Hyang-Won Lee, Kayi Lee, and Eytan Modiano

1 Introduction

The layered architecture of modern communication networks takes advantage of the flexibility of upper layer technology, such as IP, and the high data rates of lower layer technology, such as WDM. In particular, the WDM technology available today can support up to several terabits per second over a single fiber [9], making networks vulnerable to failures, because a failure for even a short period of time can result in a huge loss of data. The main theme of network survivability is to prevent such data loss by provisioning spare resources for recovery. In this chapter, we focus on the impact of layering on network survivability.

In the layered network, a logical topology is embedded onto a physical topology such that each logical link is spanned by using a path in the physical topology. This is often referred to as *lightpath routing*. Obviously, a single fiber cut can lead multiple logical links sharing the fiber to fail. Due to this correlation between logical link failures, the layered network survivability problem exhibits vastly different characteristics from the single-layer counterpart.

*Based on "Cross-Layer Survivability in WDM-based Networks," by K. Lee, E. Modiano and H. Lee which appeared in IEEE/ACM Transactions on Networking, vol. 19, no. 4, Aug. 2011, and "Reliability in Layered Networks with Random Link Failures," by K. Lee, H. Lee and E. Modiano which appeared in IEEE INFOCOM 2010, Mar. 2010.c 2011 IEEE.

H.-W. Lee (✉)
Konkuk University, Seoul 143-701, Republic of Korea
e-mail: leehw@konkuk.ac.kr

K. Lee
Google Inc, Cambridge, MA, USA
e-mail: kylee@mit.edu

E. Modiano
Massachusetts Institute of Technology, Laboratory for Information and Decision Systems,
77 Mass Ave, Cambridge, MA 02139, USA
e-mail: modiano@mit.edu

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*,
Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_12,

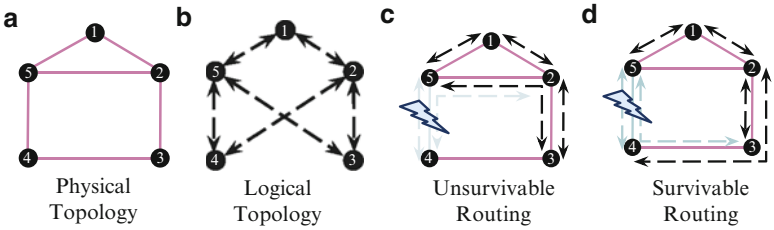


Fig. 12.1 Different lightpath routings can affect survivability. (**a**) Physical Topology, (**b**) Logical Topology, (**c**) Unsurvivable Routing, (**d**) Survivable Routing

The survivability of a layered network is dictated by the underlying lightpath routing. As an example, consider the physical and logical topologies shown in Fig. 12.1a,b. The lightpaths in the logical topology are routed over the physical topology in two different ways in Fig. 12.1c,d. In Fig. 12.1c, a failure of physical fiber (4, 5) would cause lightpaths (4, 5) and (2, 4) to fail. Consequently, node 4 will be disconnected from other nodes in the logical topology. On the other hand, in Fig. 12.1d, the logical topology will remain connected even if one of the fibers fails. This example demonstrates that to design a survivable layered network, one should carefully take into account the network structure across both the logical and physical layers. This is typically referred to as the *cross-layer survivability* problem.

In [2, 6, 7], the impact of physical layer failures on the connectivity of the logical topology was studied in the context of WDM-based networks. The authors proposed heuristic lightpath routing algorithms that minimize the number of disconnected logical node in the presence of a single physical link failure. The work of [15] was the first to introduce the notion of Survivable Lightpath Routing, which is defined to be a lightpath routing such that the logical topology remains connected in the event of a single fiber failure, and developed a mathematical formulation for finding a survivable lightpath routing. These results have been improved with more efficient formulation and extended to account for multiple physical failures [8, 11, 12, 18].

Most works in the literature consider the survivability as a constraint, however this chapter takes a more fundamental approach to addressing cross-layer survivability. In particular, in Sect. 12.2, we study connectivity parameters of a layered network, and observe that they exhibit vastly different properties compared to their single-layer counterparts. This observation motivates a new survivability metric called Min Cross Layer Cut (MCLC). The MCLC quantifies the connectivity of a layered network and is used to develop survivable lightpath routing algorithms. Simulation results show that these algorithms can find a better survivable layered network. Going beyond connectivity, a new survivability metric is introduced and analyzed in order to design a layered network that uses minimal spare capacity for protection against single-fiber failures. In Sect. 12.3 we study cross-layer survivability in the presence of random physical link failures, and in Sect. 12.4 we discuss future directions for cross-layer survivability.

2 Connectivity of Layered Networks

In this section we study key connectivity structures such as flows and cuts in multi-layer graphs in order to develop insights into cross-layer survivability. We will highlight the key differences in combinatorial properties between multi-layer graphs and single-layer graphs. In particular, it turns out that fundamental survivability results, such as the "Max-Flow Min-Cut Theorem", are no longer applicable to multi-layer networks. Consequently, metrics such as "connectivity" have significantly different meaning in the layered setting. This motivates us to revisit fundamental issues such as quantifying and maximizing survivability in the layered setting.

2.1 Max Flow vs. Min Cut

For single-layer networks, the Max-Flow Min-Cut Theorem [1] states that the maximum number of disjoint paths between two nodes s and t is always the same as the minimum number of edges that need to be removed from the network in order to disconnect the two nodes. Let MaxFlow_{st} and MinCut_{st} be integral $s-t$ Max Flow and Min Cut respectively, and let MaxFlow_{st}^{R} and MinCut_{st}^{R} be their fractional (relaxed) values. The Max-Flow Min-Cut Theorem for single-layer networks can then be stated as follows:

$$\text{MaxFlow}_{st} = \text{MaxFlow}_{st}^{R} = \text{MinCut}_{st}^{R} = \text{MinCut}_{st}.$$

Consequently, the term *connectivity* between two nodes can be used unambiguously to refer to different measures such as the maximum number of disjoint paths or the minimum size cut, and this makes it a natural choice as the standard metric for measuring network survivability. The equality among these values has profound implications on survivable network design for single-layer networks. Because all these survivability measures take on the same value, it can naturally be used as the standard survivability metric that is applicable to measuring both disjoint paths or the minimum cut. Another consequence of this equality is that linear programs, which are polynomial time solvable, can be used to find the minimum cut and disjoint paths in the network.

Because of its fundamental importance, it is crucial to understand the Max-Flow Min-Cut relationship in layered networks. The following is a generalization of *Max Flow* and *Min Cut* to the layered setting:

Definition 1 In a multi-layer network, the *Max Flow* between two nodes s and t in the logical topology is the maximum number of physically disjoint $s-t$ paths in the logical topology. The *Min Cut* between two nodes s and t in the logical topology

is the minimum number of physical links that need to be removed in order to disconnect the two nodes in the logical topology.

We model the physical topology as a network graph $G_P = (V_P, E_P)$, where V_P and E_P are the nodes and links in the physical topology. The logical topology is modeled as $G_L = (V_L, E_L)$ in a similar fashion. The light path routing is represented by a set of binary variables f_{ij}^{st}, where a logical link (s, t) uses physical fiber (i, j) if and only if $f_{ij}^{st} = 1$. Let $\mathscr{P}_{st}$ be the set of all $s - t$ paths in the logical topology. For each path $p \in \mathscr{P}_{st}$, let $L(p)$ be the set of physical links used by the logical path p, that is, $L(p) = \cup_{(s,t)\in p}\{(i,j)|f_{ij}^{st} = 1\}$. Then the Max Flow and Min Cut between nodes s and t can be formulated mathematically as follows:

$$\begin{aligned} &\text{MaxFlow}_{\text{st}}: \quad \text{Maximize} \sum\nolimits_{p\in\mathscr{P}_{st}} f_p, \quad \text{subject to}: \\ &\sum\nolimits_{p:(i,j)\in L(p)} f_p \le 1 \qquad \forall (i,j) \in E_P \\ &\qquad f_p \in \{0,1\} \qquad \forall p \in \mathscr{P}_{st} \end{aligned} \tag{12.1}$$

$$\begin{aligned} &\text{MinCut}_{\text{st}}: \quad \text{Minimize} \sum\nolimits_{(i,j)\in E_P} y_{ij}, \quad \text{subject to}: \\ &\sum\nolimits_{(i,j)\in L(p)} y_{ij} \ge 1 \qquad \forall p \in \mathscr{P}_{st} \\ &\qquad y_{ij} \in \{0,1\} \qquad \forall (i,j) \in E_P \end{aligned} \tag{12.2}$$

The variable f_p in the formulation $\text{MaxFlow}_{\text{st}}$indicates whether the path p is selected for the set of (s, t)-disjoint paths. Constraint (12.1) requires that no selected logical paths share a physical link. Similarly, in the formulation $\text{MinCut}_{\text{st}}$, the variable y_{ij} indicates whether the physical fiber (i, j) is selected for the minimum (s, t)-cut. Constraint (12.2) requires that all logical paths between s and t traverse some physical fiber (i, j) with $y_{ij} = 1$.

Note that the above formulations generalize the Max Flow and Min Cut for single-layer networks. In particular, the formulations model the classical Max Flow and Min Cut of a graph G if both G_P and G_L are equal to G, and $f_{ij}^{st} = 1$ if and only if $(s, t) = (i, j)$. Let us redefine $\text{MaxFlow}_{\text{st}}$ and $\text{MinCut}_{\text{st}}$ to be the optimal values of the above Max Flow and Min Cut formulations. We also denote $\text{MaxFlow}_{\text{st}}^{\text{R}}$ and $\text{MinCut}_{\text{st}}^{\text{R}}$ to be the optimal values to the linear relaxations of above Max Flow and Min Cut formulations.

First, it is easy to verify that the linear relaxations for the formulations $\text{MaxFlow}_{\text{st}}$ and $\text{MinCut}_{\text{st}}$ maintain a primal–dual relationship, which, by the Duality Theorem [5], implies that $\text{MaxFlow}_{\text{st}}^{\text{R}} = \text{MinCut}_{\text{st}}^{\text{R}}$. In addition, since any feasible solution to an integer program is also a feasible solution to the linear relaxation, the following relationship holds:

Observation 1 $\text{MaxFlow}_{\text{st}} \le \text{MaxFlow}_{\text{st}}^{\text{R}} = \text{MinCut}_{\text{st}}^{\text{R}} \le \text{MinCut}_{\text{st}}$.

Therefore, just as with single-layer networks, the maximum number of disjoint paths between two nodes cannot exceed the minimum cut between them in a multi-layer network. However, unlike the single-layer case, the values of $\text{MaxFlow}_{\text{st}}$,

MaxFlow_{st}^{R} and MinCut_{st} are not always identical, as illustrated in the following example. In our examples throughout the section, we use a logical topology with two nodes s and t that are connected by multiple parallel lightpaths. For simplicity of exposition, we omit the complete lightpath routing and only show the physical links that are shared by multiple lightpaths. In fact, it can be shown that for a two-node logical topology, any arbitrary fiber–sharing relationship can be realized by reconstructing a physical topology and lightpath routing [14]. Therefore, in the following discussion, we omit the details of the lightpath routing and only show the fiber–sharing relationship of our two-node logical topology.

In Fig. 12.1, the two nodes in the logical topology are connected by three lightpaths. The logical topology is embedded on the physical topology in such a way that each pair of lightpaths shares a fiber. It is easy to see that no single fiber can disconnect the logical topology, and that any pair of fibers would. Hence, the value of MinCut_{st} is 2 in this case. On the other hand, the value of MaxFlow_{st} is only 1, because any two logical links share some physical fiber, so none of the paths in the logical network are physically disjoint. Finally, the value of MaxFlow_{st}^{R} is 1.5 because a flow of 0.5 can be routed on each of the lightpaths without violating the capacity constraints at the physical layer. Therefore, Fig. 12.1 is an example where all three quantities differ.

It was shown in [14] that the gap between $\text{MaxFlow}_{st}$and MaxFlow_{st}^{R} is $O(|E_L|)$, and the gap between MinCut_{st} and MinCut_{st}^{R} is $O(\log|E_L|)$. Thus, the gaps among the three values are not bounded by any constant. Therefore, a multi-layer network with high connectivity value (i.e. that tolerates a large number of failures) does not necessarily guarantee the existence of physically disjoint paths. This is in sharp contrast to single-layer networks where the number of disjoint paths is always equal to the minimum cut.

It is thus clear that network survivability metrics across layers are not trivial extensions of the single layer metrics. New metrics need to be carefully defined in order to measure cross-layer survivability in a meaningful manner. In Sect. 12.2.3, we introduce two new metrics that can be used to measure the connectivity of multi-layer networks.

2.2 Computational Complexity

In single-layer networks, because the integral Max Flow and Min Cut values are always identical to the optimal relaxed solutions, these values can be computed in polynomial time [1]. However, computing and approximating their cross-layer equivalents turns out to be much more difficult. Theorem 1 describes the complexity of computing the Max Flow and Min Cut for multi-layer networks.

Theorem 1 ([14]) *Computing Max Flow and Min Cut for multi-layer networks is NP-hard. In addition, both values cannot be approximated within any constant factor, unless P=NP.*

In summary, the notion of survivability in multi-layer networks bears subtle yet important differences from its single-layer counterpart. Because of that, many new issues arise in the layered setting, including defining, measuring and optimizing survivability metrics. In what follows, our discussion will be focused on appropriate metrics for layered networks and efficient algorithms to maximize the cross-layer survivability.

2.3 Metrics for Cross-Layer Survivability

The previous section demonstrates some fundamental challenges in designing survivable layered network architectures. In particular, choosing the right metric to quantify survivability becomes an important and non-trivial question. Although the right metric will depend on the particular survivability requirement (e.g., disjoint paths or minimum cut), any reasonable metric must be *consistent* in that a network with a higher metric value should be more resilient to failures, *monotonic* in that any addition of physical or logical links to the network should not decrease the metric value, and *compatible* in that the metric should generalize the connectivity metric for single-layer networks.

Next, we introduce two metrics that measure the ability of the network to withstand multiple physical failures, while meeting the above criteria. Although the two metrics appear to measure different aspects of network connectivity, they are in fact closely related, as will be shown later.

2.3.1 Min Cross Layer Cut

The *Min Cross Layer Cut (MCLC)* is a natural generalization of Min Cut in single-layer networks. Similar to the way MinCut_{st} is defined in Sect. 12.2.1 between two given nodes s and t in the network, the Min Cross Layer Cut of a layered network is defined to be the smallest set of physical links whose removal will *globally* disconnect the logical network. A lightpath routing with high Min Cross Layer Cut value implies that the network remains connected even after a relatively large number of physical failures. It is also a generalization of the survivable lightpath routing definition in [15], since a lightpath routing is survivable if and only if its Min Cross Layer Cut is greater than 1.

Let S be a subset of the logical nodes V_L, and $\delta(S)$ be the set of the logical links with exactly one end point in S. Let H_S be the minimum number of physical links failures required to disconnect all links in $\delta(S)$. The Min Cross Layer Cut can be defined as follows:

$$MCLC = \min_{S \subset V_L} H_S.$$

For each S, computing H_S amounts to finding the Min Cut between the two partitions S and $V_L - S$. Therefore, by Theorem 1, computing H_S is also NP-Hard. Computing the Min Cross Layer Cut, which is defined to be the minimum among all H_S values, is therefore a difficult problem. However, for practical purposes, the MCLC of a large multi-layer network (e.g. 100 nodes) can be computed reasonably fast by solving the following integer linear program.

Given the physical and logical topologies (V_P, E_P), and (V_L, E_L), let f_{ij}^{st} be binary constants that represent the lightpath routing, such that logical link (s, t) uses physical fiber (i, j) if and only if $f_{ij}^{st} = 1$. The *MCLC* can be formulated as the integer program below [14]:

$$\begin{aligned} &\mathrm{M_{MCLC}}: \text{Minimize} \sum\nolimits_{(i,j)\in E_P} y_{ij}, \quad \text{subject to}: \\ &d_t - d_s \le \sum\nolimits_{(i,j)\in E_P} y_{ij} f_{ij}^{st} \qquad \forall (s,t) \in E_L \end{aligned} \tag{12.3}$$

$$\begin{aligned} &\sum_{n \in V_L} d_n \ge 1, d_0 = 0 \\ &d_n, y_{ij} \in \{0,1\} \qquad \forall n \in V_L, (i,j) \in E_P \end{aligned} \tag{12.4}$$

The integer program contains a variable y_{ij} for each physical link (i, j), and a variable d_k for each logical node k. Constraint (12.3) maintains the following property for any feasible solution: if $d_k = 1$, the node k will be disconnected from node 0 after all physical links (i, j) with $y_{ij} = 1$ are removed. To see this, note that since $d_k = 1$ and $d_0 = 0$, any logical path from node 0 to node k contains a logical link (s, t) where $d_s = 0$ and $d_t = 1$. Constraint (12.3) requires that such a logical link traverse at least one of the fibers (i, j) with $y_{ij} = 1$. As a result, all paths from node 0 to node k must traverse one of these fibers, and node k will be disconnected from node 0 if these fibers are removed from the network. Constraint (12.4) requires node 0 to be disconnected from at least one node, which ensures that the set of fibers (i, j) with $y_{ij} = 1$ forms a global Cross Layer Cut.

In Sect. 12.2.4, we will discuss several lightpath routing algorithms to maximize the MCLC value.

2.3.2 Weighted Load Factor

Another way to measure the connectivity of a layered network is by quantifying the "impact" of each physical failure. The *Weighted Load Factor (WLF)*, an extension of the metric *Load Factor* introduced in [10], provides such a measure of survivability. The WLF can be formulated as follows:

$$\mathrm{M_{WLF}}: \text{Maximize} \quad \frac{1}{z}, \quad \text{subject to}:$$

$$z \cdot \sum_{(s,t)\in\delta(S)} w_{st} \geq \sum_{(s,t)\in\delta(S)} w_{st} f_{ij}^{st} \qquad \forall S \subset V_L, (i,j) \in E_P$$

$$\sum_{(s,t)\in\delta(S)} w_{st} > 0 \qquad \forall S \subset V_L$$

$$0 \leq z, w_{st} \leq 1 \qquad \forall (s,t) \in E_L,$$

where $\delta(S)$ is the cut set of S, i.e., the set of logical links that have exactly one end point in S.

The variables w_{st} are the weights assigned to the lightpaths. Over all possible logical cuts, the variable z measures the maximum fraction of weight carried by a fiber within a single cut. Intuitively, if we interpret the weight to be the amount of traffic in the lightpath, the value z can be interpreted to be the maximum fraction of traffic across a set of nodes disrupted by a single fiber cut. The Weighted Load Factor formulation, defined to maximize the reciprocal of this fraction, thus tries to compute the logical edge weights that minimize the maximum fraction. This effectively measures the best way of spreading the weight across the fibers for the given lightpath routing. A lightpath routing with a larger Weighted Load Factor value is more capable of spreading its weight within any cut across the fibers.

Recall that in [15], a lightpath routing is defined to be survivable if the resulting layered network survives any single physical link failure. Hence, the survivable lightpath routing ensures that not all of the links in a logical cut share the same fiber, which in turn implies that the Weighted Load Factor $\frac{1}{z}$is greater than one. Therefore, the Weighted Load Factor captures the connectivity of a layered network, generalizing the survivable lightpath routing. In fact, the Weighted Load Factor is closely related to Min Cross Layer Cut. Given a lightpath routing, let $\mathrm{M_{MCLC}}$ be the ILP formation for its Min Cross Layer Cut, and let *MCLC* and $MCLC^R$ be the optimal values for $\mathrm{M_{MCLC}}$ and its linear relaxation respectively. In addition, let *WLF* be the Weighted Load Factor of the lightpath routing. Then we have the following relationship [14]:

Theorem 2 $MCLC^{\mathrm{R}} \leq WLF \leq MCLC$

Therefore, although the two metrics appear to measure different aspects of network connectivity, they are inherently related. In fact, the two values are often identical as shown in Sect. 12.2.5.

2.4 Lightpath Routing Algorithms for Maximizing MCLC

A natural approach to maximizing the survivability of a layered network is to design a lightpath routing that maximizes the number of failures the network can

withstand, i.e., maximizes Min Cross Layer Cut (MCLC). All the lightpath routing algorithms introduced in this section try to maximize the MCLC value. They are all based on multi-commodity flows, where each lightpath is considered a commodity to be routed over the physical network. Given the physical network $G_P = (V_P, E_P)$ and the logical network $G_L = (V_L, E_L)$, the multi-commodity flow for a lightpath routing can be generally formulated as follows:

$$\begin{aligned} \text{MCF}_{\mathscr{X}}: \quad & \text{Minimize } \mathscr{X}(f), \quad \text{subject to:} \\ & f_{ij}^{st} \in \{0,1\} \\ & \{f_{ij}^{st} : (i,j) \in E_P\} \text{ forms an } (s,t)\text{-path}, \forall (s,t) \in E_L, \end{aligned}$$

where f is the variable set that represents the lightpath routing, such that $f_{ij}^{st} = 1$ if and only if lightpath (s, t) uses physical fiber (i, j) in its route. The objective function $\mathscr{X}(f)$ depends on the lightpath routing f. Ideally, $\mathscr{X}(f)$ should be the MCLC value, however it turns out to be difficult to express the MCLC as a tractable function. For this reason, it is desired to develop a tractable objective that approximates the MCLC value [14].

2.4.1 Integer Programming Formulations

Let w be a weight assigned to each lightpath. The objective function ρ_w measures the maximum *load* of the fibers, where the *load* is defined to be the total lightpath weight carried by the fiber. The intuition is that the multi-commodity flow formulation will try to spread the weight of the lightpaths across multiple fibers, thereby minimizing the impact of any single fiber failure.

Such an objective can be formulated as an integer linear program as follows:

$$\begin{aligned} \text{MCF}_{w}: \quad & \text{Minimize } \rho_w, \quad \text{subject to:} \\ & \rho_w \geq \sum\nolimits_{(s,t)\in E_L} w(s,t) f_{ij}^{st} \ \forall (i,j) \in E_P \\ & f_{ij}^{st} \in \{0,1\} \\ & \{f_{ij}^{st} : (i,j) \in E_P\} \text{ forms an } (s,t)\text{-path}, \forall (s,t) \in E_L \end{aligned}$$

The routing strategy of the algorithm is determined by the weight function w, and with a careful choice of the weight function w, the value $\frac{1}{\rho_w}$ gives a lower bound on the MCLC. Therefore, a lightpath routing with a low ρ_w value is guaranteed to have a high MCLC. For example, if w is set to 1 for all lightpaths, the integer program will minimize the number of lightpaths traversing the same fiber. Effectively, this will minimize the number of disconnected lightpaths in the case of a single fiber failure.

It is conceivable that if the weight function somehow reflects the connectivity structure of the logical graph, then it may lead to a better objective toward

maximizing the MCLC of the solution. Consider a different weight function w_{MinCut} such that for each edge $(s, t) \in E_L$, the weight $w_{MinCut}(s, t)$ is defined to be $\frac{1}{|MinCut_L(s,t)|}$, where $MinCut_L(s, t)$ is the minimum (s, t)-cut in the logical topology. Therefore, if an edge (s, t) belongs to a smaller cut, it will be assigned a higher weight. The algorithm will therefore try to avoid putting these small cut edges on the same fiber.

If w_{MinCut} is used as the weight function in MCF_w, Lee et al. [14] shows the following relationship between the objective value ρ_w of a feasible solution to $MCF_w$and the Weighted Load Factor of the associated lightpath routing:

Theorem 3 ([14]) *For any feasible solution f of* MCF_w with w_{MinCut} as the weight function, $\frac{1}{\rho_w} \leq WLF$.

As a result of Theorems 2 and 3, the MCLC of a lightpath routing is lower bounded by the value of $\frac{1}{\rho_w}$, which the algorithm will try to maximize.

2.4.2 An Enhanced Multi-Commodity Flow Formulation

As we discussed in Sect. 12.2.3.2, the Weighted Load Factor provides a good lower bound on the MCLC of a lightpath routing. Here we discuss another multi-commodity flow based formulation whose objective function approximates the Weighted Load Factor of a lightpath routing. The formulation, denoted as MCF_{LF}, can be written as follows:

$$
\begin{aligned}
&MCF_{LF}: \quad \text{Minimize } \gamma, \quad \text{subject to:} \\
&\gamma|\delta(S)| \geq \sum_{(s,t)\in\delta(S)} f_{ij}^{st} \ \forall (i,j) \in E_P, S \subset V_L \\
&f_{ij}^{st} \in \{0,1\} \\
&\{f_{ij}^{st} : (i,j) \in E_P\} \text{ forms an } (s,t)\text{-path}, \forall (s,t) \in E_L
\end{aligned}
$$

Essentially, the formulation optimizes the *unweighted*Load Factor of the lightpath routing, (i.e., all weights equal one), by minimizing the maximum fraction of a logical cut carried by a single fiber. As this formulation provides a constraint for each logical cut, it captures the impact of a single fiber cut on the logical topology in much greater detail. The following theorem shows that for any lightpath routing, its associated Load Factor value $\frac{1}{\gamma}$ gives a tighter lower bound than $\frac{1}{\rho_w}$, given by the MCF_w formulation.

Theorem 4 ([14]) *For any lightpath routing, let* ρ_w*be its associated objective value in the formulation* MCF_w*with* w_{MinCut}*as the weight function, and let* γ *be its associated objective value in the formulation* MCF_{LF}*. In addition, let WLF be its Weighted Load Factor. Then:*

$$\frac{1}{\rho_w} \leq \frac{1}{\gamma} \leq WLF.$$

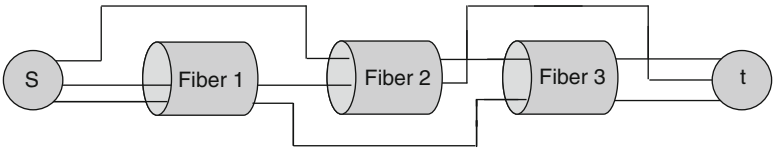


Fig. 12.2 A logical topology with 3 logical links where each pair of links shares a fiber in the physical topology

Therefore, the formulation MCF_{LF}gives a lightpath routing that is optimized for a better lower bound on the MCLC. An ILP is generally difficult to solve, and the above two formulations may not scale to large networks. Nonetheless, randomized rounding technique has been successfully used to solve multi-commodity flow problems lightpath [4, 17]. The formulations $\text{MCF}_w$and MCF_{LF}can also be solved via randomized rounding, and more details can be found in [14].

2.5 Simulation

In order to evaluate the performance of the algorithms introduced in the previous section, the NSFNET (Fig. 12.2) is augmented to have connectivity 4, and used as the physical topology. For logical topologies, 350 random graphs are generated such that each of them has connectivity 4 and its size ranges from 6 to 15 nodes. The MCLC values of the lightpath routings generated by the algorithms introduced in Sect. 12.2.4 will be compared as a measure of their survivability performance.

2.5.1 Survivability Performance of Different Lightpath Routing Formulations

We first study the survivability performance of the lightpath routings generated by the different formulations introduced in Sect. 12.2.4.1. Specifically, the following three algorithms are compared:

1. MinCut: Multi-Commodity Flow MCF_w, using the weight function w_{MinCut}
2. LF: Enhanced Multi-Commodity Flow MCF_{LF}.
3. SURVIVE: Survivable lightpath routing algorithm in [15] which computes the lightpath routing that minimizes the total fiber hops, subject to the constraint that the MCLC must be at least two.

Figure 12.4 compares the average MCLC values of the lightpath routings computed by the four different algorithms. Overall, the formulations introduced in this paper achieve better survivability than SURVIVE. This is because these formulations try to maximize the MCLC in their objective functions, whereas SURVIVE minimizes the physical hops. Therefore, even though SURVIVE does well in finding a survivable routing (i.e. MCLC ≥ 2), a more specialized formulation is required to achieve even higher MCLC values.

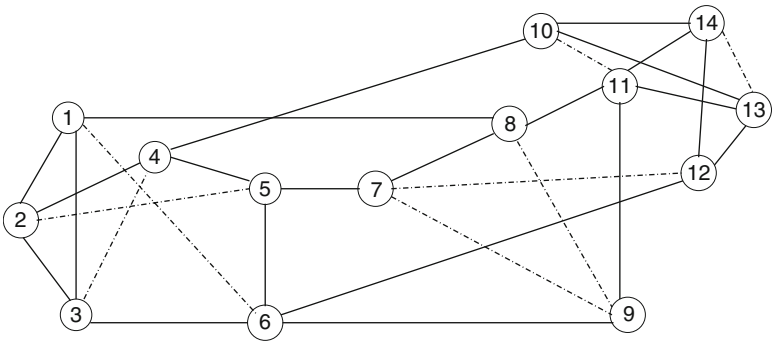


Fig. 12.3 The augmented NSFNET. The *dashed lines* are the new links

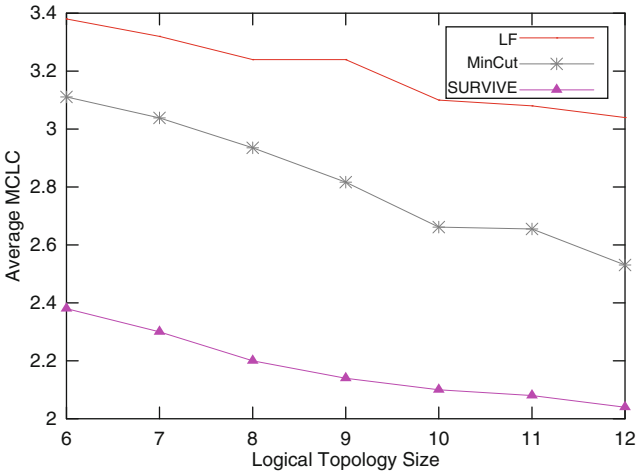


Fig. 12.4 MCLC performance of different lightpath routing formulations

The quality of the lightpath routing also depends on the graph structures captured by the formulations. Compared with $\mathrm{MCF}_{\mathrm{MinCut}}$, the formulation $\mathrm{MCF}_{\mathrm{LF}}$ captures the connectivity structure of the logical topology in much greater detail, by having a constraint to describe the impact of a physical link failure to each logical cut. As a result, the algorithm based on this enhanced formulation is able to provide lightpath routings with higher MCLC values.

2.5.2 Comparison Among Metrics

Recall the lower bounds on the Min Cross Layer Cut in Theorem 4. In this section, we study 350 different lightpath routings and measure these lower bound values for each of the lightpath routing. As Fig. 12.5 shows, the Weighted Load Factor is a

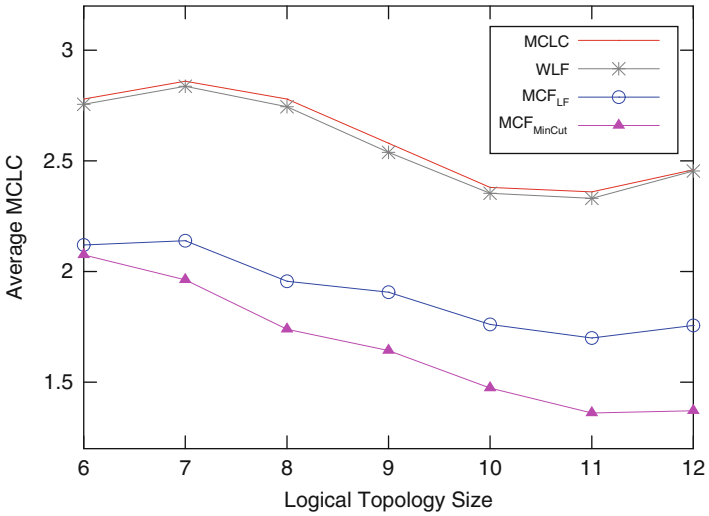


Fig. 12.5 Comparison among Min Cross Layer Cut (MCLC), Weighted Load Factor (WLF), and the optimal values of MCF_{LF} and $\text{MCF}_{\text{MinCut}}$

very close approximation of the Min Cross Layer Cut. Among the 350 routings studied, the two metrics are identical in 308 cases. This suggests a very tight connection between the two metrics, and the strong correlation between them also justifies the choice of such metrics as survivability measures.

The figure also reveals a strong correlation between the MCLC performance and the tightness of the lower bounds given by the multi-commodity flow formulations in Sect. 12.2.4.1. Compared to MCF_w, the formulation MCF_{LF} provides an objective value that is closer to the actual MCLC value of the lightpath routing. This translates to better lightpath routings, as we saw in Fig. 12.4. Since there is still a large gap between the MCF_{LF} objective value and the MCLC value, this suggests room for further improvement with formulations that give a better MCLC lower bound. A good formulation that properly captures the cross-layer connectivity structure is essential for generating lightpath routings with high survivability, and it gives a powerful tool for designing highly survivable layered networks.

2.6 Beyond Connectivity

So far, we have assumed that the survivability of a layered network is guaranteed as long as the logical topology remains connected after a failure. Implicitly assumed here is that there is sufficient capacity in the network, so that the disrupted traffic can always be supported over available alternative paths. However, the capacity of the network is finite, and thus it may not always be possible to support the disrupted traffic. Therefore, it is important to take into account spare capacity as

well as connectivity. To address this issue, we redefine survivability to meet two conditions: (1) the logical topology remains connected after any physical link failure and (2) there is sufficient capacity in the resulting network to support the traffic requirement.

As mentioned above, the design of survivable lightpath routing has focused only on the connectivity of the logical topology after a physical link failure. While there are a number of works dealing with spare capacity allocation, they assume single-link failures, and hence they are not applicable to the layered network where upon a single physical link failure, multiple logical links can fail simultaneously. In this section, joint survivable lightpath routing and capacity assignment problems are considered for layered networks. We discuss a new metric, first introduced in [10], that can measure the efficiency of spare capacity allocation. This metric is a generalization of the connectivity metrics discussed earlier to account for spare capacity, and can be used to formulate the problem of finding lightpath routings that guarantee efficient use of link capacity for protection.

The new metric, called *Load Factor*[10], quantifies the fraction of working capacity and spare capacity over each logical link. Assume that each link has capacity C. Let $\alpha(\in [0, 1])$ be the fraction of working capacity, i.e., αC is used for working paths, and subsequently, $(1 - \alpha)C$ is reserved for backup paths. Without loss of generality, we assume $C = 1$. For a given pair of logical and physical topologies and its lightpath routing, we define the load factor of the layered network to be the maximum value of α such that the two network survivability conditions mentioned above are satisfied. Clearly, the load factor measures the efficiency of capacity utilization, and it is desirable to find a lightpath routing with maximum load factor. In [10], a necessary and sufficient condition on the load factor was identified and used to develop an MILP formulation for finding a lightpath routing that maximizes the load factor. The results from [10] are discussed below.

Recall that N_L is the set of logical nodes. Denote by $CS(S)$ the cut set corresponding to a partition $\langle S, N - S\rangle$ of N_L. Given a routing of the logical topology denoted by $[f_{ij}^{st}, (i, j) \in E_P, (s, t) \in E_L]$, the following theorem gives a necessary and sufficient condition on the load factor.

Theorem 5 ([10]) *A network is survivable if and only if for every cut-set CS(S) of the logical topology and every physical link failure (i,j), the load factor satisfies the following inequality:*

$$\sum_{(s,t)\in CS(S)} (f_{ij}^{st} + f_{ji}^{st})\alpha \leq \sum_{(s,t)\in CS(S)} [1 - (f_{ij}^{st} + f_{ji}^{st})](1 - \alpha). \tag{12.5}$$

Rearranging the inequality (12.5), it can be shown that the load factor α is given by

$$\alpha = \min_{\substack{S\subset N_L \\ (i,j)\in E_P}} \frac{|CS(S)| - \sum_{(s,t)\in CS(S)} (f_{ij}^{st} + f_{ji}^{st})}{|CS(S)|}. \tag{12.6}$$

If there exists a cut-set $CS(S)$ and fiber (i, j) such that all the links in $CS(S)$ share (i, j), then the value of α is zero. Clearly, in this case, the logical topology is disconnected upon the failure of (i, j), meaning that the given lightpath routing cannot survive a single failure. Note that survivable routing algorithms in the literature such as [15] only guarantee $\alpha > 0$, i.e., the network remains connected after a single fiber failure. Therefore, the problem of finding a lightpath routing with maximum load factor α can be viewed as a generalization of finding a lightpath routing with connectivity guarantee.

The above result can be used to derive an optimal lightpath routing that maximizes the load factor. Let $\{f_{ij}^{st}\}^*$ be the routing that maximizes the load factor, and R denote the set of all possible routings. Then, it follows that

$$\{f_{ij}^{st}\}^* = \underset{f_{ij}^{st} \in R}{\arg\min} \max_{\substack{S \subset N_L \\ (i,j) \in E_P}} \frac{\sum_{(s,t) \in CS(S)} (f_{ij}^{st} + f_{ji}^{st})}{|CS(S)|}. \tag{12.7}$$

Note that the load factor is a special case of the weighted load factor discussed in the previous section where the weight function $w_{st} = 1$ for every logical link (s, t). The ratio in the above optimization is the fraction of logical links in a cut-set that will fail in the event of a fiber failure. Hence, a high ratio implies that the fiber is shared by many logical links, and it can be interpreted as the load on a fiber. The above formulation minimizes the maximum load on each fiber. Intuitively, this will minimize the amount of disrupted traffic in the event of a fiber cut, thereby reducing the demand for spare capacity. Using the representation of an optimal routing in (12.7), an MILP can be formulated for finding a lightpath routing with maximum load factor [10].

3 Extension to Random Failures

So far we considered single physical link failures, and discussed survivability metrics that account for a worst-case failure event. Failures in communication networks can be modeled as random events. It is thus important to understand the impact of random failures on the survivability of layered networks. In this section, we study the relationship between cross-layer connectivity metrics discussed in the previous section and the survivability of a layered network with random physical link failures. Interestingly, maximizing the MCLC value has the effect of maximizing cross-layer reliability in the low failure probability regime, and this observation can be used to develop reliable lightpath routing algorithms.

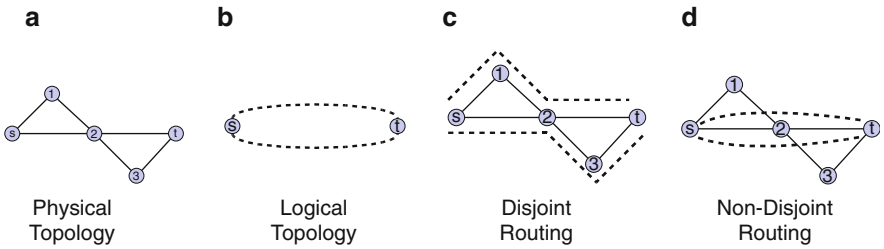


Fig. 12.6 Non-disjoint routings can sometimes be more reliable. (**a**) Physical Topology (**b**) Logical Topology (**c**) Disjoint Routing (**d**) Non-Disjoint Routing

3.1 Cross-Layer Reliability under Random Failures

Consider a layered network that consists of the logical topology $G_L = (V_L, E_L)$ built on top of the physical topology $G_P = (V_P, E_P)$ through a lightpath routing $f = [f_{ij}^{st}, (i, j) \in E_P, (s, t) \in E_L]$. If a physical link (i, j) fails, all of the logical links (s, t) carried over (i, j) (i.e., (s, t) such that $f_{ij}^{st} = 1$) also fail. A set S of physical links is called a *cross-layer cut* if the failure of the links in S causes the logical network to be disconnected. We also define the *network state* to be the subset S of physical links that failed. Hence, if S is a cross-layer cut, the network state S represents a *disconnected* network state. Otherwise, it is a *connected* state.

Each physical link fails independently with probability p. This probabilistic failure model represents a snapshot of a network where links fail and are repaired according to some Markovian process. Hence, p represents the steady-state probability that a physical link is in a failed state. The reliability of a multi-layer network is defined to be the probability that the logical network remains connected. We call this *cross-layer reliability*, and it is a natural survivability metric when the physical topology experiences random failures.

It is important to note that the cross-layer reliability depends on the underlying lightpath routing. For example, in Fig. 12.6, the logical topology consists of two parallel links between nodes s and t. Suppose every physical link fails independently with probability p. The first lightpath routing in Fig. 12.6c routes the two logical links using link-disjoint physical paths $(s, 1, 2, t)$ and $(s, 2, 3, t)$. Under this routing, the logical network will be disconnected with probability $(1 - (1 - p)^3)^2$. On the other hand, the second lightpath routing in Fig. 12.6d, which routes the two logical links over the same shortest physical route $(s, 2, t)$, has failure probability $2p - p^2$. While disjoint path routing is generally considered to be more reliable, it is only true in this example for small values of p. For large values of p(e.g. $p > 0.5$), the second lightpath routing is more reliable. Therefore, whether one lightpath routing is better than another depends on the value of p.

3.2 Cross-Layer Failure Polynomial

In single-layer networks, with random failures, reliability can be expressed as a polynomial in the failure probability p[3]. In [13], this polynomial expression was extended to the layered setting. It turns out that this expression provides important insights to the design of reliable lightpath routings.

Assume that there are m physical links, i.e.,$|E_P| = m$. The probability associated with a network state S with exactly i physical link failures (i.e., $|S| = i$) is $p^i(1-p)^{m-i}$. Let N_i be the number of cross-layer cuts S with $|S| = i$, then the probability that the network is disconnected is simply the sum of the probabilities over all cross-layer cuts, i.e.,

$$F(p) = \sum_{i=0}^{m} N_i p^i (1-p)^{m-i}. \tag{12.8}$$

Therefore, the failure probability of a multi-layer network can be expressed as a polynomial in p. The function $F(p)$ is called the *cross-layer failure polynomial* or simply the *failure polynomial*. The vector $[N_0, \ldots, N_m]$ plays an important role in assessing the reliability of a network. In particular, given the N_i values the reliability of the network can be computed using (12.8) for any value of p.

Clearly, if $N_i > 0$, then $N_j > 0, \ \forall j > i$, because any cut of size i will still be a cut with the addition of more failed links. The smallest i such that $N_i > 0$ is of special importance because it represents the Min Cross Layer Cut (MCLC) of the network, i.e., it is the minimum number of physical link failures needed to disconnect the logical network. Let d be the MCLC value of the network, and assume that it is a constant independent of the physical network size. Note that $N_i = 0, \ \forall i < d$, and the term $N_d p^d (1-p)^{m-d}$ in the failure polynomial dominates all other terms for small values of p. It was shown in [13] that there exists a region of probability p over which a lightpath routing with higher MCLC is more reliable than any lightpath routing with lower MCLC. Consequently, if a lightpath routing maximizes MCLC, i.e., make d as large as possible, it will achieve optimal reliability in the low failure probability regime.

Notice that we already discussed such lightpath routing algorithms in the previous section. In the following, we verify that these algorithms yield good reliability in the low failure probability regime.

3.3 Simulation

We used the augmented NSFNET (Fig. 12.3) as the physical topology, and generated 350 random logical topologies with size from 6 to 12 nodes and connectivity at least 4. We compare the reliability performance of the three lightpath

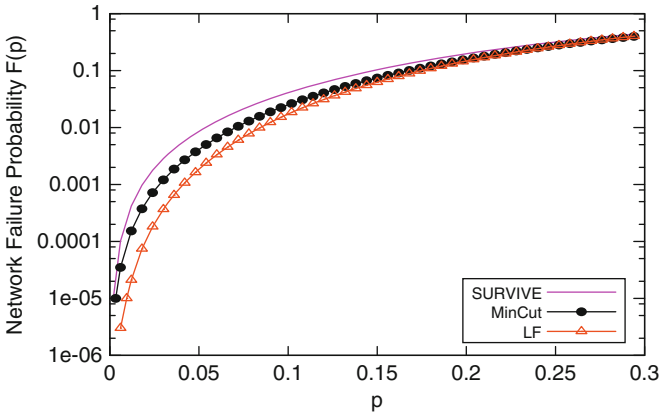


Fig. 12.7 Network failure probabilities of three different lightpath routing algorithms

routing algorithms SURVIVE, MinCut and LF presented in the previous section. For each lightpath routing generated by the algorithms, the failure polynomial is computed and compared.

The network failure probabilities of the three different lightpath routing algorithms are shown in Fig. 12.7, where for each algorithm, network failure probabilities were averaged over 350 different scenarios. When p is small, the two routings MinCut and LF which attempt to maximize the MCLC value are clearly more reliable than the SURVIVE algorithm. Note that in the lower failure probability regime, the algorithm LF whose MCLC value is higher finds a more reliable lightpath routing than the algorithm MinCut. This verifies that maximizing MCLC is a good strategy for maximizing reliability in the low failure probability regime.

4 Future Directions

In this chapter, we reviewed recent advances in cross-layer survivability. In particular, we introduced several metrics that measure the survivability of a layered network, and discussed survivable lightpath routing algorithms based on these metrics. These metrics capture the fundamentals of cross-layer survivability. We believe that many results are yet to be discovered in this context, and envision that the metrics discussed in this chapter will play an important role toward a theory of cross-layer survivability.

While this chapter focused on the role of lightpath routing in cross-layer survivability, the survivable network design problem in a layered setting consists of three components: logical topology design, physical topology design, and lightpath routing algorithm design. Obviously, the connectivity performance of a

layered network is limited by the logical and physical topology. For instance, the MCLC value of a layered network is no greater than the min-cut value of either the logical or physical topology. Therefore, for survivable layered network design, it is necessary to have logical and physical topologies that allow *good* light path routing. Note, however, that logical and physical topologies with better connectivity do not necessarily guarantee a more survivable layered network because there may not exist a mapping of the logical topology to physical topology that leads to better survivability. Therefore, when designing a physical topology, the logical topology should be taken into account and vice versa. As a consequence, the results in the survivable single layer network design may not be applicable to the survivable logical and physical topology design problem. This makes the topology design problem an interesting problem for future research.

Indeed, addressing the topology design problem in the layered setting has been largely unexplored. In [16], necessary conditions on physical topologies were developed to ensure that a ring logical topology can be embedded and survive a single fiber failure. These conditions are then used to find lower bounds on the number of physical links needed for such an embedding to exist. Despite this work, the problem of topology design remains largely unexplored.

Acknowledgements This work was supported by NSF grants CNS-0626781 and CNS-0830961 and by DTRA grants HDTRA1-07-1-0004 and HDTRA-09-1-0050. Hyang-Won Lee was supported in part by the Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A1012610).

References

1. Ahuja RK, Magnanti TL, Orlin JB (1993) Network flows: theory, algorithms, and applications. Prentice-Hall, Englewood Cliffs, NJ
2. Armitage J, Crochat O, Boudec JYL (1997) Design of a survivable WDM photonic network. In: Proceedings of the sixteenth IEEE International Conference on Computer Communications, Kobe, Japan
3. Ball M (1980) Complexity of network reliability computations. Networks 10:153–165
4. Banerjee D, Mukherjee B (1996) A practical approach for routing and wavelength assignment in large wavelength-routed optical networks. IEEE J Sel Area Comm 14(5):903–908
5. Bertsimas D, Tsitsiklis JN (1997) Introduction to linear optimization. Athena Scientific, Belmont, Massachusetts
6. Crochat O, Boudec JYL (1998) Design protection for WDM optical networks. IEEE J Sel Area Comm 16(7):1158–1165
7. Crochat O, Boudec JYL, Gerstel O (2000) Protection interoperability for WDM optical networks. IEEE ACM Trans Netw 8(3):384–395
8. Deng Q, Sasaki G, Su CF (2002) Survivable IP over WDM: a mathematical programming problem formulation. In: Proceedings of the 40th Allerton conference on communication, control and computing, Monticello, IL
9. Elhanany I, Kahane M, Sadot D (2001) Packet scheduling in next-generation multiterabit networks. Computer 34(4):104–106

10. Kan DDJ (2003) Design of survivable IP-over-WDM networks: Providing protection and restoration at the electronic layer. Master's Thesis, Massachusetts Institute of Technology
11. Kurant M, Thiran P (2006) Survivable routing in IP-over-WDM networks in the presence of multiple failures. In: EuroNGI workshop on traffic engineering, protection and restoration for NGI, Kraków, Poland
12. Kurant M, Thiran P (2007) Survivable routing of mesh topologies in IP-over-WDM networks by recursive graph contraction. IEEE J Sel Area Comm 25(5):922–933. DOI 10.1109/JSAC.2007.070606
13. Lee K, Lee HW, Modiano E (2010) Reliability in layered networks with random link failures. In: IEEE INFOCOM. San Diego, CA
14. Lee K, Modiano E, Lee HW (2011) Cross-layer survivability in WDM-based networks. In: IEEE/ACM Trans. on Networking 19(4):1000–1013
15. Modiano E, Narula-Tam A (2002) Survivable lightpath routing: A new approach to the design of WDM-based networks. IEEE J Sel Area Comm 20(4):800–809
16. Narula-Tam A, Modiano E, Brzezinski A (2004) Physical topology design for survivable routing of logical rings in wdm-based networks. IEEE J Sel Area Comm 22(8):1525–1583
17. Raghavan P, Tompson CD (1987) Randomized rounding: a technique for provably good algorithms and algorithmic proofs. Combinatorica 7(4):365–374. DOI http://dx.doi.org/10.1007/BF02579324
18. Todimala A, Ramamurthy B (2004) Survivable virtual topology routing under Shared Risk Link Groups in WDM networks. In: BROADNETS '04: Proceedings of the first international conference on broadband networks, pp. 130–139. IEEE Computer Society, Washington, DC, USA. DOI http://dx.doi.org/10.1109/BROADNETS.2004.81

Chapter 13
Photonic Grids and Clouds

Georgios Zervas and Chinwe Abosi

Photonic Grids and Clouds represent a global system composed of multiple infrastructure providers that own the networking infrastructure and the heterogeneous large-scale distributed IT infrastructure that it interconnects at its edges, to provide global value-added services to its users [1]. The structure of this global system is a result of the explosive growth and diversity of data traversing the Internet [2] and the growing penetration in terms of number and diversity of users, which has led to the growth in geographical distribution, heterogeneity and number of networked devices [1]. A significant number of users in this environment are only concerned with the proper delivery of the service and the quality of the service itself and are oblivious to the location of the resource or owner of the resource infrastructure. However, they place high expectations on the quality of service in terms of predictability, consistency, determinism and reliability [3].

Grid and Cloud computing applications refer to a class of applications that require a range of distributed IT services, such as computational and storage services, over a network. To enable these applications, IT resources need to be interconnected by dynamic high-speed and flexible networks. In addition, the underlying infrastructure needs to address unique characteristics of these applications such as the geographic distribution and the heterogeneity of resources required (network and different types of IT resources). Furthermore, it needs to support the variable usage patterns and real-time characteristics imposed by these network-centric applications. The introduction of Cloud computing has seen the number of users demanding distributed IT resources increased. These users are increasingly demanding deterministic performance in terms of consistency and repeatability.

G. Zervas (✉)
University of Bristol, Bristol, UK
e-mail: georgios.zervas@bristol.ac.uk

C. Abosi
University of Essex, Colchester, UK
e-mail: ceabosi@gmail.com

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*,
Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_13,

The photonic core network, enabled by its recent developments and technologies, is vital to fulfilling the requirements of Grid and Cloud computing, in particular wavelength-division multiplexing (WDM), multi-granular flexible technologies and dynamic network provisioning. WDM enables large numbers of wavelengths to be carried over a physical fiber, thus increasing its available bandwidth. Multi-granular technologies allow for flexible bandwidth allocations. Dynamic control planes facilitate dynamic network provisioning. These advancements in the photonic core allow Grid and Cloud applications to share network infrastructure flexibly, dynamically and in an on-demand manner.

This chapter introduces network-centric applications including Grid and Cloud computing applications and presents the diverse underlying infrastructure required by these network-centric applications. Traditionally, resource management for Grid and Cloud computing considers IT resources as the main element in delivering services. However, the need to move towards a collaborative network and IT service-oriented approach has been proposed to improve service delivery. Orchestration of the network and IT resources required for Grid and Cloud computing services provides a more optimum solution since the network infrastructure is treated as a first-class resource. In turn, this influences the performance of the application.

In order to manage the services in this infrastructure environment, it is important to look at an architectural model involving different layers such as the optical transport, control and service/middleware and the application layer. In addition, this chapter presents a description of service delivery architectural models stretching from the traditional to collaborative solutions. Collaboration can be implemented either on the service/middleware layer or at the control plane level. At the service layer, the implementation can be addressed in two ways. The first can be through middleware extensions and the second through a separate service layer, which encompasses the traditional middleware and a novel service plane. The service plane interfaces northbound with the middleware and southbound with the network control plane. Both implementations consider network and IT resources as equally as first-class resources. Two middleware extensions, G-Lambda and EnLIGHTened, and one service layer implementation, SOAFI, are reported. At the control plane layer, control plane extensions required to deliver network services need to be aware of the IT resources required in the Grid and Cloud computing environment. G^2MPLS is reported in terms of its architecture as well as the protocol extensions required to support and deliver a Grid-enabled control plane.

Finally, the heterogeneity and diversity of Grid and Cloud computing application requirements demand a flexible optical transport network able to provide flexible network services. Thus, this chapter introduces flexible multi-granular, transport solutions as well as sub-wavelength transport networks to cover the required diverse transport requirements.

1 Applications and Infrastructure

1.1 Applications: Grids, Clouds and Other Network-Centric Applications

Emerging applications tend to be highly distributed and resource intensive, addressing a wide base of user communities with different requirements [1, 4–6]. These applications are characterised by the exchange of massive amounts of data, high level of interactivity, remote high- and ultrahigh-definition visualisation, intensive distributed computations and high-capacity distributed storage [2, 3, 7]. They have been identified as network-centric distributed applications and include Grid and Cloud computing applications and are driving the evolution of Internet. Three major categories of these applications are:

1. *High-performance distributed scientific applications* (Grid, e-Science) [4, 8]: Grid computing [4] and e-Science applications [8] are two representative examples of this category. Grid computing is a paradigm that allows users to access remote computing clusters/supercomputers, high-capacity (petabyte) storage, remote visualisation and other high-end scientific and IT resources. e-Science applications include high-end physics such as the Large Hadron Collider (LHC) experiments at CERN [9, 10]; radio astronomy such as the very long baseline interferometry (VLBI) application [11], AstroGrid [12]; computational biology [13]; collaborative visualisation and data-mining applications such as OptIPuter project [14].
2. *On-demand, transparent, distributed applications* (Cloud computing applications) [15, 16]: Cloud computing applications [16, 17] are transparent applications that service a large number of simultaneous user requests across a network. These applications do not require users to know the physical location and configuration of the IT resource or system that delivers the services it requires. In Cloud computing, resources such as databases and servers are transparently provided to end users through the Internet. Cloud computing incorporates software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS).

    - Software as a service (SaaS) provides access to already developed, ready-to-use computer applications over the Internet, without the need for users to install and run the application on their own computers, as is done traditionally [18].
    - Platform as a service (PaaS) provides all the necessary entities required to develop, run, deploy, host and test applications over the Internet [18]. It includes hardware, operating systems, databases, middleware, Web servers and other software.
    - Infrastructure as a service (IaaS) allows organisations that own physical resources such as storage, computing and networking resources to provide virtualised instances of their resources to clients as on-demand services [18, 19].

The Cloud computing paradigm shares similar visions of resource sharing in dynamic, distributed environments [20] with Grid computing. However, one way in which they differ is in the way requests are made and the type and frequency of job requests. Cloud computing mostly deals with large amounts of small-to-medium on-demand job requests (small-to-medium-sized data sets) from everyday users, while Grid computing deals with small amounts of large data sets, which may be requested on demand or in advance by scientific users. While Cloud computing may not be as data intensive as Grid computing, they address a large number of users, thus requiring significant amounts of aggregated bandwidth and IT resources [20, 21].

3. *Networked media applications* [22, 23]: These are expected to be one of the dominant applications of the future Internet [6, 24] and include any type of media that needs to traverse a network environment to be processed, distributed, shared, managed and/or consumed [22, 24, 25]. Examples include ultrahigh-definition TV broadcasting that requires more than 33M pixels (7,680 × 4,820) of data [6, 26]. Depending on format and compression characteristics, applications might require up to 72 Gbps bandwidth per view [6, 27]. Other applications include high-definition (HD) interactive TV (1,920 × 1,080) and super HD TV (4,096 × 2,160 pixels) that are used for digital cinema, 3D telepresence and 3D media [22].

2 Infrastructure: IT and Network Infrastructure

The transfer of high volumes of data required by these applications to distributed sites for storage, processing, visualisation or streaming as well as the potentially huge number of users poses new and significant challenges to the underlying infrastructure [28, 29]. Furthermore, these resource-hungry applications tend to require dynamism and flexibility. In addition, users are demanding higher level of service predictability and reliable network and IT resource performance [3]. Consequently, there is a need for flexible high-capacity networks that are capable of supporting the diverse network requirements interconnecting the distributed heterogeneous IT resource sites [28–30].

The term infrastructure refers to the physical resources (network bandwidth, computational power and storage space), as well as the equipment and devices that host and support these resources. The equipment, devices and, hence, the resources may be located at geographically different, distant locations and interconnected by network links. We can characterise the resources as follows:

- *Physical IT infrastructure*: These refer to all nonnetwork elements (resources as well as equipment and devices) within the Internet that are required and used by applications. These include storage, computational, visualisation and sensors, which have traditionally been used in Grid computing as well as media-related infrastructure.

- *Physical optical network infrastructure*: These refer to transmission, switching, grooming and general processing technologies including optical as well as optoelectronic subsystems and devices required to form WDM network system. It also includes different types of bandwidth connectivity services, as well as the bandwidth capacity that the connectivity services can provide.

3 Frameworks: Photonic Grids and Clouds

3.1 Traditional Architectural Models

Figure 13.1 illustrates an IT architecture typical of Grid environments. IT resource management systems are a key component of distributed IT environments. Dynamic IT resource management systems include the local resource management system (LRMS) and, in the case of resources that are part of the Grid, the Grid middleware. Grid middleware allows IT systems belonging to different administrative domains to interoperate by performing co-allocation of resources [31]. They manage and coordinate a pool of resources, such as processing and storage, which may belong to different providers. Grid middleware supports the management of heterogeneous resources, while LRMS only performs domain-level management of homogeneous resources.

Examples of IT management resource systems include the Globus [4, 32] and Condor-G [4, 33]. Other examples of IT resource management systems include UNICORE (Uniform Interface to Computing Resources) [34], the virtual Grid [35] and Legion [36]. Currently, none of these dynamic IT resource management systems support optimal joint scheduling of optical network and IT resources [37, 38] or the management of resources required for networked media-streaming applications. Most Grid middleware only manages Grid-related IT resources such as computing and storage and do not consider network resources when matching jobs to IT resources. In these cases, once the middleware has located the required IT resources for a job, it sends a connectivity request to the network control plane to reserve the required network resources needed to deliver the end-to-end service.

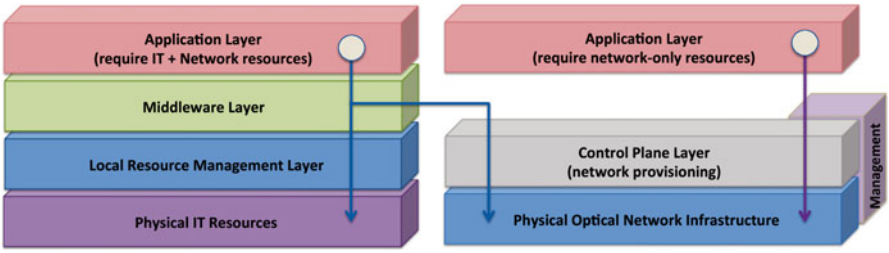


Fig. 13.1 Traditional Grid architecture

Moreover, interoperability between Grid middleware has been a major issue for the Grid community [39].

The description of each layer as shown in Fig. 13.1 is provided below:

- The top layer, the *application layer,* is where users make requests for specific and unique capabilities to satisfy different application requirements. This plane shields the user from the sharing and managing of resources.
- On the other hand, the bottom layer, the *infrastructure layer,* provides the resources that satisfy the application requirements. It can be either (a) an IT infrastructure, which is the set of computing clusters and storage discs that process or store data as required by users or (b) an optical network infrastructure, which is the set of physical network resources that are responsible for receiving, sending and switching user data through the network.
- The middle layers provide the dynamic provisioning of the infrastructure. The IT infrastructure is managed and provisioned by the LRMS and the middleware as discussed above. The network is managed and provisioned with the help of the network control plane and the management plane.
  - The network *control plane layer* is responsible for the routing, path computation, wavelength assignment and signalling functions. It also maintains topology information, which is used when abstracting information.
  - The local resource *management layer* is responsible for monitoring and managing the underlying infrastructure. In the case of the IT infrastructure, the management plane is composed of the LRMS of each domain. It is responsible for mapping a user's task graph to individual IT resource nodes that exist within each IT resource site. In the case of the network infrastructure, this is done by the network management system (NMS). It is responsible for monitoring and managing the network including controlling configurations, managing faults and maintaining and troubleshooting the network.
  - The *middleware* is unique to the IT Infrastructure. It provides a unified platform as a single point of reference where resources of various providers can be stored and/or accessed in an automated way. It is responsible for locating, coordinating and selecting resources from various providers. Thus, it acts as a "matchmaker" between users and the multiple IT resource providers that may have the resources they require.

3.2 Network and IT Service Delivery Frameworks

The quality of service achieved by service delivery solutions depends strongly on the resource management technique used particularly so in the future Internet, which is characterised by a networked IT infrastructure. It is expected that service delivery techniques using the traditional architecture, which implements separate resource management techniques, will result in less optimal performance. In the

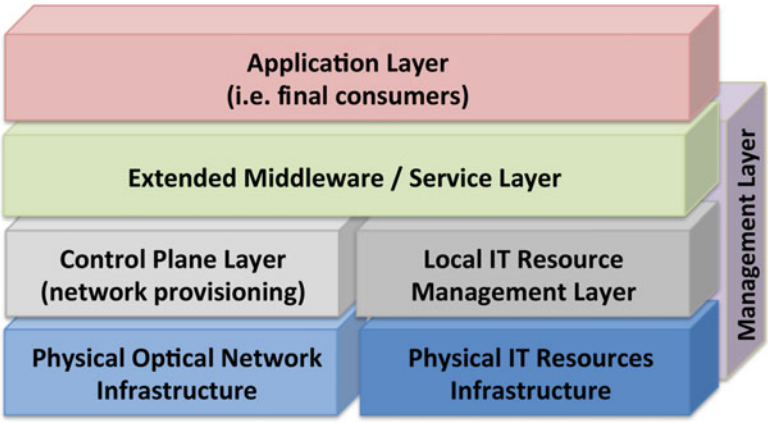


Fig. 13.2 Collaborative network and IT service delivery architecture

traditional architecture, each provider with the required resources is contacted individually. To elaborate, in a networked IT environment, if a user requires IT resources located at a remote site, the user first queries the IT resource management system, for example, a Grid middleware, for a set of available IT resources located at a particular site. The network provisioning system is then contacted to allocate network connectivity to the discovered IT resource site as shown in Fig. 13.1. However, the network provisioning system may not be able to realise the required connectivity for a number of reasons including insufficient bandwidth capacity or network link failures. In some cases, the user may repeat the request until it finds a suitable network IT configuration. Although this repetition may increase the probability of a user's request being satisfied, it does not allow for an optimised service delivery [40].

Advances in literature introduced the concept of managing network resources as a first-class resource into the IT resource management system [31, 41]. This collaborative method of service delivery considers all resource types required by a user in a single step. Collaborative solutions include a separate layer, which is either an extended middleware or a service layer where the collaboration and orchestration of the network and IT resources take place. This layer interacts with the management and control planes of the underlying network and IT infrastructure. A typical collaborative architecture is shown in Fig. 13.2. The description of the application, infrastructure, control and management layers are as described in Sect. 3.1. The service layer or extended middleware extends the traditional middleware layer to jointly consider network and IT resources and depend on the information it receives from the management and control planes. Thus, it creates a collaborative environment that facilitates the interaction between both network and IT resources and the providers that provide, control and manage these resources.

4 Service Layer: Extended Middleware and Service-Oriented Solutions

4.1 Extended Middleware Solutions

In extended middleware solutions, standard middleware such as the Globus Toolkit [32] is extended to consider network resources as a first-class resource. With this approach, network and IT resources are co-allocated in the middleware environment. Examples of extended middleware solutions include the EnLIGHTened [42, 43] and G-Lambda [42, 44] solutions.

The EnLIGHTened middleware [42, 43] focuses on the scheduling and reservation of both network and IT resources across multiple domains. In EnLIGHTened, resources can be requested in advance or on demand. EnLIGHTened uses a collaborative resource management scheme that includes an EnLIGHTened Resource Broker (ERB), which contains the Highly-Available Resource Co-allocator (HARC) system to manage resource co-allocation. The HARC allows users to reserve multiple heterogeneous, distributed resources in a single step. A user makes a request to the ERB, which gets availability from each of the local managers of the IT and network resources. Using information stored in its resource registry and the dynamic resource information collected by its monitoring and discovery system, it uses its meta-scheduler to find appropriate resources that are able to satisfy the request. The HARC then schedules, activates and deactivates the selected IT resources and network lightpaths through the local managers that provide interfaces to the chosen resources.

G-Lambda middleware [42, 44] introduces a Grid Resource Scheduler (GRS) that acts as a meta-scheduler to collaboratively manage network and IT resources as requested by users. The GRS is implemented using the Globus Toolkit 4 [32]. One of the main goals of the G-Lambda middleware is the definition of a standard Web interface between the GRS and the network resource management system (similar to the ERB and domain network scheduler of the EnLIGHTened middleware). The Web interface, called the Grid Network Service–Web Service Interface (GNS–WSI), interacts via messages with the network resource manager to manage lightpaths (activate, deactivate and query status) as required. It also defines a Web service interface for users to interact with the GRS.

G-Lambda middleware and EnLIGHTened middleware are implemented using the Globus Toolkit [32], which is extended to take network resource information into account as a first-class resource. It co-allocates all the necessary types of resources within the middleware environment to achieve a more efficient allocation of resources.

4.2 Service-Oriented Solution: The SOAFI Framework

Service-oriented solutions leverage on the service orientation principles as defined by the Organization for the Advancement of Structured Information Standards (OASIS) [45]. The SOAFI framework [46] introduces a service-oriented solution

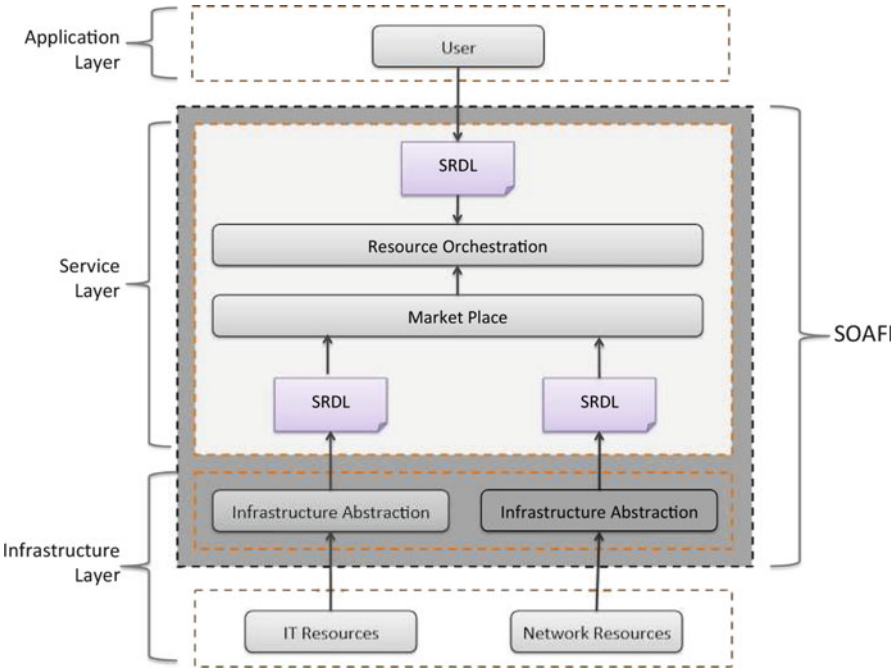


Fig. 13.3 SOAFI framework

that unifies and encapsulates the multiple providers that provide resources as well as the users that use these resources, under a single architecture to enable autonomic management of the heterogeneous resources that belong to the different providers. Similar to the extended middleware, it includes a separate platform for the co-selection and co-scheduling of resources. This separate generic platform, the service layer, consists of the traditional middleware and a novel service plane. The service plane coordinates and orchestrates different types of resources (network, IT and/or media) as required to achieve optimal provisioning of services. Within the photonic Grid and Cloud computing environment, SOAFI interacts with and coordinates the IT resources through the Grid/Cloud middleware, the network resources through the GMPLS control plane as well as the users that require services across the optical network.

The basic structure of the framework is shown in Fig. 13.3.

4.2.1 SOAFI Design Requirements

The SOAFI framework aims to address the following challenges of Grid and Cloud computing environments: (a) the resources belong to different independent infrastructure providers, (b) the resource types are heterogeneous and (c) the coordination

of the different infrastructure providers through orchestration of their resource is vital for optimality. Thus, the SOAFI framework is designed with the following requirements:

- Provide a means to share information about resources between providers and users
- Provide mechanisms that enable the viewing, scheduling and monitoring of network and IT resources to satisfy a user's request
- Facilitate the provisioning of services across the disparate providers
- Exploit the distributed nature of the underlying IT infrastructure by being able to satisfy requests on remote IT resource sites
- Allow for and support the independent advancement and development of each of the entities involved in service provisioning (i.e. users and providers) in an independent manner

Characteristics of the SOAFI Framework

Based on these requirements, the SOAFI framework was designed with five primary characteristics:

1. It encapsulates the actors within the photonic Grid and Cloud environment, which are the users and infrastructure providers, into an independent, collaborative service platform. This supports the integration of multiple providers and achieves better performance for users and better utilisation of resources by facilitating the interaction between both network and IT resources and the providers that provide, control and manage these resources.
2. It provides a mechanism that facilitates and encourages the sharing of infrastructure resource information by allowing providers to summarise and hide details of their infrastructures.
3. It introduces a semantic resource/request description language to ensure that there is uniformity in the way resources and requests are described to facilitate automatic and autonomic matching of needs and capabilities.
4. It provides a marketplace, which acts as a global database to store the descriptions of the resources from multiple providers through their middleware/control plane.
5. It provides mechanisms through orchestration algorithms that use the global information in the marketplace to optimally select and match the best providers for each request along with their respective resources. The aim of the orchestration mechanism is to optimise network utilisation and minimise blocking probability.

These characteristics of the SOAFI framework are encompassed within two key entities, namely, (a) infrastructure abstraction entity and (b) the service plane.

The infrastructure abstraction entity provides a mechanism for providers to confidently share structurally abstracted versions of their infrastructure. Infrastructure

abstraction addresses resource heterogeneity, scalability and confidentiality issues that may impede the sharing of distributed resources [47]. It provides a means for infrastructure providers to represent their resources in a summarised, scalable manner, suitable for sharing across multiple, geographically distributed resource sites. This process of summarising of domain-level information into a reduced approximation of the physical resources also allows infrastructure providers to hide the technology details and complexity of their infrastructure resources. In this way, infrastructure abstraction allows infrastructure providers to confidently expose structurally abstract, simplified versions of their infrastructure in a scalable manner. The resulting abstract representation should provide enough information to the service plane, enabling it to compose services over the multiple infrastructure domains at a targeted quality of service.

The service plane encompasses the semantic resource/request description language, the resource orchestration module and the marketplace:

- The semantic request/resource description language represents user's requirements as well as network and IT resources in a uniform format.
- The resource marketplace stores global resource information from multiple providers.
- The resource orchestration entity is concerned with locating and orchestrating resources from multiple providers in order to compose a service required by a user's request. It is responsible for coordinating providers by selecting resources stored in the market place. It aims to select the "best" or most optimal IT resource site and network path to the selected resource site to satisfy a job request.

The SOAFI Workflow

The workflow of the SOAFI framework consists of a number of steps as shown in Fig. 13.4. First, the infrastructure providers furnish the marketplace with their abstracted resources information. To make the service framework understand and reason about the available resources, the abstracted resource information is translated into a standard semantic resource/request description language [48], which is based on the Web Service Modelling Ontology [49], before being stored into the marketplace.

To make a request, a user sends its requirements to the SOAFI framework. The user can either make a detailed request or an abstract request. In the case of a detailed request, the user specifies detailed information of the resources it requires—for example, the user can indicate the source end points; the destination end point, which is either a storage, computational or server site; the duration of the connection; the start time of the connection; the required storage or processing capacity; and the duration required for storage and/or processing. In the case of an abstract request, some information about the request is omitted. For example, a user may choose not to specify the destination end point or the duration of the request

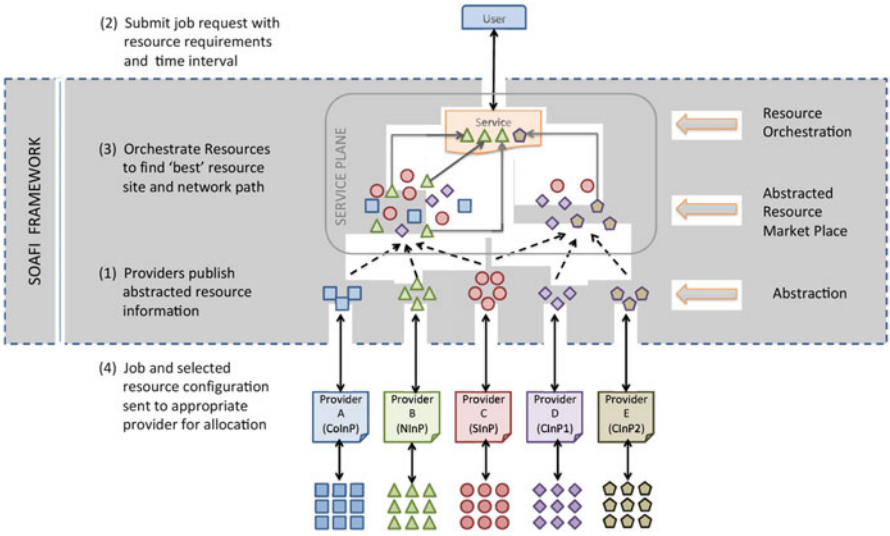


Fig. 13.4 Procedure of orchestrating a user's request in the SOAFI framework

but may choose to specify the amount of data to be transferred. The SOAFI framework also translates the request using the SRDL.

The resource orchestration module performs a feasibility analysis and determines a resource configuration that can optimally address the user's request. SOAFI then decomposes its solution and invokes the middleware and control plane of the appropriate providers to provision the request using the selected resources.

The provisioning then advances in the traditional manner: the control plane receives a simple connection request, while the middleware receives an IT (processing/storage) request. Through the resource orchestration algorithm, the SOAFI framework is capable of optimally selecting resources in order to maximise users' satisfaction by minimising blocking and to maximise provider's satisfaction by optimising resource utilisation.

5 Control Plane: Grid-Enabled GMPLS Control Plane

In the IST Phosphorus project (http://www.ist-phosphorus.org/), the solution adopted for the implementation of the Grid-enabled network control plane (NCP) considers a Grid evolution of GMPLS protocols, namely, Grid-GMPLS (G^2MPLS). G^2MPLS procedures serve Grid jobs by co-allocating and provisioning network and Grid resources in a single step. The G^2MPLS architecture is expected to expose interfaces specific for Grid computing and is made of a set of extensions to the standard ASON/GMPLS architecture. ASON/GMPLS specifies the architecture applicable to SDH transport networks and Optical Transport Networks. As such,

G^2MPLS built on top of ASON/GMPLS is primarily focused on the issues of seamless automatic control of Grid and network services in optical networks. Therefore, G^2MPLS results in a more powerful NCP solution than the standard ASON/GMPLS, because it complies with the needs for enhanced network and Grid services required by network-centric users/applications. The enrichment of G^2MPLS is driven by procedures, languages and schemas standardised by the Open Grid Forum (OGF) and OASIS and thus is not conceived to be an application-specific architecture. Nevertheless, the requirements of standard users that only require the automatic setup and resiliency of their connections across the transport network are still supported by the backward compatibility of G^2MPLS with standard GMPLS. G^2MPLS provides part of the functionalities related to the selection, co-allocation and maintenance of both Grid and network resources in the same tier, guaranteeing service availability and tailoring to the user requirements. As such, G^2MPLS as a single horizontal entity is able to manage and control both network and IT resources to deliver seamless Grid network services.

The G^2MPLS NCP is innovative due to availability of well-established procedures for traffic engineering, resiliency and crankback and uniform interface (G.OUNI) for the Grid user to trigger Grid and network transactions not natively dependent on a specific Grid middleware to deliver fast service delivery. Moreover, the compliance of G^2MPLS to the ASON/GMPLS architectures fosters for possible integration of Grid computing in real operational networks, by overcoming the current limitation of Grid computing operating as stand-alone networks with their own administrative ownership and procedures.

5.1 Phosphorus Network and Service Architectural Model

The Phosphorus framework identifies different layering solutions as illustrated in Fig. 13.5 with respect to the positioning between Grid service layer and network control plane. Layers involved are the following:

- Grid layer
- Network control plane
- Transport plane

The Grid layer comprises Grid users/applications, Grid resources and Grid middleware. Within the Phosphorus perspective, the associated functionalities of the Grid layer are delivered to/by the underlying network control and management planes. Finally, the transport plane is the basic layer comprising all the data bearing equipment and their configuration interfaces.

The functionalities associated to each layer and the relationship between them constitute the type of Phosphorus model either overlay or integrated. In the G^2MPLS overlay model, the Grid layer has both Grid and network routing knowledge. G^2MPLS provides automatic configuration for the network service part only; moreover, it acts as an information bearer of network and Grid resources.

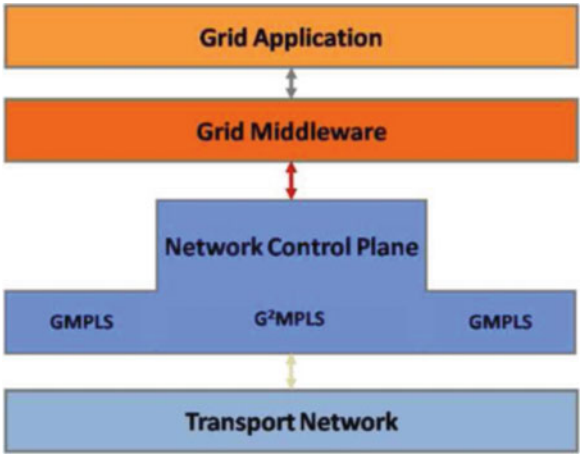


Fig. 13.5 Grid network-layered architecture

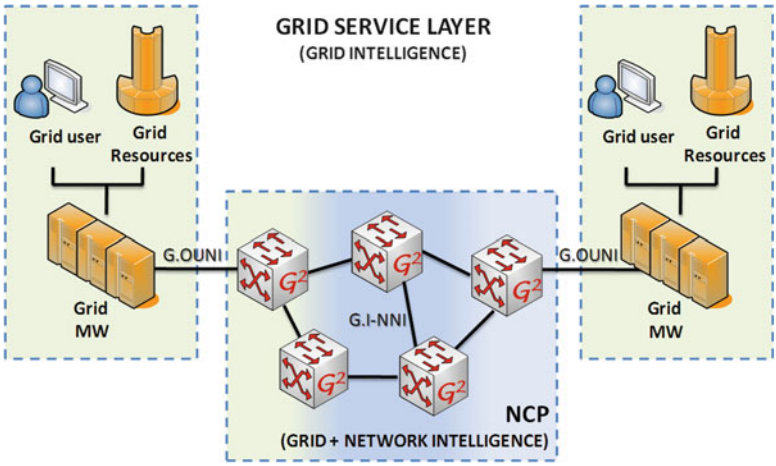


Fig. 13.6 G^2MPLS integrated model

The integrated model refers principally to the positioning between the Grid service layer and the NCP and requires specific capabilities of the G^2MPLS NCP.

In the G^2MPLS integrated model, most of the co-allocation functionalities are moved to the NCP (Fig. 13.6). G^2MPLS is responsible for scheduling and configuring all the job parts, those related to the Grid sites and those related to the network. The Grid scheduler functionality is still needed to coordinate workflow services, because G^2MPLS NCP is capable of managing just the workflow elementary unit, that is, the Grid job. In this model, the role of the network interfaces is scoped to Grid Network Service creation, which implies that Grid information concerning Grid resource availability (routing) and job description data (signalling)

becomes transparent at those interfaces in which a decision process needs to be provided: these are the G.OUNI (defined in Sect. 5.3), by which a G^{2}MPLS is entered, and the G.E-NNI, by which the border between domains is traversed.

The G^{2}MPLS NCP architecture sets analogous reference points with respect to the ASON/GMPLS, with evolved network interfaces capable of managing and advertising the semantics of both Grid and network resources. Network interfaces in the scope of Phosphorus are detailed in Sect. 5.3.

5.2 *Phosphorus Service Provisioning Defined to Address Grid Network Requirements*

The implementation of Grid Network Services (GNS) [50] for the G^{2}MPLS integrated model poses a number of requirements on the control planes of the underlying network infrastructure. To match these requirements, the Phosphorus control plane proposes a number of Grid-enabled GMPLS procedures, a comprehensive description of which is discussed in the remainder of this section.

5.2.1 GNS Discovery Procedures

Grid service discovery is essential for any NCP solution supporting GNS. The NCP provides mechanisms for the negotiation of Grid and network services, configurable across the interface between the Grid user/site and the network. The service discovery mechanism includes network-specific resources and operation modes (e.g. types of signals, protocols, routing diversity (point-to-point, point-to-multipoint, multipoint-to-point, multipoint-to-multipoint), amount of bandwidth) as well as Grid-specific capabilities and resources (e.g. amount and types of CPU, storage, OS)

In G^{2}MPLS, the functional entity responsible for discovering Grid and network capabilities between a Grid site attached to the G.OUNI (client side) and the network control plane is the GNS Discovery Agent (G-SDA). Discovery of capabilities is generally referred to as service discovery, and this functional entity is conceived as an extension of the standard service discovery functionality described in [51]. The main actions of the G-SDA can be classified into two categories. The standard service discovery includes negotiation of the signalling protocol and its version to be used across the G.OUNI, correlation of the service attributes of all the transport links connecting the Grid site to the network (e.g. encoding type, signal types) and discovery of network capabilities (e.g. transparency in case of SONET/SDH, routing capability). In addition, an auto-discovery mechanism allowing the quick discovery of available resources or acquiring updates when resources become available or unavailable in the domain is an important feature supported by G^{2}MPLS. The result of auto-discovery is the identification of connectivity between the client and the network and the available network services.

5.2.2 Path Computation Issues

Enhanced routing is an essential functionality in the G^{2}MPLS control plane allowing the consideration of both Grid computing and network parameters into the path computation process. The Phosphorus control plane is based on a distributed path computation approach in which the path computation element (PCE) [52] is responsible for providing inter- and intra-domain routes based on specified constraints (e.g. network, Grid application requirements). The PCE is a functional entity implementing the routing algorithm on the stored and updated topology view of Grid and network resources. The computed path scope may range from a portion of a route to the full end-to-end path across a chain of domains. The detailed information of the topology stored in the PCE and consequently of the computed routes depends on the adopted routing policy, that is, the amount of information that each network operator configures and publishes internally (i.e. in its domain) and towards the neighbouring domains.

5.2.3 GNS Setup and Restoration

GNSs require rapid setup and restoration mechanisms to efficiently allocate Grid computing and network resources. In G^{2}MPLS control plane, the GNS setup and restoration uses the well-established GMPLS signalling procedures including crankback. In this case, crankback is essential to cope with the scenarios of missing resources in two possible time spaces. The time space of immediate reservations (as in standard GMPLS) and the time space of advance reservation with or without resource calendars are advertised by OSPF. The amount of calendar information flooded by OSPF determines the amount of control plane traffic in the network as well as the size of the link-state database (LSDB) and respectively influences the crankback that is expected to occur.

5.2.4 Advance Reservations and "Grid-Fast" Circuit Setup

Advance reservation is a reservation scheduled for future (non-instantaneous) execution [53]. Advance reservations in Grids are needed to cope with a guaranteed service at the time of execution of the job. A set of demanding Grid computing applications requires advance reservations to reserve Grid and network resources for future (non-instantaneous) execution by specifying start time and duration of the required service. This can be provided in two ways, depending on access interface:

- Start time and duration of the reservation, which guarantees that reservation will be available for specified period of time
- Start time and end time, which will make the system to tear down the circuit at specified time in the future

In order to provide feasibility of advance reservations in G^{2}MPLS and to cope with a guaranteed service at the time of execution of the job, partitioning of the transport network resources is needed to distinguish the resources to be used for bookings from those that could be used for immediate reservations (e.g. by standard ASON/GMPLS users). As reservations are scheduled for the future, a calendar instance is needed for the maintenance of resource bookings. Calendars may be kept centralised or distributed for single- or even multidomain environments and must be scalable and open solutions in order to schedule various types of resources, including not only network parameters (like bandwidth, VLAN ids, SDH time slots) but also typical Grid attributes (number of processors, amount of memory, etc.).

Apart from advance reservation setup procedures, there are also on-the-fly path-establishing procedures. These mechanisms have to create a path between Grid resources as fast as possible taking into consideration the switching time associated with the optical data plane. The "Grid-fast" circuit setup is done by the standard GMPLS exchange of the path setup messages between ingress and egress nodes.

5.2.5 Beyond Point-to-Point Services

In Grid networks, typical point-to-point connections might not be sufficient to satisfy the wide spectrum of Grid applications. For example, there are specific applications where a number of distributed users are simultaneously sending a large amount of data to a single computation point for hardware correlation. Also Grid tasks that are executed on separated cluster environments may require high-bandwidth connections between each other to synchronise computation data. In the G^{2}MPLS architecture, three types of connections are identified representing upcoming enhancements for optimal route provisioning with more than two end points: (a) point-to-multipoint, (b) multicast and (c) anycast. In anycast, data is routed to the "nearest" or "best" destination as viewed by the routing topology. The requestor does not include the destination(s) of a service request.

5.3 G^{2}MPLS Interfaces

The deployment of the enhanced G^{2}MPLS NCP sets analogous reference points (Fig. 13.7) with respect to the ones defined by ASON/GMPLS. The resulting network interfaces are a Grid-aware evolution of the standard interfaces (UNI, I-NNI, E-NNI), with a set of procedures that maintains the backward compatibility with the original ASON references but also provides a seamless and single-step control of both Grid and network resources.

Five network interfaces are identified in the G^{2}MPLS NCP:

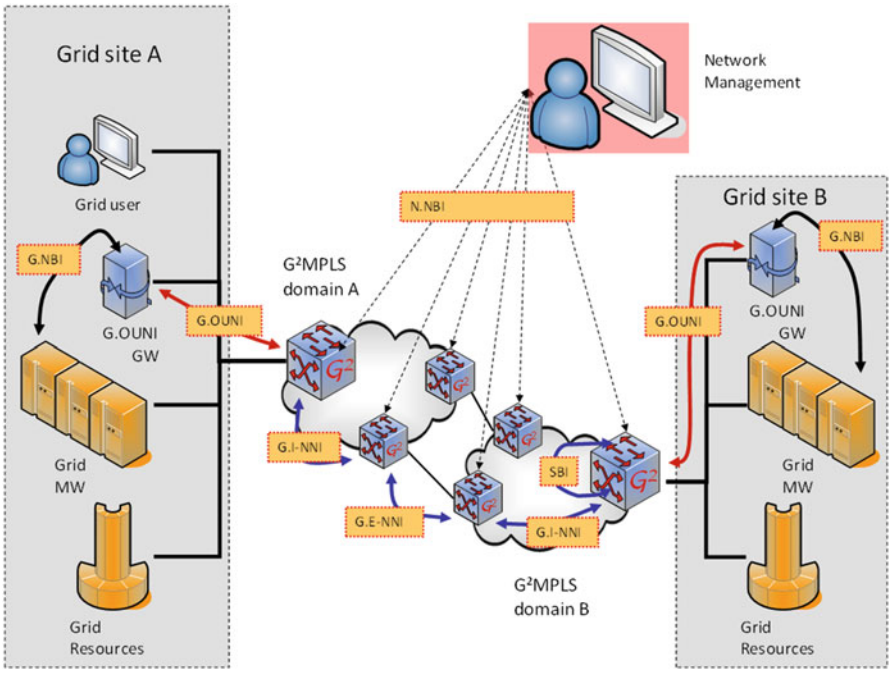


Fig. 13.7 G²MPLS interfaces

- G.OUNI, that is, the Grid Optical User Network Interface that supports Grid and network signalling and discovery between the Grid site and the G²MPLS domain. In G.OUNI, Job Submission Description Language (JSDL) [54] documents sent by the user using WS-Agreement procedures towards the network are mapped to signalling messages (RSVP-TE) at G.OUNI-C. Similarly, GLUE schema [55] used to describe IT resources (both capability and availability) is translated to routing messages (OSPF-TE) in order to publish resource information through G²MPLS NCP.
- G.I-NNI, that is, the Grid Internal Node–Node Interface (G.I-NNI) that supports the routing and signalling procedures between adjacent nodes.
- G.E-NNI, that is, the Grid External Network–Network Interface that propagates Grid and network topology information across different NCP domains and supports the inter-domain signalling mechanisms.
- SBI, that is, the southbound interface that retrieves resource status from the specific transport plane and translates NCP actions into appropriate configurations of those resources.
- NBI, that is, the northbound interface that groups two interfaces towards upper layers: one towards the Grid layer (G.NBI) and one towards the network service plane (including NRPS, N.NBI).

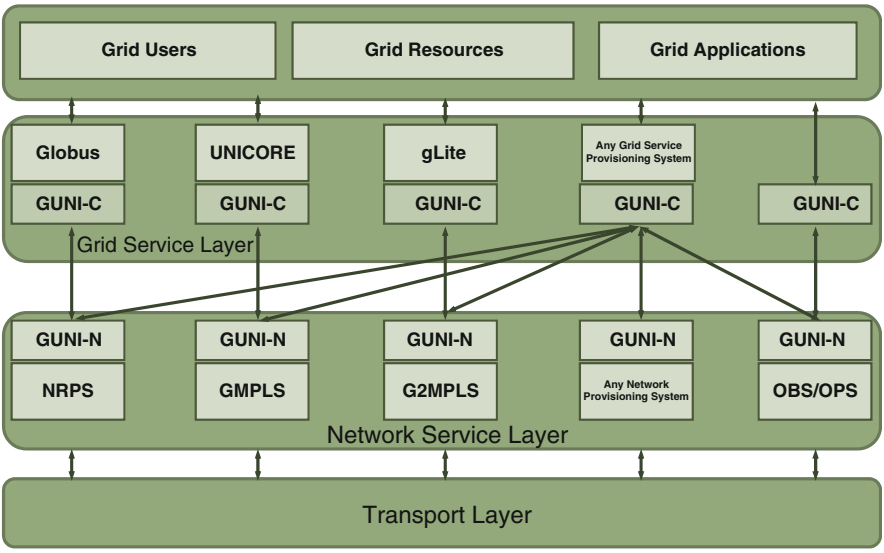


Fig. 13.8 Grid User Network Interface with Grid end points as well as Grid middleware with network provisioning systems

5.3.1 Grid/Cloud Service to Network Interface

A more visionary approach reported in OGF [56] describes a generic G.OUNI that could connect Grid users/applications/resources through any type of middleware (e.g. Globus, UNICORE, gLite) or service layer (based on SOAFI framework) with any type of service provisioning system (e.g. NRPS, GMPLS, OBS) as illustrated in Fig. 13.8. The G.OUNI describes requirements driven from GNS use cases (i.e. Phosphorus, Enlightened, G-Lambda, 3TNET [57]) and in turn provides specific G.OUNI capabilities to meet these requirements. Further developments on such interfacing aspects are addressed within the Network Service Interface Working Group (NSI-WG) in OGF [58].

5.4 G²MPLS NCP Message Flow and Protocol Extensions

Implementation of G²MPLS NCP based on the integrated model requires protocol extensions [59, 60], both signalling and routing, to support Grid services and resources. The formation of such extensions is driven by Grid service layer mechanisms (e.g. JSDL, GLUE), and the goal is to provide a seamless Grid and network service system. WS-Agreement (WS-AG) [61] is deployed to encapsulate JSDL and GLUE information and also support service-level agreement (SLA), a key element to provide and maintain a compliant interdisciplinary service system.

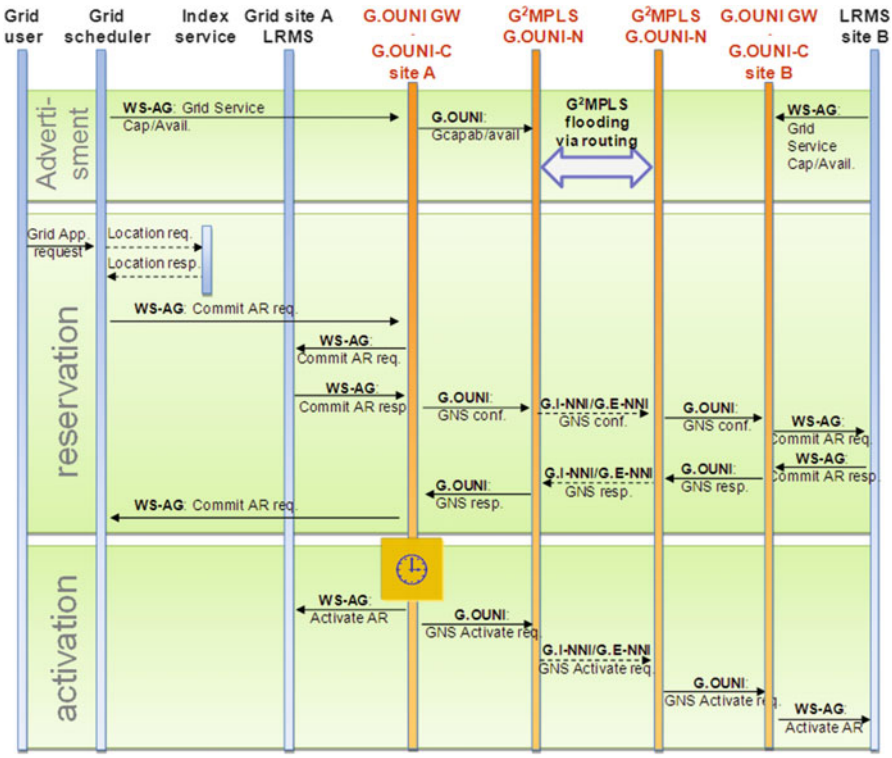


Fig. 13.9 End-to-end message flow for Grid network service provisioning over **G²MPLS** NCP

Figure 13.9 Illustrates the message flow for all required information exchange between the Grid service layer and Grid end points with G²MPLS NCP. These messages are then translated to protocol extensions (OSPF and RSVP-TE) described in the following section.

The message flow is divided into three main phases—the advertisement, reservation and finally activation. During the advertisement phase, service and resource capability and availability information is advertised from Grid sites (e.g. Grid scheduler, LRMS) to the G²MPLS NCP. Such information provided by GLUE schema is mapped to WS-AG and propagated to G.OUNI-GW (G.OUNI Gateway). The gateways aim to provide the needed bridging functionality between the two layers (Grid and NCP) and preserve the core G²MPLS /GMPLS signalling and routing procedures by concentrating in single points the adaptation functions. The G.OUNI-GW is responsible for identifying and passing such information to G.OUNI-C (G.OUNI Client) where the translation to OSPF update message containing the Grid opaque LSAs is realised. In the second phase, the GNS service discovery and reservation occur. An advance reservation (AR) application request is first initiated by the Grid user towards the Grid scheduler. A location request is then sent to the index service, and after a positive response, the Grid scheduler can initiate the AR by deploying the WS-AG. In this case, JSDL information

is mapped to the WS-AG protocol and sent to the G.OUNI-GW. The G.OUNI-GW identifies and passes it to the G.OUNI-C where the translation to extended RSVP-TE messages happens. This is propagated through various G.I-NNI or even G.E-NNI over the G^2MPLS NCP until the other end (G.OUNI-C) is reached. The path message used for reservation is then translated back to WS-AG message and forwarded to the LRMS. The LRMS sends a response to the G.OUNI-GW, which is then encapsulated to a Resv message and sent back to the initiating G.OUNI-C and, after translation to WS-AG message, to the Grid Scheduler.

6 Optical Transport and Switching Plane: Flexible Optical Network Solutions

The success of Grid and Cloud computing and the overall performance of the application(s) developed on it also depend on the physical transport network and infrastructure that interconnects the distributed resources. In fact, Cloud computing needs proper consideration for network dynamics, such as capacity, quality of service (QoS), reservation and complete integration and federation of the Cloud environment. Hence, it is vital to architect and redefine the role of optical networks, not just to carry information created by the Cloud but also to be an active and integral element of it. As such, the design of a flexible and scalable optical metro and core network to deliver ever-demanding and diverse services to users and applications is apparent.

6.1 Sub-wavelength Transparent Optical Metro Networks

The need and justification of sub-wavelength optical switching technologies in metro networks are driven by two main factors. The first one relates to evolutionary trend/growth of network traffic and emerging technologies, and the second relates to the emergence of network-centric services that require sub-wavelength bandwidth (BW) granularity, short-lived connections (secs to mins) and as such fast service delivery (few msecs) to achieve the highest BW utilisation, low end-to-end delay and multiple levels of guaranteed QoS. Such services include Video on Demand (VoD), storage area network (SAN) and a number of Cloud services, such as VirtualPC [62].

Several techniques have been proposed to deliver sub-wavelength services at metro region such as optical packet switching (OPS) and optical burst switching (OBS) [63]. However, these techniques do not provide guaranteed bandwidth services. Recently, time-driven-switched optical network [64] has been proposed to deliver sub-wavelength-switched synchronous virtual pipes. To guarantee such services, this approach uses a synchronous global common time reference from

Galileo or GPS. Also, Labelled OBS with Home Circuits (LOBS-H) [65] is an alternative solution that allocates wavelength sharable home circuits for each source–destination pair. These circuits provide guaranteed bandwidth for conforming traffic that originates from the same source to different destinations and also allow for non-guaranteed statistically multiplexed nonconforming traffic from any source to any destination. Furthermore, there has been considerable effort on routing, wavelength and time assignment (RWTA) algorithms to calculate two-way reserved time division multiplexed (TDM) wavelength services [66]. It is also worth noting that these approaches consider ring solutions [63, 67] for metro. This introduces additional complexity on how to interconnect all nodes (e.g. interconnected rings) and deliver the bandwidth services over topologies that have an inherit mesh multi-degree connectivity.

An alternative optical network solution developed within the EU FP7 MAINS project [68, 69]—the Time Shared Optical Network (TSON) [70]—delivers both highly flexible statistically multiplexed optical network infrastructure and on-demand guaranteed contention-free time-shared sub-wavelength services. It supports traffic flows from any source to any destination in transparent optical networks for the metro region supporting physical interconnection requirements. The architecture is based on user/application-driven bandwidth service requests, centralised RWTA calculation and one-way tree-based provisioning that allow for flexible symmetric/asymmetric multi-granular bandwidth services with the use of either fixed or tuneable transceivers. It delivers contention-free optical switching and transport of contiguous and noncontiguous time slices across one or multiple wavelengths per service. It also does not require global synchronisation, optical buffering and wavelength conversion, thus reducing implementation complexity. Finally standardisation efforts have been reported on a framework for GMPLS and path computation to support of sub-wavelength-switched optical network.

6.2 Multi-granular Optical Core Nodes and Networks

Colourless, directionless and contentionless reconfigurable optical add–drop multiplexers (CDC-ROADM) architectures have been proposed [71–73], and some also commercialised to deliver wavelength switching. On such nodes, ports are not associated with a specific wavelength or node degree, and multiple ports can simultaneously add/drop different channels at the same wavelength.

To extend switching flexibility, the principle of multi-granular optical cross-connects (MG-OXCs) that can switch traffic at fiber, waveband and wavelength granularities [74, 75] has been proposed to reduce the cost and complexity of traditional OXCs. The MG-OXC is a key element for routing high-speed WDM data traffic in a multi-granular optical network. A three-layer switching fabric consisting of a fiber cross-connect (FXC), a band cross-connect (BXC) and a wavelength cross-connect (WXC) was presented in [76, 77], and the application

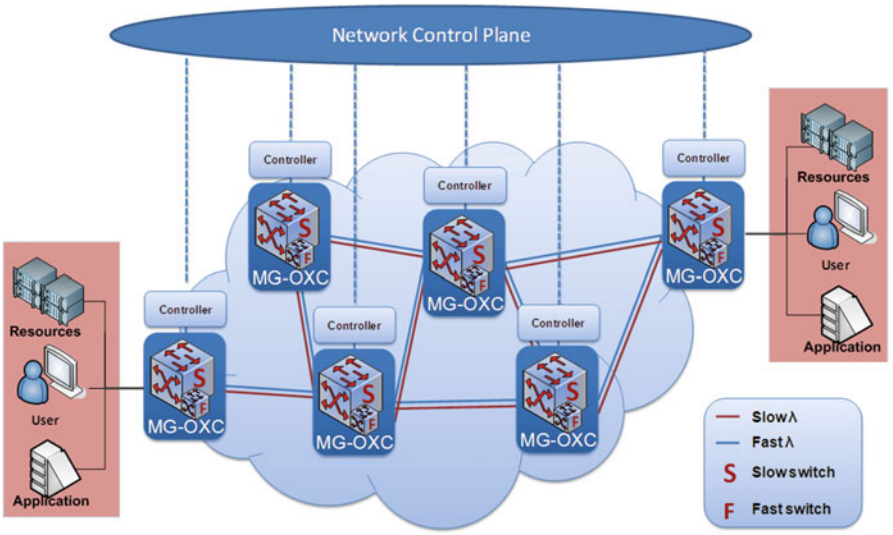


Fig. 13.10 Multi-granular optical network

of such three-layer MG-OXC architectures to metro area networks was demonstrated in [78]. Multi-granular OXCs able to support waveband and wavelength switching are commercially available [73].

Based on such nodes, multi-granular optical networks (MGON) (Fig. 13.10) [79–81] that are able to support dynamic wavelength and sub-wavelength granularities with different QoS levels have been introduced. By combining the two extensions, the full range of bandwidth granularities can be supported by a single OXC design. Moreover, it allows fast reconfigurability and flexibility on the electronic control of switching technologies while offering good cost–performance balance. However, MG-OXCs have fixed spectra associated with wavebands and wavelengths, which restrict the resource allocation flexibility and spectral efficiency of the solution. By combining the two extensions, the full range of bandwidth granularities can be supported by a single OXC design.

To design OXCs, one should keep in mind the following features:

- Switching capability of any channel to any unused channel
- Variable switching and add/drop percentage up to 100%
- Dynamic reconfiguration supporting switching speeds appropriate to transport technologies
- Transparency or bit tailored configuration
- Scalable architecture in a modular fashion
- Minimum performance degradation in terms of noise, crosstalk, filtering etc. for add, drop and switched paths and ideally uniform for all the channels
- Strictly non-blocking connectivity between input and output ports
- Simplified control

The extra specification goals when designing MG-OXCs are listed below:

- Each design must support multi-granular optical switching and consists of slow- and fast-switching fabrics.
- The architectures must be non-blocking in the sense that any input wavelength (either slow or fast) can be connected to any output fiber.
- A number of wavelengths of each fiber should be able to access the fast switch, although only a limited fiber–wavelength pairs can do so simultaneously.
- The set of fiber–wavelength pairs that have access to the fast switch should be configurable.
- The design should be able to scale in a modular fashion towards fast and slow switching on additional network requirements (new wavelength(s) and link(s)).
- The design should limit the number of fast switch elements and fast ports and still maintain high level of slow–fast reconfigurability.
- Variable slow/fast switching, which implies variable reconfigurability.

A multi-granular optical network populated with MG-OXCs is able to support dynamic wavelength and sub-wavelength bandwidth granularities with different QoS that are based on performance metrics such as throughput, latency, jitter and loss. Such network could support the three basic switching technologies in WDM networks: optical circuit switching (OCS), optical packet switching (OPS) and optical burst switching (OBS).

References

1. Anicic D, Stojanovic N (2009) Future Internet collaboration workflow. In: Domingue J, Fensel D, Traverso P (eds) Future Internet FIS 2008. Lecture Notes in Computer Science, vol 5468. Springer, Berlin, pp 141–151
2. Simeonidou D, Nejabati R, Arnaud BS, Beck M, Clarke P, Hoang DB, Hutchison D, Karmous-Edwards G, Lavian T, Leigh J, Mambretti J, Sander V, Strand J, Travostino F (2004) Optical network infrastructure for grid, global grid forum standardisation document, GFD-I.036, October 2004
3. Lehman T, Sobieski J, Jabbari B (March 2006) DRAGON: a framework for service provisioning in heterogeneous Grid networks. IEEE Commun Mag 44(3):84–90
4. Foster I, Kesselman C (eds) (2004) The grid 2—blueprint for a new computing infrastructure, 2nd edn. Elsevier/Morgan Kaufmann Publishers, Boston
5. Wosinska L, Simeonidou D, Tzanakaki A, Raffaelli C, Politi C (2009) Optical networks for the future Internet: introduction. IEEE/OSA J Opt Commun Netw 1(2):FI1–FI3
6. Sato K-I, Hasegawa H (2009) Optical networking technologies that will create future bandwidth-abundant networks [invited]. IEEE/OSA J Opt Commun Netw 1(2):A81–A93
7. O'Mahony MJ, Politi C, Klonidis D, Nejabati R, Simeonidou D (December 2006) Future optical networks. J Lightwave Technol 24(12):4684–4696
8. E-Science Institute at the University of Washington. [Online]. Available: http://escience.washington.edu/
9. European Organisation for Nuclear Research (CERN). [Online]. Available: http://public.web.cern.ch/public/
10. Large Hadron Collider (UK). [Online]. Available: http://www.lhc.ac.uk/

11. European Very-Long Baseline Interferometry (e-VLBI). [Online]. Available: http://www.evlbi.org/
12. AstroGrid. [Online]. Available: http://www.astrogrid.org/
13. Genomic science. [Online]. Available: http://genomicscience.energy.gov
14. Smarr LL, Chien AA, DeFanti T, Leigh J, Papadopoulos PM (November 2003) The OptIPuter. ACM Commun 46(11):58–67
15. Cherry S (September 2009) Forecast for cloud computing: up, up, and away. IEEE Spectrum 46(10):68–68
16. Baldine I, Xin Y, Mandal A, Renci C, Chase U, Marupadi V, Yumerefendi A, Irwin D (2010) Networked cloud orchestration: a GENI perspective. In: Proceedings of IEEE global telecommunications conference (GLOBECOM), December 2010, Miami, USA, pp 573–578
17. Aoun R, Doumith E, Gagnaire M (2010) Resource provisioning for enriched services in cloud environment. In: IEEE second international conference on cloud computing technology and science (CloudCom), December 2010, Indianapolis, USA, pp 296–303
18. Hilley D (2009) Cloud computing: a taxonomy of platform and infrastructure-level offerings. Georgia Institute of Technology, Technical Report GIT-CERCS-09-13, April 2009
19. GEYSERS, Generalised Architecture for Dynamic Infrastructure Services. [Online] http://www.geysers.eu/
20. Keahey K, Laforenza D, Reinefeld A, Ritrovato P, Thain D, Wilkins-Diehr N (2010) In: Ambra PD, Guarracino M, Talia D (eds) Grid, cluster and cloud computing. In: Euro-Par 2010 – parallel processing. Lecture Notes in Computer Science, vol 6271. Springer, Berlin, pp 341–342
21. Foster I, Zhao Y, Raicu I, Lu S (2008) Cloud computing and Grid computing 360-degree compared. In: Proceedings of grid computing environments workshop (GCE), Austin, Texas, USA, pp 1–10
22. Daras P, Alvarez F (2009) A future perspective on the 3D media Internet. In: Tselentis G, Domingue J, Galis A, Gavras A, Hausheer D, Krco S, Lotz V, Zahariadis T (eds) Towards the future internet—a European research perspective. IOS Press, pp 303–312
23. Simeonidou D, Hunter D, Ghandour M, Nejabati R (2008) Optical network services for ultra high definition digital media distribution. In: Fifth international conference on broadband communications, networks and systems (BROADNETS), London, UK, pp 165–168
24. Rodriguez-Rosello L, Ballesteros IL (October 2007) Network media of the future. EC Publication, Brussels
25. Amar A, Peter A, Alex C, John C, Gustavo G, Ahola J, Samuel K, Marco M, Costis K, Catherine L-B, Alberto L, Andreas M, Telma M, Mirco N, Antonio N, Olimpiu N, Cabral PF, Beilu S, Christian T, Emmanuel T, Theodore Z (2008) Multimedia delivery in the future Internet – a converged network perspective. White paper – media delivery platforms cluster, October 2008
26. Kubota K (2006) Beyond HDTV-ultra high-definition television system. In: Second multimedia conference, November 2006, London, UK
27. Sato K-I (2000) Role and opportunities of photonics in future networks (invited). In: IEEE photonics society winter topicals meeting series (WTM), Nagoya, Japan, pp 87–88
28. Orphanoudakis T, Leligou H-C, Kosmatos E, Stavdas A (2009) Future Internet infrastructure based on the transparent integration of access and core optical transport networks. IEEE/OSA J Opt Commun Netw 1(2):A205–A218
29. Saleh A, Simmons J (2011) Technology and architecture to enable the explosive growth of the Internet. IEEE Commun Mag 49(1):126–132
30. Piero Castoldi FB, Martini B (2009) Challenges for enabling cloud computing over optical networks. In: Optical fiber communication conference and exposition and the national fiber optic engineers conference (OFC/NFOEC), March 2009
31. Krauter K, Buyya R, Maheswaran M (February 2002) A taxonomy and survey of Grid resource management systems for distributed computing. Softw Pract Exp 32(2):135–164

32. Foster I (2005) Globus toolkit version 4: software for service-oriented systems. In: Jin H, Reed D, Jiang W (eds) Network and parallel computing. Lecture Notes in Computer Science, vol 3779, Springer, Berlin, pp 2–13
33. Frey J, Tannenbaum T, Livny M, Foster I, Tuecke S (2001) Condor-G: a computation management agent for multi-institutional Grids. In: Proceedings of 10th IEEE international symposium on high performance distributed computing, August 2001, pp 55–63
34. Almond J, Snelling D (October 1999) UNICORE: uniform access to supercomputing as an element of electronic commerce. Future Gen Comput Syst 15:539–548
35. Huang R, Casanova H, Chien A (2006) Using virtual Grids to simplify application scheduling. In: Proceedings of the 20th international conference on parallel and distributed processing symposium (IPDPS 2006), April 2006, p 10
36. Grimshaw AS, Wulf WA, C. The Legion Team (January 1997) The legion vision of a worldwide virtual computer. ACM Commun 40(1):39–45
37. Koseoglu M, Karasan E (April 2010) Joint resource and network scheduling with adaptive offset determination for optical burst switched Grids. Future Gen Comput Syst 26(4):576–589
38. Tachibana T, Kogiso K, Sugimoto K (2008) Dynamic management of computing and network resources with PID control in optical Grid networks. In: Proceeding of the IEEE international conference on communications (ICC), May 2008, pp 396–400
39. Kacsuk P, Kiss T, Sipos G (July 2008) Solving the Grid interoperability problem by P-GRADE portal at workflow level. Future Gen Comput Syst 24(7):744–751
40. Abosi C, Nejabati R, Simeonidou D (2010) A novel service-oriented resource allocation model for future optical Internet. In: Proceedings of the 12th international conference of transparent optical networks (ICTON), June 2010, pp 1–4
41. Volckaert B, Thysebaert P, De Leenheer M, De Turck F, Dhoedt B, Demeester P (2004) Network aware scheduling in Grids. In: Proceedings of European conference on networks and optical communications (NOC), June 2004, Einhoven, Netherlands
42. Thorpe SR, Battestilli L, Karmous-Edwards G, Hutanu A, MacLaren J, Mambretti J, Moore JH, Sundar KS, Xin Y, Takefusa A, Hayashi M, Hirano A, Okamoto S, Kudoh T, Miyamoto T, Tsukishima Y, Otani T, Nakada H, Tanaka H, Taniguchi A, Sameshima Y, Jinno M (2007) G-lambda and EnLIGHTened: wrapped in middleware co-allocating compute and network resources across Japan and the US. In: Proceedings of the first international conference on networks for grid applications (GridNets), Lyon, France, pp 5:1–5:8
43. Battestilli L, Hutanu A, Karmous-Edwards G, Katz D, MacLaren J, Mambretti J, Moore J, Park S-J, Perros H, Sundar S, Tanwir S, Thorpe S, Xin Y (2007) EnLIGHTened computing: an architecture for co-allocating network, compute, and other Grid resources for high-end applications. In: Proceedings of the international symposium on high capacity optical networks and enabling technologies, November 2007, Dubai, UAE, pp 1–8
44. Takefusa A, Hayashi M, Nagatsu N, Nakada H, Kudoh T, Miyamoto T, Otani T, Tanaka H, Suzuki M, Sameshima Y, Imajuku W, Jinno M, Taki-Gawa Y, Okamoto S, Tanaka Y, Sekiguchi S (2006) G-lambda: coordination of a Grid scheduler and lambda path service over GMPLS. Future Gen Comput Syst 22(8):868–875
45. OASIS. [Online]. http://www.oasis-open.org/
46. Abosi CE, Nejabati R, Simeonidou D (December 2011) A service oriented framework for efficient resource orchestration in future optical networks. J Commun 6(9):711–723
47. Ravindran R, Ashwood-Smith P, Zhang H, Wang G-Q (2004) Multiple abstraction schemes for generalized virtual private switched networks. In: Proceedings of the Canadian conference on electrical and computer engineering, May 2004, pp 519–522
48. Abosi CE, Nejabati R, Simeonidou D (November 2010) Design and development of a semantic information modelling framework for a service oriented optical Internet. J Netw 5 (11):1300–1309
49. WSMO. [Online]: http://www.wsmo.org/

50. Clapp G, Ferrari T, Hoang DB, Jukan A, Leese MJ, Mealor P, Travostino F (2005) Grid network services. draft-ggf-ghpn-netserv-2, Grid High Performance Networking Research Group working draft, May 2005
51. Optical Interworking Forum. UNI 1.0 signaling specification, release 2: common part. http://www.oiforum.com/public/documents/OIF-UNI-01.0-R2-Common.pdf
52. Farrel A, Vasseur J-P, Ash J (2006) A path computation element (PCE)-based architecture. In: IETF RFC 4655, August 2006
53. Phosphorus deliverable D2.1 (2007) The Grid-GMPLS control plane architecture. http://www.ist-phosphorus.org/files/deliverables/Phosphorus-deliverable-D2.1.pdf. Mar 2007
54. Savva A (ed) (2005) Job submission description language (JSDL) specification, version 1.0, GFD.056. http://forge.gridforum.org/projects/jsdl-wg. Nov 2005
55. Andreozzi S (ed) (2007) GLUE schema specification – version 1.3. Draft 3-16. http://glueschema.forge.cnaf.infn.it/Spec/V13. Jan 2007
56. Zervas G et al (2007) Grid optical user network interface (GOUNI). Open grid forum draft. https://forge.gridforum.org/sf/docman/do/listDocuments/projects.ghpn-rg/docman.root.current_drafts. Feb 2007
57. Jukan A, Karmous-Edwards G (2007) Optical control plane for the grid community. Commun Surv Tutorials IEEE 9(3):30–44 (Third Quarter 2007)
58. Network Service Interface Working Group (NSI-WG). Open Grid Forum (OGF). http://forge.gridforum.org/sf/projects/nsi-wg
59. Zervas G, Escalona E, Nejabati R, Simeonidou D, Carrozzo G, Ciulli N, Belter B, Binczewski A, Poznan M, Tzanakaki A, Markidis G (June 2008) Phosphorus grid-enabled GMPLS control plane (GMPLS): architectures, services, and interfaces. IEEE Commun Mag 46(6):128–137
60. Phosphorus deliverable D2.2 (2007) Routing and signalling extensions for the Grid-GMPLS control plane. http://ist-phosphorus.org/files/deliverables/Phosphorus-deliverable-D2.2.pdf. Sep 2007
61. Andrieux A, Czajkowski K, Dan A, Keahey K, Ludwig H, Nakata T, Pruyne J, Rofrano J, Tuecke S, Xu M (2004) Web service agreement specification (WS-Agreement). GFD-R-P.107. http://www.ogf.org/documents/GFD.107.pdf. Mar 2004
62. Sun Virtual Desktop Infrastructure 3.0, Deployment Guide. http://wikis.sun.com/display/VDI3/Deployment
63. Chiaroni D, Urata R, Gripp J, Simsarian JE, Austin G, Etienne S, Segawa T, Pointurier Y, Simonneau C, Suzaki Y, Nakahara T, Thottan M, Adamiecki A, Neilson D, Antona JC, Bigo S, Takahashi R, Radoaca V (2010) Demonstration of the interconnection of two optical packet rings with a hybrid optoelectronic packet router. In: The proceedings of the 36th European conference and exhibition on optical communication (ECOC), September 2010, pp 1–3
64. Vismara F, Musumeci F, Tornatore M, Pattavina A (2010) A comparative blocking analysis for time-driven-switched optical networks. In: The proceedings of the 15th international conference on optical network design and modeling (ONDM), February 2010, pp 1–6
65. Gonzalez-Ortega MA, Qiao C, Suarez-Gonzalez A, Liu X, Lopez-Ardao J-C (2010) LOBS-H: an enhanced OBS with wavelength sharable home circuits. In: The proceedings of the IEEE international conference on communications (ICC), May 2010, Capetown, South Africa, pp 1–5
66. Wen B, Shenai R, Sivalingam K (September 2005) Routing, wavelength and time-slot-assignment algorithms for wavelength-routed optical WDM/TDM networks. J Lightwave Technol 23(9):2598–2609
67. Dunne J, Farrell T, Shields J (2009) Optical packet switch and transport: a new metro platform to reduce costs and power by 50% to 75% while simultaneously increasing deterministic performance levels. In: Proceedings of the sixth international conference on broadband communications, networks, and systems (BROADNETS), September 2009, Madrid, Spain, pp 1–5
68. EU FP7 MAINS project. http://www.ist-mains.eu/

69. Fernandez-Palacios J, Gutierrez N, Carrozzo G, Bernini G, Aracil J, Lopez V, Zervas G, Nejabati R, Simeonidou D, Basham M, Christofi D (2010) Metro architectures enabliNg subwavelengths: rationale and technical challenges. In: Future network and mobile summit, Florence, Italy
70. Zervas GS, Triay J, Amaya N, Qin Y, Cervelló-Pastor C, Simeonidou D (2011) Time shared optical network (TSON): a novel metro architecture for flexible multi-granular services. In: Proceedings of the European conference on optical communication (ECOC), September 2011
71. Savi M, Zervas G, Qin Y, Martini V, Raffaelli C, Baroncelli F, Martini B, Castoldi P, Nejabati R, Simeonidou D (2009) Data-plane architectures for multi-granular OBS network. In: Proceedings of the optical fiber communication conference (OFC), May 2009
72. Ji PN, Aono Y (2010) Colorless and directionless multi-degree reconfigurable optical add/drop multiplexers. In: Proceedings of the 19th annual wireless and optical communications conference (AWOPCC), July 2010, Shanghai, China
73. LAMBDA OpticalSystems. http://www.lambdaopticalsystems.com/
74. Cao X, Xiong Y, Anand V, Qiao C (July 2002) Wavelength band switching in multi-granular all-optical networks. Proc SPIE 4874:198–210
75. Ho PH, Mouftah HT (August 2002) Routing and wavelength assignment with multi-granularity traffic in optical networks. J Lightwave Technol 20(8):1292–1303
76. Noirie L, Vigoureux M, Dotaro E (2001) Impact of intermediate grouping on the dimensioning of multi-granularity optical networks. In: Proceedings of optical fibre communication conference (OFC), March 2001
77. Noirie L, Blaizot C, Dotaro E (2000) Multi-granularity optical cross-connect. In: Proceedings of the European conference of optical communications (ECOC), September 2000
78. Noirie L, Dorgeuille F, Bisson A (2002) 32×10 gbit/s DWDM metropolitan network demonstration with 10 waveband – ADMs and 155 km teralight metro fiber. In: Proceedings of the optical fiber communication conference, March 2002, Anaheim, USA
79. Politi CT, Matrakidis C, Stavdas A, Gavalas D, O'Mahony MJ (December 2006) Single layer multigranular OXCs architecture with conversion capability and enhanced flexibility. J Opt Netw 5(12):1002–1012
80. Zervas GS, De Leenheer M, Sadeghioon L, Klonidis D, Qin Y, Nejabati R, Simeonidou D, Develder C, Dhoert B, Demeester P (June 2009) Multi-granular optical cross-connect: design, analysis and demonstration. J Opt Commun Netw 1(1):69–84
81. Nejabati R, Zervas G, Zarris G, Qin Y, Escalona E, O'Mahony M, Simeonidou D (November 2008) A multi-granular optical router for future networks. J Opt Netw 7(11):914–927

Chapter 14
Bringing Optical Network Control to the User Facilities: Evolution of the User-Controlled LightPath Provisioning Paradigm

Sergi Figuerola, Eduard Grasa, Joan A. García-Espín, Jordi Ferrer Riera, Victor Reijs, Eoin Kenny, Mathieu Lemay, Michel Savoie, Scott Campbell, Marco Ruffini, Donal O'Mahony, Alexander Willner, and Bill St. Arnaud

1 Introduction

To date, the design and management of optical networks have largely been focused around two architecture initiatives for the signalling and setup of optical circuits and/or virtual private networks (VPNs)—Generalised Multi-Protocol Label Switching (GMPLS) [1] and Automatically Switched Optical/Transport Network (ASON/ASTN) [2]. In addition, the Optical User-Network Interface

S. Figuerola (✉) • E. Grasa • J.A. García-Espín • J.F. Riera
Fundació i2CAT, Barcelona, Spain
e-mail: sergi.figuerola@i2cat.net; jage@i2cat.net; jordi.ferrer@i2cat.net

V. Reijs • E. Kenny
HEAnet Limited, Ireland's Education and Research Network, Dublin, Ireland
e-mail: victor.reijs@heanet.ie

M. Lemay
Inocybe Technologies Inc., Ottawa, Canada
e-mail: mathieu.lemay@inocybe.ca

M. Savoie • S. Campbell
Communications Research Centre, Broadband Applications and Optical Networks, Ottawa, Canada
e-mail: michel.savoie@crc.ca

M. Ruffini • D. O'Mahony
Department of Computer Science and Statistics, University of Dublin, Trinity College, Dublin, Ireland
e-mail: marco.ruffini@tcd.ie

A. Willner
Institute of Computer Science 4, University of Bonn, Bonn, Germany
e-mail: willner@cs.uni-bonn.de

B. St. Arnaud
Independent Green IT Consultant, New York, USA
e-mail: bill.st.arnaud@gmail.com

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*, Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_14,

(O-UNI) [3] has also shown considerable promise as a client interface to request the setup of an optical circuit or VPN using either GMPLS or ASON/ASTN. These technologies are well suited to traditional, centrally managed, hierarchical networks that are prevalent in today's telecom wide area network environment. However, a new type of network wide area architecture often referred to as "customer-controlled or customer-managed networks" [4] is becoming increasingly common among large enterprise networks, university research networks and government departments. Customer-controlled and customer-managed networks are radically different from the traditional networks in that the institution manipulates not only its own local/campus area network but also its own wide area optical network, assuming responsibility for direct peering and interconnection with other networks. As a consequence, traditional management and hierarchical optical network technologies, which are premised on central provisioning of optical VPNs to customers, are largely unsuitable for customer management of their own optical network.

1.1 Customer-Owned Networks

Many schools, hospitals and government departments are acquiring their own metro dark fiber. In Canada, for example, most universities have acquired dark fiber to provide their own metro network connectivity. Most of these institutions have participated in condominium [5] dark-fiber networks so that they can better manage and control their connectivity and bandwidth requirements. In these networks, cable installation and management companies, sometimes called alternate distribution companies (ADco) [6], build and maintain dark-fiber networks for a multitude of clients.

The big advantage of customer-owned metro dark-fiber networks is that the traditional dollars-per-megabit business model for bandwidth is largely replaced by the much lower cost for the one-time capital cost for the dark-fiber and initial equipment outlay [7]. Many carriers are now selling or leasing point-to-point wavelength services to large enterprise and university research networks.

Customer-owned networks provide a second indirect cost savings through reduced Internet costs via remote peering and transit. Large enterprise or research networks can use customer-owned and customer-managed LightPaths to do direct peering with each other and, more importantly, set up LightPaths to popular no-cost peering exchanges [8]. These networks also provide significant technical advantages, particularly in support of end-to-end LightPaths and quality of service (QoS) for large file transfers, storage area networks (SANs) and Grid services [9]. Since these applications require substantial bandwidth links, in the order of gigabits, they require from services provisioned across multiple independently managed optical networks.

1.2 Application-Specific Networks

As commented before, a new set of bandwidth-intensive applications is emerging. Ranging from e-Science applications, the science that uses cyber infrastructure and high-performance networks as basic instruments to perform the research, to Grid applications and high-definition digital media streaming, these new applications infer new stress conditions to the network. As examples, the reader can consider the large amounts of data to be transmitted and the particular traffic patterns to be supported. Hence, these applications should be able to configure the network in the way it better suits their needs, for a more efficient resource usage.

A meaningful example can be found in Grid applications. These are usually compute- and/or data-intensive applications being executed in a Grid environment. Grid technology [10, 11] allows secure, reliable sharing of distributed resources in a heterogeneous environment divided into multiple administrative domains. Grid applications often request a number of resources to the Grid middleware, which uses a meta-scheduler to find and reserve a suitable set of resources that match the application's requirements. As the availability period start time arrives, the application is executed in a distributed fashion.

Increasingly, more and more Grid applications have to move large amounts of data between storage and computing resources; therefore, they require high-bandwidth connectivity between the sites contributing resources to the Grid. To let the Grid efficiently manage the network resources (i.e. only setup the connections that are required by the applications that are currently being executed), network resource availability information and reservation have to be part of the Grid management infrastructure. To achieve this goal, the network has to be able to provide availability information as well as to reserve network resources in advance.

2 User-Controlled LightPaths

The user-controlled LightPaths (UCLP) research programme was put in place in September 2002 through a request for proposals (RFP) by CANARIE, a non-profit organisation funded by the Canadian government to promote the development of advanced networking. The Communications Research Centre (CRC), along with the University of Ottawa (UofO) as a partner, led one of the international implementation proposals, namely, UCLPv1. The Optical Communications Group (*Grup de Comunicacions Òptiques*, GCO) of the Technical University of Catalonia (*Universitat Politècnica de Catalunya*, UPC) and the Fundació i2CAT joined this team in April 2004.

2.1 UCLP Version 1

The UCLP [12] software allows a physical network to be partitioned into several independent management domains and exposes the network resources belonging to each partition as software objects or services. These objects can be put under the control of different users, being humans or sophisticated applications, so that they can create their own IP network topologies.

The first implementation of the UCLP, named UCLPv1, enabled the user to request end-to-end LightPaths over a single domain. UCLPv1 was successfully deployed over circuit-switched networks of different technologies: CA*net 4, based on SONET/SDH technology; i2CAT's metro area experimental dense wavelength division multiplexing (DWDM) network; and HEAnet's hybrid Ethernet/Multi-Protocol Label Switching (MPLS) network [13]. Successful circuit provisioning by a Grid application over a transcontinental LightPath was also demonstrated [14], showcasing the potential of integrating the network as a first-class Grid resource.

However, not all the objectives of the UCLP research programme were accomplished with the UCLPv1 implementation. In particular, the physical network could only be partitioned on per-connection granularity, that is, users could not fully control dedicated network element or parts of them. In order to address the research programme goals, CANARIE put in place the UCLPv2 programme.

2.2 UCLP Version 2

The UCLPv2 software allows users to define their own packet- or circuit-switched-based network architecture including topology, routing, virtual routers, switches, virtual machines and protocols based on the concept of many separate, concurrent and independently managed Articulated Private Networks (APNs). APNs can be considered as a next-generation virtual private network where a user can create a complex network topology by binding together layers one through three network links, computers, time slices and virtual or real routing/switching nodes.

2.2.1 System Architecture

The UCLPv2 system is divided into organisations. An organisation is an entity that owns a group of physical resources (network elements) and/or virtual resources (LightPaths and interfaces). Organisations also have groups of users that can perform operations on the resources owned by an organisation, based on their user role. The following are the user roles defined in the system:

- *Physical network administrator* (PN-Admin): It is the owner of a physical infrastructure. It can partition the network into multiple LightPath and Interface

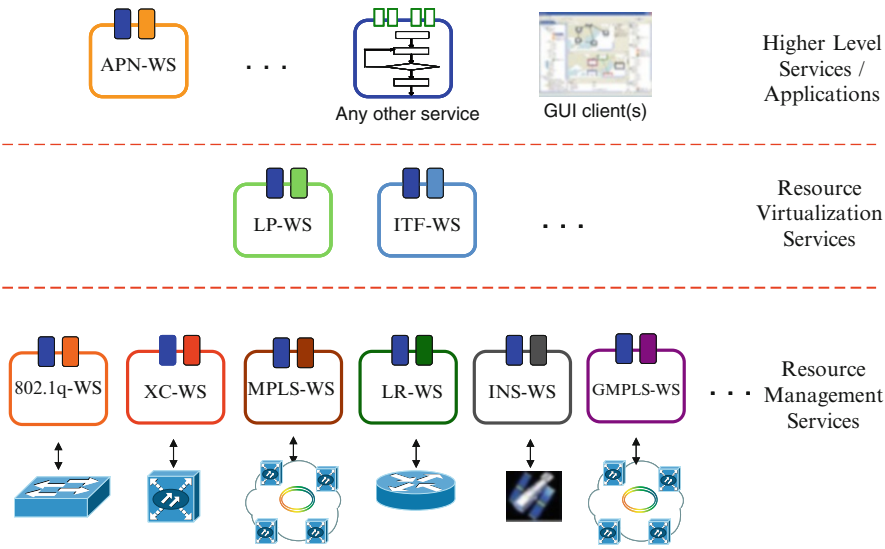


Fig. 14.1 The service-oriented architecture in UCLPv2

Web Services and either use them for its own purposes or lease them to other users. It manages the user accounts of its organisation.

- *Virtual network administrator* (VN-Admin): It does not own any physical infrastructure. It can harvest LightPath and Interface Web Services from different users (either PN or VN-Admins) and use them for its own purposes or lease them to other users. It manages user accounts of its organisation.
- *End user*: This user is only allowed to access the higher level services, but is not able to harvest or lease virtual resources.

The UCLPv2 service-oriented architecture is depicted in Fig. 14.1. Each box represents a different service, as explained in the following subsections. Each single Web service deals with a different technology. Higher-level services or applications exploit the virtualization capability to build complete end-user solutions without having to deal with underlying network complexities.

Resource Management Services

Resource management services are a family of Web services that control physical devices based on the technology they use. Each member of the family behaves like a proxy that sits on top of one or more network elements and partitions it into several management domains that can be handed off to different users. The service interface is divided in two parts: the former contains methods to add, delete and modify the configuration parameters of the equipment to be managed, whereas the latter provides different operations to control the hardware depending on the technology

of the underlying physical devices. Resource management services provide the following features to the system:

- *Network element abstraction*: The specifics of each physical device are hidden inside the network element Web service, and only a set of high-level capabilities (create connection, query status, get neighbours) is exposed, motivated by quick and easy user interface applications prototyping.
- *Service reusability*: Users; software clients, such as other services; and other network management/control applications can use the functionality offered by the resource management services.
- *Enhanced security*: Resource management services act as brokers between the users/services and the physical equipment, providing a layer of functional isolation between services and resources.

Resource Virtualisation Services

The fundamental building blocks in the system are the LightPath Web Service (LP-WS) and the InTerFace Web Service (ITF-WS). These services provide a layer of abstraction between higher-level, user-defined services and the resource management services. A LightPath is defined as a reserved, private, communication link between two network interfaces of two network elements. Examples of LightPaths can be a wavelength, a SONET circuit, an Ethernet VLAN or an MPLS label-switched path (LSP). The LightPath Web Service provides the following operations to manipulate the status of the LightPath services: create, delete, query, partition, bond and lease.

An interface is defined as a single network port on a network element. It can be either an add/drop port or a network port. LP-WS and ITF-WS provide a virtualised view of the network segments: users do not see network elements, but a set of network links (LightPaths) and end points (interfaces) with high-level parameters, like bandwidth, delay, jitter or reservation start/end time, but no technology specific information. The potential of this approach resides in the possibility of coordinating the low-level services in order to create complex services, ranging from a connection service up to a full-fledged network scheduling system. Another benefit of using LP-WS and ITF-WS is that resources (LightPaths and interfaces) can be assigned to different owners, enabling resource trading between different organisations.

Higher-Level Services

A number of services that provide different functionality to end users can be built on top of the UCLPv2 core software architecture. Two examples of these services are the APN service and the bandwidth reservation service.

- *APN service*. Enables a user to specify one or more fixed configurations of a set of network resources and to atomically activate or deactivate them.

- *Bandwidth reservation service*. Enables a user to reserve a certain bandwidth between two interfaces for a certain period of time. The reservation can be made in advance. This service is targeted to supporting e-Science and Grid applications. A prototype of this service was developed in the context of the European PHOSPHORUS project [15].

3 Argia

Argia is the evolution of the UCLPv2. It is an ongoing effort towards creating a commercial product that can be deployed in production optical networks.

3.1 System Architecture

Argia's software architecture, depicted in Fig. 14.2, is based on the Globus Toolkit 4 [16]. Argia software modules are the Optical Switch WS (a device controller service), the Connection WS and the APN Scenarios WS (end-user services).

The Optical Switch WS can interact with one or more optical switch devices. The physical device state (inventory, including physical and logical interfaces, list of cross-connections, alarms and configuration) is exposed as a former Web service

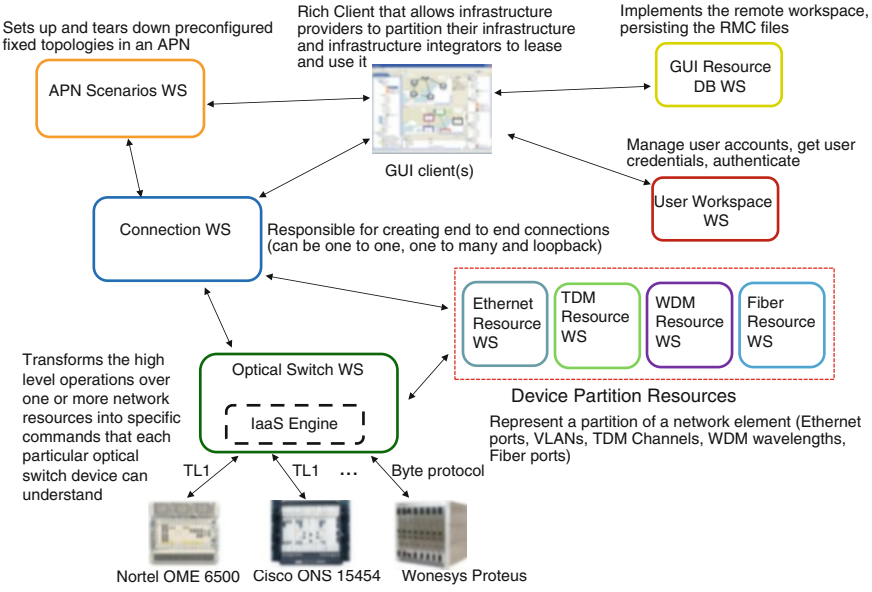


Fig. 14.2 Architecture of the Argia software

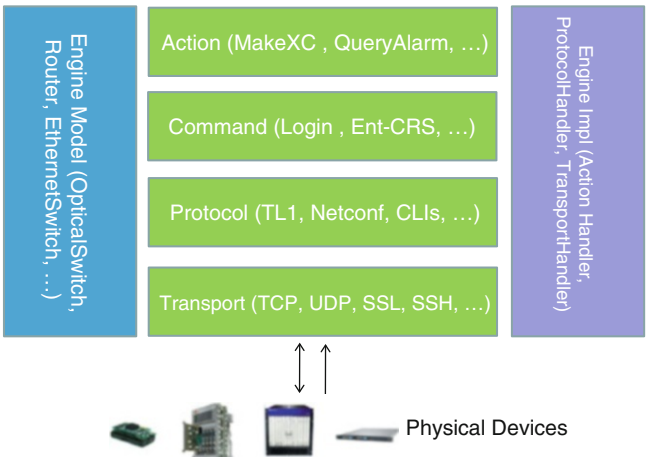


Fig. 14.3 The Engine architecture

resource so that clients of the Optical Switch WS can access the state of the physical device by querying the resource properties. The Optical Switch WS interface provides a series of high-level operations that encapsulate the physical device functionality.

Multivendor support is mainly accomplished through the use of the Engine, a Java-based framework to create drivers for physical devices. The Engine's interface provides a Java-based model of the state of the physical device that satisfies two needs:

- Engine-to-Optical Switch WS communication. The Engine fills the model attributes with the information of the physical device, allowing the Optical Switch WS to get the latest physical device information.
- Optical Switch WS-to-Engine communication. The Optical Switch WS fills some model attributes to request the Engine to perform some actions over the physical equipment, such as making a cross-connection.

The Engine, as shown in Fig. 14.3, provides abstractions to create the commands or group of commands composing an atomic action for the physical device. Additionally, protocol parsers to generate command structures and protocol transports to send and receive commands through the network are offered.

The Connection WS manages one or more connection resources. Each connection resource has pointers to the set of network resources that are connected together. To create a connection, first the Connection WS classifies all the resources belonging to the same connection per optical switch; next it extracts the relevant parameters from the network resources (like the slots/ports/channels, the bandwidth, a cross-connection description); then it issues all the required invocation messages to the Optical Switch WS; and finally it updates the state of the network resources. The procedure to undo a connection is symmetric.

Finally, the APN Scenarios WS can set up and tear down preconfigured topologies consisting in a set of connections in an APN. To achieve its goal, when the setup operation is called on the APN resource, the APN Scenarios WS calls the Connection WS to create all the connections required by the scenario. Tearing down a scenario is a similar process: the Scenarios WS calls the destroy operation on each of the connection resources that have been created in the setup operation.

3.2 Chronos

Chronos is an end-user application built on top of Argia that provides an advance reservation service for bandwidth. Advance reservations are useful network services that can not only provide guaranteed connectivity for network users but also allow service providers to more efficiently manage network resources and thus provide a better QoS [17, 18].

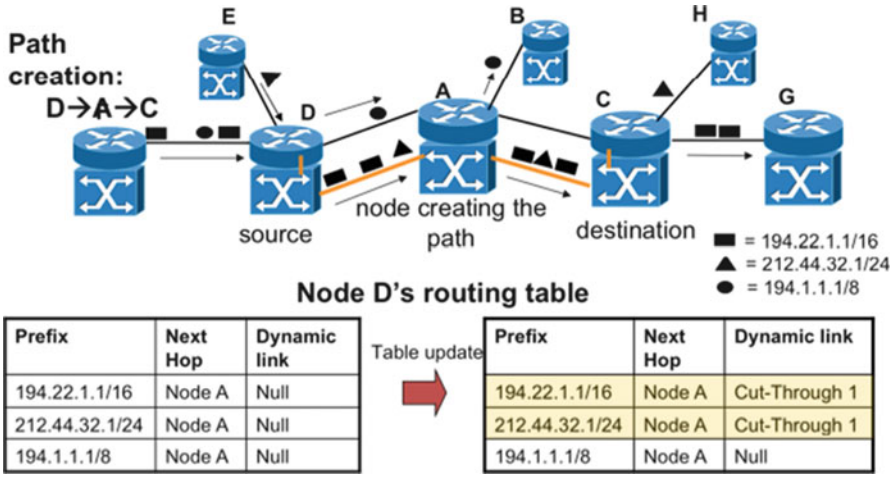
In general, two types of resource reservations in networks can be distinguished: immediate reservations, which are configured as soon as possible after a request is received, and advance reservations, which allow reserving resources a period of time before they are used. Actually, immediate reservations are a special case of advance reservations, where the start time requested is the present time. Advance reservations are useful in any application where large amounts of data have to be transmitted over a network and the time of such a transmission is known in advance. This is the case in Web caching or distributed multimedia applications where large amounts of content such as video files have to be transmitted up to a certain, predefined deadline. Another example is again Grid computing, where typical computations on the distributed parallel systems result in large amounts of data that also have to be transmitted in time between different machines.

The Chronos service interface defines operations for creating, cancelling, querying and activating reservations of end-to-end circuits.

3.2.1 Reservation Scheduling and Life Cycle

Chronos requires a scheduler as one of its building blocks because it needs to activate/cancel reservations when their start/end time arrives. The scheduler performs the following tasks:

- Activate a pending reservation when its start time arrives, in case it has to be automatically activated.
- Cancel an active reservation in any of the three following cases: the end time of the reservation arrives, the user requested its cancellation or there were problems with the network resources or devices.



Prefix	Next Hop	Dynamic link
194.22.1.1/16	Node A	Null
212.44.32.1/24	Node A	Null
194.1.1.1/8	Node A	Null

Prefix	Next Hop	Dynamic link
194.22.1.1/16	Node A	Cut-Through 1
212.44.32.1/24	Node A	Cut-Through 1
194.1.1.1/8	Node A	Null

Fig. 14.4 Procedure for creating an optical cut-through path in OIS

Argia and Chronos have been successfully demonstrated at several international events (Supercomputing, Global Lambda Integrated Facility (GLIF) [19] yearly meeting, TERENA Networking Conference) in the context of the PHOSPHORUS and HPDMnet research projects.

3.3 The OptISUP Project

The OptISUP (Optical IP Switching with UCLP interconnect) project was an experiment designed to test the interoperation of Optical IP Switching (OIS) [20, 21] and UCLP. Optical IP Switching is a technique developed at CTVR, in the University of Dublin, Trinity College, where an optical router creates and deletes optical cut-through paths in a distributed fashion in response to local analysis of IP traffic. Each optical router, composed of a modified IP router and an optical switch, typically a fiber switch or a wavelength selective switch, constantly analyses the transiting IP traffic. Suitable IP flows are aggregated based on their destination IP and rerouted into dedicated optical paths that bypass the IP layer.

The OIS network architecture operates as follows. After a statistical analysis of the IP traffic, suitable IP flow aggregates are singled out and redirected to an optically switched path, which is established between selected upstream and downstream neighbours. As shown in Fig. 14.4, only three nodes take part to the optical cut-through path creation initially, as optical paths are created in a step-by-step basis. The node that proposes the cut-through path signals to its upstream neighbour the list of IP addresses that should be redirected over the new path. Based on this information the upstream node updates its IP routing table and starts injecting the appropriate packets into the newly provisioned optical path. The advantage of

bypassing the IP layer is that node A can switch the flows at the optical layer, without consuming expensive router resources.

Path extension is the mechanism by which an initial cut-through path involving three nodes can be extended to additional neighbours. Similarly to path creation, it is based on local analysis; hence, each node can only extend the path by one hop, either upstream or downstream. Existing cut-through paths are deleted in order to free resources such as interfaces, optical ports and wavelength channels that have become underexploited due to changes in the traffic pattern. Every time a dynamic path is deleted, the traffic previously switched by the path returns to be routed hop by hop through the default IP links. Although such traffic increases the occupancy of routing resources, the freed optical resources can be reused to allocate new cut-through paths, bypassing additional IP traffic.

The first test-bed setup [22] involved the creation of a link between i2CAT in Barcelona, Spain; Trinity College in Dublin (TCD); and Dublin City University (DCU), Ireland. The test consisted in the establishment of cross-connections in the Glimmerglass-branded optical switches located at i2CAT, TCD and DCU, to create a VLAN, where all nodes shared the same Ethernet segment.

The second test-bed setup enabled the interconnection of the UCLP with the Optical IP Switching network architecture. The UCLP network is manually configured to form the intended topology: a resource object is instantiated for each connected node, and the list of destination prefixes is loaded into the server. The UCLP server is the central unit in charge of signalling and scheduling operations, receiving requests, calculating the path routes and signalling the network elements to provision the optical paths.

3.4 Mantychore

This European project, from the 7th Framework Programme, although it is not directly targeting the optical network itself, is leveraging the expertise behind UCLP and providing solutions for layer 3 networks. Current National Research and Education Networks (NRENs) in Europe provide connectivity services to their customers: the research and education community. The focus is being put in the automation and the ability of the community to control some characteristics of these connectivity services so that users can change some of the service characteristics without having to renegotiate with the service provider. The Mantychore [23] project consolidates this trend and allows the NRENs to provide a complete, flexible IP network service that allows research communities to have an IP network under their control, where they can configure (a) layer 1—optical, using Argia; (b) layer 2—MPLS; and (c) layer 3—IP addressing, internal routing, peering, and firewalls. To achieve its goals, Mantychore integrates and improves the tools developed in the past by the series of privately funded Mantychore projects, whose partnership was composed of HEAnet, i2CAT, NORDUnet, Telefónica, University of Essex, Cisco Systems, Juniper Networks and RedIRIS. Thus,

Mantychore follows the infrastructure as a service (IaaS) paradigm to enable NRENs and other e-infrastructure providers to enhance their service portfolio by building and deploying software and tools to serve infrastructure resources like routers, switches, optical devices and IP networks as a service to virtual research communities. It must be said that the flexibility provided by IaaS comes at the cost of increased management complexity; therefore, IaaS management solutions that keep track of permissions, infrastructure allocation, usage and monitoring are of key importance to fully realise the IaaS advantages.

Besides the introduction above, what motivates Mantychore is that IP network as a service is seen as a key enabler of the flexible and stable e-infrastructures of the future. Today a myriad of tool prototypes to provide point-to-point links to researchers have been developed (AutoBAHN [24], Harmony [25], G^{2}MPLS [26] and G-Lambda [27]). These tools, while providing high-bandwidth pipes to researchers, only address one side of the problem. Researchers that want to create a virtual community to address scientific problems are still connected to each institution's networks, and it is a hard problem to directly connect them with high-bandwidth pipes because it causes a number of issues such as security or routing integrity. One of the ways of efficiently solving this problem is to create a logically separated IP network on top of the high-bandwidth pipes.

The Mantychore project will produce an open source software toolset. This toolset will be offered both as off-the-shelf installable product (binaries format) as well as downloadable source code. Regardless of the packaging format, the software toolset will be composed at least of:

- Mantychore server. The Mantychore server component is to be installed on one or more servers and is responsible for connecting to the managed resources and the Engine logic and delivering the Web application to clients.
- Mantychore Web application for administrators and users. This will be installed on a server and accessed by the user as a regular website. It will offer a rich interface to manage the Mantychore server functionality and will be extended as additional plugging is installed on it.

4 The IaaS Framework

The previous sections outlined the different evolutions of UCLP [28–30] over the past decennium. While in the end it addressed different technologies, initially starting with SONET, then extended to support optical switching or even virtual routers, its initial premise never changed, which is to provide user-empowered networking. The enterprise information technologies (IT) and hosting industry is currently going through this user empowerment with the rise of cloud computing's popularity. UCLP and related work always have been a form of infrastructure as a service (IaaS) offering and are changing the way network pieces are being acquired. However, the initial UCLP-based software tools were never designed to efficiently

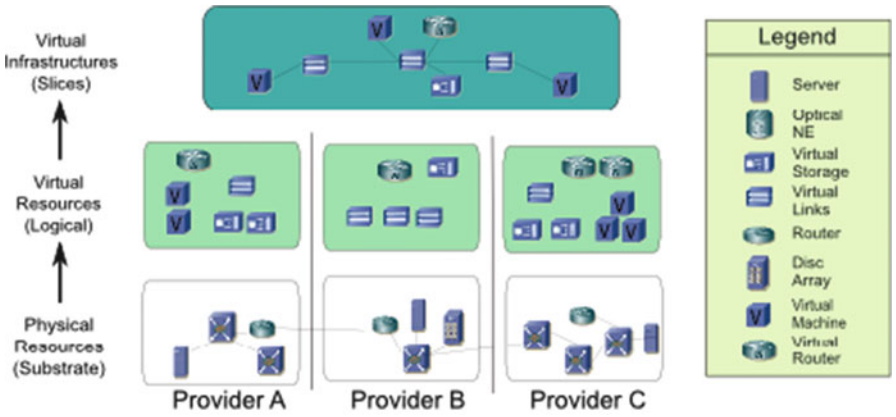


Fig. 14.5 Virtual infrastructures

support such a wide range of different technologies and concepts (from optical networks to computers and sensors). Over the different implementations of UCLP, the codebase has switched from Java-based remote technologies to Web service orchestrations and even Grid middleware frameworks such as Globus Toolkit 4 [16]. These frameworks met most requirements, but they also were bringing in additional complexity and made using these systems in production a very difficult task. This led to a need to have a unified framework using enterprise-grade tools and libraries in which new resources can quickly and easily be created. The IaaS Framework was created to serve such a purpose.

4.1 Requirements

The main requirement of the IaaS Framework is the ability to describe virtual infrastructures [31], as shown in Fig. 14.5. In order to do that, resources must be exchanged and shared between different provider domains.

As described in the related work section, several approaches make it possible to describe resources and create solutions that span multiple domains using these Grid-related approaches. However, because Grids have a fundamental difference in the way resources are consumed and described, the solution is suboptimal for use in virtual infrastructures (VI). Moreover, in Grids a limited number of users take all the available resources, whereas in virtual infrastructures, many users take a small subset of resources. This has a direct impact on the policies of the system. The creation of a new data model, architecture and security framework for use in multidomain VI management is thus required, as presented in Sect. 4.3.

4.2 Related Work

Over the past years, many efforts were made to create a unified management system for any type of resources. While some solutions are technology centric like the Java Managed Beans (MBeans), other solutions go into lengths in defining complex management systems like the Common Object Request Broker Architecture (CORBA) or IBM's Web Service Distributed Management (WSDM), now an OASIS standard. All these initiatives aimed at finding the silver bullet for a generic resource definition, but they came from different communities with different needs, and the solutions provided are optimised for their respective contexts.

4.2.1 The Grid Computing Approach

The Grid development motivated many of the cloud concepts. As envisioned by Foster et al. back in late 1990s [16], it was possible to describe computing, network and storage resources and have the possibility for a user to harvest, discover and use these at will for its experiments. This led to the creation of Grid computing where jobs are created and submitted to a scheduling system that will use messages to synchronise the calculations across multiple clusters. In the Grid approach resources are modelled using the Web Service Resource Framework (WSRF).

4.2.2 The Distributed Management Approach

In the meantime, the distributed management community backed by Microsoft took a different approach to managed resources using a framework called Web-Based Enterprise Management (WBEM) where the model used is an XML version of the Common Information Model (XML-CIM). CIM is a generic model that is used to describe different devices and is led by the Distributed Management Task Force (DMTF) group. The CIM is a relational object model in which the resource representation is done by extending special classes from the generic model and creating specific subsets of these extensions defining the resources. This provides the ability to have an object-oriented description for the resources that follows a definite tree model in the data structure for the properties.

4.2.3 Experimental Facilities

With emerging initiatives of experimental facilities in the United States and Europe, new managed resource frameworks are being created. In the Panlab II initiative under the FIREworks programme, a set of hierarchical map of properties is used to create definitions of new resources. These resources are then accessed via a Representational State Transfer (REST) interface at a specific Uniform Resource

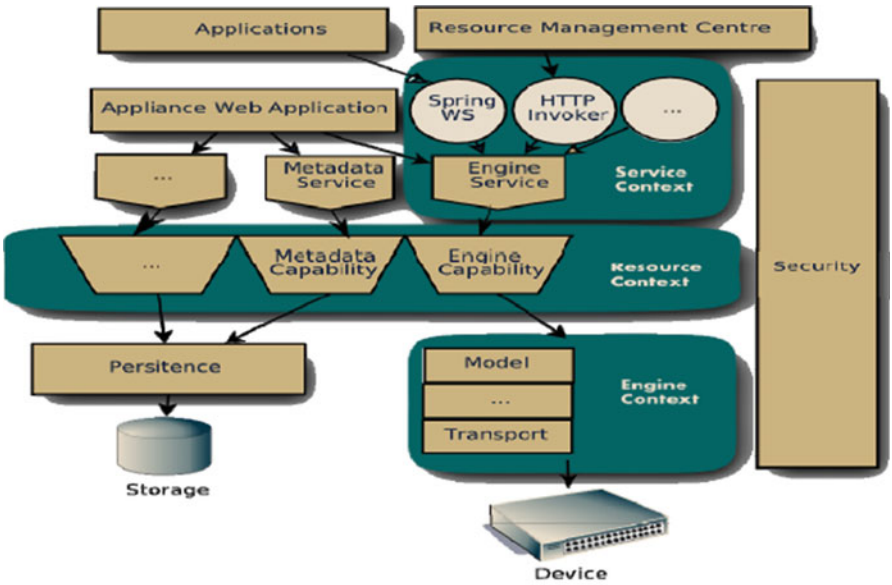


Fig. 14.6 IaaS Framework architecture

Locator (URL) pattern, making the creation of new resources easy and simple. In the United States, the ORCA/BEN initiative, one of the control frameworks from the GENI programme, uses a semantic ontology model based on the Network Description Language (NDL) schema [32]. This provides a rich and extensible data model in which complex queries can be formed with exhaustive constraints and relationships needed when creating network slices dynamically.

4.3 System Architecture

The IaaS Framework architecture is made of three fundamental tiers, as shown in Fig. 14.6:

- The mediation tier is made of Engines or database drivers that are responsible of contextualising model instances and synchronisation.
- The resource tier consists of a modular approach to resource data models and message handling.
- The messaging or service tier is used to handle communication using various remote technologies, such as SOAP, REST, XML Remote Procedure Calls, Java Messaging Services and others.

In the IaaS Framework, the Engine component is positioned at the lowest level of the architecture and maintains interfaces with physical devices. It uses services

provided by protocols and transport layers in order to achieve communication. Each Engine has a state machine that parses commands and decides to perform appropriate actions per command.

The resource component serves as an intermediate layer between Engines and services. The resources tier provides a data model and a set of interfaces for the resources it describes (for instance, optical switches, routers, computers or any manageable resource in general). Resources are modular and can be decomposed in a set of independent pieces called *capabilities*. Each capability is an independent unit that provides a data model and an interface with a limited set of functionality. Resources are assembled by aggregating several of these capabilities together, thus inheriting the aggregated data model and interface of the individual capabilities. Capabilities can be reused between different resources, therefore providing a modular design where independent modules can be reused.

4.3.1 Resource Framework

Resources are the main contribution of the IaaS Framework. The Engine concept was already developed in UCLP and Argia, while the messaging tier is mostly borrowed from the underlying libraries used to implement the framework.

The resource model must be generic and provide the ability to perform the creation of network slices that consists of heterogeneous resources. Moreover, the focus must be on the relationships and features offered by these resources and not at acquiring a certain quantity of a particular type of resource as found in Grid environments. This results in complex pathfinding problems, resource constraints and security requirements. By representing pieces of infrastructure as resources allows them to be exchanged or traded between distributed islands of management called virtual organisations (VOs). Based on Grid results and deployments in security, various techniques are being used to share the resources across many groups called VOs.

In addition to that, the resources implement the concept of capability aggregation. In order to access a resource, a client may only know how to deal with capabilities with no knowledge of the resource type. This is what makes it possible to have a heterogeneous resource environment, and this is valid as long as capabilities are homogeneous. The base resource model does not extend a modelling framework, but only defines a basic vocabulary to represent the resources.

Capabilities are not simply the data format, which gets aggregated to form the unified resource document but also consists in the associated specific behaviour and interface. When an action is performed on a resource, the message is routed to the appropriate handler, and the instance is contextualised from the mediation tier and behaviour applied accordingly.

4.4 Future IaaS Framework Work

The current framework instance provides the ability to have a good mediation, resource and messaging layer while having federated security. However, the additional work should be done to see how performance and stability compares to the related work initiatives such as ORCA and Globus Toolkit 4. The framework is currently being used in the GreenStar Network initiative [33] to create a distributed community cloud.

5 Multidomain Provisioning Systems

As described in the previous chapters, the demand for dynamic, user-controlled networks led in the past to the development of several Network Resource Provisioning Systems (NRPSs). While these systems serve their purpose well in a single-domain scenario, many cases involve connections through multiple domains.

The context of multidomain environments introduces several challenges that are not solvable by these NRPSs. Although there have been different attempts to a solution, most of them do not consider obstacles such as heterogeneity of the environment, independency and privacy of different administrative transport domains, inter-domain topology abstraction, sophisticated types of reservations or a coupled integration with a Grid middleware.

Based on this observation, the Harmony system [25] was developed within the PHOSPHORUS project, funded by the European Commission by means of the 6th Framework Programme [34]. PHOSPHORUS lasted for two and a half years, starting October 2006. It aimed to interconnect existing single-domain NRPS solutions while focusing on issues in the context of a multidomain environment.

5.1 Related Work

Traditionally, NRPSs aimed at provisioning dynamic network connections over a single, independent, administrative transport domain. Apart from UCLP/Argia, several other projects pursued this goal. Two well-known examples among the whole set of related projects are the following systems. On the one hand, Nortel developed a proof-of-concept middleware called Dynamic Resource Allocation Controller (DRAC) [35] that allows for an application-initiated configuration of transport network resources on an end-to-end basis. On the other hand there is the allocation and reservation of Grid-enabled optical networks (ARGON) [36] system that enables the integration of metro and wide area network resources into a Grid environment.

There are others projects aiming at achieving the challenges of multidomain dynamic circuit provisioning. Originating from the GÉANT2 project, the automated bandwidth allocation across heterogeneous networks (AutoBAHN) [37] architecture targets at the needs of a multidomain, multi-technology research community. Based on the so-called Inter-Domain Manager (IDM), the AutoBAHN architecture defines an inter-domain network reservation mechanism based on a decentralised architecture for peer domain signalling.

As an achievement of the DANTE-Internet2-CANARIE-ESnet (DICE) collaboration [38], a Web-based, inter-domain control plane was developed where Inter-Domain Controllers (IDCs) communicate in a decentralised way to provision end-to-end multidomain network paths.

In Japan, the main goal of the G-lambda project [39] was to establish a standardised Web service interface between Grid resource management systems and the network resource management systems that also supported advance reservations.

The EnLIGHTened Computing [40] project focused on dynamic optical LightPaths between supercomputing sites that were created upon application needs. A domain manager allocated network resources by setting up circuits using GMPLS. Similarly, the Dynamic Resource Allocation in GMPLS Optical Networks (DRAGON) [41] project aimed at both the intra- and inter-domain provisioning of packet and circuit-switched network resources in response to user requests for high-performance e-Science applications. DRAGON used a peer inter-domain model supporting abstracted topology information sharing and inter-domain path computation, equivalent to the Path Computation Element Architecture [42].

Although the previous citations represent a set of projects aiming at similar challenges, in real environments, network resources will often be heterogeneous in type and independently controlled and administrated. To integrate malleable advance reservations and co-allocation into such an environment was not addressed by any of them.

5.2 The Harmony System

The Harmony system is a multidomain, multivendor and multi-technology network resource broker architecture with advance reservation features. The main objective of the Harmony system is to provide users or Grid applications/middleware the ability to create point-to-point connections using network resources from several domains in a transparent way. The PHOSPHORUS project demonstrated solutions that facilitate vertical and horizontal communication among application middleware, NRPSs and extended GMPLS control plane: the Grid-GMPLS (G^2MPLS). The project addressed some of the key technical challenges to enable on-demand, end-to-end network services across multiple domains. The concept and test bed made applications aware of a Grid resource environment (computational

and networking), including their capabilities, and allowed to make dynamic, adaptive and optimised use of heterogeneous network infrastructures connecting various high-end resources.

PHOSPHORUS' assessment relied on experimental activities on a distributed test bed interconnecting European and worldwide optical infrastructures spread over more than ten countries. Specifically, the test bed involves European NRENs and national test beds, as well as international resources (GÉANT2, Internet2, CANARIE, Cross Border Dark Fiber infrastructures and GLIF virtual facility).

The Harmony system was designed to meet a set of requirements. First, the system had to be multidomain and capable of creating end-to-end optical paths in a seamless environment for the scientific personnel at the end points. The domains should be considered as a set of independent administrative transport domains controlled by different NRPSs within a heterogeneous environment. Also, they may or may not accept the same policies when provisioning paths. Secondly, there are privacy and confidentiality reasons that force not to share the internal topology to the other providers or to the public, also to avoid business disadvantages. Finally, considering again heterogeneity of the different involved NRPSs, there exists the requirement of making their signalling interoperable in order to provide on-demand, multidomain circuit provisioning, given that each one of the NRPSs offers a different communication interface.

The proposed solution, Harmony, allows the creation of complex resource reservations with in advance booking features, involving several NRPSs and/or GMPLS control plane. A common Network Service Plane (NSP) for signalling is defined, and hence, interoperability between NRPSs (DRAC, UCLP/Argia and ARGON), GMPLS control plane and the Grid applications/middleware is seamlessly achieved. The validity of the definition, design and implementation of the Harmony system was demonstrated in several international events in the period from 2007 to 2010, proving the feasibility to provide services across multidomain and multivendor transport network test beds for research. The Harmony test bed involved up to ten independent domains.

The Harmony system, apart from controlling multidomain scenarios, introduces the network as a manageable resource in the Grid by means of the Harmony Service Interface (HSI). The Harmony system implements resource co-allocation and scheduling capabilities (reservation service), is able to reduce the probability of resource blocking and provides inter-domain topology awareness services (topology service) by restricting the intra-domain topology information.

Due to the successful tests and public demonstrations performed so far, some of the work of the HSI was taken into consideration within the NSI (Network Service Interface) Working Group of the OGF (Open Grid Forum) [43].

5.2.1 System Architecture

Harmony's integration of application middleware and optical transport networks has been successfully achieved based on an architecture with three planes:

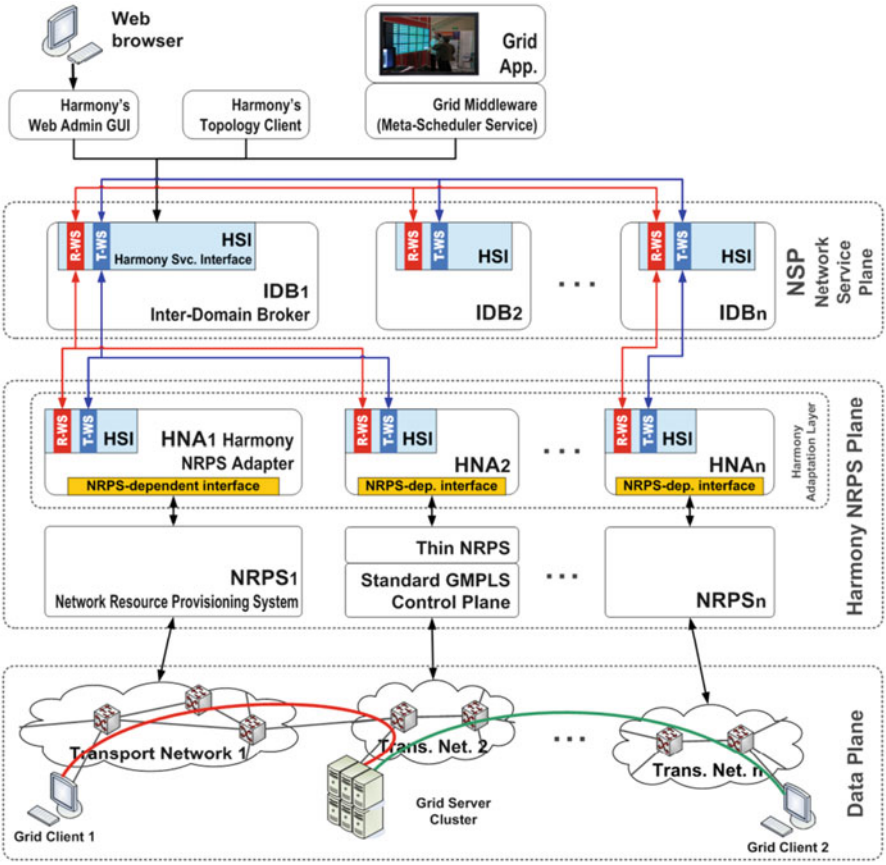


Fig. 14.7 The architecture of Harmony comprised of three planes

- *Network Service Plane*. It contains middleware APIs to expose network and Grid resources and to create network connectivity services with advance reservations and policy mechanisms, in a global hybrid network infrastructure.
- *Network Resource Provisioning Plane*. It contains the NRPSs and standard GMPLS control plane. An adaptation layer is required for interacting with the NSP and allowing multidomain interoperability.
- *Data Plane*. All the transport network devices, Grid resources and links between all of them compose the data plane.

In Harmony, these planes are built over a service-oriented architecture (SOA), which is compliant with the Web Service Resource Framework (WSRF) version 1.2 [44]. Figure 14.7 shows the plane separation in Harmony, as well as the Reservation Web Service (R-WS) and the Topology Web Service (T-WS) used for signalling between entities in the different planes.

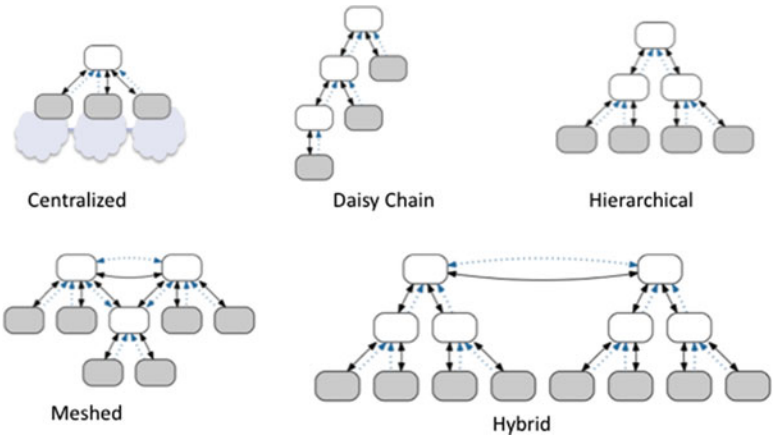


Fig. 14.8 Different NSP architectures supported by the Harmony system

The Network Service Plane

The Network Service Plane (NSP) is the highest plane in the architecture of Harmony. Administrative users, Grid middleware or applications access Harmony by invoking the services offered by this plane. The NSP is populated with inter-domain broker (IDB) software entities. Each IDB is responsible of managing and brokering the network resources offered by its underlying administrative domains. Thus, one IDB may control one or more Network Resource Provisioning Systems, also called simply *domains* in Harmony.

Since each IDB has exclusive control over its underlying domains, a multidomain path reservation request requires cooperation of several IDBs. Several options exist to perform this cooperation: centralised, hierarchical, peering, daisy chain or hybrid setups (Fig. 14.8). The Harmony system is designed so that the NSP operation mode can be elected and configured at system deployment time.

Moreover, Harmony was successfully tested using a mixed-role IDB operating mode. The NSP was set up with both hierarchical and peering relationships between entities. Therefore, a single IDB could act as a peer in a federation of several IDBs and as parent at the same time. This complex architecture allows addressing security and policy issues in the NSP while taking advantages of both peering and hierarchical configurations. The hierarchical approach adds flexible abstraction and federation capabilities for domains, while its lack of scalability is corrected with the peering approach. A use case for the mixed-role IDB mode can be found when two or more international research groups share an experiment with a collaboration application, which is high-bandwidth consuming.

Since the peering approach is based on a distributed pattern, its implementation required enhancing the original, hierarchical IDB entities with a protocol for the dissemination of topology information and database consistency mechanisms.

Topology information is distributed by a flooding protocol similar to Open Shortest Path First (RFC 2328) [45] where each peer has a limited list of known IDBs for flooding its inter-domain topology knowledge.

The Network Resource Provisioning Plane

The NRPSs compatible with Harmony at the current implementation stage are Argia, ARGON, DRAC and GMPLS. The latter requires an intelligent adapter named *Thin NRPS*, in order to support advance reservations and topology sharing through the Optical User-to-Network Interface (O-UNI). All these NRPSs have a Web service interface. Nevertheless, none of them are directly interoperable, because each NRPS has its own interface specification and data type schemas. Harmony enables the interoperability between NRPSs, assuming that none of them can significantly be altered on neither its internals nor its interfacing. This constraint led to the design and implementation of the Harmony NRPS Adapters (HNA). The HNA performs the translation from the NPRS-dependent Web service interface to the common Harmony Service Interface. Each NRPS requires an HNA to be registered by the NSP and be able to operate as a domain in Harmony.

The Data Plane

The Harmony system was created to work over data plane infrastructures based on the circuit-oriented paradigm. For this reason, the network technologies considered are SONET/SDH, WDM or optical Ethernet for optical transmission systems, and Ethernet and MPLS for electrical systems.

5.2.2 Services

Malleable, Multidomain Network Resource Allocation

Harmony's NSP provides different types of advance reservations so that the service is provided on top of the inter-domain topology abstracted by the service plane. A basic form of an advance reservation request is described in detail in [25] and is defined as follows: the request is received at t_{arrival}, is admitted and starts at t_{start}. Furthermore, the usage phase (duration) is limited by t_{end}.

- *Fixed advance reservation*: A fixed advance reservation request for a single connection is depicted on the left side in Fig. 14.9 and is defined as tuple (t_{start}, t_{end}, s, d, C) where $t_{\text{start}} < t_{\text{end}}$. The reservation starts at t_{start} and ends at t_{end}. The end points of the connections are specified by s and d. C represents additional resource constraints, which are the required capacity and delay.

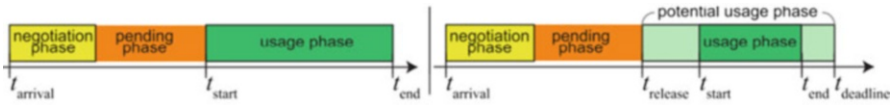


Fig. 14.9 Life cycles of a fixed (*left*) and a deferrable (*right*) advance reservations

- *Deferrable advance reservation*: A deferrable advance reservation request has a degree of freedom in the time domain. In particular, time-related parameters define a range of possible values to establish the reservation. The life cycle of a deferrable reservation is given on the right side in Fig. 14.9 and is defined as tuple (t_{release}, t_{deadline}, Δt, s, d, C) where $t_{\text{release}} + \Delta t < t_{\text{deadline}}$. The reservation can start at t_{release} and must end before t_{deadline}. The usage phase is specified by the duration $\Delta t > 0$. Compared to a fixed advance reservation, the parameters t_{start} and t_{end} ($\Delta t = t_{\text{end}} - t_{\text{start}}$) can be determined by the NRPS.
- *Malleable advance reservation*: A specification of the exact transmission rate can be omitted when a fixed amount of data has to be transmitted. By joining the time and resource constraints, the reservation system can find the most efficient solution for the requested transmission. A malleable reservation request is defined as tuple (t_{release}, t_{deadline}, s, d, S, C) where $t_{\text{release}} < t_{\text{deadline}}$. The end points of the connections are specified by s and d. S determines the data size (transmission rate and time product) and C represents additional constraints. Typical constraints are lower/upper boundaries for the transmission rate.

Multidomain Path Computation

The Network Service Plane in Harmony provides multidomain path computing over the abstracted topology based in border points (or end points) of the network domains and the inter-domain links. The path computation element in Harmony implements Dijkstra's algorithm in order to find the shortest end-to-end path between the involved resources, basing on this abstracted view.

It must be noted that the approach chosen is suboptimal, since Harmony only seeks the shortest path in the abstracted topology view of the multidomain scenario, regardless of the internal topology or limitations in each domain. In case an NRPS reports any resource as being not available, this resource is blocked in the service plane, and the path computing algorithm tries to find another path.

Topology Abstraction and Security

One of the main concerns in networking is determining how to partition, abstract or virtualise network resources and how to distribute the network intelligence among network control and service management planes and the Grid middleware.

In Harmony, the abstraction of the network resources is mainly performed by the NRPSs, but as each NRPS controls an autonomous transport network, privacy and security issues arise in two fronts: user-to-network and network-to-network. Harmony has faced these two issues from different perspectives.

On the one hand, the fact that Harmony enables interoperability between autonomous systems potentially threatens the privacy of network resources. Specially, Harmony fetches information about inter-domain links: end points allocated for the inter-domain links and generic domain information, such as identifier, description or reachable local user namespaces. End points are highly sensitive information in the knowledge of the network. Harmony classifies end points in two different categories: *user end points* (connected to user premises) and *border end points* (connected to inter-domain links). While border end points knowledge must be propagated through signalling along the NSP for allowing path computation at any IDB entity, user end point information must be kept partly secret. Harmony applies an IPv4-like addressing schema for naming all end points. This allows an IDB to locate user or border end points from any NRPSs by simply applying a mask to the end point identifier, named Transport Network Address (TNA). This allows Harmony to treat a group of end points' TNAs belonging to the same domain as an IPv4 subnet. TNA subnets' addresses are periodically flooded in the NSP.

On the other hand, security issues are addressed in Harmony, thanks to the implementation of a security infrastructure, the Authentication and Authorization Infrastructure (AAI). Harmony's AAI enables the possibility of performing secure cross-layer operations. In this case, cross-layer must be understood from end user/application middleware down to the Network Resource Provisioning Plane, across the different layers defined in Harmony.

5.2.3 Harmony Service Interface

Services in Harmony are offered by means of its northbound interface, the Harmony Service Interface (HSI). The HSI enables interoperability between the capabilities and services being offered by resource brokers to external entities such as Grid middleware or applications. HSI also enables the communication between entities within the service plane. The functionalities provided by HSI have been considered within the Network Service Interface (NSI) working group of the Open Grid Forum, for standardisation purposes.

The design of the interface takes into account the modularity and the services nature in order to build an easily maintained module. Thus, the HSI is composed of three main modules: Reservation Web Service (R-WS), Topology Web Service (T-WS) and Notification Web Service (N-WS). Moreover, there is an extra module within the HSI that contains the common data types.

The Reservation-WS enables the Grid middleware to create, cancel and query advance reservations, both malleable and fixed types, across one or multiple domains. Moreover, the Harmony system has to know which resources are under

its control, in order to create the multidomain paths. Therefore, the system must store, retrieve, modify and delete the resource-related information according to the topology of the controlled network domains. Topology-WS enables the system to carry out all these tasks. Finally, the Notification-WS is the component of the HSI responsible of the event notification management and signalling. N-WS eliminates the need for the system to be polled periodically. When a connection is aborted, the higher layer entity that created the corresponding reservation is notified.

5.2.4 Harmony NSP Evaluation: Performance and Scalability

The implemented NSP is used to provide the network as a first-class Grid resource and integrate it in the multidomain scheduling process. The co-allocation of the network as a resource requires specific quality of service (QoS) features that the underlying infrastructure must support, and an agreement-based resource management system. The challenge is to analyse the scalability of such NSP considering both job workloads and different available network topologies.

As stated in [46] by Degilà et al., performance measurement analysis has become an important tool in order to decide the best network topology and must, therefore, be carefully chosen. In general, the topology of a network seriously affects its reliability, throughput or even traffic patterns [47]. Likewise, the topology of the NSP directly influences the reliability, performance and scalability of the service plane itself. Thus, an inadequate choice of the NSP deployment model may lead to the non-desirable situation where lower layers are underperforming. Consequently, the problem evaluation of the performance and scalability of the different architectures supported by the NSP acquires importance [48].

Parameter selection is not a straightforward task, since the selected parameters should characterise the system behaviour. Considering that the Harmony NSP behaves as a transactional system, where a path creation request is an atomic event, the different test cases performed have mainly focused on collecting information about the following parameters: service provisioning time and average rate of successful reservation requests, although others have been considered in order to complement the evaluation.

The tests executed have been divided into three basic test cases, each one corresponding to each of the NSP supported basic topologies. Advanced configurations of the NSP such as hybrid topologies have been also configured and tested, although the initial evaluation does not consider advanced topologies in the analysis. Furthermore, an analytical model was created in [49], and the results of the tests were compared to the ones obtained in the model. Figure 14.10 depicts the comparison of the results obtained for the service provisioning time over different NSP topologies with the results obtained from the analytical model.

Figure 14.10 shows the service provisioning times as a function of the involved transport domains in the path request (abscissa). The obtained results present slightly higher or lower values than the expected from the analytical model,

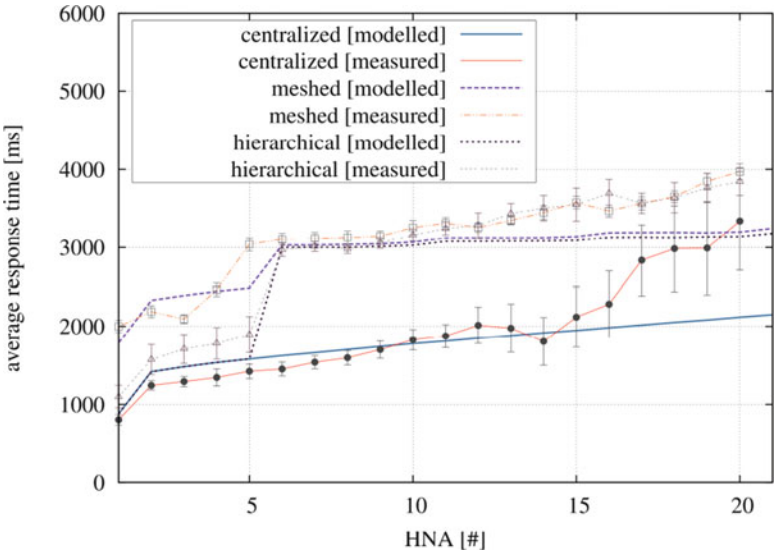


Fig. 14.10 Centralised, hierarchical and meshed measurements validating the expected analytical model of the service provisioning time

although the bounds of the standard deviation obtained remain close enough to the analytical model.

Based on the results obtained in all the tests executed [27, 41, 53], the centralised approach provides a faster service provisioning time when controlling a small- or medium-sized data plane. However, due to the IDB nature, where the path computing and forwarding process depend on the number of controlled transport domains, the scalability of the centralised approach deteriorates in favour of the meshed model, which maintains lower response times when controlling larger data planes. On the other hand, the service provisioning time of the hierarchical model is critically affected by the delay introduced due to the communication process between hierarchy levels. Nevertheless, the hierarchical model is the most suitable in order to ensure privacy and security in real environments due to its nature and the strict control of the relationships between entities it can perform.

5.3 The Fenius project

The Fenius project is a parallel project to Harmony with similar motivation and objectives. Like Harmony, Fenius is a multidomain, multivendor and multitechnology network resource provisioning system with the objective of being able to create point-to-point connections using network resources from many different domains. In many ways the two systems are competing solutions but have been

implemented in different ways and provide different functionality, which sets them apart from one another. These differences can be seen in the sections below.

One of the primary reasons to develop a common interface to be able to create connections across domains managed by different NRPSs came about from the many activities being carried out within the GLIF community.

The GLIF encourages and supports the establishment of GLIF Open LightPath Exchanges (GOLEs) around the world and the partner contribution of high-capacity transport links to interconnect them [50]. A GOLE is an interconnection point between two or more high-speed networks. GOLEs are usually operated by GLIF participants and are comprised of equipment that is capable of terminating lambdas and performing LightPath switching. GOLEs provide *open* peering policies unmediated by the host organisation—that is, policy free cross-connect capability. Within the GLIF community, there is no policy that requires R&E network service providers to use a specific NRPS to provision their resources. Each member has the flexibility to use whatever system best fits their own needs or to develop their own. Until a standard such as the developing Network Service Interface (NSI) can be developed and pushed through the formal standardisation process, there is a need to develop an interim solution to be able to communicate with each of the NRPSs to create connections across these heterogeneous domains.

The Fenius project was initiated under the GNI API task force. The goal of the task force was to assess the interfaces between the various NRPSs and to keep only the bare essentials that could be provided by all systems [51]. The Fenius project was developed as an interim solution until actual standards emerged by ESnet. Fenius has been deployed and is being used as the translation software for the Automated GOLE Pilot launched by the GLIF in 2010.

The Automated GOLE Pilot is a project initiated to create an infrastructure of multiple GOLEs that allow automated user agents to request VLAN connections from an end point at any one of the GOLEs, across the multidomain GOLE fabric, to another edge attached to some other GOLE. Automated agents within the networks, communicating to establish the end-to-end connection, perform the entire connection process across GOLEs. Participating Automated-GOLEs (A-GOLEs) are Ethernet switching nodes, each controlled by a local NRPS that reconfigures the GOLE switches along the selected path to establish a dedicated VLAN between the two end points. This VLAN can be reserved in advance for a specified time and is provisioned with dedicated capacity and performance characteristics guaranteed between the two end points.

5.3.1 System Architecture

Fenius is a bare-bones virtual circuit creation service. It is a many-to-one-to-many translator: from one or more external Web service APIs to an internal ad hoc Java API, and from the internal API to an NRPS-specific API. By allowing translators to be created on both sides of the common internal API, it allows both external API

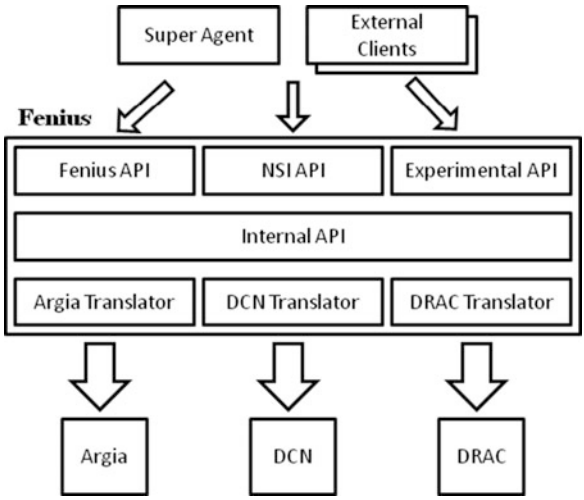


Fig. 14.11 Fenius architecture, many-to-one-to-many

and NRPS translator developers to write a single translation that can be used by all, avoiding the O(N!) many-to-many translation problem [52].

As shown in the diagram of Fig. 14.11, there is a single layer in the Fenius architecture. There are no intermediate layers for path computation, resource brokering or other related services. This is where Fenius is different from Harmony in that it does not offer any extra features above simple connections.

At the top level, there is what is called the Fenius Super Agent that acts as an external client to the different Fenius instances. It is the user-facing service in which clients can use to select end points and to schedule connections to be made. Because Fenius does not do topology discovery, the Super Agent must be configured with the domain end points for each NRPS.

5.3.2 How Fenius Works

When launched for a specific NRPS, the translator code is responsible to launch an instance of Jetty, an HTTP server, and to bind one or more external APIs to be published to it. It is also responsible for wiring things together so that any incoming request to an external API is translated to the internal API, which is forwarded to the local NRPS-specific translator and then pushed to the local NRPS. The response from the local NRPS is then translated again and eventually returned to the original Web service client.

In the user interface, when a reservation is made, the user is able to select the source and destination end points for the desired connection, and the Super Agent will use Dijkstra's algorithm to compute the shortest path between them.

There are currently translators for Harmony, Argia, DCN/OSCARS, DRAC, AutoBAHN and G-lambda. The Fenius Service Interface consists of an API including a service implementation that is used by the Super Agent called the connection

service. Its methods, similarly to Harmony Service Interface, are *reserve* (creates a new connection between the selected end points), *release* (removes an existing connection), *query* (requests details about a specific connection), *list* (lists all connections) and *reservable* (checks the availability of the selected resource for the connection).

6 Conclusions

Over the years, traditional control planes have not been able to address all the new emerging requirements of customer-operated networks. As mentioned before, at the same time those customer-operated networks appeared, a new bunch of bandwidth-demanding applications emerged. In this chapter we have shown how the UCLP concept appeared as a response to the new environment. It was born in order to allow a network to be partitioned into several management domains and let users control the network elements as software entities or services.

The first implementation of the UCLP concept was the UCLPv1 system. It successfully enabled the user to request end-to-end LightPaths over a single domain. It was also deployed and demonstrated over several network technologies. However, this first version failed to provide full support to the user over network elements or parts of them, since it only allowed partitioning the physical network on per-connection granularity.

As a consequence, UCLPv2 was launched to overcome the drawbacks and limitations of its predecessor. This second version, built as a service-oriented architecture, introduced the concept of Articulated Private Networks. Such APNs are considered as next generation VPNs, where the user creates complex network topologies including computers, time slices and virtual or real routing/switching devices. UCLPv2 enabled the construction of high-level services on top of its core resource management services. Well-known examples of these services comprehend the APN creation service and the bandwidth reservation service. However, UCLPv2 was still not mature and robust enough to be deployed and operated into production environments, since it was designed as a proof of concept for APNs. A new evolution was required in order to extend the resource provisioning system.

Therefore, Argia, which means *light* in the Basque language, appeared in the arena. The Argia system is the natural evolution of the UCLPv2 system, and it is a work in progress. Its end-user services, built on top of the Globus Toolkit suite, can be summarised into the Optical Switch WS, the Connection WS and the APN Scenarios WS. Although definitely enhancing its predecessor, Argia did not deal with an innovative feature required for the customer-controlled networks: advance reservations. Thus, Chronos system, leveraging on the work done for the Harmony system in the European project FP6 PHOSPHORUS, represents an end-user application built on top of Argia that provides such advance reservation capabilities for bandwidth scheduling.

Related to UCLP and its implementations, we have also shown the OptISUP project that aimed at testing the interoperation of Optical IP Switching and the UCLP implementations. Additionally, the Mantychore project, leveraging on UCLP expertise, has been presented. It provides solutions for layer 3 technologies and automates IP management protocols configuration, based on users' needs.

Furthermore, with the expertise acquired during the UCLP programme and its different implementations, and the current trends in the ICT realm, the IaaS Framework was built. It provides a unified framework using enterprise-grade tools and libraries in which new resources can quickly and easily be created and managed. This is crucial in an ICT world where the IT side is gaining huge momentum with cloud computing.

In parallel, the need of multidomain resource provisioning appears as a key requirement of the bandwidth-demanding applications. Multidomain network provisioning was already studied in the context of the Grid, but such a context introduces several challenges that are not solvable by these Network Resource Provisioning Systems. Most of them do not consider barriers such as heterogeneity of the environment, independency and privacy of different administrative transport domains, inter-domain topology abstraction, sophisticated types of reservations or coupled integration with a Grid middleware. Due to these limitations, the Harmony system was implemented within the PHOSPHORUS project. Harmony is a multidomain, multivendor and multi-technology network resource broker architecture with advance reservation features, which provides a Network Service Plane to the applications. The Harmony system solved the multidomain provisioning problem. In this chapter, we have presented the interface for provisioning network services and the system architecture and also how the advance reservations scheduling was handled in the system. With similar objectives and challenges, we have also presented the Fenius project. Although solving the multidomain interfacing problem, Fenius does not provide the service plane architecture and recommendations for managing network services with advanced features.

Finally, it is worth to comment that despite the long journey so far in the field of automated optical network provisioning and multidomain network services, there is still some way to go, especially given the new challenges brought by the emerging applications and also by the user communities, whose necessities are constantly evolving and whose execution environments are changing, from the well-structured Grid to the full abstracted cloud.

References

1. Berger L (ed) (2003) Generalized multi-protocol label switching (GMPLS) signaling functional description. IETF, RFC 3471, January 2003
2. Architecture for the automatic switched optical networks (ASON) (2001) ITU-T Recommendation G.8080, 2001
3. UNI 1.0 signaling specification OIF-UNI-01.0—user network interface (UNI) 1.0 signaling specification (2004) Optical internetworking forum. [Online] 2004 Available: www.oiforum.com

4. St. Arnaud B (2002) Customer owned networks. [Online] Available: http://www.lightreading.com/spc/document.asp?doc_id=20448&site=serviceprovidercircle
5. St. Arnaud B (2002) Frequently asked questions about customer owned dark fiber, condominium fiber, community and municipal fiber networks. [Online] Available: http://whitepapers.techrepublic.com.com/abstract.aspx?docid=12105
6. Beard TR, Ford GS, Spiwak LJ (2001) Why ADCo? Why now? An economic exploration into the future of industry structure for the "last mile" in local telecommunications markets. Phoenix Center Policy Paper No. 12. [Online] Available at SSRN: http://ssrn.com/abstract=503442 or doi:10.2139/ssrn.503442. Nov 2001
7. Waldron D (2000) Canadian School Board investments in private fiber optic networks, a cost benefit analysis. [Online] Available: http://www.canarie.ca/conferences/advnet2000/presentations/waldron.pdf. Communications Magazine, IEEE, vol.49, no.8, pp. 101–109, Aug. 2011
8. Norton WB (2001) Internet service providers and peering. [Online] Available: http://www.ecse.rpi.edu/Homepages/shivkuma/teaching/sp2001/readings/nortonpeering.pdf
9. DeFanti T, de Laat C, Mambretti J, Neggers K, St. Arnaud B (2001) TransLight: a global scale lambda grid for e-science. Commun ACM 46(11):34–41 (Special Issue on Blueprint for the Future of High Performance Networking)
10. Foster I, Kesselman C (2001) The grid 2: blueprint for a new computing infrastructure. Morgan Kauffman, San Francisco
11. Erl T (September 2005) Service oriented architecture: concepts, technology and design. ISBN: 0131858580
12. Wu J, Savoie M, Zhang H, Campbell S, Bochmann Gv, St. Arnaud B (2003) (September/October 2005) Customer-managed end-to-end light path provisioning. Int J Netw Manag, San Francisco, USA, 15(5):349–362
13. Sanchez A, Figuerola S, Junyent G, Kenny E, Reijs V, Ruffini M (2007) A user provisioining tool for EoMPLS services based on UCLPv1. In: TERENA networking conference, Denmark. [Online] Available: http://tnc2007.terena.org/programme/presentations/show.php?pres_id=97
14. Grasa E, Figuerola S, Recio J, López A, de Palol M, Ribes L, Díaz V, Sangüesa R, Junyent G, Savoie M (October 2006) Video transcoding in a Grid network with user controlled light paths. Future Gen Comput Syst 22(8):920–928
15. Figuerola S, Ciulli N, de Leenheer M, Demchemko Y, Ziegler W, Binczewski A et al (2007) PHOSPHORUS: single-step on-demand services across multi-domain networks for e-Science. In: Wang J, Chang G-K, Itaya Y, Zech H (eds) Network architectures, management, and applications V. Proc SPIE, Denmark, 6784:67842X
16. Kesselman C, Foster I (1997) Globus: a metacomputing infrastructure toolkit. Int J Supercomput Appl 115–128
17. Zheng J, Mouftah HT (2001) Supporting advance reservations in wavelength-routed WDM networks. In: Tenth international conference on computer communications and networks, South Carolina
18. Figueira S, Kaushik N, Naiksatam S, Chiappari SA, Bhatnagar N (2006) Advance reservation of light paths in optical-network based grids. In: Proceedings of ICST/IEEE gridnets, San Jose
19. Global Lambda Integrated Facility (2010) Retrieved March 2011, from http://www.glif.is/
20. Ruffini M, O'Mahony D, Doyle L (August 2010) Optical IP switching: a flow-based approach to distributed cross-layer provisioning. IEEE/OSA J Opt Commun Netw, San Jose (California) 2(8):609–624
21. Ruffini M (2008) Optical IP switching. Ph.D. Thesis. Computer Science Department, University of Dublin, Trinity College
22. Sánchez A, Figuerola S, Junyent G, Kenny E, Reijs V, Ruffini M (2007) A user provisioning tool for EoMPLS services based on UCLPv1.5. In: Proceedings of the TERENA networking conference, Denmark
23. Mantychore project website (2010) [online]. Available: http://www.mantychore.eu/
24. AutoBAHN: GEANT2 Bandwidth on Demand (BoD) User and Application Survey (DJ.3.2.1) (2005) Available online: http://www.geant2.net/upload/pdf/GN2-05-086v11.pdf

25. Willner A, Barz C, Garcia JA, Ferrer J, Figuerola S, Martini P (2009) Work in progress: harmony—advance reservations in heterogeneous multi-domain environments. In: IFIP TC6 networking congress, Aachen
26. Zervas G, Escalona E, Nejabati R, Simeonidou D, Carozo G, Ciulli N, Belter B, Binczewski A, Stroinski M, Tzanakaki A, Marikdis G (2010) Phosphorus Grid-enabled GMPLS control plane (G2MPLS): architectures, services, and interfaces. In: IEEE communications magazine, multi-domain optical network issues and challenges
27. G-Lambda (2006) Website online available, Vienna (Austria) at: http://www.g-lambda.net/
28. Wu J, Savoie M, Campbell S, Zhang H, St. Arnaud B (2006) Layer 1 virtual private network management by users. Commun Mag IEEE, Azores (Portugal), 44(12):86–93
29. Wu J, Zhang H, Campbell S, Savoie M, Bochmann GV, St Arnaud B (2005) A Grid oriented light path provisioning system. In: Global telecommunications conference workshops, Dallas, 2004. GlobeCom workshops 2004. IEEE, pp 395–399
30. Grasa E, Figuerola S, Forns A, Junyent G, Mambretti J (2009) Extending the argia software with a dynamic optical multicast service to support high performance digital media. Opt Switch Netw 6(2):120–128. Recent trends on optical network design and modeling—selected topics from ONDM 2008, Aachen (Germany)
31. Figuerola S, Lemay M (July 2009) Infrastructure services for optical networks [invited]. IEEE/OSA J Opt Commun Netw 1(2):A247–A257
32. Dijkstra F, van der Ham J, Grosso P, de Laat C (2006) A path finding implementation for multi-layer networks. Future Gen Comput Syst 25(2):142–146
33. Despins C, Labeau F, Labelle R, Ngoc TL, McNeil J, Leon-Garcia A, Cheriet A, Cherkaoui O, Lemieux Y, Lemay M, Thibeault C, Gagnon F (2011) Leveraging green communications for carbon emission reductions: techniques, test beds and emerging carbon footprint standards. Commun Mag IEEE 49(8):101–109
34. Available Dallas (Texas), online: http://ec.europa.eu/research/fp6/index_en.cfm
35. Travostino F, Keates R, Lavian T, Monga I, Schofield B (2005) Project DRAC: creating an applications-aware network. Nortel Tech J
36. Barz C, Bornhauser U, Martini P, Pilz M, de Waal C, Willner A (2008) ARGON: reservation in grid-enabled networks. In: Proceedings of the 1st DFN-forum on communications technologies, Kaiserslauten
37. Campanella M, Krzywania R, Sevasti A, Thomas S-M (2008) Generic domain-centric bandwidth on demand service manager: deliverable GN2-08-129, Géant, August 2008. [Online]. Available: http://www.geant2.net/upload/pdf/GN2-08-129-DS3-3-4_Functional_Specification_and_Design_of_Generic_Domain-centric_BoD_Service_Manager.pdf
38. GÉANT2: DICE (dante-internet2-canarie-esnet) collaboration. [Online]. Available: http://www.geant2.net/server/show/conWebDoc.1308
39. Takefusa A, Hayahsi M, Nagatsu N, Nakada H, Kudoh T, Miyamoto T, Otani T, Tanaka H, Suzuki M, Sameshima Y et al (2006) G-Lambda: coordination of a Grid scheduler and lambda path services over GMPLS. Future Gen Comput Syst 22:868–875
40. Battestilli L, Hutanu A, Karmous-Edwards G, Katz D, MacLaren J, Mambretti J, Moore J, Park S, Perros H, Sundar S, et al (2007) EnLIGHTened computing: an architecture for co-allocating network, compute, and other grid resources for high-end applications. In: Grid computing, high-performance and distributed applications (GADA'07), Algarve
41. Lehman T, Sobieski J, Jabbari B (2006) DRAGON: a framework for service provisioning in heterogeneous Grid networks. Commun Mag IEEE 44:84–90
42. Farrel A, Vasseur JP, Ash J (2006) A path computation element (PCE)-based architecture. IETF RFC 4655
43. Network Service Interface WG (NSI-WG). ogf.org (2010) Retrieved March 2011, Nortel Technical Journal, February 2005, from http://www.ogf.org/gf/group_info/view.php?group=nsi-wg

44. OASIS Web Service Resource Framework Technical Committee (2006) [Online], Kaiserslauten (Germany) Available: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsrf
45. Moy J (1998) OSPF Version 2. Retrieved March 2011, from www.faqs.org/rfc/rfc2328.txt
46. Degilà J et al (2009) A survey of topologies and performance measures for large-scale networks. IEEE Commun Surv Tutorials 6:18–31
47. Kamiyama N et al (2008) Network topology design using analytic hierarchy process. In: Comm.: Proceedings of the IEEE international conference on communications ICC, pp 2048–2054
48. Willner A, Ferrer Riera J, Garcia-Espin JA, Siguerola S, De Leenheer M, Develder C (2010) An analytical model of the service provisioning time within the harmony network service plane. In: IEEE Globecom 2010 workshop on management of emerging networks and services (IEEE MENS 2010), Miami, FL, December 2010, pp 514–519. ISBN: 978-1-4244-8863-6, Algarve (Portugal), doi:10.1109/BLO-COMW.2010.5700373
49. Figuerola S, García-Espín JA, Ferrer Riera J, Willner A (2009) Performance analysis of harmony: an optical, multi-domain network resource broker. In: 11th international conference on transparent optical networks, 2009 (ICTON'09), Azores, June–July 2009, pp 1–5. doi:10.1109/ICTON.2009.518032
50. Sobieski J, Volbrecht J, Chaniotakis E (2011) Automated GOLE pilot project overview. In: Optical communications and networking. Presented at the APAN conference, Hong Kong, March 2011
51. Chaniotakis E (2011) Fenius—interoperability framework for virtual circuit provisioning systems. Retrieved March, 2011, from http://code.google.com/p/fenius/
52. Automated GOLE Pilot Project. wiki.glif.is. (2010) Retrieved March 2011, from http://wiki.glif.is/index.php/Automated_GOLE_Pilot_Project
53. Figuerola S, García-Espín JA, Ferrer Riera J, Willner A (2009) Scalability analysis and evaluation of the multi-domain, optical network service plane in harmony. In: 35th European conference on optical communication (ECOC '09), Vienna, September 2009, pp 1–2, 20–24

Chapter 15
Cross-Layer Network Design and Control Testbeds

Masahiko Jinno and Yukio Tsukishima

1 Introduction

Cross-layer design and control in optical networks include a wide range of techniques from the service layer to physical layer. In order to develop these interdisciplinary techniques, significant cooperative research efforts involving users of research and educational networks, application developers, network operators, and network equipment suppliers are needed. Once these techniques are developed, it is very important to build testbeds to evaluate the feasibility and effectiveness of networks employing these interdisciplinary techniques as well as to determine issues that need to be addressed before these networks can be deployed in the real world. Chapter 15 mainly presents cross-layer network design and control testbeds developed by Nippon Telegraph and Telephone Corporation (NTT) in close cooperation with other research institutes over the last few years, showcasing cross-layer network design and control techniques from service-layer aware to physical-layer aware techniques. Exhaustive enumeration of testbeds in this area will not be the focus of this chapter.

Before describing various testbeds, it may be useful to clarify the classification of cross-layer network design and control techniques used in this chapter. Figure 15.1 illustrates the network design and control model. At the top of this model is the application layer, which includes users and all types of applications such as scientific calculations, visualization, and a computer-supported collaborative work environment. The layer below is the network layer where Layer 2/3 and Layer 1 are

M. Jinno (✉)
NTT Network Innovation Laboratories, Yokosuka, Kanagawa, Japan

Kagawa University, Takamatsu, Kagawa, Japan
e-mail: jinno@eng.kagawa-u.ac.jp

Y. Tsukishima
NTT Network Innovation Laboratories, Yokosuka, Kanagawa, Japan
e-mail: tsukishima.yukio@lab.ntt.co.jp

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*,
Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_15,

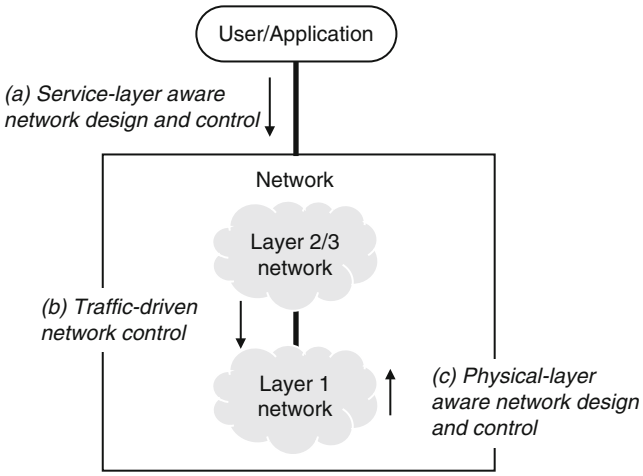


Fig. 15.1 Network design and control model

explicitly illustrated. In this chapter, cross-layer network design and control testbeds using (a) service-layer aware, (b) multilayer traffic-driven, and (c) physical-layer aware network design and control techniques are presented.

Using service-layer aware network design and control techniques, users or applications can request network resources that meet their bandwidth and latency requirements through the management plane. Coordinated computer/network resource allocation demonstrations across several network domains are presented in Sect. 2. Traffic-driven network control techniques improve the utilization efficiency of network resources as a whole. Testbed demonstrations of server-layer path provisioning as well as coordinated UNI-link failure recoveries by using the control plane are covered in Sect. 3. Physical-layer aware network design and control techniques overcome the limitations of transparent and translucent optical networks. Experimental demonstrations of impairment-aware management and control plane are introduced in Sect. 4.

In Sect. 5, we will focus our attention to the middle term future. Thanks to the recent remarkable advances in digital coherent detection followed by sophisticated Digital Signal Processing (DSP), the spectral efficiency of the cutting edge 100 Gb/s Dense Wavelength Division Multiplexing (DWDM) systems employing dual-polarization, quadrature-phase-shift-keying (QPSK) modulation is reaching as a high level as 2 b/s/Hz. Unfortunately, it is well known that bit loading higher than that for QPSK causes a rapid increase in the optical signal-to-noise ratio (OSNR) penalty, while further increase in the launched signal power results in serious impairment due to nonlinear effects in optical fibers. Therefore, it is becoming widely recognized that we are rapidly approaching the physical capacity limit of conventional optical fiber. Considering these challenges, there is growing anticipation toward the implementation of elastic optical path networks where the right-sized spectral resource is adaptively allocated to an optical path according to the actual client-layer traffic volume and/or network physical conditions in the fully optical domain. Section 5 provides a brief overview of

elastic optical path networks, together with testbed demonstrations in terms of service-layer and physical-layer aware network design and control in elastic optical networks.

2 Service-Layer Aware Network Design and Control Testbeds

2.1 Network Management Models

In most current and future bandwidth-hungry applications such as a computer-supported collaborative work environment, large-scale scientific calculation, and high-definition video conferencing, computing resources are geographically distributed and need to be connected across multiple network domains. In general, each independently managed network domain is controlled using different technology and managed under different operational policies. A network resource manager (NRM) manages the current and future available network resources in its domain and reserves those resources for future use. NRMs bridge network domains via a vertical or horizontal interface to harmonize provisioning of an end-to-end connection. There are two models for service-layer aware network management across multiple network domains, namely, the tree model (centralized management approach) and chain model (distributed management approach), as shown in Fig. 15.2.

In the tree model, a user or application issues a request to a resource coordinator (RC) for an end-to-end connection. The RC communicates with each NRM to provide the required end-to-end connection. In the chain model, a user or application sends a connection request message to an NRM. NRMs pass the message to an adjacent NRM in sequence on a one-by-one basis.

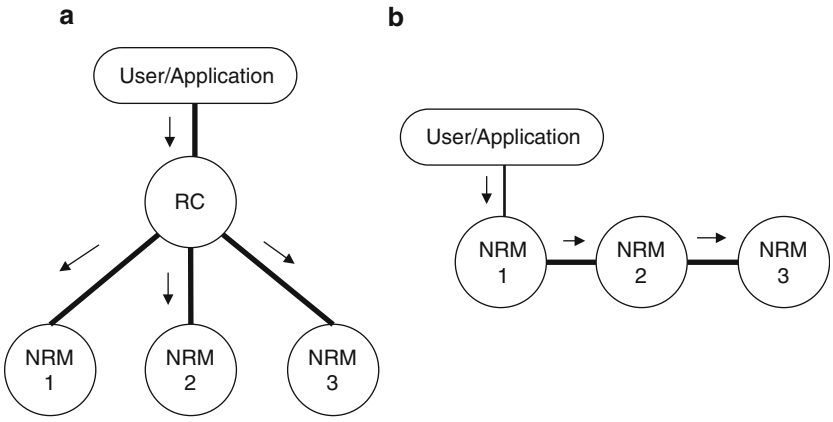


Fig. 15.2 Two models for application-driven network resource control across multiple domains. (**a**) Tree model (centralized management approach), (**b**) chain model (distributed management approach). *RC* resource coordinator, *NRM* network resource manager

2.2 Tree Model Network Control Testbeds

2.2.1 G-Lambda/EnLIGHTened Application-Driven Network Control Testbed

The G-lambda/EnLIGHTened project is an ambitious joint testbed to demonstrate in-advance reservation of coordinated computing and network resources between Japan and the USA [1]. Japan's G-lambda project is a joint collaboration among NTT, the National Institute of Advanced Industrial Science and Technology (AIST), KDDI R&D Laboratories (KDDI Labs), and the National Institute of Information and Communications Technology (NICT) that was started in 2004 [2]. The goal of the project is to establish a standard web services interface between an RC and an NRM. USA's EnLIGHTened computing project is an interdisciplinary effort among the Microelectronics Center of North Carolina (MCNC), Louisiana State University (LSU), North Carolina State University (NCSU), and several other organizations [3]. The project began in 2005 to design an architectural framework that would allow e-science applications to request dynamically in advance or on demand any type of grid resource.

The joint testbed consists of multiple domains with each domain managed by its own NRM, as shown in Fig. 15.3. On the Japan side, the testbed is built on the JGN

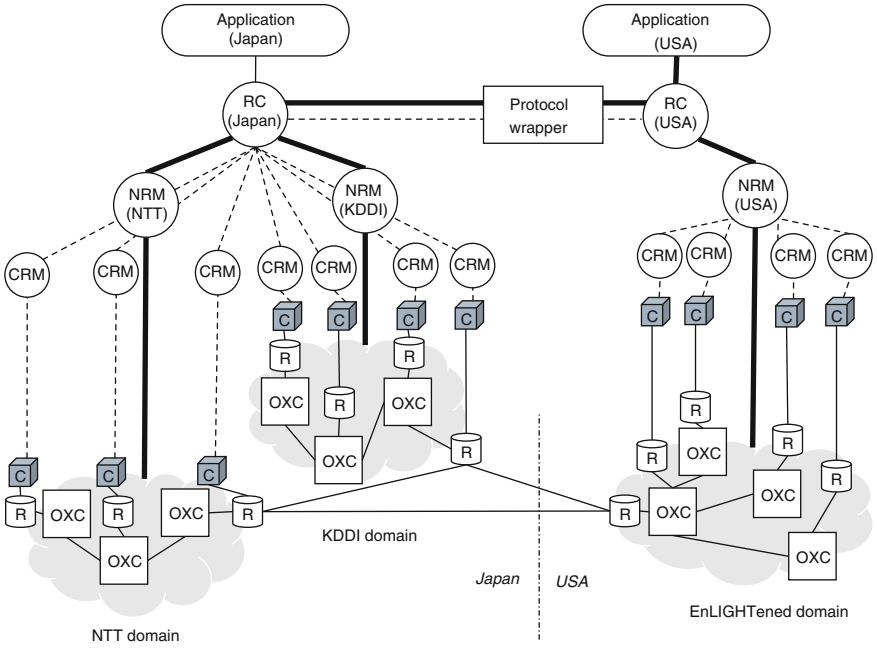


Fig. 15.3 G-lambda/EnLIGHTened application-driven network control testbed (tree model). *RC* resource coordinator, *NRM* network resource manager, *CRM* computing resource manager, *OXC* optical crossconnect, *R* router/switch, *C* computer cluster. Adapted from [4], © 2010 IEICE

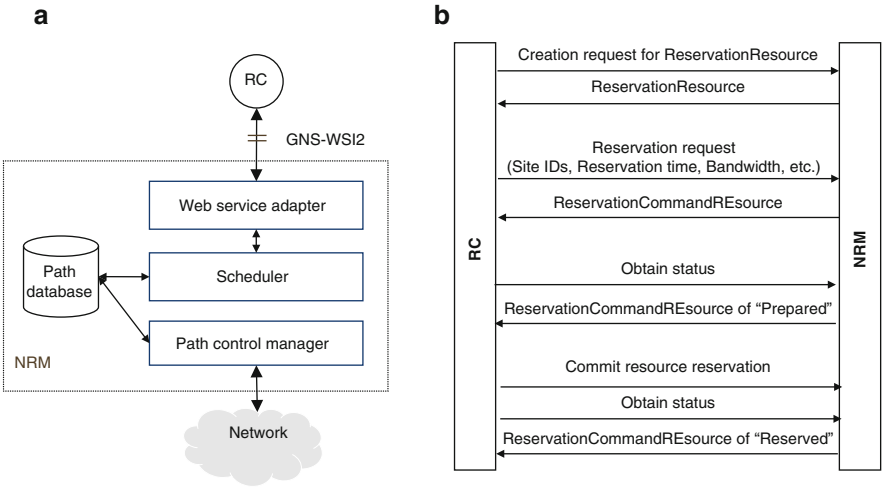


Fig. 15.4 NRM architecture example and GNS–WSI2 reservation scheme. (**a**) NRM architecture developed by NTT, (**b**) GNS–WSI2 sequence diagram defined by G-lambda project. Adapted from [4], © 2010 IEICE

2 research network, which includes two Generalized Multi-protocol Label Switching (GMPLS) administrative domains managed by NTT and KDDI Labs, and the Tokyo–Chicago intermediate transmission line. Each domain consists of multiple optical crossconnects (OXCs) and L2/3 devices controlled using GMPLS. Seven sites each with a computing cluster are each managed by a computing resource manager (CRM). The NRMs and CRMs communicate with an RC on a one-by-one basis to reserve computing and optical path resources. On the USA side, the testbed is built on the National Lambda Rail (NLR) and local Regional Optical Networks using four OXCs controlled using GMPLS. Since the RCs, NRMs, and CRMs were independently developed and had different interfaces between the two projects, a set of protocol wrappers was developed to bridge the different interfaces. A distributed numerical simulation and experimental distributed visualization were conducted using reserved computing and optical path resources.

Figure 15.4a illustrates an example of the NRM architecture. For the interface between the RC and the NRMs, the G-lambda project defined grid network service–web services interface version 2 (GNS–WSI2), a stateful web service protocol, in which a state is managed by an entity called an End Point (EP) [4]. The NRM consists of a GNS–WSI2 server module, a scheduler, a path database, and a network control and management module. GNS–WSI2 employs polling-based operations and a two-phase commit sequence, as shown in Fig. 15.4b, in order to enable a generic, non-blocking, and secure in-advance reservation process based on distributed transaction. According to the reservation request from a user or application, the RC sends the resource reservation request to NRMs containing site IDs, bandwidth, start and end time of service, etc. Each NRM calculates the necessary network resources and confirms the availability of the network resources.

If network resources that meet the request are available during the service period, the NRMs return the "prepared" message. The RC/CRM interface also employs a web service-based two-phase commit protocol similar to that of GNS–WSI2. The RC sends the "commit reservation" message only after all the NRMs and CRMs return the "prepared" message.

The major accomplishments of this testbed are the demonstration of the in-advance reservation of global scale coordinated computing and optical path resources across multiple administrative domains. One important finding obtained through the demonstration was that developing protocol wrappers to bridge different interface requires considerable effort, which could easily increase with the number of different interfaces [1]. In order to achieve interoperability among network service provisioning systems, standardization is under way to develop a common network provisioning service interface and a universal network service interface proxy in the Global Lambda Integrated Facility (GLIF) (http://code.google.com/p/fenius) and network service interface working group (NSI-WG) in the Open Grid Forum (OGF) (http://forge.gridforum.org/sf/projects/nsi-wg). In late 2010, a collaborative team comprising KDDI Labs, NICT, and AIST demonstrated dynamic circuit provisioning using a web service interface proxy for lambdas and Ethernet-based administrative domains on JGN2 plus and Internet 2 [5].

Another important finding is that since user/application traffic volume ranges from hundreds of megabits per second to tens of gigabits per second, bandwidth stranding occurs when the user/application traffic volume is not sufficient to fill the entire capacity of an optical path. One promising solution is to employ multilayer traffic control with the help of the emerging QoS-guaranteed packet transport technology.

2.2.2 Multilayer Lambda Grid Testbed

Figure 15.5 shows the multilayer lambda grid testbed built by NTT and AIST, which has two network domains connected via the JGN 2 research network and GEMnet 2 (Global Enhanced Multifunctional Network for NTT R&D testbed) [6]. The testbed comprises OXCs for optical path provisioning, Domain Edge Routers (DERs) for sub-lambda packet path provisioning, an RC, NRMs for each domain, and CRMs for each site. When a user sends a request message that contains the number of CPUs, bandwidth between computer sites, and the duration period, the RC reserves computers with the CRMs and a sub-lambda or lambda between the allocated computers with the NRM via GNS–WSI2. In order to ensure that the reserved sub-lambda packet path accommodates packets only from the allocated computers, the CRM advertises the information of the allocated computers, such as the MAC address or IP address, to the RC. The RC sends a message that contains the source and destination addresses of packets to the NRM. At the set time, the NRM configures the DER filtering conditions and establishes a sub-lambda with the reserved bandwidth between the allocated computers and discards the best effort traffic. In the demonstration, three sub-lambda packet paths with the bandwidth of 200 Mb/s for scientific calculation with the grid message passing interface (MPI), 500 Mb/s for super high-definition video (SHD), and 100 Mb/s for FTP were established over a 1 Gb/s optical path.

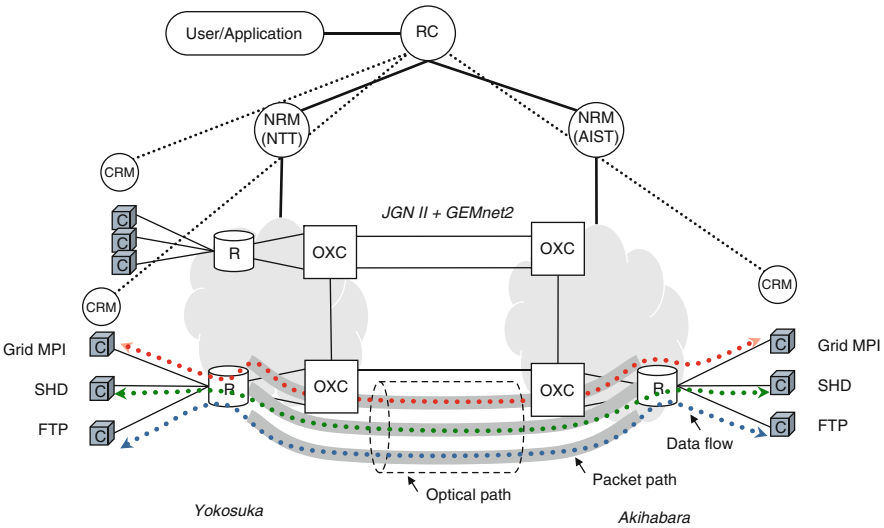


Fig. 15.5 Multilayer lambda grid testbed (tree model). *RC* resource coordinator, *NRM* network resource manager, *CRM* computing resource manager, *OXC* optical crossconnect, *R* router/switch, *C* computer cluster. Adapted from [6], © 2009 IEEE

2.3 Chain Model Network Control Testbeds

2.3.1 NTT–EVL Service-Layer Aware Network Control Testbed

NTT and the Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago (UIC) build a joint testbed in Chicago metropolitan area in order to demonstrate the establishment of a connection of skew-minimized parallel optical paths that possibly traverse multiple domains for latency-deviation sensitive, parallel visualization applications [7]. The scalable adaptive graphic environment (SAGE) developed by the EVL is a middleware system for managing visualization and high-definition video streams for viewing on ultrahigh-definition tiled displays. The SAGE application accommodates 55 LCD displays driven by a 30 node cluster of PCs with the graphics rendering capacity approaching nearly a terabit per second. The testbed has two administrative domains each having two OXCs, where NRMs are connected based on the chain model, as shown in Fig. 15.6. The SAGE middleware sent an advanced reservation request. Optical virtual concatenation (OVC) is implemented in the OXCs in order to de-skew the optical paths. The SAGE proxy sends reservation requests to NRM 1 with the node-IDs of the SAGE transmitter and SAGE receiver systems and the bandwidth of 1.6 Gb/s. NRM 1 allocates one GbE optical path for its domain and another GbE optical path for the other domain and sends a reservation request to NRM 2.

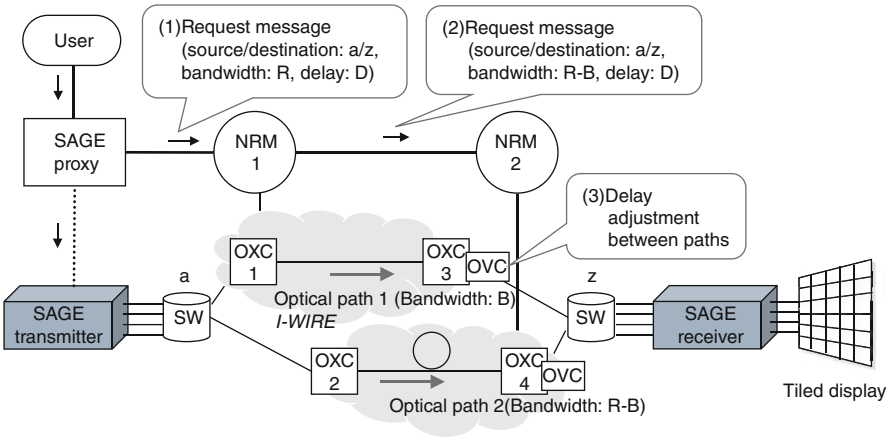


Fig. 15.6 NTT–EVL service-layer aware network control testbed (chain model). *NRM* network resource manager, *SAGE* scalable adaptive graphic environment, *OXC* optical crossconnect, *OVC* optical virtual concatenation

This testbed demonstrates the feasibility of coordinated advanced reservation over multiple domains in the chain model and de-skewing of parallel optical paths with OVC, which is required for high-end visualization applications.

2.3.2 UvA-Nortel Heterogeneous Domains Testbed

The experiment demonstrated during the SC 2004 Conference by Universiteit van Amsterdam, Nortel networks, and other organizations is another example of chain model service-layer aware network control [8]. NetherLight, StarLight, and OMNINet were used as three optical domains. Each domain was managed by Nortel's Dynamic Resource Allocation Controller (DRAC), which was given responsibility for the setup of intra- and inter-domain connections and was implemented with the grid network service agent. Another important key element in the service plane was the authentication, authorization, and accounting (AAA) subsystem.

3 Multilayer Traffic-Driven Network Control Testbeds

3.1 NTT–EVL Traffic-Driven Network Control Testbed

NTT and EVL also designed and built a testbed to demonstrate two scenarios of multilayer traffic-driven network control. The first scenario is the IP-routing-stable link failure restoration achieved through the coordinated control of Layers 1 and 3 [9].

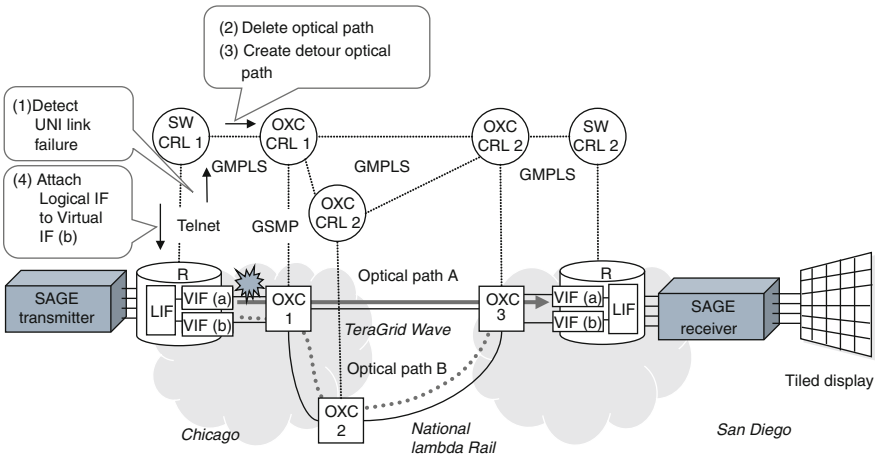


Fig. 15.7 NTT–EVL traffic-driven network control testbed. *OXC* optical crossconnect, *OXC CRL* OXC controller, *R CRL* router controller, *SAGE* scalable adaptive graphic environment, *LIF* logical interface, *VIF* virtual interface

The testbed consists of two OXCs at EVL in Chicago and an OXC at the University of California at San Diego (UCSD), which are connected by 10 Gb/s links via the NRL and TeraGrid Wave facilities, as shown in Fig. 15.7. A logical interface and proprietary virtual interfaces were implemented into edge Layer 3 switches to avoid reconfiguring the Layer 3 level network for both UNI and NNI link failures. A switch controller periodically monitors the health and traffic volume of the links. If there is a UNI link failure, for example, the virtual interface sends a failure message to the switch controller but pauses the IP-routing restart process for a certain guard time. Meanwhile, the switch controller localizes the failure, determines a detour route, requests the OXC controller to delete the optical path on the failed route, and establish an optical path on the detour route. Subsequently, the switch controller tears off the logical IF from the virtual IF (a) and attaches it to the virtual IF (b). In the demonstration, a link failure between Switch 1 and OXC 1 was simulated by disconnecting the fiber, and successful Layer 1/Layer 3 harmonized restoration was confirmed within a few seconds without any IP-routing instability.

The second scenario of multilayer traffic-driven network control is adaptive control of parallel optical paths according to time-varying demands of high-end, bandwidth-hungry applications [10, 11]. A proprietary GMPLS extension for controlling each Layer 2 link, which is virtually bundled with other links using the IEEE 802.3ad link aggregation technique, was implemented into switch controllers. In the demonstration, as the initial condition, a small streaming application on SAGE was operating over a 10 Gb/s Ethernet (10GE) connection established between the SAGE transmitter and receiver via optical path A traversing through the NRL. Subsequently, two additional streams were launched. A router controller that monitors the traffic from the SAGE transmitter detects when

the traffic volume exceeds the wavelength capacity, and then sends an optical path setup request to an OXC controller. After the second optical path was established, the switch controller activated another 10GE port and bundled the two 10GE connections into a 20 Gb/s virtual link using the link aggregation technique. The harmonized Layer 1 and Layer 2 network reconfiguration procedure enabled smooth bandwidth adjustment of the end-to-end Layer 2 connection over multiple nationwide optical paths.

4 Physical-Layer Aware Network Design and Control Testbeds

4.1 DICONET Impairment-Aware Network Planning and Operation Testbed

In order to operationalize transparent/translucent optical networks, which have the minimum number of expensive optical–electrical–optical regenerators, monitoring physical-layer impairments and optical performance incorporated with impairment-aware routing algorithms will be key. The goal of the DICONET (Dynamic Impairment Constraint Optical Networking) project, which is funded by the European Commission, is to design and develop an intelligent network planning and operation tool (NPOT), which considers the impact of physical-layer impairments in the planning and operation phases of optical networking.

Figure 15.8 illustrates the DICONET testbed located in Barcelona [12]. The testbed consists of a configurable signaling communications network (SCN) running over wavelength selective switch (WSS)-based OXC emulators. In the SCN, optical connection controllers (OCCs) are interconnected by 100 Mb/s Ethernet links, resembling the same physical topology of the emulated optical transport plane. Each OCC implements the full GMPLS protocol stack and is interconnected to the respective OXC. The NPOT consists of network description repositories, a physical-layer performance evaluator, impairment-aware routing and wavelength assignment engines, and component placement modules. The network management system (NMS) allows global supervision of the network active optical path state, the current configuration in each network node, and the requests for soft-permanent optical paths. Centralized and distributed control plane integration schemes are implemented in the testbed. The experimental evaluation in terms of the setup delay reveals that the distributed approach outperformed the centralized one especially for high traffic loads.

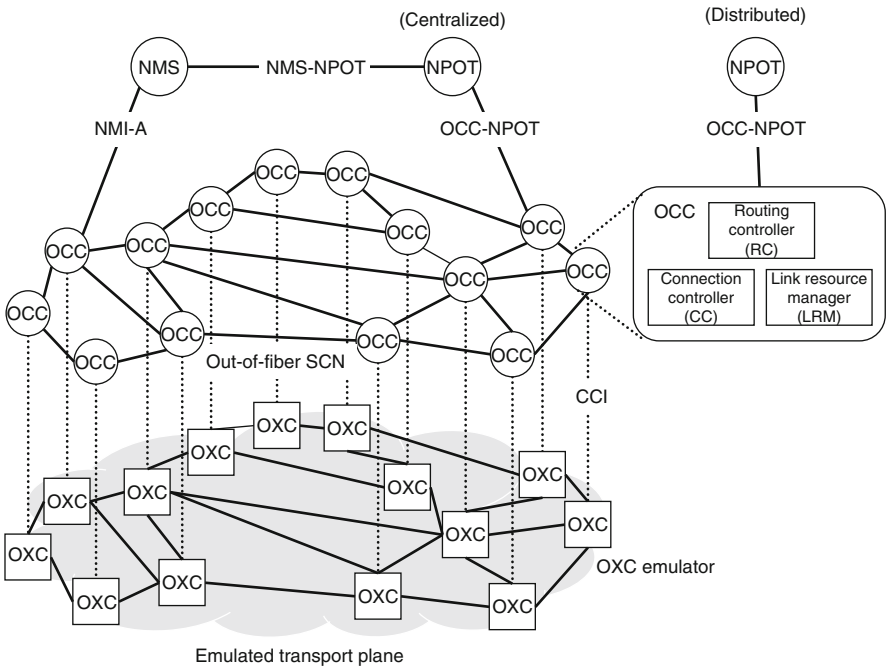


Fig. 15.8 DICONET impairment-aware network planning and operation testbed. *OXC* optical crossconnect, *SCN* configurable signaling communications network, *OCC* optical connection controller, *NPOT* intelligent network planning and operation tool, *NMS* network management system. Adapted from [12], © 2010 IEEE

4.2 Heterogeneous Translucent Optical Network Testbed

A multi-vender interoperability experiment in heterogeneous translucent optical network was demonstrated by a collaborative team comprising the NEC Corporation, Mitsubishi Electric Corporation, NTT, and KDDI Labs [13]. The testbed incorporated transparent and translucent domains. The transparent domain consists of three Reconfigurable Optical Add/Drop Multiplexers (ROADMs) and four Wavelength Crossconnects (WXCs), and both types had a colorless and directionless node architecture. The translucent domain consists of two OXCs connected with DWDM links. In order to achieve optical path setup across the multi-vender multi-domain translucent optical networks, the testbed implements GMPLS protocol extensions in terms of wavelength availability distribution and lambda label exchange. The major accomplishment of the testbed was the demonstration of second-order highly resilient multiple failure recovery in the multi-vender translucent optical network.

5 Spectrally Efficient Elastic Optical Path Network Testbeds

5.1 Elastic Optical Path Network Overview

In the current optically routed network, optical channels are aligned on the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) G.694.1 frequency grid. In such traditional optical networks, client-layer traffic flows are aggregated and groomed in the electrical domain to the level of limited kinds of line rates, for example, approximately 2.5, 10, and 40 Gb/s. In the future 100 G era and beyond, we may face difficulties in increasing the capacity in electrical aggregation and grooming switches while keeping cost, footprint, and power consumption at an acceptable level. In the mean time, the pace of spectral efficiency in WDM transmission systems is slowing down due to the vulnerability of higher-order multilevel modulation formats to OSNR degradation and nonlinear effects. One promising strategy to resolve the incoming capacity crunch is to boost the capacity at the network level by introducing "elasticity" and cross-layer "adaptation" into the optical domain [14–16]. This is the reason why elastic optical path networks, where the right-sized spectral resource is adaptively allocated to an optical path according to the client-layer actual traffic volume and/or network physical conditions in the fully optical domain, have been fostering growing anticipation in the past few years. This subsection provides a brief overview of elastic optical path networks, together with testbed demonstrations in terms of the service-layer and physical-layer aware network design and control.

Figure 15.9 illustrates how elastic optical path networks provide spectrally efficient transport of 100 Gb/s services and beyond. If based on the conventional design philosophy, every optical path is aligned on an ITU-T fixed grid (Fig. 15.9a) regardless of the path length, bit rate, or actual client traffic volume (Fig. 15.9b). By taking advantage of spectral-efficiency-conscious adaptive signal modulation and elastic channel spacing, elastic optical path networks yield significant spectral savings. For shorter optical paths, which suffer from less OSNR degradation, a more spectrally efficient modulation format such as 16 Quadrature Amplitude Modulation (QAM) is employed instead of a more SNR tolerant but less spectrally efficient QPSK format. For client traffic that does not fill the entire capacity of a wavelength, the elastic optical path network provides right-sized intermediate bandwidth, such as 200 Gb/s (Fig. 15.9c). Combined with elastic channel spacing based on the "frequency slot" concept (Fig. 15.9d), where the required minimum guard band is assigned between channels, elastic optical path networks accommodate a wide range of traffic in a spectrally efficient manner without any intermediate electrical grooming switches (Fig. 15.9e).

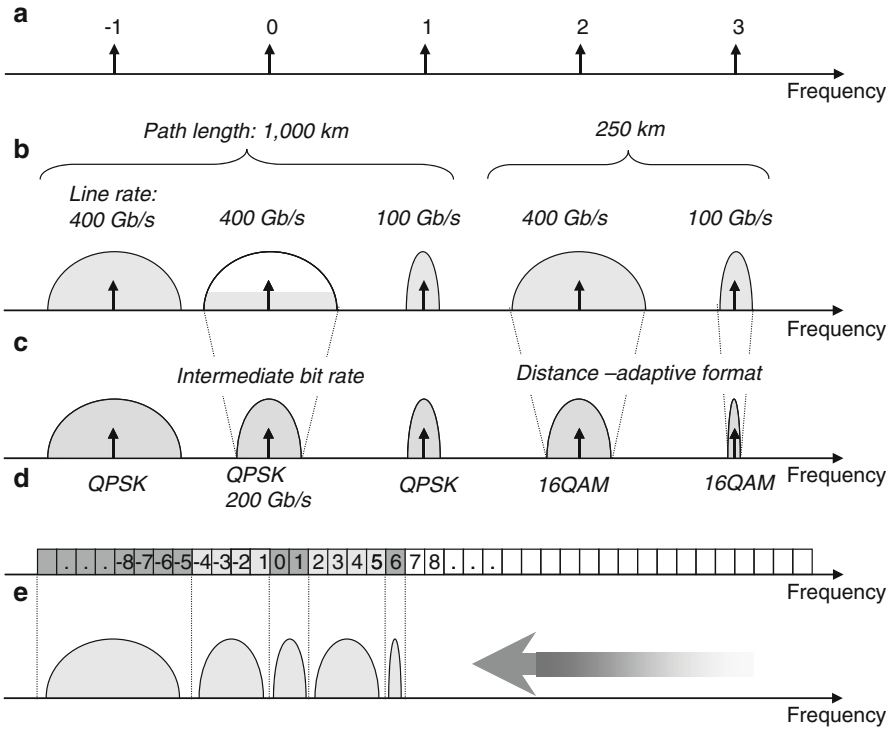


Fig. 15.9 User-rate and distance-adaptive spectrum resource allocation in elastic optical path network. (**a**) ITU-T G.694.1 frequency grid, (**b**) conventional channel plan, (**c**) optimized modulation considering actual user bandwidth and path distance, (**d**) frequency slot related to current ITU-T grid, (**e**) spectrally efficient elastic channel plan

5.2 Service-Layer Aware Elastic Optical Path Network Testbed

5.2.1 SLICE Service-Layer Aware Network Control Testbed

The SLICE (*s*pectral s*lic*ed *e*lastic optical path network) testbed built by NTT is schematically depicted in Fig. 15.10. Key building blocks of the elastic optical network are a rate and format flexible optical transponder and a bandwidth agnostic WXC [17]. Introduction of coherent detection followed by DSP will yield a novel degree of freedom in designing transponders. By optimizing three parameters, the symbol rate, the number of modulation levels, and the number of subcarriers, the required data rate and optical reach can be provided while minimizing the spectral width. For example, the flexible optical reach can be achieved by changing the number of bits per symbol with a high-speed digital-to-analog converter and In-phase/Quadrature (IQ) modulator. Optical Orthogonal Frequency Division

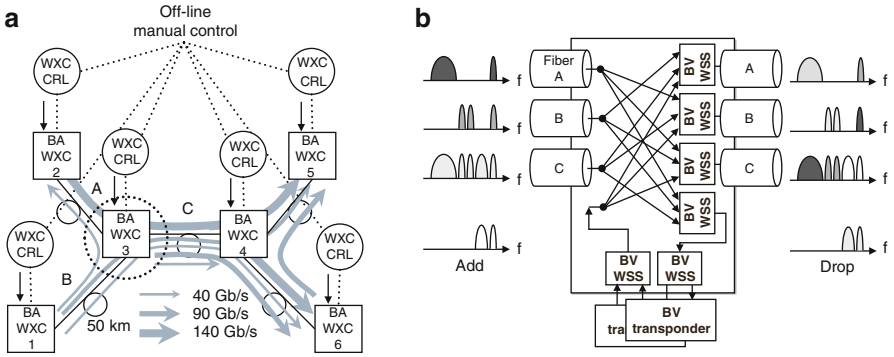


Fig. 15.10 SLICE service-layer aware network control testbed. (**a**) Testbed configuration with multiple rate elastic optical paths, (**b**) bandwidth agnostic wavelength crossconnect (BA-WXC) architecture and signal spectra at BA-WXC 3. *WXC CRL* wavelength crossconnect controller, *BV-WSS* bandwidth-variable wavelength selective switch. Adapted from [17], © 2010 OSA

Multiplexing (OFDM) is a spectrally overlapped orthogonal subcarrier modulation scheme and allows a flexible rate transmitter by customizing the number of subcarriers of the OFDM signal. Bandwidth agnostic WXCs can be achieved by using a continuously bandwidth-variable WSS based on, for example, liquid crystal on silicon (LCoS) technology. In a bandwidth-variable WSS, the incoming optical signals with different optical bandwidths and center frequencies can be routed to any of the output fibers. These technologies allow opening of the required minimum spectrum window at every node along the optical path.

The SLICE testbed incorporates six bandwidth agnostic WXCs connected by 50 km spans of optical fiber to establish a mesh configuration (Fig. 15.10a). The bandwidth agnostic WXCs employ a broadcast-and-select architecture, where $1 \times N$ optical splitters and bandwidth-variable $N \times 1$ WSSs are arranged at the input and output, respectively. This architecture enables forwarding of channels with arbitrary spectral widths to arbitrary output ports, and add-and-drop functions as well as broadcasting functionality. Optical OFDM signals having a bandwidth ranging from 40 to 440 Gb/s are generated according to a user/application request and demodulated using a multicarrier light source synchronized with the data signal, an optical multiplexer, an optical demultiplexer, and optical gate as an equivalent of the electrical inverse First Fourier Transform (FFT) and FFT method. An example of the optical channel assignment in terms of the optical spectrum is illustrated in Fig. 15.10b with the spectra of optical channels at the input, output, and the add/drop ports of bandwidth agnostic WXC 3. In addition to the multiple rate optical path setup and transmission according to the user/application demand described so far, the SLICE testbed demonstrated bandwidth scaling of a single elastic optical path from 40 to 440 Gb/s in accordance with demand change. In both cases, the Q-factor performance of elastic optical paths were measured and confirmed above the Forward Error Correction (FEC) Q limit of 9.1 dB.

5.3 *Physical-Layer Aware Elastic Optical Path Network Testbed*

5.3.1 SLICE Physical-Layer Aware Network Design Testbed

Another important accomplishment of the SLICE testbed is the demonstration of adaptive spectrum allocation design with modulation format/baud rate and filter passband adjustment according to physical conditions of the network. In the testbed, multiple WDM optical paths with QPSK and 16 amplitude phase-shift keying (APSK) formats are transmitted over a bandwidth agnostic WXC placed in a recirculating loop mimicking the characteristics of a ring network, as shown in Fig. 15.11 [18]. The 16APSK format employs four phase levels and four intensity levels in a cross-shaped constellation, which results in constellation point spacing similar to that of 64-QAM. As a result, the 16APSK format provides half the spectral width at the expense of an approximate 8 dB OSNR penalty when compared to those for the QPSK format. A multicarrier source generates a 50 GHz-spaced optical wavelength comb, and the following WSS directs the comb output into one of two modulator branches to produce 100 GHz-spaced 21.4 Gbaud, 42.7 Gb/s QPSK signals for the long-reach paths and 10.7 Gbaud, 42.7 Gb/s 16APSK signals for the short reach paths. A separate, single 16APSK channel using a narrow linewidth optical source is also generated to test the performance of the multilevel format. All branches are multiplexed in a WSS. Six QPSK and 10 16APSK signals are transmitted in a recirculating loop containing 40 km of single mode fiber (SMF), a dispersion compensation fiber (DCF), a bandwidth agnostic WXC, and a gain equalizing filter.

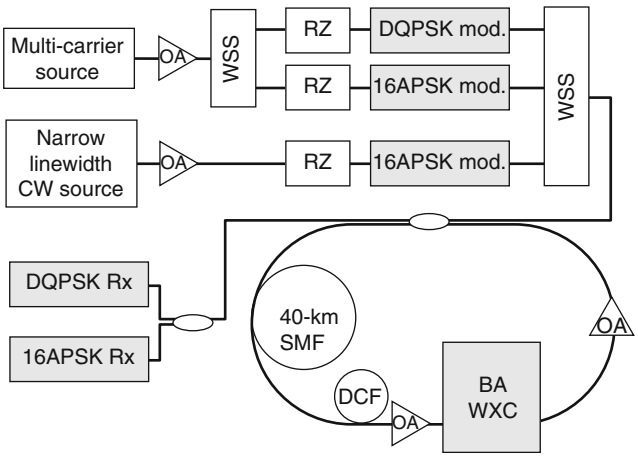


Fig. 15.11 SLICE physical-layer aware network design testbed. *CW* continuous wave, *OA* optical amplifier, *WSS* wavelength selective switch, *RZ* return to zero, *DQPSK mod.* differential quadrature phase-shift keying modulator, *16APSK* 16 amplitude phase-shift keying modulator, *OSW* optical switch, *SMF* single mode fiber, *DCF* dispersion compensation fiber, *BA-WXC* bandwidth agnostic wavelength crossconnect, *Rx* receiver

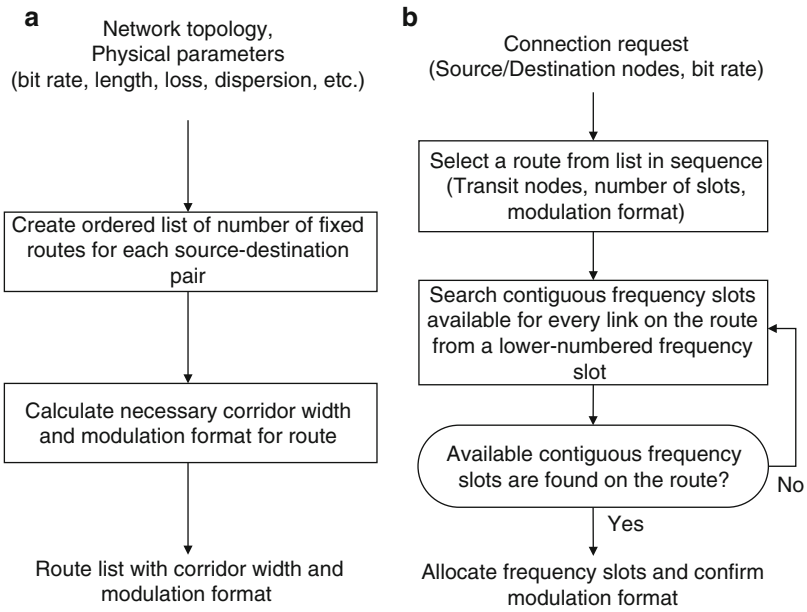


Fig. 15.12 Rate- and distance-adaptive elastic bandwidth allocation sequence diagrams. (**a**) Necessary spectrum resource calculation sequence, (**b**) routing and spectrum allocation algorithm

Performance analysis of both the QPSK and the 16APSK signals after transmission of multiple WXCs and fiber spans revealed that the distance-adaptive spectrum allocation allows transmission of 16 instead of 11 channels, which corresponds to an increase in spectral efficiency of 45%.

The problem of calculating routes for optical paths in transparent optical networks is called the routing and wavelength assignment (RWA) problem with the wavelength continuity constraint. Adaptive spectral allocation in elastic optical path networks introduces a more severe constraint on spectrum continuity, as sort of the routing and spectrum assignment (RSA) problem taking into account linear and nonlinear impairment factors. The RSA algorithm can be divided into two stages, as shown in Fig. 15.12 [15]. The first stage is to create a route list with the necessary spectral width and modulation format. Given a network topology and physical parameters, an ordered list of a number of fixed routes for each source-destination pair is created first. Subsequently, the necessary spectral width and modulation format for each route at a given bit rate can be calculated taking into account linear and nonlinear impairment factors. The second stage is to allocate a contiguous frequency slot to the route. When a connection request arrives, a route from the ordered route list created in the previous stage is selected in sequence. The number of necessary slots is found from the list according to the physical parameters of the route. Then, the available contiguous slots available for every link on the route is searched from a lower-numbered slot, and the lowest available contiguous slots is selected. If no available contiguous slots are found on the route, an alternate route is selected from the route list.

6 Summary

In this chapter, we described a wide variety of research testbeds on cross-layer network design and control in optical networks in terms of service-layer aware, multilayer traffic-driven, and physical-layer aware techniques. Table 15.1 summarizes the described testbeds and their respective objectives. These testbeds demonstrated the feasibility and effectiveness of cross-layer design in optical networks. Many issues that need to be solved before these networks can be commercialized had been recognized during the planning and operating of the testbeds.

Fortunately, such feedback has given rise to insights for the subsequent work. For example, the interoperability issue when a user/application requires optical paths across multiple administrative domains is being investigated, and a common network provisioning service interface and universal network service interface proxy are under standardization in the OGF as described in Sect. 2.2. The first interoperability demonstration was successfully conducted in late 2010. The granularity mismatch issue between an optical path and user/application traffic was first addressed by introducing coordinated packet and optical path architecture. In the

Table 15.1 Network design and control testbeds

Network design and control technique	Testbed	Institutions	References
Service-layer aware	G-lambda/ EnLIGHTened	AIST, NTT, KDDI Labs, NIST, MCNC, LSU, NCSU, etc.	[1–3]
	Dynamic circuit provisioning	KDDI Labs, NICT, AIST	[5]
	Multilayer lambda grid	NTT, AIST	[6]
	NTT–EVL service-layer aware network control	NTT, EVL at UIC	[7]
	UvA-Nortel heterogeneous domains	Universiteit van Amsterdam, Nortel networks, etc.	[8]
	SLICE service-layer aware network control	NTT	[17]
Traffic driven	NTT–EVL traffic-driven network control	NTT, EVL at UIC	[9–11]
Physical-layer aware	DICONET	UPC, ALU, Create-Net, RACTI, TELECOM ParisTech, DT, AIT	[12]
	Heterogeneous translucent optical network	NEC, Mitsubishi, NTT, KDDI Labs	[13]
	SLICE physical-layer aware network design	NTT	[18]

middle term, this issue should be addressed by introducing elasticity and adaptation into the optical domain. The results achieved through the SLICE testbed confirmed that elastic optical path networks have the potential for a more prominent role in optics in the context of a more efficient and scalable optical layer as a mission critical infrastructure to support the future Internet and services.

References

1. Thorpe SR, Battestilli L, Karmous-Edwards G, Hutanu A, MacLaren J, Mambretti J, Moore JH, Sundar KS, Xin Y, Takefusa A, Hayashi M, Hirano A, Okamoto S, Kudoh T, Miyamoto T, Tsukishima Y, Otani T, Nakada Ho, Tanaka H, Taniguchi A, Sameshima Y, Jinno M (2007) G-lambda and EnLIGHTened: wrapped in middleware co-allocating compute and network resources across Japan and the US. In: Proceedings of the international conference on networks for grid applications 2007 (Lyon), Article 5
2. Takefusa A, Hayashi M, Nagatsu N, Nakada H, Kudoh T, Miyamoto T, Otani T, Tanaka H, Suzuki M, Sameshima Y, Imajuku W, Jinno M, Takigawa Y, Okamoto S, Tanaka Y, Sekiguchi S (2006) G-lambda: coordination of a grid scheduler and lambda path service over GMPLS. Future Gen Comput Syst 22(8):868–875
3. Battestilli L, Hutanu A, Karmous-edwards G, Katz DS, Maclaren J, Mambretti J, Moore JH, Park S, Perros HG, Sundar S, Tanwir S, Thorpe SR, Xin Y (2007) EnLIGHTend computing: an architecture for co-allocated network, compute, and other grid resources for high-end applications. In: Proceedings of the international conference of high capacity optical networks and enabling technologies 2007 (Dubai), pp 1–8
4. Tsukishima Y, Hayashi M, Kudoh T, Hirano A, Miyamoto T, Takefusa A, Taniguchi A, Okamoto S, Nakada H, Sameshima Y, Tanaka H, Okazaki F, Jinno M (2010) Grid network service–web services interface version 2 achieving scalable reservation of network resources across multiple network domains. IEICE Trans Commun E93-B(10):2696–2705
5. Miyamoto T, Tanaka J, Otsuki H, Kudoh T (2010) Field demonstration of dynamic circuit provisioning by web services interface proxy for lambdas and Ethernet based administrative domains. In: Proceedings of the Asia communications and photonics conference and exhibition 2010 (Shanghai), pp 465–466
6. Tsukishima Y, Sameshima Y, Hirano A, Jinno M, Kudoh T, Okazaki F (2009) Multi-layer Lambda grid with exact bandwidth provisioning over converged IP and optical networks. J Lightwave Technol 27(12):1776–1784
7. Tsukishima Y, Hirano A, Taniguchi A, Imajuku W, Jinno M, Hibino Y, Takigawa Y, Hagimoto K, Wang X, Renambot L, Jeong B, Jagodic R, Nam S, Leigh J, DeFanti T, Verlo A (2007) The first optically-virtual-concatenated lambdas over multiple-domains in Chicago metro area network achieved through inter-working of network resource managers. In: Proceedings of the optoelectronics and communications conference 2007 (Yokohama) 12A2-5, July 2007
8. Gommans L, Dijkstra F, de Laat C, Taal A, Wan A, Oudenaarde B, Lavian T, Monga I, Travostino F (2006) Applications drive secure lightpath creation across heterogeneous domains. IEEE Commun Mag 44(3):100–106
9. Tsukishima Y, Hirano A, Nagatsu N, Ohara T, Imajuku W, Jinno M, Takigawa Y, Hagimoto K, Renambot L, Jeong B, Leigh J, DeFanti T, Verlo A, Winkler L (2006) Stable IP-routing link restoration: GUNI restoration for data link failure between routers in a nationwide photonic network. In: Proceedings of the European conference on optical communication 2006 (Cannes), We4.1.4
10. Tsukishima Y, Hirano A, Nagatsu N, Ohara T, Imajuku W, Jinno M, Takigawa Y, Hagimoto K, Renambot L, Jeong B, Leigh J, DeFanti T, Verlo A, Winkler L (2006) The first application-driven

lambda-on-demand field trial over a US nationwide network. In: Proceedings of the optical fiber communication conference and exposition and the national fiber optic engineers conference 2006 (Anaheim), PDP48
11. Tsukishima Y, Hirano A, Nagatsu N, Imajuku W, Jinno M, Hibino Y, Takigawa Y, Hagimoto K, Wang X, Renambot L, Jeong B, Leigh J, DeFanti T, Verlo A (2007) Lambda sharing demonstration via traffic-driven lambda-on-demand. In: Proceedings of the European conference on optical communication 2007 (Berlin), 06.5.3
12. Azodolmolky S, Perelló J, Angelou M, Agraz F, Velasco L, Spadaro S, Pointurier Y, Francescon A, Saradhi CV, Kokkinos P, Varvarigos E, Zahr SA, Gagnaire M, Gunkel M, Klonidis D, Tomkos I (2010) Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks. J Lightwave Technol 29(4):439–448
13. Nishioka I, Liu L, Yoshida S, Huand S, Hayashi R, Kudo K, Tsuritani T (2010) Experimental demonstrations of dynamic wavelength path control and highly resilient recovery in heterogeneous optical WDM networks. In: Proceedings of the optoelectronics and communications conference 2010 (Sapporo), PD1
14. Jinno M, Takara H, Kozicki B, Tsukishima Y, Sone Y, Matsuoka S (2009) Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies. IEEE Commun Mag 47(11):66–73
15. Jinno M, Kozicki B, Takara H, Watanabe A, Sone Y, Tanaka T, Hirano A (2010) Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network (SLICE). IEEE Commun Mag 48(8):138–145
16. Jinno M, Ohara T, Sone Y, Hirano A, Ishida O, Tomizawa M (2011) Elastic and adaptive optical networks: possible adoption scenarios and future standardization aspects. IEEE Commun Mag 49(6):164–172
17. Kozicki B, Takara H, Tsukishima Y, Yoshimatsu T, Kobayashi T, Yonenaga K, Jinno M (2010) Experimental demonstration of spectrum-sliced elastic optical path network (SLICE). Opt Express 18(21):22105–22118
18. Kozicki B, Takara H, Tanaka T, Sone Y, Hirano A, Yonenaga K, Jinno M (2011) Distance-adaptive path allocation in elastic optical path networks. IEICE Trans Commun E94-B (7):1823–1830

Chapter 16
Free Space Optical Wireless Network

Vincent W.S. Chan

1 Introduction

This chapter explores the design of an over the air optical wireless network architecture. Behavior of the physical layer of this network necessitates the creation of a new architecture construct for a reliable high data rate network that can interoperate with traditional wired and fiber networks, Fig. 16.1 [1–4]. Such networks will have significant impact for applications between ships, airborne vehicles, satellites, and land-based local and metro networks, as well as cloud computing and data centers as part of an integrated high-speed information system. Throughputs of many Tbps can be supported by such networks, which are compatible with the next generations of carrier-class Ethernet at 100 Gbps per wavelength.

The big challenge for the architecture construct is to engineer the free-space optical wireless network in the presence of atmospheric turbulence and weather, so it will behave properly when interfaced to other wire-line networks and for the protocols of the various network layers to interoperate end-to-end in a heterogeneous network setting. To engineer the free-space optical wireless network one needs to address two major components of the architecture:

1. Model the physical layer and data link layer properties of the atmospheric channel as seen by the upper network layers.
2. Design and optimize an optical wireless network (from Layer 1 to Layer 4) fully taking into account the behavior and system tunability of the physical and data link layers as well as space of possible changes that can be made to the upper layers (L3 and L4).

V.W.S. Chan (✉)
Joan and Irwin Jacobs Professor, Department of Electrical Engineering and Computer Science, Claude E. Shannon Communication and Network Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: chan@mit.edu

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*, Optical Networks 15, DOI 10.1007/978-1-4614-5671-1_16,

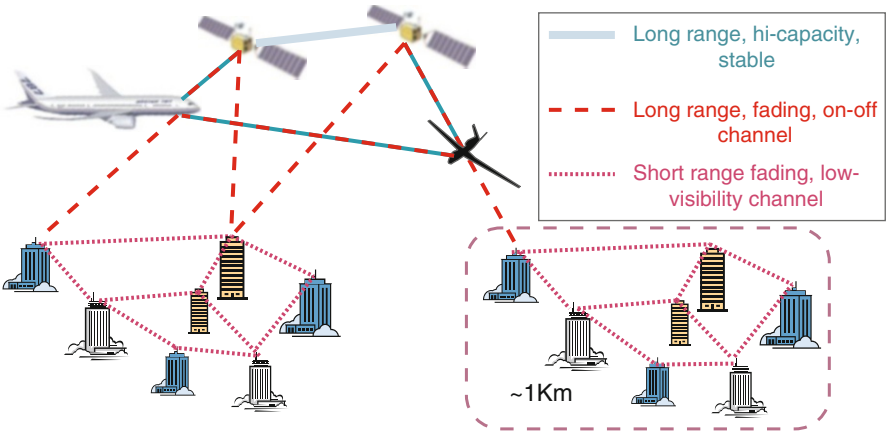


Fig. 16.1 Optical wireless network concept

This chapter will start with a brief summary of the statistical model of the atmospheric turbulent optical channel and discuss the design of all the optical wireless network layers. To make easy appreciation of the subtle issues involved in such architecture, an urban network is used as an example to guide the reader through the new architecture constructs. Similar techniques can be used for longer range systems such as aircraft to ground and satellite channels.

2 Transmission Model for the Turbulence Optical Atmospheric Channel

Atmospheric optical communications are susceptible to fading due to refractive index fluctuations induced by air turbulence. As predicted by the Kolmogorov turbulence model, temperature variations in the air on the order of 1 K cause refractive index variations on the order of several parts per million, Fig. 16.2. The Extended Huygens–Fresnel Principle models the short-term fading of a wave due to passage through air turbulence [1–4]. Depending on the crosswind velocity, it is typical for deep fades to last approximately 1–100 ms, which, when a link of a network is operating at multi-gigabits per second, results in the loss of potentially up to $>10^9$ consecutive bits. This motivates the need to consider schemes that minimize the probability that the receiver "sees" a significant fade [5–10]. Figure 16.2 illustrates the intensity pattern of a laser beam transmitted over the atmosphere. The highs and lows of the pattern move over the receiver aperture (typically in a small area of the pattern) as transverse winds blow across the beam. Thus, intensity fluctuations are seen at the receiver with a coherence time proportional to the scale of the turbulence divided by the transverse wind velocity.

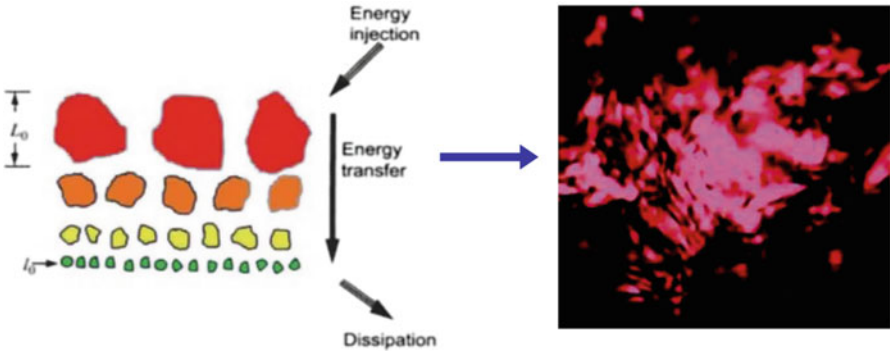


Fig. 16.2 Free-space optical communication in turbulent atmosphere

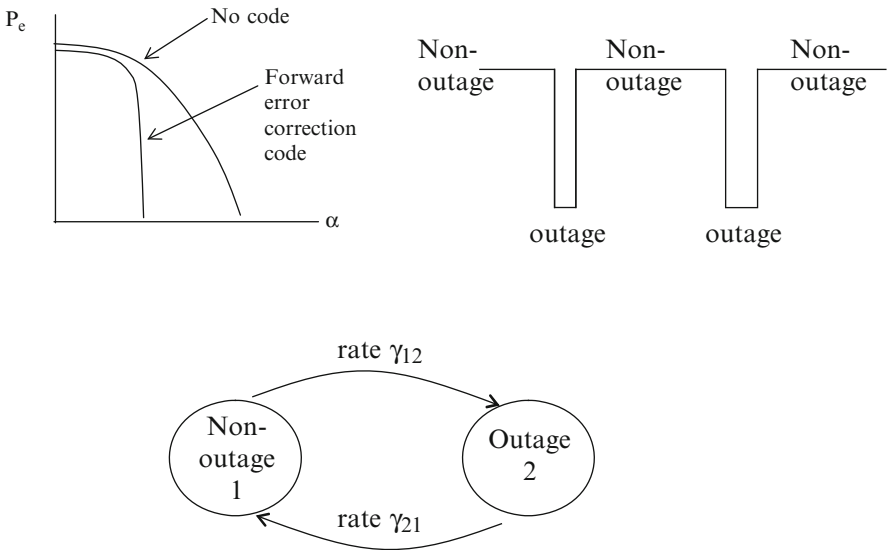


Fig. 16.3 Markov channel model of optical atmospheric transmission

If the communication system uses an error correction code (which almost always will be the case), the system is either communicating effectively if there is enough power received or off if the signal is faded. Thus, the channel can be modeled as a two-state "on–off" Markov Process, Fig. 16.3. The channel can be easily measured and characterized by a terrestrial optical link with single or multiple transmit and receive apertures. Figure 16.4 shows the statistics of a single transmit and two receivers link over a horizontal path of ~500 m. The one time statistics of the outage and non-outage states are approximately exponential, and the correlation function of the intensity is also approximately exponential [11, 12]:

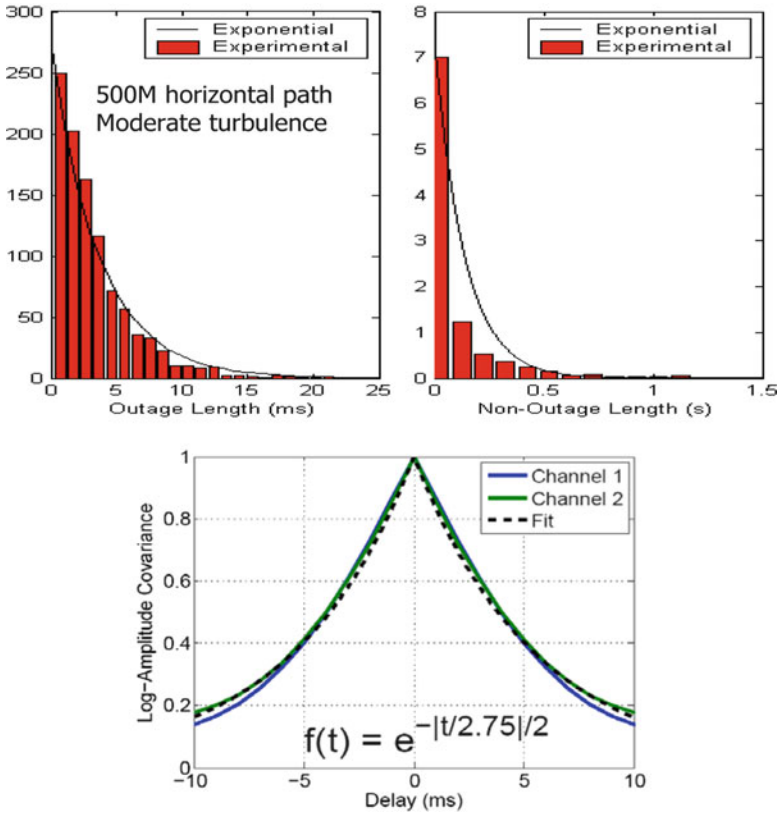


Fig. 16.4 Outage and non-outage experimental data fit to two-state atmospheric fading Markov model [7]

$$R_{hh}(t) = \exp\left\{-\frac{1}{2}\left(\frac{v_\perp t}{\rho_0}\right)^{5/3}\right\} \sim \exp\left\{-\frac{t}{t_0}\right\}$$

where $v_\perp$ is the transverse wind velocity, ρ_0 is the coherence length, and t_0 is the coherence time of the receiver intensity.

The usual performance metric in analyzing communication systems is the probability of bit error. However, when analyzing systems for optical communication through atmospheric turbulence, the average probability of error is not the best metric. This is because errors resulting from signal fades are no longer independent and large strings of data can be lost. Using the performance metric of outage probability indicates how often the system is below performance threshold. Outage probability is defined as the probability that the short-term (over a duration less than the channel coherence time) bit error rate P_e is above a determined required value P_e^*.

$$Pr(outage) = Pr(P_e > P_e^*) = Pr(\alpha < \alpha^*)$$

where α is the power fading factor and α^* is the fading power factor value that yields P_e^*. For an optical wireless network, fades are typically very deep (~10 dB). Packets transmitted during an outage are assumed to be lost, and packets transmitted during a non-outage are assumed to be received correctly. The length of time spent in states 1 and 2 (the non-outage and outage lengths, respectively) is exponentially distributed (being Markov processes by approximation). Typical outage duration ranges from a hundred ms to a fraction of an ms when diversity transmission and/or reception is used. The inter-arrival times of these outages are of the order of 100 ms. These are important parameters for the link to be operated as part of a data communication network particularly for the design of the transport and network layers. Significant improvements of the values of these parameters can be realized with spatial and temporal diversity for direct detection receivers, and the gains will be much more for coherent systems allowing much higher data rates.

3 Multi-Aperture Incoherent and Coherent Transmitter/Receiver—Statistical Model and Communication Performance Enhancements

Spatial diversity is an attractive technique to mitigate fades in the received signal. In contrast to spatial diversity for wireless systems, spatial diversity for atmospheric optical systems can be readily implemented in a compact fashion since the coherence length is of the order of centimeters, i.e., multiple transmitters or receivers only need to be placed centimeters apart to see approximately independent channel fades. On the other hand, an error correcting code alone is not an attractive technique to mitigate fades. It would require a huge interleaver on the order of 100 Gb for a 100 Gbps channel, resulting in, among other impediments, excessive link delays for the upper layer protocols, severely affecting the performance of most congestion control algorithms, yielding very low throughputs and poor efficiencies.

The available diversity equals the product of the number of independent (in the sense of being in different phase coherence cells) transmitter elements and receiver elements. If a low rate (~10 kbps) feedback link is available between the receiver and the transmitter, transmitter phase predistortion can be used to help focus the optical energy on the receiver array, enhancing energy delivery efficiency of the free-space channel. Receiver phase tracking element by element either via local-oscillator tuning or signal path phase modulation can be used to estimate the atmospheric optical channel state and the data fed back to the transmitter for phase and amplitude modulation of the individual array elements, Fig. 16.5.

Typical outage durations are shortened from a few ms to a fraction of an ms when diversity is used. Figure 16.6 shows typical expected outage probabilities, Fig. 16.7 outage lengths and Fig. 16.8 good channel durations over a short

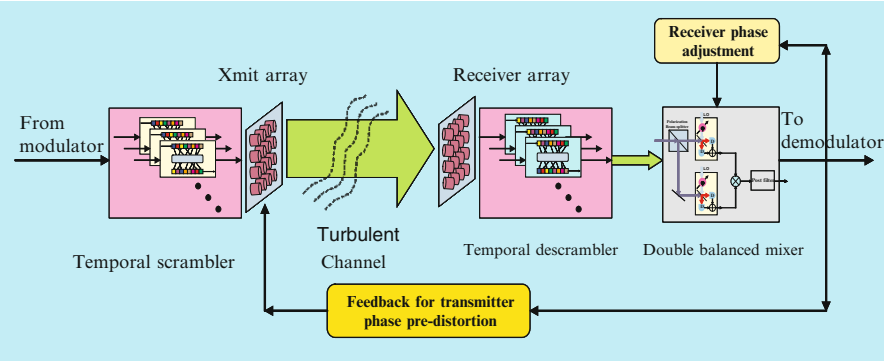


Fig. 16.5 Transmission system with dynamically and adaptively reconfigurable receiving plane field patterning via transmitter signal predistortion

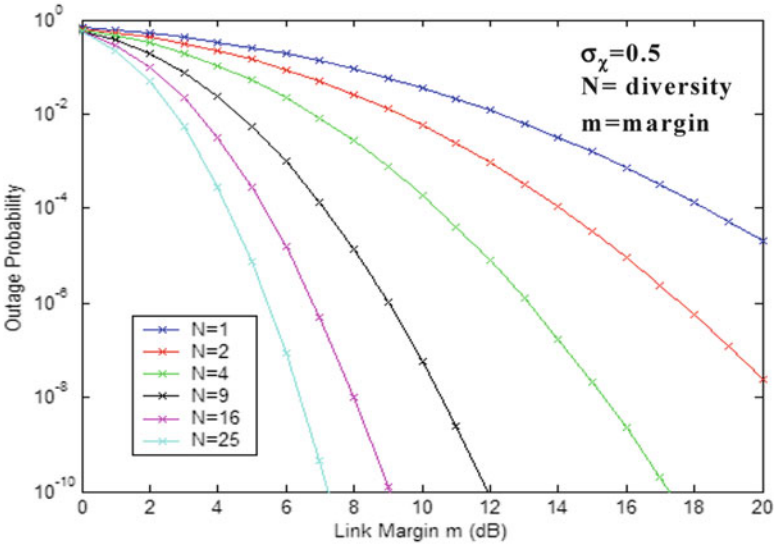


Fig. 16.6 Outage probability for different degrees of diversity; log-amplitude standard deviation, $\sigma_\chi = 0.5$ (strong turbulence), transverse wind speed of 10 km/h (moderate), and cut-off wind speed of 100 km/h (nominal) [7]

horizontal range (~500 m) for spatial diversity direct detection receivers for fairly strong turbulence. Results for longer ranges have similar behaviors. It is evident that a modest diversity yields substantial gains.

For large diversity degree N and link margin m, the outage probability is given by

$$P_{outage} \sim c_3 \, \mathrm{e}^{-c_2 N (\ln m)^2}$$

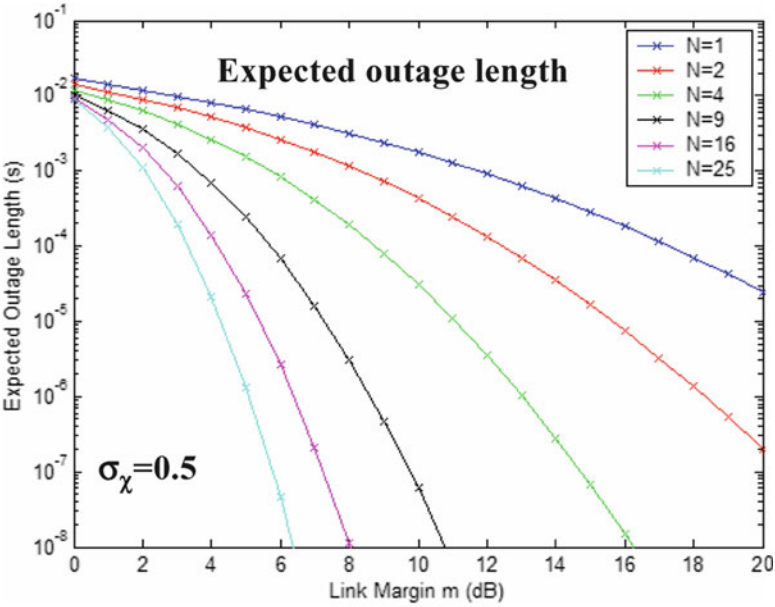


Fig. 16.7 Expected outage length for different degrees of diversity; $\sigma_\chi = 0.5$ (strong turbulence), transverse wind speed of 10 km/h (moderate), and cut-off wind speed of 100 km/h (nominal) [7]

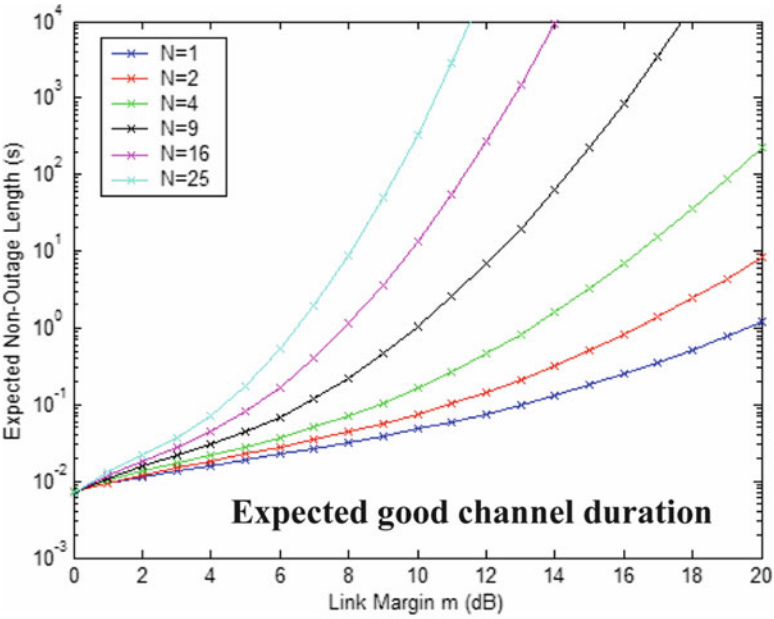


Fig. 16.8 Expected good channel durations for different degrees of diversity; $\sigma_\chi = 0.5$ (strong turbulence), transverse wind speed of 10 km/h (moderate), and cut-off wind speed of 100 km/h (nominal) [7]

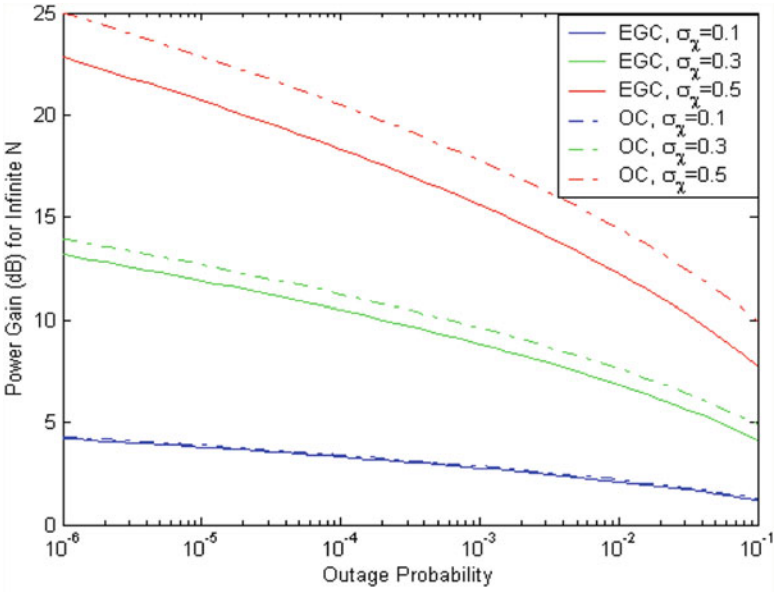


Fig. 16.9 Power gain of direct detection diversity optical communication systems for moderate turbulence and optimum receivers and equal gain combining receivers—infinite number of diversity receivers assumption [7]. *EGC* equal gain combining, *OC* optimum combining

The expected outage length is given by

$$E[\textit{outage length}] \sim c_1 \, \mathrm{e}^{-c_2 N (\ln m)^2}$$

The c's are constants and the exact analytical expressions are available in [7, 8, 10]. The important observation is the trend that as the degree of diversity (spatial, temporal, or frequency) increases the performance improves rapidly.

We define the power gain of a diversity system to be the fractional decrease in required transmitted power in a diversity system compared to a non-diversity system to achieve the same specified outage probability. This definition of power gain provides us with a useful means to evaluate the performance gain of diversity systems. Although for a full performance comparison, the outage lengths and good channel durations statistics must be considered together with the transport layer protocol used. Figure 16.9 shows the magnitude of the power gain of a diversity direct detection system for moderate turbulence. Substantial gains (~10 dB) can be realized using moderate to high degree of diversity. Some form of diversity of this type is critical for the space to ground optical atmospheric links especially for the last mile in near terrestrial applications where atmospheric turbulence is strongest. Otherwise the transmitter power required will be prohibitively large ($\gg$100 W) and expensive.

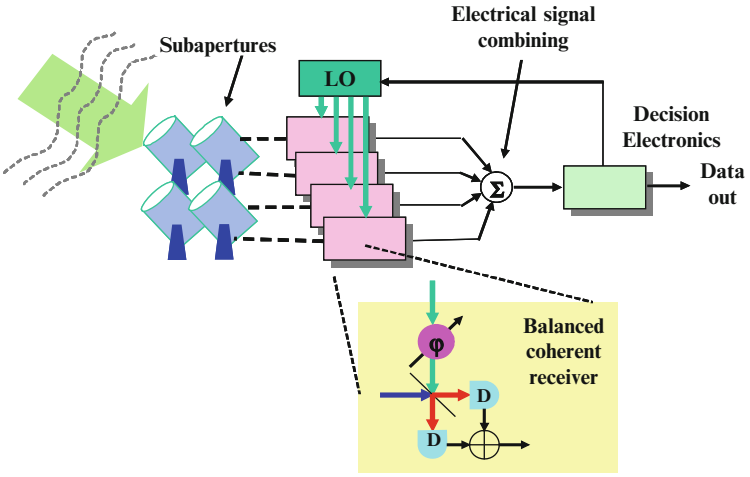


Fig. 16.10 Multi-aperture diversity coherent receiver with subapertures more than one coherence length apart [13]

Direct detection systems with photon counting receivers are preferred in vacuum channels with little or no background noise. Background noise in the receiver field of view couples into the receiving detection system. Direct detection systems are phase insensitive systems and each subaperture will each admit one mode of background noise resulting in multimode noise in the receiver. Whereas in a coherent system, Fig. 16.10, the received signals of the subapertures are coherently combined, and thus only one spatial mode of background noise couples into the detection process. Coherent systems thus have two useful properties that can provide significant system performance enhancements in some application scenarios: (1) spatial–temporal selectivity and (2) "clean" heterodyne gain with little device impairments [12–16]. The transform-limited spatial selectivity property is useful for applications where background noise, multiple access, and intentional interference are factors limiting system performance rather than fundamental quantum effects usually only encountered in a vacuum channel.

For a multi-aperture coherent transmitter and receiver system, the atmospheric channel model needs further refinement [11, 12]. The input/output relationship between the n_{tx}, transmit apertures, and n_{rx}, receive apertures, is given by a random Green's function H as shown in Fig. 16.11, where r, ρ, z are the output/input plane spatial coordinates and the propagation distance, respectively, and the $U's$ and Σ are the source/output fields and source aperture, respectively.

For small discrete apertures such as those assumed here, the input/output relationship can be summarized by a random matrix **H** and input and output vectors x and output y, Fig. 16.12. **H** has the singular value decomposition as shown with singular values Γ's. The optimum transmission scheme uses the input eigenvector

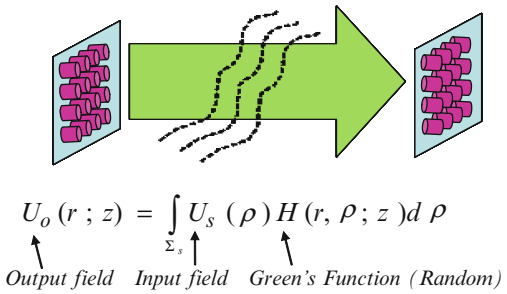


Fig. 16.11 Input/output random Green's function describing the relationship of multi-aperture transmitter and receiver communication system [13]

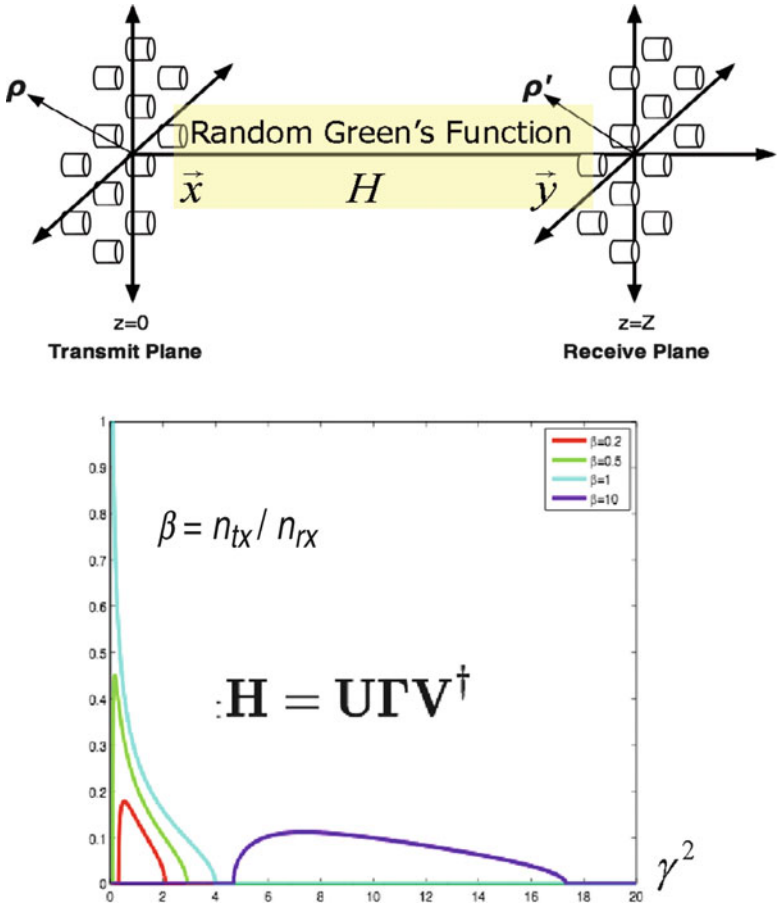


Fig. 16.12 Discrete channel model for the optical turbulent atmospheric channel and the eigenvalue distribution of the singular value decomposition of the random Green's function: Marcencko–Pastur density for large number of subapertures N [11, 12]

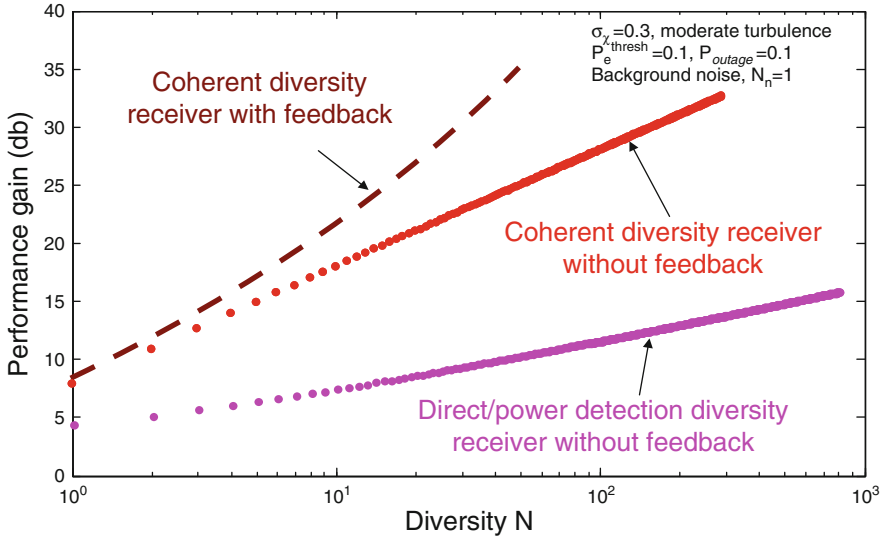


Fig. 16.13 Performance gains of coherent diversity systems with feedback over diversity systems without feedback [13]

corresponding to the largest eigenvalue for maximum power transfer and the receiver uses the corresponding output vector as a spatial matched filter [11, 12].

$$\begin{aligned} &\vec{y} = H\vec{x} + \vec{w} \\ &H = U\Gamma V^{+} \\ &\vec{x}^{*} = a\vec{v}_{\max} \\ &\phi = Re\{\vec{u}^{+}_{\max}\vec{y}\} \end{aligned}, \qquad \Gamma = \begin{bmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \gamma_m \end{bmatrix}$$

The gain over no feedback is

$$\sim \left(1 + \sqrt{\beta}\right)^2, \quad \beta = \frac{n_{tx}}{n_{rx}}$$

The outage probability and expected outage length for coherence time $t_{0,}$ large degree of diversity N and link margin m. is given by

$$P_{outage} \sim c_3 \exp\{-c_2 N(\ln m)^2\}$$

$$E[outage\ length] \sim c_1 \frac{t_0}{\sqrt{n_{tx} n_{rx}}} = c_1 \frac{t_0}{N} \quad \text{for } \beta = \frac{n_{tx}}{n_{rx}} = 1$$

Figure 16.13 compares the performance of coherent and incoherent systems with and without feedback. Direct detection system with no feedback suffers the classical

incoherent combining losses of such a system, ~ order of $\sqrt{N}$, where N is the degree of diversity. The performance gain of a coherent system is a factor of N due to coherent addition, which is substantial. By tracking the largest eigenmode of the random transfer function between input and output, not only can the bit error and dropout rate be reduced but the duration and inter-arrival times of these dropouts are also reduced [11, 12]. The gain of low rate feedback is another ~10 dB over that of diversity systems with no feedback.

The gain of coherent system is magnified when the network receivers are operating in the multiple access mode running a MAC protocol. The improvements realized for these parameters have first order significant positive implications on network performance. Figure 16.14 shows more clearly the performance comparison for a 100 aperture transmit/receive system. Even with a simple arrangement of selecting the best subaperture to transmit yields 3 dB of gain. With full wavefront predistortion, the gains are over 10 dB.

For coherent transmitters, the power and beam pointing of the multiple apertures can be allocated to form a dynamic physical connection topology, Fig. 16.15. This availability of dynamic physical topology has significant implications on reliability of the network and also the performance of Layer 3, the routing layer, and Layer 4, the transport layer. It is worthwhile to find the optimum dynamic topology using adaptive algorithms to optimize the performance of the optical wireless network.

4 Free-Space Dynamically Reconfigurable Optical Wireless Network—An Architecture Construct

When used in a free-space optical network as shown in Fig. 16.16, dropouts due to fading will significantly drives network architecture. During a fade event, many packets from the same user could be dropped triggering TCP window closing resulting in very bad throughput. If the fade is particularly long, TCP session time-out may occur terminating the session prematurely. Thus, the network architecture at multiple layers will have to be redesigned. We will assume, in general, the network has capacities >1 Gbps/link so there will be plenty of signal power for channel estimation. Moreover, we will consider coherent systems here with the note that incoherent systems can only use a subset of the techniques given in the following due to the lack of phase information. In a new network architecture construct, we will use a single highly reliable control plane to communicate across all network layers. This control network can be done via RF or even optical means with rich diversity to make fading practically non-existent. Optical links can also be used for this network, but due to dropouts in these links, each control packet transmission must used diversity paths to guarantee high packet delivery without loss. This can be done via erasure coding over diversity paths or network coding.

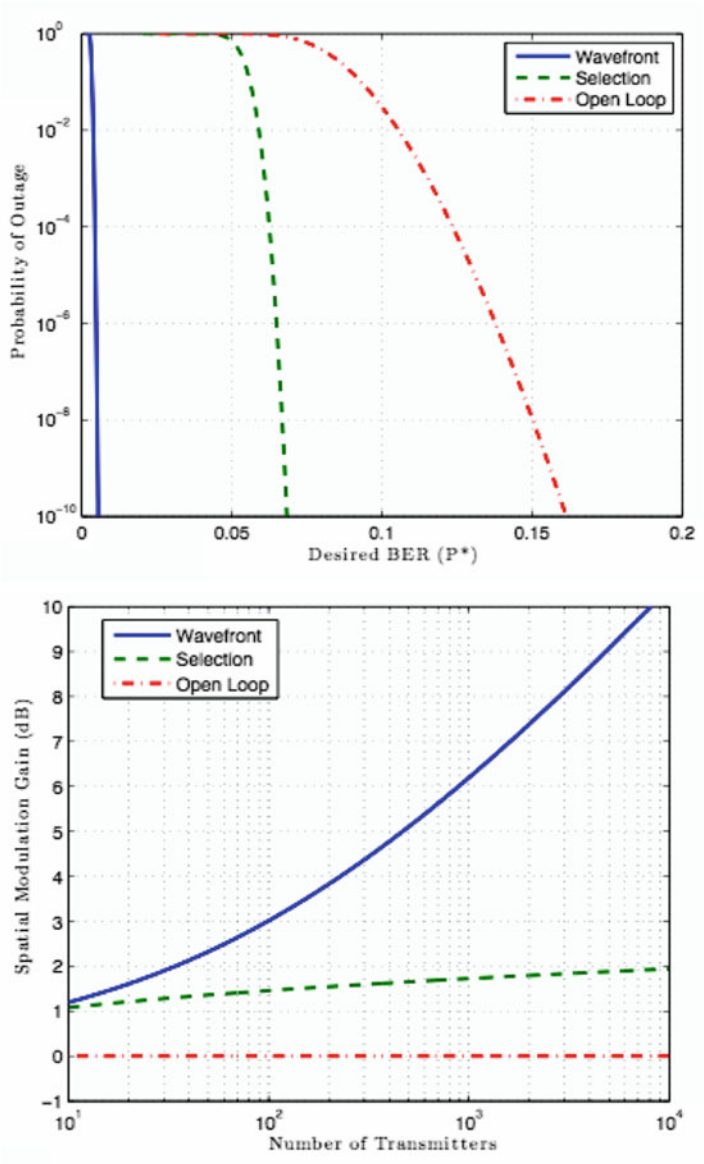


Fig. 16.14 Performance gains of coherent diversity systems with feedback over diversity systems without feedback: probability of outage for open loop with no feedback and feedback with full adaptation and selection of best subaperture for transmission [11, 12]

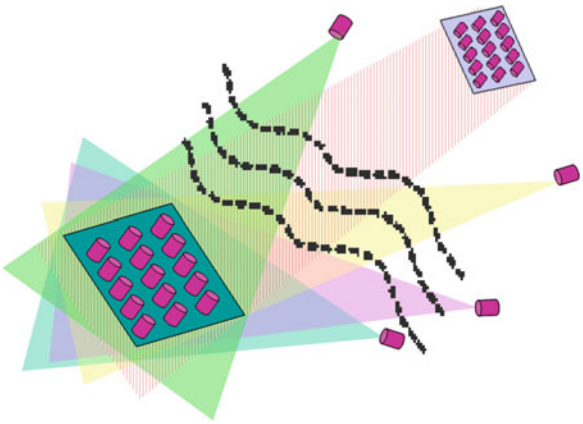


Fig. 16.15 Free-space dynamically reconfigurable optical network [13]

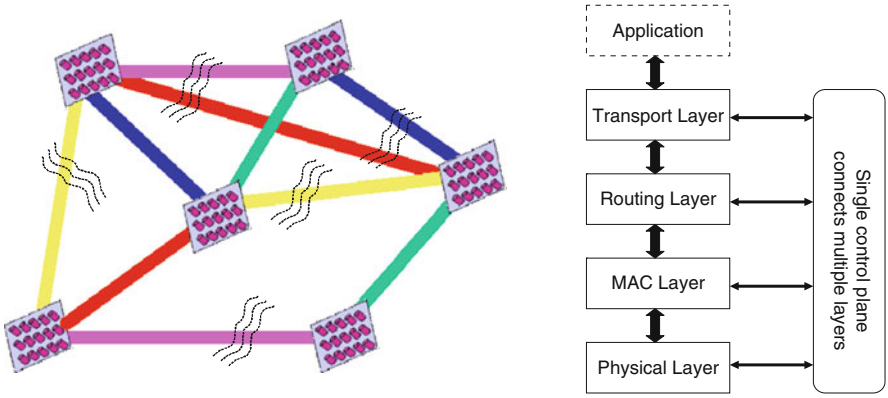


Fig. 16.16 Free-space dynamically reconfigurable optical network with single reliable control plane [13]

4.1 Physical Layer

It is possible to dynamically configure the array transmitters and array receivers at the nodes of a coherent optical wireless network to provide any desirable physical connection by appropriate phase and amplitude selection of each array element. Since there is no fiber and only free space, there is no connectivity constraint. In fact, multiple simultaneous transmit and receive beams can be used. With tuning speeds of <1 nanoseconds, the reconfiguration can be essentially as fast as it needs to be and definitely possible at packet rates, changing from packet to packet if need be. With reliable receiver feedback, the following actions can possibly benefit the network performance:

1. Dropout links can be reported to the routing (network) layer and temporarily deleted from the routing table and re-established after the fades have subsided.
2. More transmitter and receiver array elements, power, and duration (in case of time division multiplexing) can be devoted to links with high turbulence and/or longer distances.
3. Hold off sending data on badly faded links until link recovery (this necessitates a fix on transport layer time-out policies), and retransmit per intermediate nodes at the link layer, faded packets.
4. Provide optimum diversity routing over multiple paths with erasure coding [5,6].

We can use a combination of the above techniques to make the network much friendlier to the routing and transport layers.

4.2 Multiple Access, Media Access Control, MAC, Layer

If the receiver is in the near field of the transmitter, it is possible to modulate the received spatial mode of the signal and via feedback to the transmitter, select a spatial mode with as small an overlap (small inner product) with the interfering field as possible. In a time scale faster than 1 mS, the turbulence can be considered frozen, and the transmitter with the aid of receiver feedback can select amplitudes and phases of each element so that the received field ψ and the other user's field ξ have inner product $|\langle \xi, \psi \rangle|^2 \sim 0$, so that receiver array can be used in a multiple access scheme in the Media Access Control, MAC, Layer. As a means to guard against rogue transmitters or intentional interference, spatial mode hopping according to a crypto-keyed hopping sequence can substantially increase immunity to interference by an order of magnitude equal to the number of possible selectable spatial modes.

4.3 Transport Layer Impediments and the Necessity for a New Design

The dropouts in the links of optical wireless networks have a significant impact on the performance of the higher layer protocols. We will illustrate the effects on Layer 4, here and later we will address the problems as seen in Layer 3, the routing layer. The transport layer protocol (TCP) is widely used in commercial networks today. The main problem of TCP is that when there is a packet loss, the algorithm interprets the phenomenon as network congestion and begins closing its window to hunt for the "right" window size by slowly increasing or decreasing the window every roundtrip. For very high rate atmospheric optical links (those with large bandwidth-delay products), TCP's throughput is impaired since the number of packets in flight over one round trip time, RTT, is large and the time it takes to

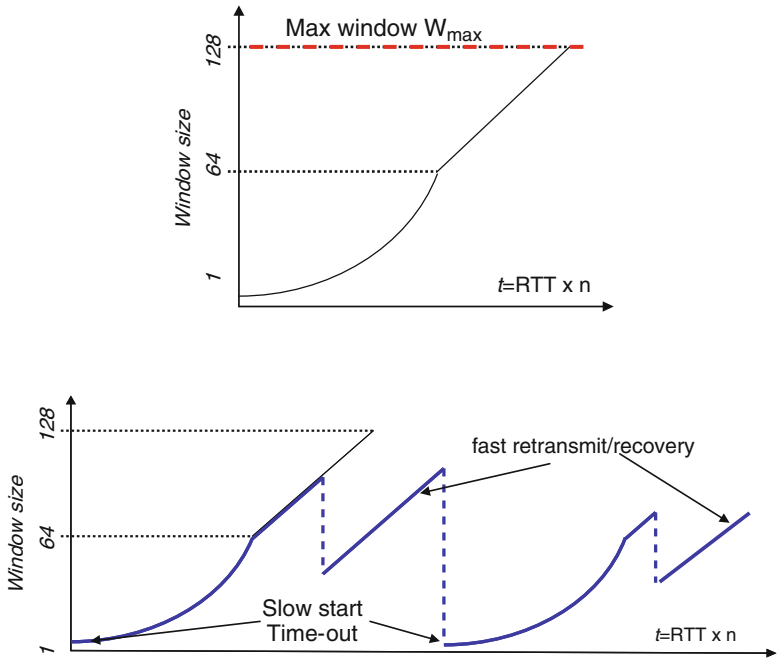


Fig. 16.17 TCP window size increase over time as a function of the number of roundtrip time RTT, and a typical TCP window size excursion over time with network congestions and/or packet loss due to fades, indicating slow start, fast transmit/recovery, and time-out

increase the window is long after an outage causes window closing, Fig. 16.17. The loss in throughput due to slow start, the slow rate at which the window is ramped up after an outage, heavily impacts the performance. In many nominal cases the throughput is less than 10 %, and sometimes even less than 1 %! TCP's throughput can be made higher if a moderate amount of diversity and/or high link margin is used to decrease the frequency of occurrence and durations of dropouts. Thus, the window is reduced less frequently. Though the time it takes to increase the window to a point where the link is fully utilized is still rather long, a reasonable amount of diversity and link margin improves the throughput.

Over moderate to long range (1–1,000 km) and very high data rate (>10 Gbps) atmospheric links, TCP's throughput is not high for direct detection even for large diversity and link margins. The throughput gain of a coherent system over that of a direct detection system is about $\sqrt{N}$ better, where N is the total diversity available in the system. Thus, we will use coherent system models for rest of the network architecture. Coherent systems are needed also for our dynamic routing adaptation approach. To quantify the potential of a diversity system (especially with coherent detection technology) on TCP performance, we can use as an approximation a Markov model for the states of the TCP window size, Fig. 16.18. The state n is the packet window size. The transition probability is given by the turbulent atmospheric channel model [11, 12].

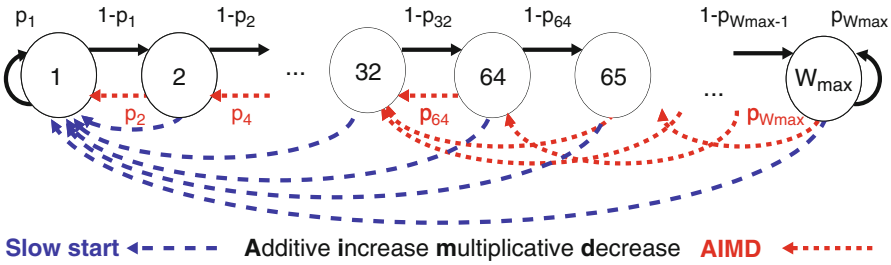


Fig. 16.18 Discrete Markov model for TCP states during transient and steady state [13]

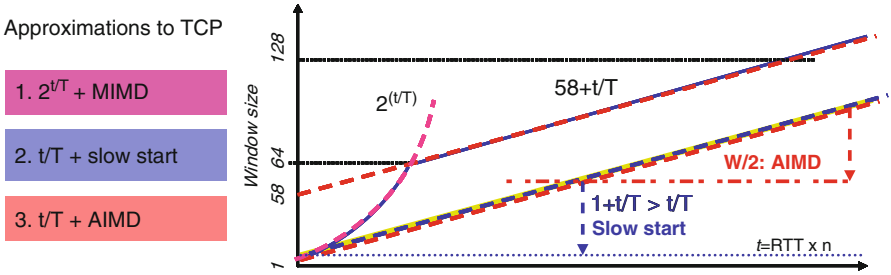


Fig. 16.19 Upper and lower bounds and approximations to TCP [13]

The transition probability of the discrete state continuous time Markov model in Fig. 16.18 is given by the outage probability, the mean outage duration and the inter-arrival time of outages as given in Figs. 16.6, 16.7, 16.8, and 16.9.

The essential feature of the TCP algorithm and most of its variants runs as follows:

- For every RTT:
  - No loss $\rightarrow W = W + 1$
  - Loss $\rightarrow W = W/2$ if 3 duplicate feedback of same RN – fast retransmit-recover, AIMD
  - Time-out when no ack or nak for RTT + $4\sigma \rightarrow$ goto $W = 1$, slow start

The performance of TCP in transient and in steady state can be well approximated by a set of upper and lower bounds which readily yield analytical performance expressions, Fig. 16.19. Depending on the application scenarios, one of these bounds will be the best approximation:

1. Upper bound 1, MIMD (multiplicative increase and multiplicative decrease) is very aggressive and is a good approximation for short transactions where TCP seldom or never gets out of its exponential increase region and also in the case of frequent consecutive packet losses triggering slow start and preventing TCP from getting out of the exponential increase region.

2. Lower bound 2 which uses linear increase but decreases to one packet in flight for all TCP states, is good for scenarios where the window has a good chance of being in the linear increase region (sometimes with modified larger maximum window sizes) particularly for large delay-bandwidth product links and heavy (consecutive) packet losses.
3. Approximation 3 with linear increase and multiplicative decrease, AIMD, is good for large delay-bandwidth product links and light packet loss.

It is then possible to upper and lower bound the performance of TCP as follows for TCP both in transient and in steady state [9]:

$$\text{TCP throughput} \approx \frac{E[\text{window size}]}{\text{roundtrip time RTT}}$$

When a free-space optical wireless network session is long enough for TCP to have reach steady state for a long time with steady state probability distributions π's [9], the TCP performance is as follows:

$$\begin{array}{c} \text{linear window increase gives lower bound} \\ \downarrow \\ \dfrac{\sum_{n=1}^{n_{\max}} n\pi_n^{linear}}{\text{RTT}} \le \text{TCP throughput} \le \dfrac{\sum_{n=1}^{n_{\max}} 2^{n-1}\pi_n^{\exp}}{\text{RTT}} \\ \uparrow \\ \text{exponential window increase gives upper bound} \end{array}$$

where

$$p_1^{linear} = 1 - (1 - P_{outage}) \exp\left(- \frac{KP_{outage}}{R_{\max}(1 - P_{outage})E[\text{outage length}]} \right)$$

$$\begin{aligned} p_n^{linear} = {} & 1 - (1 - P_{outage}) \exp\left(- \frac{nKP_{outage}}{R_{\max}(1 - P_{outage})E[\text{outage length}]} \right) \\ & - P_{outage} \exp\left(- \frac{-\text{RTT} + \frac{K}{R_{\max}}[n(1 - P_{outage}) - 1]}{(1 - P_{outage})E[\text{outage length}]} \right) \quad \text{for } n \ge 2 \end{aligned}$$

and K is the packet size.

The probability of being in the state **n** is

$$\pi_1 = \left(1 + \left(\sum_{n=2}^{n_{\max}-1} \prod_{j=1}^{n-1} (1 - p_j) \right) + \frac{1 - p_{n_{\max}-1}}{p_{n_{\max}}} \prod_{j=1}^{n_{\max}-2} (1 - p_j) \right)^{-1}$$

$$\pi_n = \prod_{j=1}^{n-1} (1 - p_j)\pi_1 \quad \text{for } n = 2,3,\ldots,n_{\max} - 1$$

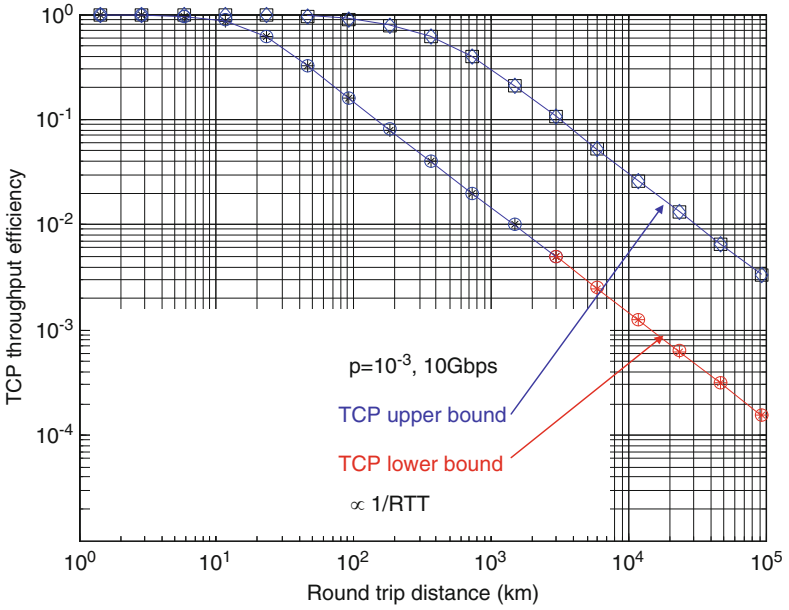


Fig. 16.20 Steady state transport layer throughput efficiency of 10 Gbps optical wireless systems for various distances and packet loss probability p [9]

$$\pi_{n_{\max}} = \frac{1 - p_{n_{\max}-1}}{p_{n_{\max}}} \prod_{j=1}^{n_{\max}-2} (1 - p_j)\pi_1$$

In most cases, the lower bound is actually a close approximation of the actual throughput. Figure 16.20 shows the end-to-end throughput of an ultra high-speed 100 Gbps link due to transport layer protocol mismatch to the channel, exhibiting the two detrimental behavior "window closing" and "slow start," substantially decreasing throughput [9]. For moderate to long distances, the throughput ranges from poor to practically nil. This problem of TCP is a major impediment of the high-speed free-space optical wireless network. There are several proposed possible "fixes" such as the use of PEP, Performance Enhancement Proxies, but none is totally satisfactory since they create other mismatches to problems in the network that the unmodified protocol were designed for. Coherent detection will contribute a long way towards the alleviation of this impairment. However, coherent detection alone will not totally eliminate dropouts, and thus this effect persists albeit to a much lesser extent. Thus, an additional technique in Layer 3, the routing layer, is needed as shown later.

For very large delay-bandwidth product links, TCP may converge very slowly or never reach steady state. If $\mathbf{P}$ is the transition matrix of the TCP Markov Model, the steady state distribution is the solution of the equation $\pi = \pi\mathbf{P}$, i.e., π is the left

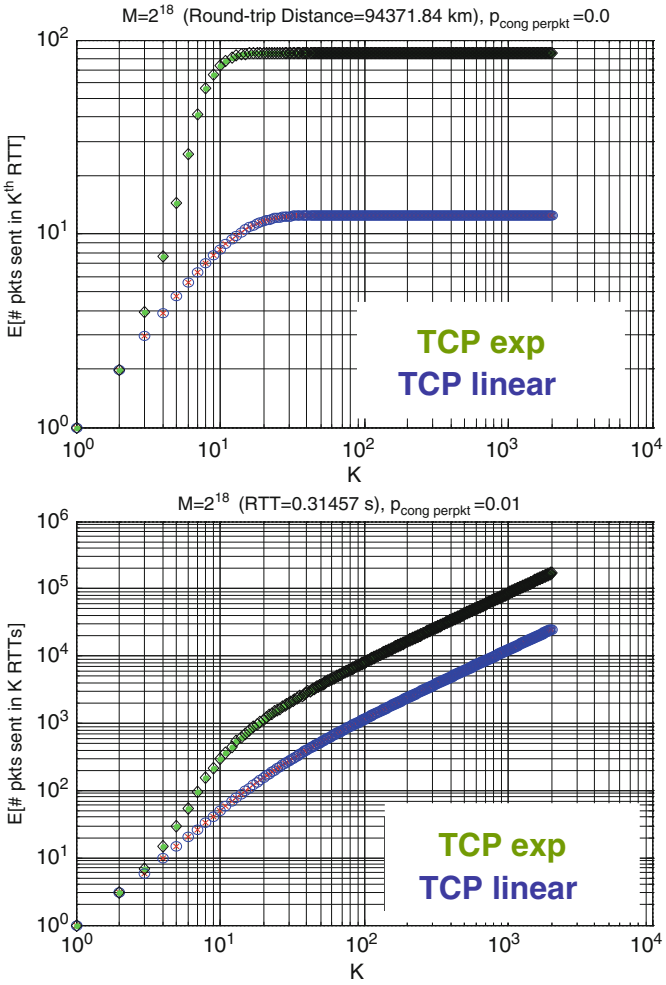


Fig. 16.21 TCP performance during transient and steady state in packets/RTT and total packets sent after K RTT, with unlimited upper limit to window size [9]

eigenvector of $\mathbf{P}$ with eigenvalue $= 1$. The transient state probability distribution $p(n)$ after n RTT is given by, $p(n) = p(n-1)\mathbf{P} = p(0)\mathbf{P}^n$. Figure 16.21 shows a typical example of the performance of TCP without an upper limit to the window size. Note that it takes 10's of RTT to reach steady states which can be several seconds for some aircraft/ground to space links, and for moderate file sizes, the protocol never reaches steady states. To estimate the sizes of large transactions that TCP cannot reach steady state before the session is over, it is possible to estimate the convergence rate of the algorithm to the steady state [14]. Using the L_1 norm for the distance between the transient state probability distribution and the steady state

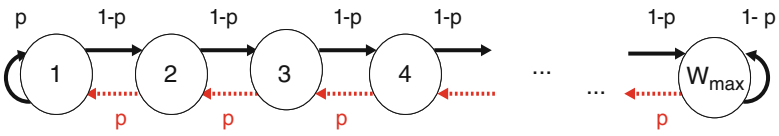


Fig. 16.22 Modified TCP for dealing with atmospheric fades: linear increase and decrease

distribution, the convergence rate is upper bounded using the second largest eigenvalue of **P** as

$$||p(t) - \pi||_1 = \sum_{k=1}^{J} |p_k(t) - \pi_k|$$

$$||p(t) - \pi||_1 \leq ||p(0) - \pi||_1 r^t, \text{ Markov chain with J states}$$

$$\text{where } r = 1 - \sum_{j=1}^{J} \min_i(p_{ij}) \geq \lambda_2, \text{ second largest eigenvalue of } \mathbf{P}$$

The transition matrices of the Markov model of TCP are not symmetrical and to get the eigenvalues of **P**, singular value decomposition of **P** can be used: $\mathbf{P} = U\Sigma V^{\mathrm{T}}$ = (orthogonal matrix)(diagonal matrix)(orthogonal matrix), where Σ is a diagonal matrix with eigenvalues from $\mathbf{P}^{\mathrm{T}}\mathbf{P}$, and U and V are orthogonal matrices; the columns of U are eigenvectors of $\mathbf{P}\mathbf{P}^{\mathrm{T}}$ and columns of V are eigenvectors of $\mathbf{P}^{\mathrm{T}}\mathbf{P}$; and the singular values λ_i on the diagonal of Σ are the square roots of the eigenvalues, λ_i^2, of the matrices $\mathbf{P}\mathbf{P}^{\mathrm{T}} = U\Sigma\Sigma^{\mathrm{T}}U^{\mathrm{T}}$ and $\mathbf{P}^{\mathrm{T}}\mathbf{P} = V\Sigma^{\mathrm{T}}\Sigma V$.

As an example which is a good approximation for large files and fades that last longer than the window size (occurs frequently at high link rates $>$ Gbps), consider a modified TCP with linear increase and linear decrease as shown in Fig. 16.22. The corresponding transition matrix **P** is shown in Fig. 16.23. The window, W, of transmitted packets will either collide with the burst of atmospheric fade with probability p or survive with probability $1 - p$. For this simple case, it is easy to find the second largest eigenvalue.

The largest eigenvalue is 1 and corresponds to the steady state distribution π. A simple bound can be found for the second largest eigenvalue. The matrix $\mathbf{P}\mathbf{P}^{\mathrm{T}}$ is shown in Fig. 16.24, and the trace of that matrix is the sum of all the eigenvalues:

$$Tr\{\mathbf{P}\mathbf{P}^{\mathrm{T}}\} = \sum_{i=1}^{W} \lambda_i^2 = W[p^2 + (1-p)^2]$$

Therefore for large W,

$$\lambda_2^2 \geq \frac{W[p^2 + (1-p)^2] - 1}{W - 1} \sim p^2 + (1-p)^2 \approx (1-p)^2$$

$$\mathbf{P} = \begin{bmatrix} p & 1-p & 0 & 0 & 0 & 0 & 0 & 0 \\ p & 0 & 1-p & 0 & 0 & 0 & 0 & 0 \\ 0 & p & 0 & 1-p & 0 & 0 & 0 & 0 \\ 0 & 0 & p & 0 & 1-p & 0 & 0 & 0 \\ 0 & 0 & 0 & p & 0 & 1-p & 0 & 0 \\ 0 & 0 & 0 & 0 & p & 0 & 1-p & 0 \\ 0 & 0 & 0 & 0 & 0 & p & 0 & 1-p \\ 0 & 0 & 0 & 0 & 0 & 0 & p & 1-p \end{bmatrix}$$

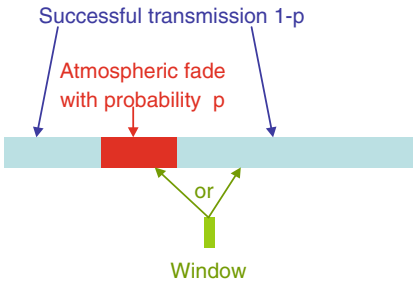


Fig. 16.23 Transition matrix for the modified TCP algorithm considered and the dropping of entire window of packets due to atmospheric fades

$$\mathbf{PP^T} = \begin{bmatrix} p^2+(1-p)^2 & p^2 & p(1-p) & 0 & 0 & 0 & 0 & 0 \\ p^2 & p^2+(1-p)^2 & p^2 & p(1-p) & 0 & 0 & 0 & 0 \\ p(1-p) & p^2 & p^2+(1-p)^2 & p^2 & p(1-p) & 0 & 0 & 0 \\ 0 & p(1-p) & p^2 & p^2+(1-p)^2 & p^2 & p(1-p) & 0 & 0 \\ 0 & 0 & p(1-p) & p^2 & p^2+(1-p)^2 & p^2 & p(1-p) & 0 \\ 0 & 0 & 0 & p(1-p) & p^2 & p^2+(1-p)^2 & p^2 & p(1-p) \\ 0 & 0 & 0 & 0 & p(1-p) & p^2 & p^2+(1-p)^2 & p^2 \\ 0 & 0 & 0 & 0 & 0 & p(1-p) & p^2 & p^2+(1-p)^2 \end{bmatrix}$$

Fig. 16.24 $\mathbf{PP}^{\mathrm{T}}$ for the modified TCP algorithm

The convergence rate is,

$$||p(t) - \pi||_1 \leq ||p(0) - \pi||_1 r^t \sim ||p(t) - \pi||_1 \lambda_2^t; \quad \frac{||p(t) - \pi||_1}{||p(0) - \pi||_1} \sim (1-p)^t$$

This a good approximation of TCP indicating an initial rate of increasing the window at a slope of $1 - p$ and then slowing down before reaching steady state at the rate $(1-p)^t$.

4.4 New Coupled Transport and Routing Layers

Irrespective of whether TCP has a chance to reach equilibrium or not, the algorithm misinterprets packet loss due to fades as congestion and reacts inappropriately by slowing down substantially reducing network efficiency. The necessity to redesign is clear, but it will involve both Layer 3, the routing layer, and Layer 2, the link layer. We will start from an understanding of what the network and the network user want to experience and create new designs for all four layers to satisfy those needs, as well as mindful of being interoperable with conventional wired networks. The following are necessary properties for a good transport layer protocol for the free-space optical wireless network.

From the network point of view:

1. High network utilization and efficiency
2. Fairly allocated rate across users (according to its demand)
3. Avoid congestion collapse

From the user point of view:

1. High user throughput, achieve maximum available rate quickly
2. Correct reception of data
3. Small delay (within delay requirements)

The issues the architect has to contend with are:

1. Atmospheric turbulence (fades/outages)
2. Long RTT + high data rates
3. Dynamically changing users, traffic, and network state
4. Adverse network responses:

   (a) Window halving due to packet lost in outage
   (b) Time-out due to delay variance bound exceeded
   (c) Linear window increase (too slow for long distances, high data rate sessions)
   (d) Trying to give equal rates to all users (some users only want small rate)

Figure 16.25 lists a number of possible new added actions that can be taken by the users, receivers and routers that can help alleviate the low efficiency problem. A major revision of the observables and the parameters to feedback is required for the network to behave properly.

A set of good pathways for the redesign of the Transport Layer are:

1. Users can distinguish congestions from outages and react accordingly—outage observations made by routers and feedback to users
2. Attain available rate quickly with faster rate increase, taking into account senders' demand—done by routers rather than by senders but in a scalable way without per-flow tracking
3. Stop increasing rate when near congestion, with routers providing more feedback information to end users than mere packet drop.

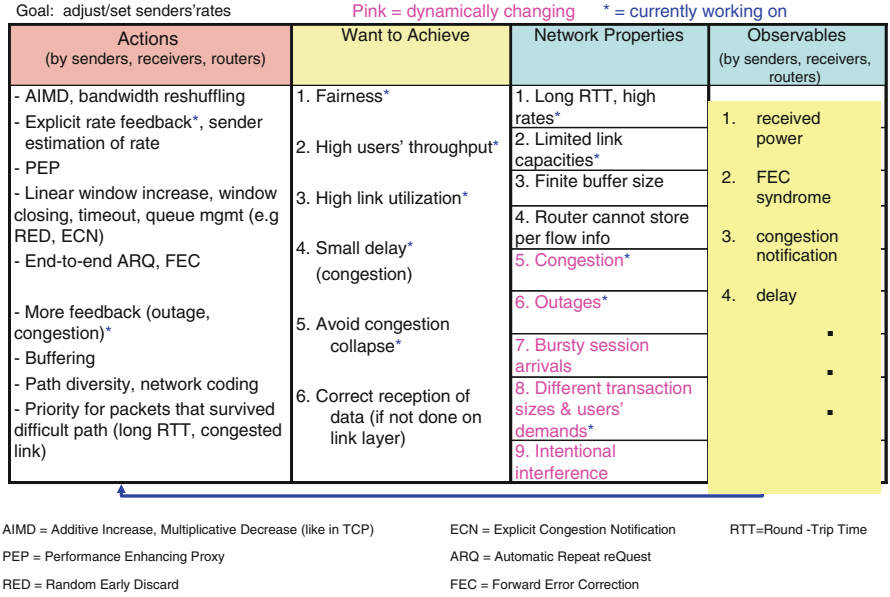


Fig. 16.25 Different possible actions (cures) with observable feedback to combat free-space optical wireless network properties that may help achieve high transport layer efficiency [9]

4. New types of responses to outages (per link or end-to-end, e.g., ARQ/error correction/monitor outages and have router buffer incoming packets and hold off transmission during outages)
5. Diversity routing over multiple independent paths with optimum combining at the transport layer and the physical layer [13].

Rather than discussing the merits of numerous modifications of TCP, we will provide a single design that can perform well as an example here.[1] The design uses both single and multiple path routing, congestion and outage sensing and feedback, and physical topology design that yield maximum diversity for a given constraint on transmitter and receiver array resources. It is possible to dynamically configure the array transmitters and array receivers at the nodes of a coherent optical wireless network to provide any desirable physical connection, Figs. 16.5 and 16.16. Since there is no fiber and only free space, there is no connectivity constraint. With tuning speeds of <1 ns, the reconfiguration can be essentially as fast as it needs to be and at packet rates if needed to.

[1] For illustrative purposes we will consider the simpler example of a homogeneous optical wireless network with a diameter of 1 km. Other scenarios will need scaling of the numerical parameters given here.

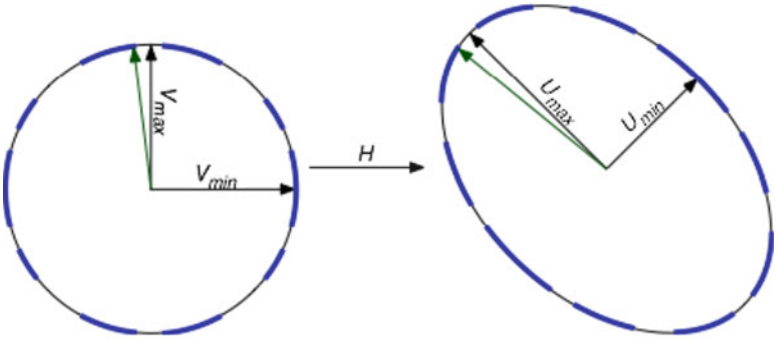


Fig. 16.26 Input and output eigenvalues of the random Green's function are the axes of a sphere (input) and ellipsoid (output). Maximum power transfer is provided by the input eigenvector corresponding to $V_{\max}$ [11]

When multipath routing is used, each path will fade independently. With many paths, the probability of all paths being faded at the same time is extremely small though not zero. Thus, the architect may want to maximize the number of independent paths consistent with the constraints on the number of transmitter and receiver array elements and available power. This is similar to the use of phased array antenna systems in RF systems (the physics is somewhat different), except there have been little or no application of phased array antennas as coupled to the network layer in microwave systems in the past.

Given that the network is in free space and not confined to installed fiber, a much richer set of connection topologies are accessible. Moreover, they can be changed rapidly as the situation calls for. While mesh topologies have been studied for fiber networks, for this application, the objective will be different: which is "maximize the number of independent paths within the constraint of the transmitter and receiver resources". One can find the answer to this design problem using graph theory techniques [3]. For example, there are different types of graphs with the same degree and number of edges, but some have more independent paths than others. In fact, the problem is analogous to designing a graph with low probability of disconnection given independent link failures (drop-outs). Typical good candidates are Moore graphs and their generalized cousins such as deBruijn graphs, Circulants and Harary graphs [3].

With receiver feedback, the following actions can benefit the network performance:

1. Track the state of the physical link and use the input eigenstate that yields the maximum power transfer to the receiver, Fig. 16.26.
2. Dropped out links can be deleted from the routing table temporarily and reestablished after the fades have subsided.
3. More transmitter and receiver array elements, power, and duration (in case of time division multiplexing) can be devoted to links with high turbulence.

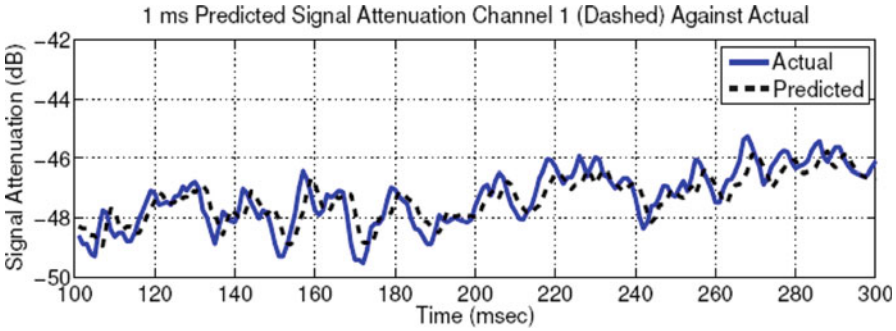


Fig. 16.27 Prediction of signal power level using an autoregressive model [12]

4. Hold off sending data on badly faded links until link recovery (necessitates a fix on transport layer time-out policies).
5. Provide optimum diversity routing over multiple paths with erasure coding [5, 6], or network coding.

In the physical layer the receiver tracks all the eigenstates via signature signals sent from each transmitter element. The channel state is represented by an N-dimensional ellipsoid, Fig. 16.26. The longest axis of the ellipsoid corresponds to the largest eigenvalue and the input spatial signal that provides the maximum power transfer between transmitter and receiver [12, 13].

The receiver can predict power level received 1 s ahead to <1 dB, except for dropouts due to fading (~50 μs),[2] Fig. 16.27. It can also track the derivatives with respect to time of the eigenvalues γ' and γ''. The receiver reports the link state via a low rate feedback (RF) channel to the transmitter (~1 μs) and transmit the largest eigenvalue state plus the next largest eigenvalue state with $\gamma' > 0$. The transmitter can send data over this second link and also send on an alternative path with no common link, as shown in Fig. 16.28.

$$\vec{x}^* = \begin{bmatrix} \vec{x}_1^* \\ \vec{x}_1' \\ \vec{x}_2^* \end{bmatrix} = \begin{bmatrix} a_1^* \ \vec{v}_1^* \\ a'_1 \ \vec{x}_1' \\ a_2^* \ \vec{v}_2^* \end{bmatrix}$$

The transmitter adjusts the coefficients a_i based on channel state predictions. When a link is reported to be faded the transmitting node will retransmit (~10 μs for fade durations with coherent diversity systems) buffered data on switch-over to a new link/s. This diversity/retransmission scheme results in almost no dropouts since the retransmission of buffer cover losses, Fig. 16.28. However, this scheme will

[2] Fading is a result of cancellation of two or more fields and the time scale of the action is approximately ten times faster than that of the nominal power fluctuations.

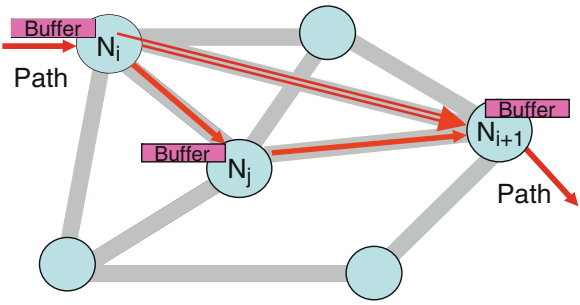


Fig. 16.28 Routing based on prediction, diversity combining, dynamic route switching and retransmissions of buffered data upon fade notifications [13]

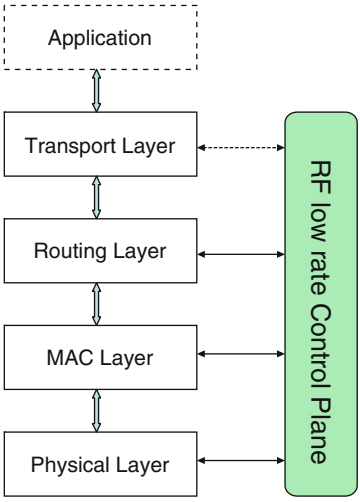


Fig. 16.29 Congestion control and end-to-end reliable data delivery as a multi-layer function using a single control plane

often result in the destination node receiving duplicate packets. A new transport layer protocol will purge duplications.

A reliable low rate control plane, Fig. 16.29 (low-frequency RF or wired links), with 1 μS timing, is used for:

1. Feedback of predistortion information from receiver in L1/2:
2. $\Delta\theta$'s (incremental rotation of ellipsoid)—determines array phase adjustments
3. $\Delta\gamma$'s (change in eigenvalue)—determines transmitter rate and notification of dropouts from receiver to transmitter ($<$10 μs)
4. Signal L2/3 to transition to new eigenmodes/paths and terminate old ones
5. Signal L2/3 to retransmit buffer (~10 μs) upon dropouts and increase channel rate for catch-up

With feedback, the coherent diversity transceivers throughput will be much higher, though any dropouts will still cause slow start with TCP. A new transport

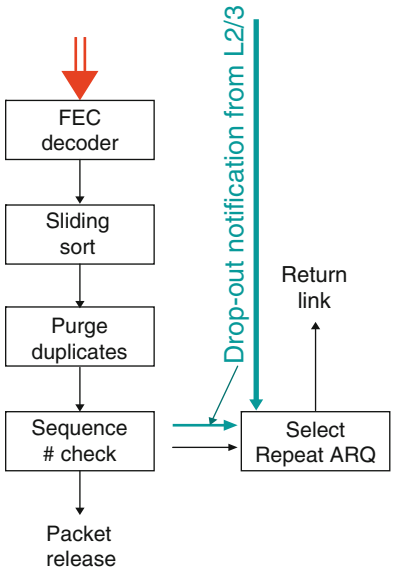


Fig. 16.30 New transport layer protocol functional blocks [13]

layer protocol will be much more efficient, Fig. 16.30. There are three major functions of Layer 4:

1. Transmitter/receiver rate matching
2. End-to-end reliability
3. Congestion control—admission control + speed-up

The first two functions can be taken care of by ARQ and forward error correction. Congestion control is provided via link state sensing in the physical layer, access rate control, and speedup upon congestion at each node and with fast rate allocation mechanisms. Sometimes when buffer retransmissions occur, duplicate packets are received and can also be out of sequence. The new transport layer protocol will use a sliding sort, purge, and check sequence number algorithm to recover missing data via Select Repeat ARQ. This transport layer protocol also performs the following functions to avoid window closing when packet loss is due to dropouts:

1. Receiver (L2/3) observe dropouts (rare); tag ACK before and after dropout notifying sender with <50 ms delay.
2. Transmitter: if dropout is due to fading continue to transmit at previous window size.
3. Otherwise interpret dropped packets are due to congestion; trigger window closing.

This architecture construct makes use of a single reliable control plane to communicate between user terminals and nodes at all layers of the network. The

key is to pass minimal essential information across layers to trigger the proper actions for congestion control and fade mitigations. Conventional features of error recovery due to noise in the receivers are still being taken care of by ARQ in Layer 4. However, compare to the indiscriminate use of TCP/IP in the optical wireless network over the atmosphere, this construct will have much better efficiency and faster response to link variations.

5 Summary

The new optical wireless network architecture described makes use of cross-layer techniques to substantially improve network efficiency. The most important attributes are the use of the best spatial modulation with feedback for each link and the recovery of lost packets upon fading using diversity transmission over alternative eigenmodes and links. Since packets can be duplicated and out of order, a new transport layer protocol is needed to perform the tasks of end-to-end reliable data delivery and congestion control. The behavior of this network will appear similar to that of a fiber network with delays comparable to moderately congested long haul fiber networks.

Acknowledgments Research partially supported by DARPA and ONR and the manuscript is a much expanded version of [13].

References

1. Chan VWS (2006) Free-space optical communications. J Lightwave Technol 24 (12):4750–4762 (Invited paper)
2. Chan VWS (2003) Optical satellite networks. J Lightwave Technol 21(11):2811–2827 (Invited)
3. Chan VWS (2000) Optical space communications. IEEE J Sel Topics Quantum Electron 6 (6):959–975 (Millennium Issue; Invited)
4. Chan VWS (1987) Space coherent optical communication systems—an introduction. J Lightweight Technol LT-5(4):633–637 (Invited)
5. Lee EJ, Chan VWS (2004) Part 1: Optical communication over the clear turbulent atmospheric channel using diversity. IEEE J Sel Areas Commun Opt Commun Netw 22(9):1896–1906
6. Chan VWS (1982) Coding for the turbulent atmospheric optical channel. IEEE Trans Commun COM-30(1)
7. Lee E, Chan VWS (2007) Performance of diversity coherent and diversity incoherent receivers for optical communication over the clear turbulent atmosphere in the presence of an interferer. In: Proceedings of the SPIE conference on optics and photonics, August 2007, San Diego, CA
8. Lee E, Chan VWS (2007) Diversity coherent receivers for optical communication over the clear turbulent atmosphere. In: Proceedings of the IEEE international conference on communications (ICC 2007), Glasgow, pp 2485–2492

9. Lee E, Chan VWS (2005) Performance of the transport layer protocol for diversity communication over the clear turbulent atmospheric optical channel. In: Proceedings 2005 IEEE International Conference on Communications (ICC), pp 333–339, Seoul, Korea
10. Shin E, Chan VWS (2002) Optical communication over the turbulent atmospheric channel using spatial diversity. In: IEEE GLOBECOM 2002, vol 3, Taipei, pp 2055–2060
11. Puryear A, Chan VWS (2009) Optical communication through the turbulent atmosphere with transmitter and receiver diversity, wavefront predistortion, and coherent detection. In: IEEE Globecom, Honolulu
12. Puryear A, Chan VWS (2010) Optical communication over atmospheric turbulence with limited channel state information at the transmitter. In: IEEE Globecom, Miami
13. Chan VWS (2011) Free space optical networks. In: WOCC, Newark, April 2011, and ICTON, Stockholm, July 2011
14. Gallager R (2012) Discrete stochastic processes. Cambridge University Press, Cambridge (in press)
15. Wen Y, Chan VWS (2005) Ultra-reliable communications over vulnerable all-optical networks via lightpath diversity. IEEE J Sel Areas Commun Opt Commun Netw 23(8):1572–1587
16. Chan VWS, Chan AH (1997) Reliable message delivery via unreliable networks. In: IEEE international symposium on information theory (MS-12104), Ulm, 29 June–4 July 1997

Index

S. Subramaniam et al. (eds.), *Cross-Layer Design in Optical Networks*, Optical Networks 15, DOI 10.1007/978-1-4614-5671-1,