Optical Networks
Series Editor: Biswanath Mukherjee

Jane M. Simmons

Optical Network Design and Planning

Second Edition

Springer

Optical Networks

Series editor Biswanath Mukherjee
University of California
Davis Dept. Computer Science
Davis
California
USA

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

Jane M. Simmons

Optical Network Design and Planning

Second Edition

Springer

Jane M. Simmons
Monarch Network Architects
Holmdel
New Jersey
USA

ISSN 1935-3839 ISSN 1935-3847 (electronic)
ISBN 978-3-319-05226-7 ISBN 978-3-319-05227-4 (eBook)
DOI 10.1007/978-3-319-05227-4
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014933574

Printed on acid-free paper

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

To my beautiful mother, Marie

Foreword

The huge bandwidth demands predicted at the start of the millennium have finally been realized. This has been sparked by the steady growth of a variety of new broadband services such as high-speed Internet applications, residential video-on-demand services, and business virtual private networks with remote access to huge databases. In response, carriers are undergoing widespread upgrades to their metro and backbone networks to greatly enhance their capacity. Carriers are demanding WDM optical networking technologies that provide both low capital expenses and low operational expenses. This need has been satisfied by automatically reconfigurable optical networks that support optical bypass. Automatic reconfigurability enables the carriers, or their customers, to bring up new connections and take down existing ones to meet fluctuating bandwidth requirements in near real-time. It also enables rapid automatic restoration from network failures. The optical-bypass property of the network, coupled with long-reach WDM optics, greatly reduces the need for optical-electrical-optical conversion, thus resulting in huge savings in capital and operational expenses.

This book provides a timely and thorough coverage of the various aspects of the design and planning of optical networks in general, with special emphasis on optical-bypass-enabled networks. While the reality of such networks today is somewhat different from the earlier research visions of a purely all-optical network that is transparent to signal format and protocol, the goals of greatly improved economics, flexibility, and scalability have been realized. The optical-bypass networking paradigm has been adopted by many of the major carriers around the world, in both metro and backbone networks. Moreover, efficient optical networking algorithms have emerged as one of the critical components that have enabled this technology to work in practice.

This book provides broad coverage of the architecture, algorithms, and economics of optical networks. It differs from other books on this general subject in that it focuses on real-word networks and it provides good perspective on the practical aspects of the design and planning process. The book serves as a valuable guide to carriers, vendors, and customers to help them better understand the intricacies of the design, planning, deployment, and economics of optical networks. The book also

provides practitioners, researchers, and academicians with a wealth of knowledge and ideas on efficient and scalable optical networking algorithms that are suitable for a broad range of optical networking architectures and technologies.

Jane Simmons has been actively working in this area since the mid 1990s. In this time-frame, there was much activity covering all aspect of optical networking—technology, architecture, algorithms, control, and applications. A particularly influential research effort that started in the United States around this time, in which Jane participated, was the government-supported Multiwavelength Optical Networking (MONET) consortium among the telecommunications giants AT&T, Lucent, Verizon and SBC. Just a short time later, much of the vision generated by this research was turned into reality. In the 2000 time-frame, Corvis Corporation became the first company to commercialize the "all-optical" backbone-network vision when it introduced a product with 3,200-km optical reach and the associated optical switching equipment. Jane played a key role at Corvis, where she developed efficient and scalable networking algorithms to support and exploit this technology. This culminated in the first commercial deployment of the "all-optical" vision with Broadwing's backbone network, in 2001. Jane performed the network design for the Broadwing network, from link engineering to network architecture. Jane also performed network designs for a broad array of North American and European carriers. She successfully showed in these diverse and real environments that "all-optical", or more accurately, optical-bypass-enabled networks are architecturally viable in terms of achieving high network efficiency.

She has continued to work on optical network architecture and algorithms, as a founding partner of Monarch Network Architects, which provides network design expertise to carriers and system vendors. More recently, she has worked as the Subject Matter Expert on the DARPA-sponsored Core Optical Networks (CORONET) program, which investigated highly dynamic and highly resilient multi-terabit optical networks. With this vast experience, and being in the right place at the right time, Jane has developed a unique perspective in the field of optical networking, which she brings forward in this book. I thoroughly enjoyed reading it and I learned a lot from it. I am sure that the reader is in for a real treat!

Department of Electrical
and Computer Engineering,
and Institute for Energy Efficiency,
University of California, Santa Barbara

Dr. Adel A. M. Saleh
Research Professor

Preface to the Second Edition

Optical networking has greatly matured since the early 2000s. Optical bypass, where connections remain in the optical domain as they traverse network nodes, is now a well-accepted technology for reducing the amount of electronic equipment in the network. The industry emphasis has shifted from transformative innovations to technological and architectural advancements that address operational and network management challenges. Some of the more important areas of development include network configurability, energy consumption, and fiber capacity.

Carriers readily appreciate the advantages afforded by automating the network provisioning process. Remote configurability of the network equipment via software reduces the amount of required manual intervention, thereby allowing more rapid provisioning of services and quicker revenue recognition. To reap the full benefits of a configurable network, the network equipment must be as flexible as possible, within the bounds of cost effectiveness. The network equipment that enabled optical bypass was a major departure from legacy equipment, with significant capital-cost reduction being the main driver of the technology. However, the main motivators guiding network element development today are reduced operational cost and greater network flexibility, where limitations imposed by the equipment are removed.

Recent developments with regard to network-element flexibility are covered extensively in Chap. 2. For example, the so-called colorless, directionless, contentionless, and gridless properties of optical-bypass-enabling equipment are covered in detail. This includes coverage of several architectural options for achieving these properties, as well as a discussion regarding the relative importance of incorporating this flexibility in the network elements.

The next frontier in automated connection setup is dynamic optical networking, where connections can be established on the order of a second. Furthermore, the requests for shifts in bandwidth come from the networking layers that sit atop the optical layer (e.g., the Internet Protocol (IP) layer), as opposed to operations personnel. Most carriers have been slow to embrace dynamic optical networking, similar to the initial skepticism regarding optical-bypass technology.

However, as applications such as cloud computing proliferate (where enterprises migrate much of their locally situated back-office functionality to remotely located

data centers distributed throughout the network), the need will increase for a rapidly responsive network that can maintain a high quality of service as network conditions change. Furthermore, dynamic networking is one enabler of network virtualization, which provides the ability to customize the network topology and resources that are seen by a given customer. In the future, one can envision dynamic cognitive networks, where the network autonomously reacts to the current conditions, based on its knowledge of past performance. Once the machinery for increased dynamism is in place, it is likely that new revenue-generating opportunities will arise as well, which will accelerate the pace of adoption.

An entire new chapter in this second edition (Chap. 8) is devoted to discussing the merits and challenges of dynamic optical networking. Recent research results as well as standardization efforts are covered. This includes extensive discussion regarding which functions are best handled by a distributed protocol as opposed to a centralized one. The vexing topics of minimizing resource contention and implementing optical bypass in a distributed environment are discussed in detail. Dynamic optical networking across multiple domains is covered as well, where the domains may be under the control of different service providers or different administrative organizations within one service provider. The challenge is performing close to optimal routing and resource allocation without violating the security and administrative boundaries of the various entities that are involved.

Chapter 8 also covers Software-Defined Networking (SDN), which is a relatively new networking paradigm that is relevant to dynamic networking (among other aspects of future networks). One of the goals of SDN is to provide carriers and enterprises with greater control of their networks, which includes providing a centralized view of the network that extends down to the optical layer. This potentially enables dynamic multi-layer optimization, although the scalability of such an approach is still an open question.

The growing role of data centers as the repository for enterprise computing and storage resources also impacts more conventional aspects of network design, such as the algorithms used for routing traffic. For example, the concept of "manycasting" has grown in importance, where enterprises need to be connected to some subset of distributed data centers (i.e., gaining access to the desired resources is what is important not the particular data centers that provide them). Algorithms to provide this connectivity are covered in Chap. 3.

The topics mentioned thus far are proactive strategies to improve network economics and network control. However, there are at least two daunting developments that have caught the attention of the industry, driving much research in response. The first is the growing amount of energy consumed by information and communication technologies (ICT). While estimates vary, it is generally agreed that ICT energy usage represents at least 2 % of total worldwide usage. Despite its major role in network transmission, the optical layer is responsible for just a small portion of this energy consumption, due to the relative energy efficiency of optical technology. Thus, as compared with sectors of the ICT industry where reducing energy consumption is an imperative (leading to energy-saving solutions such as locating data centers near bodies of water to reduce the need for more conventional cooling

methods), the approach with optics has been to consider *expanding* its role to reduce the energy strains in other portions of the network (e.g., pushing more switching from the electronic layers into the optical layer; introducing WDM technology in data centers; using optics for off- and on-chip interconnects). However, optical technology is not a panacea. For example, it is not ideal for applications that require data buffers, time shifting, and read/write operations. Discovering how best to harness the genuine advantages of optical technology is an ongoing task.

The topic of power consumption is brought up throughout this second edition. First, optical bypass, while initially implemented to reduce capital expenditures, has proved to be equally effective at reducing power consumption in the transport layer. The larger problem now is reducing power consumption in the IP layer. In the near term, one solution may be to insert an additional layer between the IP and optical layers to offload some of the burden from the IP routers, as discussed in Chap. 6. Other strategies include routing traffic away from certain regions of a network, e.g., to avoid the higher energy costs of a particular region, or to allow a subset of the equipment to be powered down. Such proposals are also discussed in Chap. 6. One particular longer-term solution that has generated a lot of follow-on research involves grooming (i.e., "traffic packing") in the more energy-efficient optical layer as opposed to the electronic layer. This proposed scheme differs from other optical-grooming schemes in that the grooming is performed in the frequency domain as opposed to the time domain. A large portion of Chap. 9 (which is new to the second edition) is dedicated to discussing the potential benefits and the challenging realities of this scheme. This discussion is supplemented by a detailed network study in Chap. 10.

Another recent development is the realization that, at the current rate of traffic growth, the capacity limit of conventional fiber will be reached by the 2025 time-frame. For many years, fiber capacity appeared almost infinite in comparison with the level of traffic being carried. The number of wavelengths supported on a fiber and the bit-rate of each wavelength have greatly increased over the past two decades. Furthermore, these advancements have enabled a significant reduction in two key networking metrics: cost per bit/sec and power consumption per bit/sec. However, the pace of these advancements is likely to slow, as further improvements become more challenging to implement. While deploying multiple fibers can address the need for additional capacity, this approach does not provide economies of scale with respect to cost and power consumption. Better solutions are desired.

Architectures and technology aimed at addressing fiber capacity limits are covered in Chap. 9, primarily in the context of flexible optical networks. By engendering the network with more flexibility, the fiber capacity can be used more efficiently, thereby prolonging the time until the capacity limit is reached. Most of the solutions involve employing more flexible spectral grids. This includes the optical grooming scheme mentioned above for purposes of reducing power consumption, which is also being championed as a means of using capacity more efficiently (though the limitations of optical filtering technology may curb the capacity benefits that can actually be attained). In addition to these flexible schemes, which take relatively small steps towards alleviating capacity limits, Chap. 9 also discusses longer-term

solutions, such as new fibers that can cost effectively increase the capacity of a single fiber by at least an order of magnitude (e.g., multi-core fiber).

One of the outgrowths of the various trends in optical networking is that network design has become more complex. For example, higher wavelength bit-rates and the introduction of more advanced transmission formats have resulted in needing to account for more optical impairments to maintain a high level of optical bypass. Additionally, mixed line-rate (MLR) networks, where wavelengths with different bit-rates and transmission formats are routed on one fiber, pose special challenges depending on the combination of formats that are present. These impediments need to be captured in the algorithms used for network design, as described in Chaps. 4 and 5.

Furthermore, some of the architectural schemes that have been proposed to add more networking flexibility concomitantly add more algorithmic complexity. While technology advancements often make network design more challenging, effective algorithms may in some instances lessen the need for technology-based solutions. For example, some of the networking limitations that are imposed by the optical-layer equipment can be sufficiently minimized through an algorithmic approach, rather than requiring costly upgrades to the equipment. Despite the added complexity, efficient optical network design is still a manageable process, as investigated throughout this book.

These trends, and their ramifications for network design, have motivated the second edition of this text.

Major Changes from the First Edition

Each of the eight original chapters in the first edition has been updated to address the latest technology and algorithmic approaches. Where appropriate, the terminology has been updated as well. For example, in keeping with popular usage, the term "ROADM" is generally used for any reconfigurable network element that allows optical bypass, regardless of its precise properties. The first edition had distinguished between ROADMs and All-Optical Switches, depending on the element functionality. Additionally, the term "edge configurable", used in the first edition in reference to a particular ROADM property, has been replaced by the term "directionless", which is now widely used in the industry.

As noted above, Chaps. 8 and 9 are new. Chap. 8 covers dynamic optical networking and Chap. 9 examines the trend towards greater flexibility in the underlying network, where much of the research is driven by the desire to use fiber capacity more efficiently. The original Chap. 8 (on economic studies) in the first edition is now Chap. 10. Additionally, the algorithm code that had been in an appendix is now in Chap. 11.

The second edition also includes more case studies throughout the text to illustrate the concepts. Three reference networks, presented in Chap. 1, are typically used for these studies. Readable text files containing the topologies for these networks (i.e., the nodes and links) can be found at: *www.monarchna.com/topology.html*

A few of the sections that were included in the first edition of the textbook have been removed. For example, the discussion regarding early generations of ROADM technology has been removed. A few of the network studies have also been eliminated from Chap. 10. For example, one of these studies had analyzed the economics of providing two types of transponders, one that supported the full nominal optical reach for the system and one that supported a shorter optical reach. Given the recent development of programmable transponders that can support a range of reach values and bit-rates, this study was no longer relevant. An additional study, on flexible networks, has been added.

Exercises

A set of exercises has been added to the end of almost all chapters, to test the understanding of the fundamental concepts. The exercises range in difficulty from simple application of an algorithm to thought-provoking architectural questions. Many of the exercises extend the discussion presented in the chapter text; e.g., an alternative architecture may be considered, along with questions that probe its performance. The exercises also include suggestions for future research.

The exercises that require some amount of network design use small networks to enable manual solution. Alternatively, the algorithm code that is provided in the last chapter can be used in some of these exercises. The code should be portable to any standard C compiler.

Some basic queuing theory is required to solve some of the exercises; for example, knowledge of Poisson processes and the Erlang-B formula.

Acknowledgements

As with the first edition of this book, I am indebted to Dr. Adel Saleh for spending numerous hours reviewing the text. His insightful suggestions have improved the content and the readability of the book. He also served as a sounding board for many of the exercises.

I also received very helpful comments from Prof. Biswanath Mukherjee, the editor of the Springer Optical Networks Series. His comments were especially beneficial in putting various topics in their proper context.

It was once again a pleasure to work with the team from Springer. I thank Brett Kurzman, the Springer Editor, for his support in publishing a second edition. Rebecca Hytowitz, the Springer Editorial Assistant, promptly answered all of my questions and provided useful documentation that was helpful in preparing the manuscript and all of the necessary files.

Preface to the First Edition

I have been involved with the research and development of optical networks for the past 15 years. More specifically, I have worked on the architecture and algorithms of networks with optical bypass, where much of the electronic regeneration is removed from the network. These networks are referred to in this book as optical-bypass-enabled networks.

Optical bypass has progressed from a research topic to a commercial offering in a relatively short period of time. I was fortunate to be in the midst of the activity, as a member of Bell Labs/AT&T Labs Research and Corvis Corporation. There are a few key lessons learned along the way, which I hope have been successfully captured in this book.

First, algorithms are a key component of optical networks. It is not hard to produce studies where poor algorithms lead to inefficient network utilization. Conversely, armed with a good set of algorithms, one can generate efficient designs across a range of network topologies, network tiers, and traffic distributions. It is also important to stress that while replacing electronics with optics in the network poses unique challenges that require algorithms, which is often cited as a concern by the opponents of such networking technology, the design of electronic-based networks requires algorithms as well. Processes such as shared protection or subrate-traffic-grooming are complex enough that algorithms are needed regardless of the nature of the underlying technology.

Second, there should be a tight development relationship between the system engineers, hardware designers, and the network architects of any system vendor developing optical networking equipment. The mantra of many a hardware developer when dealing with the potentially messy consequences of a design decision is often "the algorithms will take care of it." While their confidence in the algorithms may be flattering, this is not always the wisest course of action. It is the responsibility of the network architects to push back when appropriate to ensure that the overall system complexity does not grow unwieldy. Based on experience, when challenged, much more elegant solutions were forthcoming. Of course, there are times when the physics of the problem, as opposed to expediency, dictates a solution; it is important to recognize the difference.

This leads to the last point in that the algorithms in a well-designed system do not need to be overly complex. Much effort has been put into algorithm development, which has been successful in producing efficient and scalable algorithms. Furthermore, it is not necessary that the algorithms take many hours or days to run. With well-honed heuristics, a design that is very close to optimal can often be produced in seconds to minutes.

The primary goal of this book is to cover the aspects of optical network design and planning that are relevant in a practical environment. The emphasis is on planning techniques that have proved to be successful in actual networks, as well as on potential trouble areas that must be considered. While the algorithms and architecture are the core of the content, the various enabling optical network elements and the economics of optical networking are covered as well. The book is intended for both practitioners and researchers in the field of optical networking.

The first two chapters should be read in order. Chapter 1 puts the book in perspective and reviews the terminology that is used throughout the book. Chapter 2 covers the various optical network elements; it is important to understand the functionality of the elements as it motivates much of the remainder of the book.

Chapters 3, 4, and 5 cover routing, regeneration, and wavelength assignment algorithms, respectively. Chapter 3 is equally applicable to O-E-O networks and optical-bypass-enabled networks; Chaps. 4 and 5 are relevant only to the latter. The first four sections of Chap. 4 (after the introduction) are more focused on physical-layer issues and can be skipped if desired.

Chapters 6 and 7 are standalone chapters on grooming and protection, respectively. Much of these chapters apply to both O-E-O networks and optical-bypass-enabled networks, with an emphasis on the latter. Finally, Chap. 8 (i.e., Chap. 10 in the second edition) presents numerous economic studies.

Acknowledgements

The nucleus of this book began as a Short Course taught at the Optical Fiber Communication (OFC) conference. I would like to thank the students for their suggestions and comments over the past 5 years that the course has been taught.

I am indebted to Dr. Adel Saleh, with respect to both my career and this book. As a leader of MONET, AT&T optical networking research, and Corvis, he is recognized as one of the foremost pioneers of optical networking. I appreciate the time he put into reading this book and his numerous helpful suggestions and encouragement.

I thank the editor of the Springer Optical Networks Series, Prof. Biswanath Mukherjee, for providing guidance and enabling a very smooth publication process. He provided many useful comments that improved the readability and utility of the book.

The team from Springer, Alex Greene and Katie Stanne, has been very professional and a pleasure to work with. Their promptness in responding to all my questions expedited the book.

Contents

Abbreviations

3WHS	3-Way Handshake
ADM	Add/Drop Multiplexer
ANSI	American National Standards Institute
AR	Advance Reservation
ASE	Amplified Spontaneous Emission
ASON	Automatically Switched Optical Network
ATM	Asynchronous Transfer Mode
AWG	Arrayed Waveguide Grating
BFS	Breadth First Search
BLSR	Bi-directional Line-Switched Ring
BN	Boundary Node
BPSK	Binary Phase-Shift Keying
BRPC	Backward-Recursive PCE-Based Computation
BVT	Bandwidth Variable Transponder (or Bandwidth Variable Transceiver)
CAGR	Compound Annual Growth Rate
CapEx	Capital Expenditure
CBR	Constant Bit Rate
CDC	Colorless, Directionless, and Contentionless
CDN	Content Distribution Network
CDS	Connected Dominating Set
CONUS	Continental United States
CO-OFDM	Coherent Optical Orthogonal Frequency-Division Multiplexing
CORONET	Core Optical Networks
COTS	Commercial Off-the-Shelf
CSP	Constrained Shortest Path
dB	Decibel
DCE	Dynamic Channel Equalizer
DCF	Dispersion Compensating Fiber
DDC	Differential Delay Constraint
DGD	Differential Group Delay
DIR	Destination-Initiated Reservation

DLP®	Digital Light Processor
DP-QPSK	Dual-Polarization Quadrature Phase-Shift Keying (or Dual-Polarization Quaternary Phase-Shift Keying)
DPSK	Differential Phase-Shift Keying
DQPSK	Differential Quadrature Phase-Shift Keying
DSE	Dynamic Spectral Equalizer
DSF	Dispersion Shifted Fiber
DSP	Digital Signal Processor
EDC	Electronic Dispersion Compensation
EDFA	Erbium-Doped Fiber Amplifier
E-NNI	External Network-Network Interface
FC	Fiber Channel (or Fibre Channel)
FCAPS	Fault, Configuration, Accounting, Performance, and Security
FEC	Forward Error Correction
FIT	Failures in 10^9 Hours
FMF	Few-Mode Fiber
FWM	Four-Wave Mixing
FXC	Fiber Cross-connect
GbE	Gigabit Ethernet
Gb/s	Gigabit per second (10^9 bits per second)
GC	Grooming Connection
GFP	Generic Framing Procedure
GHz	Gigahertz
GMPLS	Generalized Multi-Protocol Label Switching
h	hours
IAAS	Infrastructure-as-a-Service
IA-RWA	Impairment-Aware Routing and Wavelength Assignment
ICT	Information and Communication Technologies
IEEE	Institute of Electrical and Electronic Engineers
IETF	Internet Engineering Task Force
ILP	Integer Linear Programming
I-NNI	Internal Network-Network Interface
InP	Indium Phosphide
IP	Internet Protocol
IR	Immediate Reservation
ISP	Internet Service Provider
ITU	International Telecommunication Union
ITU-T	International Telecommunication Union-Telecommunication Standardization Sector
JET	Just Enough Time
JIT	Just In Time
km	kilometer
KSP	K-Shortest Paths
LCAS	Link Capacity Adjustment Scheme
LCoS	Liquid Crystal on Silicon
LP	Linear Programming

LSA	Link-State Advertisement
MAC	Media Access Control
Mb/s	Megabit per second (10^6 bits per second)
MCF	Multicommodity Flow or Multicore Fiber
MCP	Multi-Constrained Path
MCS	Multicast Switch
MEMS	Micro-electro-mechanical System
MIMO	Multiple-Input Multiple-Output
min	minute
MLSE	Maximum Likelihood Sequence Estimation
MMF	Multimode Fiber
MP	Minimum Paths
MPLS	Multi-Protocol Label Switching
ms	millisecond
MSTE	Minimum Spanning Tree with Enhancement
NDSF	Non Dispersion-Shifted Fiber
NF	Noise Figure
NFV	Network Functions Virtualization
NHOP	Next-Hop
nm	nanometer
NMS	Network Management System
NNHOP	Next-Next-Hop
NNI	Network-Network Interface
NRZ	Non-Return-to-Zero
N-WDM	Nyquist Wavelength Division Multiplexing
NZ-DSF	Non-Zero Dispersion-Shifted Fiber
OADM	Optical Add/Drop Multiplexer
OADM-MD	Multi-Degree Optical Add/Drop Multiplexer
OAM	Operations, Administration, and Maintenance
OBS	Optical Burst Switching
OC	Optical Carrier
OCh-SPRing	Optical-Channel Shared Protection Ring
ODU	Optical channel Data Unit
O-E-O	Optical-Electrical-Optical
OFDM	Orthogonal Frequency Division Multiplexing
OFS	Optical Flow Switching
OIF	Optical Internetworking Forum
OMS-SPRing	Optical Multiplex Section Shared Protection Ring
O-OFDM	Optical Orthogonal Frequency Division Multiplexing
OOK	On-Off Keying
OpEx	Operational Expenditure
OPM	Optical Performance Monitor
OPS	Optical Packet Switching
OSC	Optical Supervisory Channel
OSNR	Optical Signal-to-Noise Ratio
OSPF	Open Shortest Path First

OSS	Operations Support System
OTN	Optical Transport Network
OTU	Optical channel Transport Unit
O-UNI	Optical User Network Interface
OXC	Optical Cross-connect
PBB-TE	Provider Backbone Bridge - Traffic Engineering
PCC	Path Computation Client
PCE	Path Computation Element
PCEP	PCE Communication Protocol
PDL	Polarization Dependent Loss
PIC	Photonic Integrated Circuit
PMD	Polarization-Mode Dispersion
PON	Passive Optical Network
P-OTP	Packet-Optical Transport Platform
P-OTS	Packet-Optical Transport System
pPCE	Parent Path Computation Element
ps	picosecond (10^{-12} second)
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QoT	Quality of Transmission
QPSK	Quadrature Phase-Shift Keying (or Quaternary Phase-Shift Keying)
RFC	Request for Comments
RFI	Request for Information
RFP	Request for Proposal
RMLSA	Routing, Modulation Level, And Spectrum Assignment
ROADM	Reconfigurable Optical Add/Drop Multiplexer
ROADM-MD	Multi-degree Reconfigurable Optical Add/Drop Multiplexer
ROLEX	Robust Optical-Layer End-to-End X-Connection
RSA	Routing and Spectrum Assignment (or Routing and Spectrum Allocation)
RSP	Restricted Shortest Path
RSVP-TE	Resource ReserVation Protocol-Traffic Engineering
RWA	Routing and Wavelength Assignment
RZ	Return-to-Zero
s	second
SA	Spectrum Assignment
SDH	Synchronous Digital Hierarchy
SDM	Space Division Multiplexing
SDN	Software-Defined Networking
SIR	Source-Initiated Reservation
SLA	Service Level Agreement
SNR	Signal-to-Noise Ratio
SONET	Synchronous Optical Network
SPDP	Shortest Pair of Disjoint Paths

SPM	Self-Phase Modulation
SPRing	Shared Protection Ring
SRG	Shared Risk Group
SRLB	Selective Randomized Load Balancing
SRLG	Shared Risk Link Group
STM	Synchronous Transport Module
STS	Synchronous Transport Signal
Tb/s	Terabit per second (10^{12} bits per second)
TDM	Time Division Multiplexing
TED	Traffic Engineering Database
THz	Terahertz
TWIN	Time-Domain Wavelength Interleaved Networking
TxRx	Transmitter/Receiver Card (Transponder)
UNI	User Network Interface
VCAT	Virtual Concatenation
VSPT	Virtual Shortest Path Tree
WA	Wavelength Assignment
WDM	Wavelength Division Multiplexing
WGR	Wavelength Grating Router
WSS	Wavelength-Selective Switch
WSXC	Wavelength-Selective Cross-connect
XOR	Exclusive Or
XPM	Cross-phase Modulation

Chapter 1
Introduction to Optical Networks

1.1 Brief Evolution of Optical Networks

While the basic function of a network is quite simple—enabling communications between the desired endpoints—the underlying properties of a network can greatly affect its value. Network capacity, reliability, cost, scalability, and operational simplicity are some of the key benchmarks on which a network is evaluated. Network designers are often faced with trade-offs among these factors and are continually looking for technological advances that have the potential to improve networking on a multitude of fronts.

One such watershed development came in the 1980s as telecommunications carriers began migrating much of the physical layer of their intercity networks to fiber-optic cable. Optical fiber is a lightweight cable that provides low-loss transmission; but clearly, its most significant benefit is its tremendous potential capacity. Not only did fiber optics offer the possibility of a huge vista for transmission but it also gave rise to optical networks and the field of optical networking.

An optical network is composed of the fiber-optic cables that carry channels of light, combined with the equipment deployed along the fiber to process the light. The capabilities of an optical network are necessarily tied to the physics of light and the technologies for manipulating lightstreams. As such, the evolution of optical networks has been marked with major paradigm shifts as exciting breakthrough technologies have been developed.

One of the earliest technological advances was the ability to carry multiple channels of light on a single fiber. Each lightstream, or wavelength[1], is carried at a different optical frequency and multiplexed (i.e., combined) onto a single fiber, giving rise to wavelength division multiplexing (WDM). The earliest WDM systems supported fewer than ten wavelengths on a single fiber. Since 2000, this number has rapidly grown to over 100 wavelengths per fiber, providing a tremendous growth in network capacity.

[1] The term "wavelength" is commonly used in two different contexts: first, it refers to a channel of light; second, it refers to the specific point in the spectrum of light where the channel is centered (e.g., 1,550 nanometers). The context should be clear from its usage; however, when necessary, clarifying text is provided.

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4_1, © Springer International Publishing Switzerland 2014

A key enabler of cost-effective WDM systems was the development of the erbium-doped fiber amplifier (EDFA). Prior to the deployment of EDFAs, each wavelength on a fiber had to be individually regenerated[2] at roughly 40 km intervals, using costly electronic equipment. In contrast, EDFAs, deployed at roughly 80 km intervals, optically amplify all of the wavelengths on a fiber at once. Early EDFA systems allowed optical signals to be transmitted on the order of 500 km before needing to be individually regenerated; with more recent EDFA systems, this distance has increased to 1,500–2,500 km.

A more subtle innovation was the gradual migration from an architecture where the optical network served simply as a collection of static pipes to one where it was viewed as another networking layer. In this optical networking paradigm, network functions such as routing and protection are supported at the granularity of a wavelength, which can be operationally very advantageous. A single wavelength may carry hundreds of circuits. If a failure occurs in a fiber cable, restoring service by processing individual wavelengths is operationally simpler than rerouting each circuit individually.

The benefits of scale provided by optical networking have been further accelerated by the increasing bit rate of a single wavelength. In the mid-1990s, the maximum bit rate of a wavelength was roughly 2.5 Gb/s (Gb/s is 10^9 bits/sec). This has since ramped up to 10, 40, and 100 Gb/s. Furthermore, 400 Gb/s and 1 Tb/s rates are likely to be deployed in the 2015–2020 time frame (Tb/s is 10^{12} bits/sec).

Increased wavelength bit rate combined with a greater number of wavelengths per fiber has expanded the capacity of optical networks by several orders of magnitude over a period of 25 years. However, transmission capacity is only one important factor in evaluating the merits of a network. The cost-effectiveness and scalability of the network, typically embodied by the required amount of equipment, are important as well. While EDFAs enabled the removal of a sizeable amount of electronic equipment, each wavelength still underwent electronic processing at numerous points in the network, i.e., at each switching or traffic-generating site along the path of a wavelength. As network traffic levels experienced explosive growth, this necessitated the use of a tremendous amount of electronic terminating and switching equipment, which presented challenges in cost, power consumption, heat dissipation, physical space, reliability, deployment time, and maintenance.

This bottleneck was greatly reduced by the development of *optical-bypass* technology. This technology eliminates much of the required electronic processing and allows a signal to remain in the optical domain for all, or much, of its path from source to destination. Because optical technology can operate on a spectrum of wavelengths at once and can operate on wavelengths largely independently of their bit rate, keeping signals in the optical domain allows a significant amount of equipment to be removed from the network and provides a scalable trajectory for network growth.

Achieving optical bypass required advancements in areas such as optical amplification, optical switching, transmission formats, and techniques to mitigate optical

[2] Regeneration is performed to restore the quality of the signal.

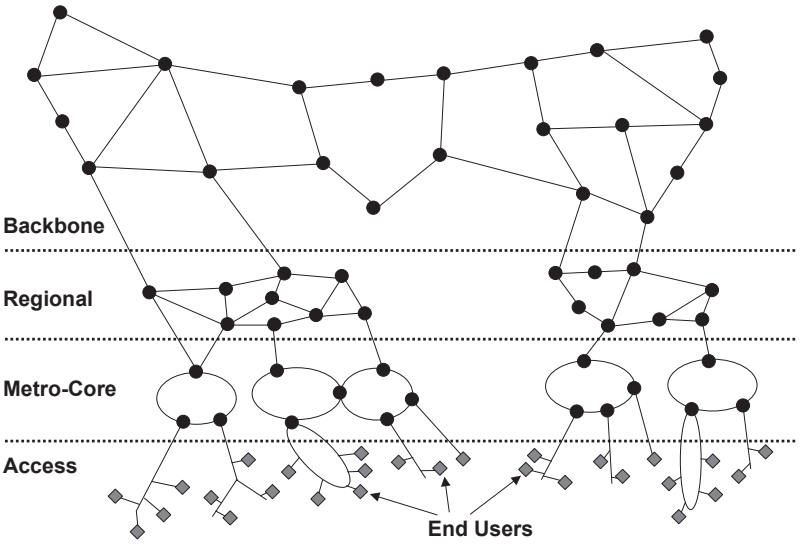


Fig. 1.1 Networking hierarchy based on geography

impairments. Commercialization of optical-bypass technology began in the mid-to-late 1990s, eventually leading to its deployment by most telecommunications carriers over the following decade. While reducing the amount of electronic processing addressed many of the impediments to continued network growth, it also brought new challenges. Most notably, it required the development of new algorithms to assist in operating the network so that the full benefits of the technology could be attained. Overall, the advent of optical-bypass technology has transformed the architecture, operation, and economics of optical networks, all of which is covered in this book.

1.2 Geographic Hierarchy of Optical Networks

When considering the introduction of new networking technology, it can be useful to segment the network into multiple geographic tiers, with key differentiators among the tiers being the number of customers served, the required capacity, and the geographic extent. One such partitioning is shown in Fig. 1.1. (In this section, the standalone term "network" refers to the network as a whole; when "network" is used in combination with one of the tiers, e.g., "backbone network," it refers to the portion of the overall network in that particular tier.)

At the edge of the network, closest to the end users, is the *access* tier, which distributes/collects traffic to/from the customers of the network. Access networks generally serve tens to hundreds of customers and span a few kilometers. (One can

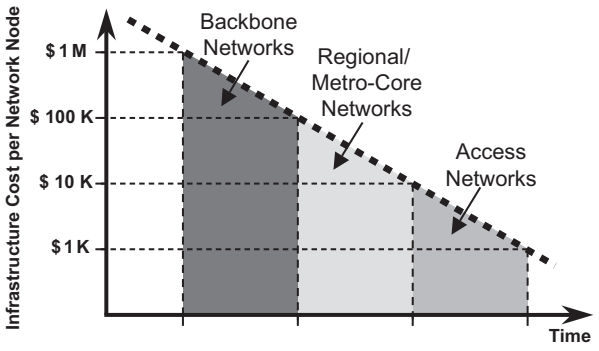


Fig. 1.2 Cost trend for introducing WDM technology in different tiers of the network. As the cost of WDM infrastructure decreases over time, it is introduced closer to the network edge. (Adapted from Saleh [Sale98b])

further subdivide the access tier into business access and residential access, or into metro access and rural access.) The *metro-core* tier is responsible for aggregating the traffic from the access networks, and typically interconnects a number of telecommunications central offices or cable distribution head-end offices. A metro-core network aggregates the traffic of thousands of customers and spans tens to hundreds of kilometers.

Moving up the hierarchy, multiple metro-core networks are interconnected via *regional* networks. A regional network carries the portion of the traffic that spans multiple metro-core areas, and is shared among hundreds of thousands of customers, with a geographic extent of several hundred to a thousand kilometers. Interregional traffic is carried by the *backbone* network.[3] Backbone networks may be shared among millions of customers and typically span thousands of kilometers.

While other taxonomies may be used, the main point to be made is that the characteristics of a tier are important in selecting an appropriate technology. For example, whereas the backbone network requires optical transport systems with very large capacity over long distances, that same technology would not be appropriate for, nor would it be cost effective in, an access network.

As one moves closer to the network edge, the cost of a network in a particular tier is amortized over fewer end users, and is thus a more critical concern. Because of this difference in price sensitivity among the tiers, there is often a trend to deploy new technologies in the backbone network first. As the technology matures and achieves a lower price point, it gradually extends closer towards the edge. A good example of this trend is the deployment of WDM technology, as illustrated in Fig. 1.2.

Even as a technology permeates a network, the particular implementation may differ across tiers. For example, with respect to WDM technology, backbone networks generally have 80–160 wavelengths per fiber, regional networks have roughly 40–80 wavelengths per fiber, metro-core WDM networks have anywhere from 8–40 wavelengths per fiber, and access networks typically have no more than 8 wavelengths.

[3] Other common names for this tier are the long-haul network or the core network. These terms are used interchangeably throughout the book.

A similar pattern has emerged with the introduction of optical-bypass technology. Appreciable commercial deployment began in backbone networks in the 2000 time frame, and has gradually been extended closer to the network edge. The capabilities of optical-bypass-based systems are tailored to the particular network tier. For example, the distance a signal can be transmitted before it suffers severe degradation is a fundamental attribute of such systems. In backbone networks, technology is deployed where this distance is a few thousand kilometers; in metro-core networks, it may be only several hundred kilometers.

While optical networking is supported to varying degrees in the different tiers of the network, the architecture of access networks (especially residential access) is very distinct from that of the other portions of the network. For example, one type of access network is based on passive devices (i.e., the devices in the field do not require power). These systems, aptly named passive optical networks (PONs), would not be appropriate for larger-scale networks. Because the topological characteristics, cost targets, and architectures of access networks are so different from the rest of the network, they are worthy of a book on their own; hence, access networks are not covered here. Access technologies are covered in detail in Lin [Lin06], Lam [Lam07], and Abdallah et al. [AbMA09]. Suffice it to say that as optics enters the access network, enabling the proliferation of high-bandwidth end-user applications, there will be increased pressure on the remainder of the network to scale accordingly.

It should be noted that there is a recent trend in the telecommunications industry to "blur the boundaries" between the tiers. Carriers are looking for technology platforms that are flexible enough to be deployed in multiple tiers of the network, with unified network management and provisioning systems to simplify operations [ChSc07; Gril12].

1.3 Layered Architectural Model

Another useful network stratification is illustrated by the three-layered architectural model shown in Fig. 1.3. At the top of this model is the applications layer, which includes all types of services, such as voice, video, and data. The intermediate layer encompasses multiplexing, transport, and switching based on electronic technology. For example, this layer includes Internet Protocol (IP) routers, Ethernet switches, Asynchronous Transfer Mode (ATM) switches, Synchronous Optical Network / Synchronous Digital Hierarchy (SONET/SDH) switches, and Optical Transport Network (OTN) switches. Each of these protocols has a particular method for partitioning data and moving the data from source to destination.

The payloads of the electronic layer are passed to the optical layer, where they are carried in wavelengths. In the model of interest, the optical layer is based on WDM technology and utilizes optical switches that are capable of dynamically routing wavelengths. Thus, the bottom tier of this particular model can also be referred to as the "configurable WDM layer."

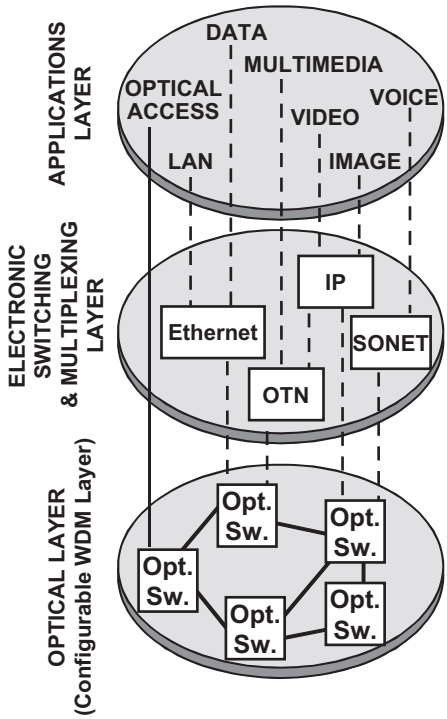


Fig. 1.3 Three-layered architectural model. In the model of interest, the optical layer is based on wavelength division multiplexing (*WDM*) technology with configurable optical switches. (Adapted from Wagner et al. [WASG96]. ©1996 IEEE)

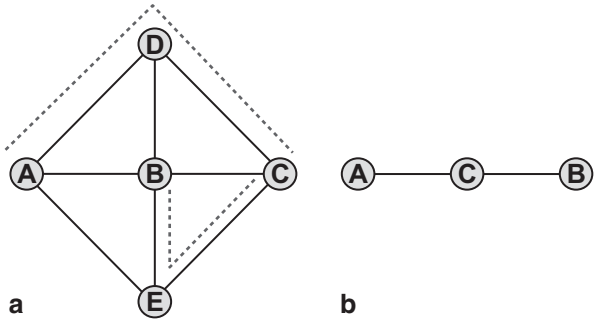


Fig. 1.4 **a** The *solid lines* represent the physical fiber-optic links and the *dotted lines* represent the paths of two routed wavelengths. **b** The two wavelength paths create a virtual topology where the *solid lines* represent virtual links. The virtual topology can be modified by establishing different wavelength paths

From the viewpoint of the electronic layer, the wavelengths form a *virtual topology*. This concept is illustrated in Fig. 1.4 by a small network interconnecting five points. In Fig. 1.4a, the solid lines represent fiber-optic cables, or the physical topology, and the dotted lines represent the paths followed by two of the wavelengths. This arrangement of wavelengths produces the virtual topology shown in Fig. 1.4b; i.e., this is the network topology as seen by the electronic layer. In contrast to the fixed physical topology, the virtual topology can be readily modified by reconfiguring the paths of the wavelengths.

Note that it is possible for the application layer to directly access the optical layer, as represented in Fig. 1.3 by the optical access services. This capability could be desirable, for example, to transfer very large streams of protocol-and-format-independent data. Because the electronic layers are bypassed, no particular protocol is imposed on the data. By transporting the service completely in the optical domain, the optical layer potentially provides what is known as *protocol and format transparency*. While such transparency has often been touted as another benefit of optical networking, thus far these services have not materialized in a major way in practical networks.

1.4 Interfaces to the Optical Layer

One difficulty with carrying services directly in wavelengths is that the network can be difficult to manage. Network operations can be simplified by using standard framing that adds overhead for management. For example, the SONET and SDH specifications, which are closely related, define a standard framing format for optical transmission, where the frame includes overhead bytes for functionality such as performance monitoring, path trace, and operations, administration, and maintenance (OAM) communication. SONET/SDH has been commonly used as the interface to the optical layer; standards exist to map services such as IP and ATM into SONET/SDH frames. In addition to using SONET/SDH for framing, it is often used for switching and multiplexing in the electrical domain, as shown in Fig. 1.3. SONET/SDH makes use of *time division multiplexing* (TDM), where circuits are assigned to time slots that are packed together into a larger frame.

SONET/SDH has been gradually superseded by the OTN standard as the interface to the optical layer [ITU01], where SONET/SDH becomes one of the services that can be carried by the OTN layer. As with SONET/SDH, OTN provides TDM switching and multiplexing capabilities, in addition to framing. Although OTN is better suited to today's optical networks (as discussed below), there is still a great deal of deployed legacy SONET/SDH-based equipment.

1.4.1 SONET/SDH

As noted above, the SONET and SDH specifications are very similar. SONET is the American National Standards Institute (ANSI) standard and is generally used in North America, whereas SDH is the International Telecommunication Union (ITU) standard and is typically used in the rest of the world.[4] The SONET/SDH standards

[4] The ITU recommendations discussed in this book have been developed by the Telecommunication Standardization Sector of the ITU, also known as ITU-T.

Table 1.1 Commonly used SONET/SDH signal rates

SONET signal	SDH signal	Bit rate
STS-1, OC-1	–	51.84 Mb/s
STS-3, OC-3	STM-1	155.52 Mb/s
STS-12, OC-12	STM-4	622.08 Mb/s
STS-48, OC-48	STM-16	2.49 Gb/s (2.5 Gb/s)
STS-192, OC-192	STM-64	9.95 Gb/s (10 Gb/s)
STS-768, OC-768	STM-256	39.81 Gb/s (40 Gb/s)

were initially developed in the 1980s with a focus on voice traffic, although features have been added to make them more suitable for data traffic.

SONET defines a base signal with a rate of 51.84 Mb/s, called synchronous transport signal level-1 or STS-1 (Mb/s is 10^6 bits/sec). Multiple STS-1 signals are multiplexed together to form higher rate signals, giving rise to the SONET rate hierarchy. For example, three STS-1 signals are multiplexed to form an STS-3 signal. The optical instantiation of a general STS-N signal is called optical carrier level-N, or OC-N. SDH is similar to SONET, although the framing format is somewhat different. The SDH base signal is defined as synchronous transport module level-1, or STM-1, which has a rate equivalent to an STS-3. Some of the most commonly used SONET and SDH rates are shown in Table 1.1. The bit rates shown in parentheses for some of the signals are the nominal rates commonly used in reference to these signals. For more details on SONET/SDH technology, see Tektronix [Tekt01], Goralski [Gora02], and Telcordia Technologies [Telc09].

1.4.2 Optical Transport Network

In the late 1990s, the ITU began work on OTN to better address the needs of optical networking and multi-service networks. The associated transport hierarchy and formats are defined in the ITU G.709 standard [ITU12a], with the basic transport frame called the *Optical channel Transport Unit* (OTU). The bit rate of the OTU hierarchy is slightly higher than that of SONET/SDH to account for additional overhead, as shown in Table 1.2. It is likely that the OTN hierarchy eventually will be extended beyond OTU4 to support higher line rates as they become standardized (e.g., 400 Gb/s, 1 Tb/s). (G.709 in fact alludes to OTU5 through OTU7.)

Each transport frame contains one or more *Optical channel Data Units* (ODUs), where the ODU is the basic unit for switching/multiplexing. The ODU hierarchy is shown in Table 1.3. Note that the ODU granularity is finer than that of the OTU, supporting service rates as low as 1.25 Gb/s. For example, a Gigabit Ethernet (GbE) connection can be mapped into an ODU0. As there is no corresponding OTU0, multiple ODU0s are multiplexed into a higher-order ODU, and then transported in an OTUk frame (k can be 1–4).

ODU-Flex is the most recent addition to the ODU hierarchy, to enable OTN to be used as an efficient transport mechanism for a wider range of data rates and services. ODU-Flex comes in two flavors. With ODU-Flex-Generic Framing Procedure

Table 1.2 OTN transport rate hierarchy

OTU type	Nominal bit rate
OTU1	2.666 Gb/s
OTU2	10.709 Gb/s
OTU3	43.018 Gb/s
OTU4	111.810 Gb/s

Table 1.3 OTN switching/multiplexing rate hierarchy

ODU type	Nominal bit rate
ODU-Flex (CBR)	~ Client signal bit rate
ODU-Flex (GFP)	N × ~ 1.25 Gb/s
ODU0	1.244 Gb/s
ODU1	2.499 Gb/s
ODU2	10.037 Gb/s
ODU3	40.319 Gb/s
ODU4	104.794 Gb/s

(GFP), an appropriate number of ODUk (k can be 2–4) tributary slots are allocated for the service, where each tributary slot corresponds to approximately 1.25 Gb/s (the exact tributary rate depends on k). For example, four ODU2 tributary slots would be allocated to carry a 4-Gb/s Fiber Channel (4G FC) connection. In addition to ODU-Flex-GFP, there is ODU-Flex-Constant Bit Rate (CBR), where the ODU overhead is wrapped around the client data to carry an arbitrary bit-rate connection.

Compared with SONET/SDH, OTN provides benefits such as more efficient multiplexing and switching of high-bandwidth services, enhanced monitoring capabilities, and stronger forward error correction (FEC). FEC allows bit errors picked up during signal transmission to be corrected when the signal is decoded. Enhanced FEC can be used to compensate for more severe transmission conditions. For example, it potentially allows more wavelengths to be multiplexed onto a single fiber, or allows a signal to remain in the optical domain for longer distances, which is important for optical-bypass systems.

OTN provides a combination of both transparency and manageability. Its framing structure, often referred to as a "digital wrapper," can carry different protocols transparently without affecting content, control channels, or timing. Its associated OAM capabilities provide a consistent managed view for a range of services. OTN potentially provides a convergence layer for optical networks, where carriers can support multiple services with a single network rather than deploying parallel networks, without compromising the resilient operations and management capabilities that carriers have come to expect from SONET/SDH.

OTN and SONET/SDH are circuit-based transport layers. As such, they are not optimized for carrying packet-based services, such as IP and Ethernet. Circuits are routed over dedicated "channels" in the network that remain active for the duration of the communication session. In contrast, packets are blocks of data that may be individually routed in the network; bandwidth is typically shared among packets

from multiple services. While packet-based routing is generally more bandwidth efficient, it also may result in packet loss, excess latency, or delay variation (jitter). These factors need to be considered when transporting packets to ensure that packet services achieve their desired *quality of service* (QoS). Thus, there have been initiatives to also develop a packet transport layer. The ITU and Internet Engineering Task Force (IETF) have jointly worked on the *Multi-Protocol Label Switching-Transport Profile* (MPLS-TP) recommendation [NBBS09; BBFL10; ITU11]; the Institute of Electrical and Electronic Engineers (IEEE) has developed the *Provider Backbone Bridge-Traffic Engineering* (PBB-TE) standard [IEEE09]. In both endeavors, there has been an attempt to transform widely used existing packet-based standards (MPLS and Ethernet, respectively) into connection-oriented transport technologies with OAM capabilities similar to SONET/SDH and OTN. Widespread support for these technologies thus far has not emerged among most carriers.

1.5 Optical Control Plane

A communications network can be viewed as being composed of three planes: the data, management, and control planes. The data plane is directly responsible for the forwarding of data, whereas the management and control planes are responsible for network operations. The management plane generally operates in a centralized manner; the control plane implements just a subset of the network operations functionality, typically in a more distributed manner.

Optical networks have historically been managed using a centralized network management system (NMS), where the NMS performs what are commonly known as the FCAPS management functions—fault, configuration, accounting, performance, and security. However, beginning in the late 1990s, control plane software was introduced into optical networks. The optical control plane, composed of a set of applications that resides on the physical network equipment, is capable of automating many of the processes related to network configuration. It enables functionality such as: discovery of the local network topology, network resources, and network capabilities; dissemination of this information throughout the network; path computation; and signaling for connection establishment and teardown.

Various organizations have developed recommendations in support of the optical control plane. For example, the ITU has developed the *Automatically Switched Optical Network* (ASON) recommendation, which focuses on the architecture and requirements for control-plane-enabled optical transport networks [ITU12c]. The IETF has developed the *Generalized Multi-Protocol Label Switching* (GMPLS) suite of protocols for routing and signaling for various network technologies, including optical networks [BeRS03; Mann04; SDIR05].

GMPLS includes three models for interacting with the optical layer: peer, overlay, and augmented. For concreteness, the discussion of these models will focus on the interaction of the IP and optical layers, but the principles apply to other electronic layers that sit above the optical layer.

In the peer (or integrated) model, the IP and optical layers are treated as a single domain, with IP routers having full knowledge of the optical topology. (A domain is a collection of network resources under the control of a single entity. The interface between domains is known as the *external network–network interface* (E-NNI), whereas the interface between networks within a domain is referred to as the *internal NNI* (I-NNI) [ITU12c].) In this model, the IP routers are capable of determining the entire end-to-end path of a connection including how it should be routed through the optical layer. In the overlay model, the IP and optical layers are treated as distinct domains, with no exchange of routing and topology information between them. The IP layer is essentially a *client* of the optical layer and requests bandwidth from the optical layer as needed. The augmented model is a hybrid approach where a limited amount of information is exchanged between layers.

Given the amount of information that needs to be shared in the peer model, and the potential trust issues between the layers (e.g., the IP and optical layers may be operated by different organizations), the overlay and augmented models are generally favored by carriers, with the overlay model being the most commonly accepted policy. The boundary between the client layer (e.g., IP) and the optical transport layer is called the *user network interface* (UNI); it is also more specifically referred to as the optical-UNI (O-UNI). Signaling specifications for the UNI have been developed by the IETF as well as the Optical Internetworking Forum (OIF; [SDIR05; OIF04]).

It should be noted that the standards activities thus far do not fully address networks with optical bypass. The technology for accomplishing optical bypass is typically specific to the particular system vendor that is providing the equipment. Each vendor has its own set of system engineering rules that impact how the traffic should be routed. This makes codifying the rules for configuring a network with optical bypass difficult. Consequently, control plane implementations for such networks generally remain proprietary to the vendor.

The optical control plane plays an important role in supporting dynamic traffic, and will be discussed in much more detail in Chap. 8.

1.6 Terminology

This section introduces some of the terminology that is used throughout the book. Refer to the small network shown in Fig. 1.5. The circles represent the network *nodes*. These are the points in the network that source/terminate and switch traffic. The lines interconnecting the nodes are referred to as *links*. While the links are depicted with just a single line, they typically are populated by one or more fiber pairs, where each fiber in a pair carries traffic in just one direction. (It is possible to carry bidirectional traffic on a single fiber, but not common.) Optical amplifiers may be periodically located along each fiber, especially in regional and backbone networks. Sites that solely perform amplification are *not* considered nodes. The portion of a

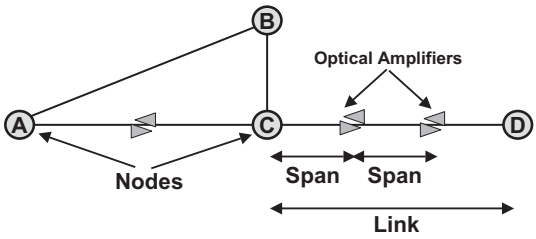


Fig. 1.5 Nodes are represented by *circles,* and links are represented by *solid lines*. Nodes *A* and *B* have a degree of two, Node *C* has a degree of three, and Node *D* has a degree of one

link that runs between two amplifier sites, or between a node and an amplifier site, is called a *span*.

A very important concept is that of *nodal degree*. The degree of a node is the number of links incident on that node. Thus, in the figure, Nodes A and B have a degree of two, Node C has a degree of three, and Node D has a degree of one. Nodal degree is very important in determining the type of equipment appropriate for a node.

The specific arrangement of nodes and links constitutes the *network topology*. Early networks were commonly based on ring topologies due to the simple restoration properties of rings. More recently, networks, especially those in the backbone, have migrated to more flexible *mesh* topologies. In mesh networks, the nodes are arbitrarily interconnected, with no specific routing pattern imposed on the traffic. In Fig. 1.1, the topologies in the metro-core tier are shown as rings, whereas the regional and backbone topologies are mesh. While it is possible to develop network design techniques that are specifically optimized for rings, the approach of this book is to present algorithms and design methodologies that are general enough to be used in any topology (with a few exceptions).

The *traffic* in the network is the collection of services that must be carried. The term *demand* is used to represent an individual traffic request. For the most part, demands are between two nodes and are bidirectionally symmetric. That is, if there is a traffic request from Node A to Node B, then there is equivalent traffic from Node B to Node A. In any one direction, the originating node is called the *source* and the terminating node the *destination*. In multicast applications, the demands have one source and multiple destinations; such demands are typically one-way only. It is also possible to have demands with multiple sources and one or more destinations, but not common.

The term *connection* is used to represent the path allocated through the network for carrying a demand. The process of deploying and configuring the equipment to support a demand is called *provisioning,* or *turning up,* the connection. The rate of a demand or a connection will usually be referred to in absolute terms (e.g., 10 Gb/s). Occasionally, OTN terminology (e.g., OTU2) or SONET terminology (e.g., OC-192) may be used in a particular example.

The optical networks of interest in this book are based on WDM technology. Figure 1.6 shows the portion of the light spectrum where WDM systems are generally based, so chosen because of the relatively low fiber attenuation in this region. (As shown in the figure, the fiber loss is typically between 0.20 and 0.25 dB/km

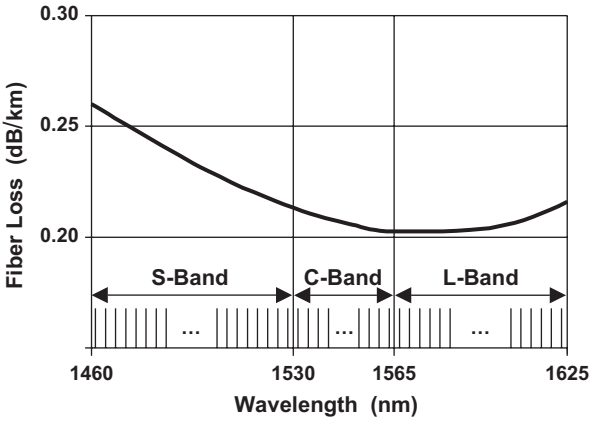


Fig. 1.6 Approximate *S, C,* and *L* wavelength bands, and the corresponding typical fiber loss

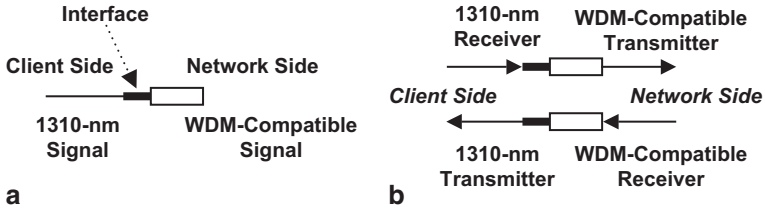


Fig. 1.7 **a** A simplified depiction of a wavelength-division multiplexing (*WDM*) transponder that converts between a 1,310 nm signal and a *WDM*-compatible signal. **b** A more detailed depiction of the *WDM* transponder, which emphasizes its bidirectional composition. There is both a 1,310 nm transmitter/receiver and a *WDM*-compatible transmitter/receiver

in this region.) This spectrum is broken into three regions: the conventional band or C-band; the long wavelength band or L-band; and the short wavelength band or S-band. Most WDM systems make use of the C-band; however, there has been expansion into the L- and S-bands to increase system capacity.

An optical channel can be referred to as operating at a particular wavelength, in units of nanometers (nm), or equivalently at a particular optical frequency, in units of terahertz (THz). The term *lambda* is frequently used to refer to the particular wavelength on which an optical channel is carried; *lambda i,* or λ_i, is used to represent the i^{th} wavelength of a WDM system. The distance between adjacent channels in the spectrum is generally noted in frequency terms, in units of gigahertz (GHz). For example, a 40-channel C-band system is achieved with 100-GHz spacing between channels, whereas an 80-channel C-band system is obtained using 50-GHz spacing.

An important transmission component is the WDM *transponder,* which is illustrated in Fig. 1.7a. One side of the transponder is termed the *client side,* which takes a signal from the client of the optical network, e.g., an IP router. The client optical signal is generally carried on a 1,310 nm wavelength. (1,310 nm is outside the

WDM region; WDM is usually not used for intra-office[5] communication.) Various interfaces can be used on the client side of the transponder, depending on how much optical loss is encountered by the client signal. For example, *short-reach interfaces* tolerate up to 4 or 7 dB of loss depending on the signal rate, whereas *intermediate-reach interfaces* tolerate up to 11 or 12 dB of loss.[6] The interface converts the client optical signal to the electrical domain. The electrical signal modulates (i.e., drives) a WDM-compatible laser such that the client signal is converted to a particular wavelength (i.e., optical frequency) in the WDM region. The WDM side of the transponder is also called the *network side*. In the reverse direction, the WDM-compatible signal enters from the network side and is converted to a 1,310 nm signal on the client side.

A single WDM transponder is shown in more detail in Fig. 1.7b, to emphasize that there is a client-side receiver and a network-side transmitter in one direction and a network-side receiver and a client-side transmitter in the other direction. For simplicity, the transponder representation in Fig. 1.7a is used in the remainder of the book; however, it is important to keep in mind that a transponder encompasses separate devices in the two signal directions.

In fixed-tuned transponders, the client signal can be converted to just one particular optical frequency. In transponders equipped with tunable lasers, the client signal can be converted to any one of a range of optical frequencies. Some architectures require that the transponder have an optical filter on the network side to receive a particular frequency (or some other methodology to pick out one frequency from a WDM signal). Tunable filters allow any one of a range of optical frequencies to be received. Since the early 2000s, most networks have been equipped with transponders with tunable lasers; transponders with both tunable lasers and filters were commercially available some time later. While there is a small cost premium for tunable transponders as compared to fixed transponders, they greatly improve the flexibility of the network as well as simplify the process of maintaining inventory and spare equipment for failure events.

The signal rate carried by a wavelength is called the *line rate*. It is often the case that the clients of the optical network generate traffic that has a lower rate than the wavelength line rate. This is referred to as *subrate* traffic. For example, an IP router may generate 10 Gb/s signals but the line rate may be 40 Gb/s. This mismatch gives rise to the need to *multiplex* or *groom* traffic, where multiple client signals are carried on a wavelength in order to improve the network efficiency. End-to-end multiplexing bundles together subrate traffic with the same endpoints; grooming uses more complex aggregation than multiplexing and is thus more efficient, though more costly. It is also possible, though less common, for the client signal rate to be higher than the wavelength line rate. In this scenario, *inverse multiplexing* is used, where the client signal is carried over multiple wavelengths.

[5] Office refers to a building that houses major pieces of telecommunications equipment, such as switches and client equipment.

[6] The loss increases with fiber distance and the number of fiber connectors; thus, these various types of interfaces determine the allowable interconnection arrangements within an office.

1.7 Network Design and Network Planning

As indicated by the title of the book, both network design and network planning are covered. Network design encompasses much of the up-front work such as selecting which nodes to include in the network, laying out the topology to interconnect the nodes, selecting what type of transmission and switching systems to deploy (e.g., selecting the line rate and whether to use optical bypass), and what equipment to deploy at a particular node. Network planning is more focused on the details of how to accommodate the traffic that will be carried by the network. For example, network planning includes selecting how a particular demand should be routed, protected, and groomed, and what wavelength(s) in the system spectrum should be assigned to carry it.

Network planning is carried out on two timescales, both of which are covered in this book. In *long-term network planning,* there is sufficient time between the planning and provisioning processes such that any additional equipment required by the plan can be deployed. In the long-term planning that typically occurs before a network is deployed, there is generally a large set of demands to be processed at one time. In this context, the planning emphasis is on determining the optimal strategy for accommodating the set of demands. After the network is operational, long-term planning is performed for the incrementally added traffic, assuming the traffic does not need to be provisioned immediately. Again, the focus is on determining optimal strategies, as there is enough time to deploy equipment to accommodate the design.

In *real-time network planning,* there is little time between planning and provisioning, and demands are generally processed one at a time. It is assumed that the traffic must be accommodated using whatever equipment is already deployed in the network. Thus, the planning process must take into account any constraints posed by the current state of deployed equipment, which, for example, may force a demand to be routed over a suboptimal path. (A related topic is *traffic engineering,* which in this context is a process where traffic is routed to meet specific performance objectives; e.g., a demand may be routed over a specific path to meet a particular latency metric. Traffic-engineering support for real-time routing has been incorporated in several protocols; e.g., see Awduche et al. and Katz et al. [ABGL01; KaKY03].)

1.8 Research Trends in Optical Networking

This chapter started with a brief summary of how optical networks have evolved. We now discuss areas of current research, which may provide insight into future optical networking advancements. Most of these topics are covered in greater depth in later chapters.

Optical bypass has become a well-accepted technology for reducing equipment costs in a network. Furthermore, it is now recognized that one of its prime benefits is greater operational efficiency, due to reduced power consumption and physical space requirements, greater reliability, and the need for less manual intervention.

To derive even greater operational benefits, the industry has turned its attention to increasing the flexibility of optical-bypass equipment, as is covered in detail in Chap. 2. This will allow the network to be more easily reconfigured. The desired new features may actually increase the cost of equipment somewhat; thus, they are likely to be implemented only if they can be justified by accompanying operational-cost reductions.

Of the various benefits afforded by optical technology, one of the most important has turned out to be its relative efficiency with respect to power consumption. As energy usage in the data and telecommunications industry continues to soar, energy-efficient optical networking may make further inroads into networking realms that are currently dominated by electronics. There are at least two lines of research in this direction. First is an attempt to reduce the amount of fine-granularity electronic switching, either by performing some of the switching in the optical domain or by utilizing architectures that obviate the need for such switching. These techniques are covered in Chaps. 6 and 9. Second, optical networking is likely to pervade closer to the network edge. The chart in Fig. 1.2 shows WDM technology being gradually introduced into various tiers of the network. The next step may be to introduce this technology *within* the premises of large end users, e.g., to interconnect the huge number of servers inside data centers.

Another benefit potentially afforded by optical technology is configurability, where wavelength connections can be established, rerouted, or torn down remotely through software. Today, despite the presence of a configurable optical layer, most networks are relatively static. Configurability does play a role in reducing the time required for operations personnel to respond to new service requests and in minimizing the number of "truck rolls" to various sites in the network. This falls far short of the rapidly responsive optical layer that had been envisioned in early research papers. However, the trend towards more "discontinuous" traffic, where there is a sudden flux of traffic in various areas of the network, has revived interest in *dynamic* optical networking. In this model, the optical layer is reconfigured in seconds (or even faster), in response to requests from higher layers of the network. This is clearly a major departure from the current mode of operation.

Dynamic optical networking is covered extensively in Chap. 8. Some of the topics covered include centralized versus distributed architectures, latency, resource contention, regeneration, pre-deployment of equipment, and multi-domain environments. Both recent research results and standardization efforts are covered.

Dynamic networking will only come to fruition if a solid business case can be made; e.g., if performance begins to suffer due to the relative rigidity of current networks, or if dynamic networking engenders a new set of revenue opportunities. This will likely come about as data centers increasingly play a prominent role in network architecture. Applications such as cloud computing (where enterprises migrate much of their computing and storage resources to distributed data centers) and network virtualization (where network resources can be dynamically reconfigured based on customer needs) will grow increasingly more reliant on having a responsive network that can deliver a consistent level of performance. Furthermore, data centers are changing the typical well-defined point-to-point routing model.

Enterprises are more interested in connectivity to resources, which are distributed among a set of data centers, rather than connectivity to particular sites. Algorithms to address this are included in Chap. 3.

One area where optics has met its expectations (or perhaps even surpassed them) is with respect to the volume of traffic that can be supported on a fiber. As discussed in Sect. 1.1, the rapid growth in network capacity has been achieved by both increasing the number of wavelengths carried on a fiber and increasing the bit rate of each wavelength. However, the maximum number of wavelengths supported on a fiber (in backbone networks) has generally stabilized around 80–100 wavelengths. (While supporting more wavelengths is possible, the trade-offs involved may not be cost effective.) In contrast, the bit rate of a wavelength has continued to increase. The state of the art in 2014 deployments is 100 Gb/s per wavelength, with an expectation that the wavelength bit rate will eventually evolve to 400 Gb/s or even 1 Tb/s.

In order to achieve increased wavelength bit rates, the transmission formats have become more complex. This has resulted in needing to account for a greater number of detrimental physical effects in order to maintain a high level of optical bypass. Additionally, mixed line-rate (MLR) networks, where wavelengths with different bit rates are routed on one fiber, pose special challenges depending on the transmission formats that are present. These factors need to be captured in the network planning algorithms, as described in Chaps. 4 and 5.

The nature of capacity evolution going forward is likely to be very different from the past. Further increases in the wavelength bit rate will probably be accompanied, at least initially, by *fewer* wavelengths on a fiber, due to the need to space the wavelengths further apart for transmission performance reasons. Thus, the pace of growth in network capacity is likely to slow. Furthermore, analysis indicates that the ultimate capacity of conventional fiber is being approached (assuming current technology trends) and may be reached in the next decade or so (depending on the rate of traffic growth). This has spurred research into new fiber types and technologies that can dramatically increase the fiber capacity. There is also an architectural push to use capacity more efficiently by eliminating the fixed spectral grid that has been in place for more than a decade. Such "flex-grid" schemes would drive the need for accompanying network design algorithms. These architectures and algorithms are covered in detail in Chap. 9.

Finally, there is ongoing research into improving network management, including management of the optical layer. This includes improved transport of IP and Ethernet services and more tightly integrated control across network layers. For the most part, these are software-based developments, as opposed to hardware innovations. One software-based approach in particular, i.e., Software-Defined Networking (SDN), is discussed in Chap. 8.

It can be difficult to introduce paradigm-changing software into the network, as "backward compatibility" is typically a high priority. This modulates the pace at which networks advance, which is understandable given the enormous investment in current networks. This investment is not just in the existing equipment, but in the operational expertise required to run the network.

As will be frequently emphasized throughout this textbook, network evolution will continue to be driven by economics.

1.9 Focus on Practical Optical Networks

This book examines the design and planning of state-of-the-art optical networks, with an emphasis on the ramifications of optical-bypass technology. It expands on the aspects of optical network design and planning that are relevant in a practical environment, as opposed to taking a more theoretical approach. Much research has focused on idealized optical-bypass systems where all intermediate electronic processing is removed; such networks are often referred to as "all optical." However, in reality, a small amount of intermediate electronic processing may still be required, for example, to improve the quality of the signal or to more efficiently pack data onto a wavelength. This small deviation from the idealized "all-optical" network can have a significant impact on the network design, as is covered in later chapters. Thus, rather than use the term "all-optical network," this book uses the term "*optical-bypass-enabled network*."

Many of the principles covered in the book are equally applicable in metro-core, regional, and backbone networks. However, it will be noted when there are significant differences in the application of the technology to a particular tier.

The foundation of today's optical networks is the network elements, i.e., the major pieces of equipment deployed at a node. Chapter 2 discusses the various network elements in detail, with a focus on functionality and architectural implications. The underlying technology will be touched on only to the level that it affects the network architecture. From the discussion of the network elements, it will be apparent why algorithms play an important role in optical-bypass-enabled networks. Chapters 3–5 focus on the algorithms that are an integral part of operating an efficient and cost-effective optical network. The goal is not to cover all possible optical networking algorithms, but to focus on techniques that have proved useful in practice. Chapter 3, on routing algorithms, is equally applicable to optical-bypass-enabled networks as well as legacy networks. Chapters 4 and 5, on regeneration and wavelength assignment, respectively, are relevant just to optical-bypass-enabled networks.

As mentioned earlier, treating the optical network as another networking layer can be very advantageous. However, networking at the wavelength level can potentially be at odds with operating an efficient network if the wavelengths are not well packed. Chapter 6 looks at efficient grooming of subrate demands, with an emphasis on various grooming architectures and methodologies that are compatible with optical bypass. Given the high cost and large power consumption associated with electronic grooming, strategies for reducing the amount of required grooming are also considered. Chapter 7 discusses protection in the optical layer. Rather than covering the myriad variations of optical protection, the discussion is centered on how protection in the optical layer is best implemented in a network with optical-bypass technology.

Chapters 8 and 9 are both new to the second edition of this text. Both address network flexibility, but from different points of view. Chapter 8 covers flexibility from a traffic perspective; i.e., it specifically addresses dynamic optical networking, including the motivation for supporting this type of traffic. Chapter 9 covers flexibility with regard to the underlying network operation, especially the assignment of

spectrum. Depending on the policies that are adopted, an extended set of associated network design algorithms may need to be developed.

From the perspective of the network operator, perhaps the single most important characteristic of a network is its cost, both capital cost (i.e., equipment cost) and operating cost. Chapter 10 includes a range of economic studies that probe how and when optical networking can improve the economics of a network. These studies can serve as a guideline for network architects planning a network evolution strategy, as well as equipment vendors analyzing the potential benefits of a new technology. The emphasis of the studies in this chapter, as well as the book as a whole, is on real-world networks. Recent research in optical networking is covered as well, to provide an idea of how networks may evolve. Other books on optical networking include Mukherjee [Mukh06], Stern et al. [StEB08], and Ramaswami et al. [RaSS09].

1.10 Reference Networks

Throughout the book, various algorithms and architectural design concepts are presented. To expand beyond a purely abstract discussion, many of these concepts are illustrated using a set of three reference networks. The three networks represent continental-scale backbone networks of varying nodal density. Backbone networks were selected to ensure that regeneration needed to be considered, as the presence of regeneration can have a significant impact on both algorithms and architecture. (As noted earlier, many studies in the literature make the simplifying assumption that networks with optical bypass are purely all-optical, where no regeneration is required. These studies typically consider networks of small geographic extent or they scale down the link sizes of actual continental-scale networks.) In some instances, the discussion is augmented by studies on metro-core networks, when greater fiber connectivity, smaller geographic extent, or pure all-optical networking are important variations to be investigated.

The node locations in the three reference networks are similar to those of existing carriers; however, none of the networks represent an actual carrier network. The largest of the networks, shown in Fig. 1.8, is the baseline continental United States (CONUS) network used in the *Core Optical Networks* (CORONET) program [Sale06]. The network is composed of 75 nodes and 99 links. The average nodal degree of 2.6 is in line with that of most US backbone networks. This topology was specifically designed to be capable of providing a high degree of protection. For example, four completely link-diverse cross-continental paths exist in this network, which is not a common feature in US carrier networks. The second network, shown in Fig. 1.9, is somewhat more representative of current carrier networks. This network has 60 nodes and 77 links, and provides three link-diverse cross-continental paths. Finally, the third network, shown in Fig. 1.10, is representative of a relatively small carrier, with only 30 nodes and 36 links, and just two link-diverse cross-continental paths. Table 1.4 summarizes the topological statistics of the three networks.

Fig. 1.8 Reference network 1, with 75 nodes and 99 links

Fig. 1.9 Reference network 2, with 60 nodes and 77 links

Fig. 1.10 Reference network 3, with 30 nodes and 36 links

Table 1.4 Summary of reference network topologies

	Network 1	Network 2	Network 3
Number of nodes	75	60	30
Number of links	99	77	36
Average nodal degree	2.6	2.6	2.4
Number of nodes with Degree 2	39	34	20
Largest nodal degree	5	5	4
Average link length (km)	400	450	700
Longest link length (km)	1,220	1,200	1,450

1.10.1 Traffic Models

The backbone traffic statistics across telecommunications carriers will clearly vary; however, we can make general observations based on the traffic of several carriers. Traffic tends to be distance dependent, where nodes that are closer exchange more traffic. In many networks, there is an exponentially decaying relationship between traffic and distance. In addition to the distance-dependent traffic, there may be a spike of traffic between some large nodes, independent of the distance between these nodes. For example, in the USA, there is often a relatively large component of

traffic between the East and West coasts. The typical average connection distance in US carriers is roughly 1,600–1,800 km. (Note that this is the routed distance, not the "as the crow flies" distance.)

The traffic set in real carrier networks is far from uniform all-to-all traffic. If one were to designate the largest 20% of the nodes (based on traffic generated) as *Large,* the next largest 30% of the nodes as *Medium,* with the remaining nodes designated as *Small,* then a more realistic traffic breakdown among node pairs is approximately: *Large/Large:* 30%; *Large/Medium:* 30%; *Large/Small:* 15%; *Medium/Medium:* 10%; *Medium/Small:* 10%; *Small/Small:* 5%.

Note that there is not necessarily a correlation between the amount of traffic generated at a node and the degree of a node. There are typically nodes that are strategically located in a network, where these nodes have several incident links but do not generate a lot of traffic. There is also usually a set of nodes that generate a lot of traffic but that have a degree of only two or three.

Again, it needs to be emphasized that while the three reference networks and their respective traffic sets are representative of actual US backbone networks, the statistics vary across carriers.

References

[ABGL01] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow, RSVP-TE: Extensions to RSVP for LSP tunnels, Internet Engineering Task Force, Request for Comments (RFC) 3209, (December 2001)

[AbMA09] S. Abdallah, M. Maier, C. Assi (ed.), Broadband Access Networks: Technologies and Deployments, (Springer, New York, 2009)

[BBFL10] M. Bocci, S. Bryant, D. Frost, L. Levrau, L. Berger, A framework for MPLS in transport networks, draft-ietf-mpls-tp-framework-12, Internet Engineering Task Force, Work In Progress, (May 2010)

[BeRS03] G. Bernstein, B. Rajagopalan, D. Saha, Optical Network Control: Architecture, Protocols, and Standards, (Addison-Wesley Professional, Reading, 2003)

[ChSc07] M. W. Chbat, H.-J. Schmidtke, Falling boundaries from metro to ULH optical transport equipment. Proceedings, Optical Fiber Communication/National Fiber Optic Engineers conference (OFC/NFOEC'07), Paper NTuA3. Anaheim, 25–29 March 2007

[Gora02] W. J. Goralski, SONET/SDH, 3rd edn. (McGraw-Hill, New York, 2002)

[Gril12] E. Griliches, Ciena wavelogic 3 Technology: "Moving the goal posts", ACG Research Technology Impact, (March 2012)

[IEEE09] IEEE, Provider Backbone Bridge Traffic Engineering, IEEE Std 802.1Qay™, (August 2009)

[ITU01] International Telecommunication Union, Architecture of Optical Transport Networks, ITU-T Rec. G.872, (November 2001)

[ITU11] International Telecommunication Union, Architecture of the Multi-Protocol Label Switching Transport Profile layer network, ITU-T Rec. G.8110.1/Y.1370.1, (December 2011)

[ITU12a] International Telecommunication Union, Interfaces for the Optical Transport Network (OTN), ITU-T Rec. G.709/Y.1331, (February 2012)

[ITU12c] International Telecommunication Union, Architecture for the Automatically Switched Optical Network (ASON), ITU-T Rec. G.8080/Y.1304, (February 2012)

[KaKY03] D. Katz, K. Kompella, D. Yeung, Traffic engineering (TE) extensions to OSPF version 2, Internet Engineering Task Force, Request for Comments (RFC) 3630, (September 2003)

[Lam07] C. Lam (ed.), Passive Optical Networks: Principles and Practice, (Academic Press, Burlington, 2007)
[Lin06] C. Lin (ed.), Broadband Optical Access Networks and Fiber-to-the-Home: Systems Technologies and Deployment Strategies, (Wiley, West Sussex, 2006)
[Mann04] E. Mannie (ed.), Generalized Multi-Protocol Label Switching (GMPLS) architecture, Internet Engineering Task Force, Request for Comments (RFC) 3945, (October 2004)
[Mukh06] B. Mukherjee, Optical WDM Networks, (Springer, New York, 2006)
[NBBS09] B. Niven-Jenkins, D. Brungard, M. Betts, N. Sprecher, S. Ueno, Requirements of an MPLS Transport Profile, Internet Engineering Task Force, Request for Comments (RFC) 5654, (September 2009)
[OIF04] Optical Internetworking Forum, User Network Interface (UNI) 1.0 Signaling Specification, Release 2, 27 February 2004
[RaSS09] R. Ramaswami, K. N. Sivarajan, G. Sasaki, Optical Networks: A Practical Perspective, 3rd edn. (Morgan Kaufmann Publishers, San Francisco, 2009)
[Sale98b] A. A. M. Saleh, Short- and long-term options for broadband access to homes and businesses, *Conference on the Internet: Next Generation and Beyond*, Cambridge, 1–2 November 1998.
[Sale06] A. A. M. Saleh, Program Manager, Dynamic multi-terabit core optical networks: Architecture, protocols, control and management (CORONET), Defense Advanced Research Projects Agency (DARPA) Strategic Technology Office (STO), BAA 06-29, Proposer Information Pamphlet (PIP), August 2006
[SDIR05] G. Swallow, J. Drake, H. Ishimatsu, Y. Rekhter, Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) support for the overlay model, Internet Engineering Task Force, Request for Comments (RFC) 4208, (October 2005)
[StEB08] T. E. Stern, G. Ellinas, K. Bala, Multiwavelength Optical Networks: Architectures, Design, and Control, 2nd edn. (Cambridge University Press, Cambridge, 2008)
[Tekt01] Tektronix, SONET telecommunications standard primer, (August 2001), www.tek.com/document/primer/sonet-telecommunications-standard-primer. Accessed 20 Mar 2014
[Telc09] Telcordia Technologies, Synchronous Optical Network (SONET) transport systems: Common generic criteria, GR–253–CORE, Issue 5, October 2009
[WASG96] R. E. Wagner, R. C. Alferness, A. A. M. Saleh, M.S. Goodman, MONET: Multiwavelength optical networking. J. Lightwave Technol. 14(6), 1349–1355, (June 1996)

Chapter 2
Optical Network Elements

2.1 Introduction

The dramatic shift in the architecture of optical networks that began in the 2000 time frame is chiefly due to the development of advanced optical network elements. These elements are based on the premise that the majority of the traffic that enters a node is being routed *through* the node en route to its final destination as opposed to being *destined for* the node. This transiting traffic can potentially remain in the optical domain as it traverses the node rather than be electronically processed. By deploying technology that enables this so-called *optical bypass,* a significant reduction in the amount of required nodal electronic equipment can be realized.

After briefly discussing some basic optical components in Sect. 2.2, we review the traditional network architecture where all traffic entering a node is electronically processed. The fundamental optical network element in this architecture is the *optical terminal,* which is covered in Sect. 2.3. Optical-terminal-based networks are examined in Sect. 2.4. The economic and operational challenges of these legacy networks motivated the development of optical-bypass technology, which is discussed in Sect. 2.5. The two major network elements that are capable of optical bypass are the *optical add/drop multiplexer* (OADM) and the *multi-degree OADM* (OADM-MD); they are described in Sect. 2.6 and Sect. 2.7, respectively. These two elements are more generically referred to in the industry as *reconfigurable OADMs,* or *ROADMs*; in large part, that terminology is adopted here.

There are three principal ROADM design architectures: *broadcast-and-select, route-and-select,* and *wavelength-selective,* all of which are covered in Sect. 2.8. The chief attributes that affect the flexibility, cost, and efficiency of ROADMs are covered in Sect. 2.9, including the *colorless, directionless, contentionless,* and *gridless* properties. A variety of designs are presented to illustrate several possible ROADM operational alternatives.

ROADMs are one type of optical switch. A more complete taxonomy of optical switches is covered in Sect. 2.10. Hierarchical, or multigranular, optical switches, which may be desirable for scalability purposes, are presented in Sect. 2.11.

In a backbone network, bypass-capable network elements must be complemented by extended *optical reach,* which is the distance an optical signal can travel

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4_2,

before it degrades to a level that necessitates it be "cleaned up," or regenerated. The interplay of optical reach and optical-bypass-enabled elements is presented in Sect. 2.12.

Integration of elements or components within a node is a more recent development, motivated by the desire to eliminate individual components, reduce cost, and improve reliability. There is a range of possible integration levels as illustrated by the discussions of Sect. 2.13 (integrated transceivers), Sect. 2.14 (integrated packet-optical platforms), and Sect. 2.15 (photonic integrated circuits, PICs).

Throughout this chapter, it is implicitly assumed that there is one fiber pair per link; e.g., a degree-two node has two incoming and two outgoing fibers. Due to the large capacity of current transmission systems, single-fiber-pair deployments are common. However, the last section of the chapter addresses multi-fiber-pair scenarios. (The related topic of fiber capacity is covered in Chap. 9.)

Throughout this chapter, the focus is on the functionality of the network elements, as opposed to the underlying technology.

2.2 Basic Optical Components

Some of the optical components that come into play throughout this chapter are discussed here. (Several of these components are illustrated in the various optical-terminal architectures shown in Fig. 2.3.) One very simple component is the *wavelength-independent optical splitter,* which is typically referred to as a *passive splitter*. A splitter has one input port and N output ports, where the input optical signal is sent to all of the output ports. Note that if the input is a wavelength-division multiplexing (WDM) signal, then each output signal is also WDM. In many splitter implementations, the input power level is split equally across the N output ports, such that each port receives $1/N$ of the original signal power level. This corresponds to a nominal input-to-output optical loss of $10 \cdot \log_{10} N$, in units of decibels (dB). Roughly speaking, for every doubling of N, the optical loss increases by another 3 dB. It is also possible to design optical splitters where the power is split nonuniformly across the output ports so that some ports suffer lower loss than others.

The inverse device is called a *passive optical coupler* or *combiner*. This has N input ports and one output port, such that all of the inputs are combined into a single output signal. The input signals are usually at different optical frequencies to avoid interference when they are combined. The nominal input-to-output loss of the coupler is the same as that of the splitter.

Another important component is the $1 \times N$ *demultiplexer,* which has one input port and N output ports. In the most common implementation, a WDM signal on the input line is demultiplexed into its constituent wavelengths, with a separate wavelength sent to each output port. The inverse device is an $N \times 1$ *multiplexer,* with N input ports and one output port, where the wavelengths on the input ports are combined to form a WDM signal.

Demultiplexers and multiplexers may be built, for example, using *arrayed waveguide grating* (AWG) technology [Okam98, RaSS09, DoOk06]; such a device is

simply referred to as an "AWG," or a *wavelength grating router* (WGR). For large N, the loss through an AWG is on the order of 4–6 dB. AWGs are generally $M \times N$ devices, where individual wavelengths on the M input ports can be directed only to one specific output port. In the typical AWG $1 \times N$ demultiplexer implementation, the number of output ports and the number of wavelengths in the input WDM signal are the same, such that exactly one wavelength is sent to each output port. Similarly, with an AWG $N \times 1$ multiplexer, where the number of input ports typically equals the number of wavelengths, each input port is capable of directing only one particular wavelength to the output port.

Throughout this chapter, various types of switches are mentioned; it is advantageous to introduce some switch terminology here. There is a broad class of switches known as *optical switches*. Contrary to what the name implies, these switches do not necessarily perform the switching function in the optical domain. Rather, the term "optical switch" is used to indicate a switch where the ports operate on the granularity of a wavelength or a group of wavelengths, as opposed to on finer granularity subrate signals.

Wavelength-selective is a term used to classify devices that are capable of treating each wavelength differently. For example, a $1 \times N$ *wavelength-selective switch* (WSS) can direct any wavelength on the one input port to any of the N output ports [MMMT03; Maro05; StWa10], thereby serving as a demultiplexer. An $N \times 1$ WSS performs a multiplexing function. More generally, an $M \times N$ WSS can direct any wavelength from any of the M input ports to any of the N output ports [FoRN12]. Note that a WSS is capable of directing multiple wavelengths to an output port. However, it is typically *not* possible to multicast a given wavelength from one input port to multiple output ports, nor is it typically possible for multiple input ports to direct the same wavelength to one output port (although in principle, a WSS could support both of these functions, depending on the technology). WSSs play a prominent role in many of the architectures discussed in this chapter.

Micro-electro-mechanical-system (MEMS) technology [WuSF06] is often used to fabricate switches with an optical switch fabric. (The switch fabric is the "guts" of the switch, where the interconnection between the input and output ports is established.) This technology essentially uses tiny movable mirrors to direct light from input ports to output ports. Note that an individual MEMS switching element is not wavelength selective; it simply switches whatever light is on the input port without picking out a particular wavelength. However, when combined with multiplexers and demultiplexers that couple the individual wavelengths of a WDM signal to the ports of the MEMS switch, the combination is wavelength-selective, capable of directing any input wavelength to any output port.

2.3 Optical Terminal

In traditional optical network architectures, optical terminals are deployed at the endpoints of each fiber link. Figure 2.1 illustrates a single optical terminal equipped with several WDM transponders. An optical terminal is typically depicted in fig-

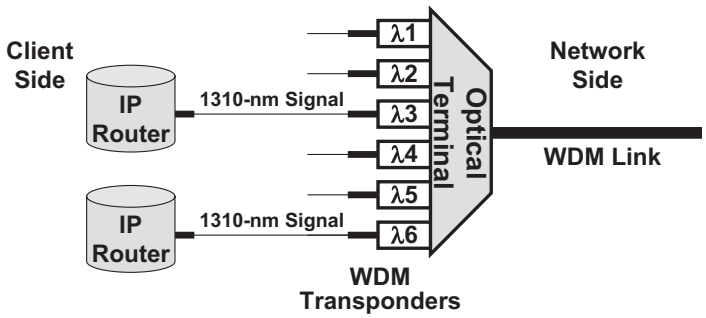


Fig. 2.1 A representation of an optical terminal equipped with wavelength-division multiplexing (*WDM*) transponders

ures as a trapezoid to capture its multiplexing/demultiplexing functionality. In most architectures, there are individual wavelengths on the client side of the terminal and a WDM signal on the network side. Unfortunately, a trapezoid is often used to specifically represent a $1 \times N$ AWG. While an optical terminal can be built using AWG technology, there are other options as well, some of which are discussed in Sect. 2.3.1. *Throughout this book, the trapezoid is used to represent a general optical terminal, or any device performing a multiplexing/demultiplexing function, not necessarily one based on a specific technology.*

In Fig. 2.1, two Internet Protocol (IP) routers are shown on the client side of the optical terminal. Tracing the flow from left to right in the figure, both IP routers transmit a 1,310-nm signal that is received by a WDM transponder. The transponder converts the signal to a WDM-compatible optical frequency, typically in the 1,500-nm range of the spectrum. The optical terminal multiplexes the signals from all of the transponders onto a single network fiber. In general, the transponders plugged into an optical terminal generate different optical frequencies; otherwise, the signals would interfere with each other after being multiplexed together by the terminal. Note that the 1,310-nm signal is sometimes referred to as *gray optics,* to emphasize that the client signals are nominally at the same frequency, in contrast to the different frequencies (or *colored optics*) comprising the WDM signal.

In the reverse direction, a WDM signal is carried by the network fiber into the optical terminal, where it is demultiplexed into its constituent frequencies. Each transponder receives a signal on a particular optical frequency and converts it to a 1,310-nm client-compatible signal.

Recall from Sect. 1.6 that each fiber line shown in Fig. 2.1 actually represents two fibers, corresponding to the two directions of traffic. Also, recall from Fig. 1.7b that the WDM transponder encompasses both a client-side receiver/network-side transmitter in one direction and a network-side receiver/client-side transmitter in the other direction. Similarly, the optical terminal is composed of both a multiplexer and a demultiplexer. Note that it is possible for the network-side signal transmitted by a transponder to be at a different optical frequency than the network-side signal received by the transponder; however, in most scenarios these frequencies are the same.

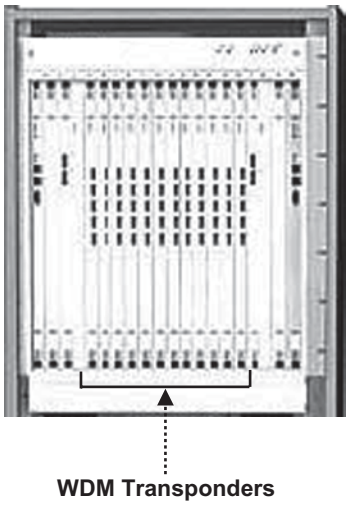


Fig. 2.2 An example of an optical-terminal shelf, with the transponder slots fully populated

2.3.1 Colorless Optical Terminal (Slot Flexibility)

An optical terminal is deployed with equipment shelves in which the transponders are inserted. A prototypical optical-terminal shelf is shown in Fig. 2.2. As depicted, the shelf is fully populated with WDM transponders (i.e., the vertically oriented circuit boards inserted in the slots at the center of the shelf). One figure of merit of an optical terminal is the density of the transponders on a shelf, where higher density is preferred. For example, if a shelf holds up to sixteen 10-Gb/s transponders, the density is 160 Gb/s per shelf. The density is typically determined by properties such as the physical size or power requirements of the transponders (there are industry-wide accepted maxima for the heat dissipation in a fully populated shelf [Telc05a]).

Another desirable feature of an optical terminal is a "pay-as-you-grow" architecture. A fully populated optical terminal may require multiple equipment racks, with multiple shelves per rack, to accommodate all of the transponders. However, ideally the optical terminal can be deployed initially with just a single shelf, and then grow in size as more transponders need to be installed at the site.

The flexibility of the individual slots in the transponder shelves is another important attribute. In the most flexible optical-terminal architecture, any slot can accommodate a transponder of any frequency. Such an architecture is referred to as *colorless*. Clearly, the colorless property simplifies network operations, as a technician can plug a transponder into any available slot. This architecture also maximizes the benefits of tunable transponders, as it allows a transponder to tune to a different frequency without needing to be manually moved to a different slot. Additionally, a colorless optical terminal is typically pay-as-you-grow; i.e., the number of slots deployed needs to be only as large as the number of transponders at the node (subject to the shelf granularity).

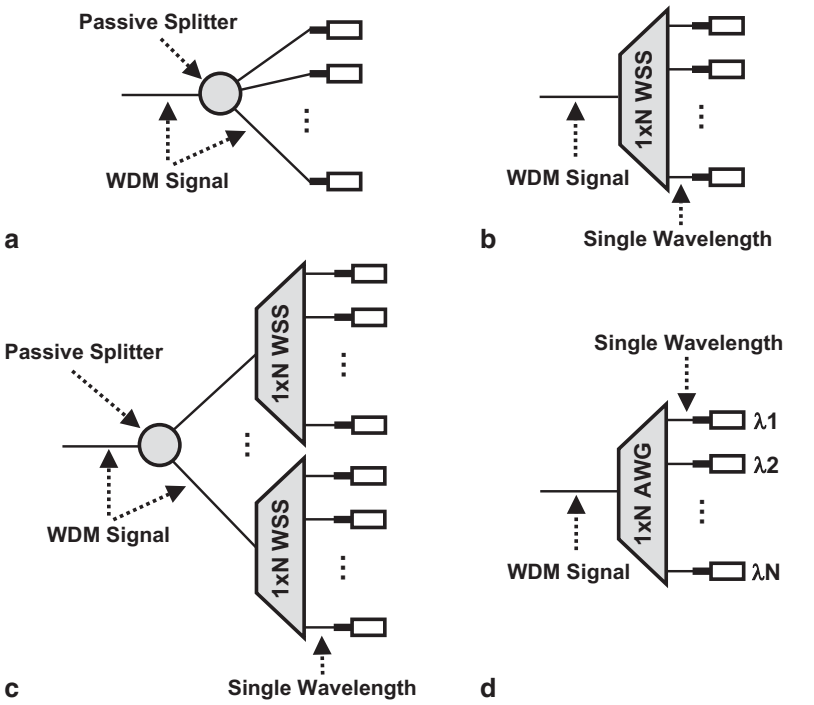


Fig. 2.3 Four optical-terminal architectures, the first three of which are colorless. Only the receive sides of the architectures are shown. **a** The passive splitter architecture has high loss and the transponders must be capable of selecting a particular frequency from the wavelength-division multiplexing (*WDM*) signal. **b** This wavelength-selective switch (*WSS*) architecture limits the number of transponders that can be accommodated to *N*. **c** A WSS tree architecture increases the number of supported transponders, but at an increased loss. **d** The architecture based on the arrayed waveguide grating (*AWG*) is not colorless; a transponder of a given frequency must be inserted in one particular slot

Four optical-terminal architectures are shown in Fig. 2.3, the first three of which are colorless, while the fourth is not. Only the receive sides of the architectures are shown; the transmit sides are similar.

Figure 2.3a depicts a colorless optical terminal based on a passive splitter in the receive direction, and a passive coupler in the transmit direction (not shown). The received WDM signal is passively split, rather than demultiplexed, and directed to each of the transponders. The transponder receiver (on the network side) is equipped with an optical filter (or other frequency-selective technology) to select the desired optical frequency from the WDM signal. For maximum transponder flexibility, this filter should be tunable. In the reverse direction, the signals from the transponders are passively coupled together into a WDM signal; again, for maximum flexibility, the transponders should be equipped with tunable lasers. Because passive splitters and couplers can result in significant optical loss if the number of supported transponders is large, this architecture often requires optical amplifiers to boost the signal level. Additionally, if the outputs of a large number of transmitters are directly

combined in the passive coupler (i.e., without any filtering to clean up the signals), there may be issues resulting from adding all of the spontaneous emission and other noises of the various lasers; the problem is exacerbated with tunable transmitters.

Another colorless optical-terminal architecture, shown in Fig. 2.3b, is based on WSSs. In the receive direction, a $1 \times N$ WSS demultiplexes the signal; it is capable of directing any wavelength from the input WDM signal to any of the N transponders. A second $N \times 1$ WSS (not shown) is used in the reverse direction to multiplex the signals from the transponders. In this architecture, the transponder receiver does not need to have an optical filter because the wavelength selection is carried out in the WSS (i.e., the transponder is capable of receiving whatever optical frequency is directed to its slot). One drawback of the WSS approach is the relatively high cost compared to the other architectures, although the cost difference is shrinking as WSS technology matures. Another drawback is the limited size of the WSS, which limits the number of transponders that can be supported by the terminal. Commercially available WSSs in the 2015 time frame have a maximum size on the order of 1×20, although they continue to increase in size. If more than N transponders need to be installed in the optical terminal, then a "tree" composed of a passive splitter/coupler and multiple WSSs can be deployed, as shown for the receive direction in Fig. 2.3c ([WFJA10]; also see Exercise 2.6).

In contrast to these colorless optical-terminal architectures, there are also *fixed* optical terminals where each slot can accommodate a transponder of only one particular frequency. This type of optical terminal is often implemented using AWG technology, as shown in Fig. 2.3d. A $1 \times N$ AWG demultiplexes the WDM signal in the receive direction; a second $N \times 1$ AWG (not shown) multiplexes the signals from the transponders in the transmit direction. The transponder receiver does not need to have an optical filter. Using current commercially available technology, an AWG can accommodate many more transponders than a WSS (i.e., much larger N). Though relatively cost effective and of low loss, this fixed architecture can lead to inefficient shelf packing and ultimately higher cost in networks where the choice of optical frequencies is very important; i.e., a new shelf may need to be added to accommodate a desired frequency even though the shelves that are already deployed have available slots. The fixed architecture also negates the automated configurability afforded by tunable transponders.

In an intermediary optical-terminal architecture, the WDM spectrum is partitioned into groups, and a particular slot can accommodate transponders only from one group [ChLH06]. This type of terminal can be architected with lower loss and/or cost than a fully colorless design, but has limited configurability.

2.4 Optical-Electrical-Optical (O-E-O) Architecture

2.4.1 O-E-O Architecture at Nodes of Degree-Two

The traditional, non-configurable, optical-terminal-based architecture for a node of degree-two is shown in Fig. 2.4. There are two network links incident on the node,

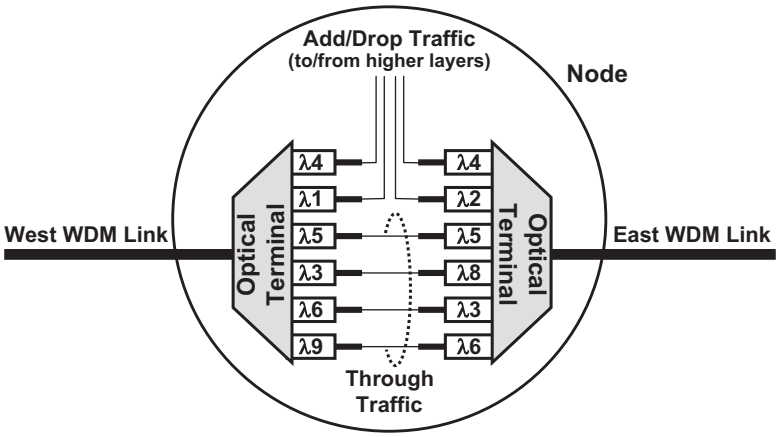


Fig. 2.4 O-E-O architecture at a degree-two node (without automated reconfigurability). Nodal traffic is characterized as either add/drop traffic or through traffic. All traffic entering and exiting the node is processed by a transponder. Note that the through traffic can undergo wavelength conversion, as indicated by interconnected transponders of different wavelengths (e.g., the bottom pair of interconnected transponders converts the signal from λ_6 to λ_9 in the East-to-West direction)

where it is common to refer to the links as the "East" and "West" links (there is not necessarily a correspondence to the actual geography of the node). As shown in the figure, the node is equipped with two optical terminals arranged in a "back-to-back" configuration. The architecture shown does not support automated reconfigurability. Connectivity is provided via a manual patch panel, i.e., a panel where equipment within an office is connected via fiber cables to one side (typically in the back), and where short patch cables are used on the other side (typically in the front) to manually interconnect the equipment as desired. Providing automated reconfigurability is discussed in the next section in the context of higher-degree nodes.

Tracing the path from right to left, the WDM signal enters the East optical terminal from the East link. This WDM signal is demultiplexed into its constituent wavelengths, each of which is sent to a WDM transponder that converts it to a 1,310 nm optical signal. (Recall that 1,310 nm is the typical wavelength of the client-side optical signal.) At this point, it is important to distinguish two types of traffic with respect to the node. For one type of traffic, the node serves as the exit point from the optical layer. This traffic "drops"[1] from the optical layer and is sent to a higher layer (the higher layers, e.g., IP, are the clients of the optical layer). The other type of traffic is transiting the node en route to its final destination. After this transiting traffic has been converted to a 1,310 nm optical signal by its associated transponder, it is sent to a second transponder located on the West optical terminal. This transponder converts it back into a WDM-compatible signal, which is then multiplexed by the West optical terminal and sent out on the West link. There are also transponders on

[1] While "drop" often has a negative connotation in telecommunication networks (e.g., dropped packets, dropped calls), its usage here simply means a signal is exiting from the optical layer.

the West terminal for traffic that is being "added" to the optical layer, from higher layers, that needs to be routed on the West link.

In the left to right direction of the figure, the operation is similar. Some of the traffic from the West link drops from the optical layer and some is sent out on the East link. Additionally, there are transponders on the East terminal for traffic that is added to the optical layer at this node that needs to be routed on the East link.

The traffic that is being added to or dropped from the optical layer at this node is termed *add/drop* traffic; the traffic that is transiting the node is called *through* traffic. Regardless of the traffic type, note that all of the traffic entering and exiting the node is processed by a WDM transponder. In the course of converting between a WDM-compatible optical signal and a client optical signal, the transponder processes the signal in the electrical domain. Thus, all traffic enters the node in the optical domain, is converted to the electrical domain, and is returned to the optical domain. This architecture, where all traffic undergoes optical-electrical-optical (O-E-O) conversion, is referred to as the *O-E-O architecture*.

2.4.2 O-E-O Architecture at Nodes of Degree-Three or Higher

The O-E-O architecture readily extends to a node of degree greater than two. In general, a degree-*N* node will have *N* optical terminals. Figure 2.5 depicts a degree-three node equipped with three optical terminals, with the third link referred to as the "South" link. The particular architecture shown does not support automated reconfigurability.

As with the degree-two node, all of the traffic entering a node, whether add/drop or through traffic, is processed by a transponder. The additional wrinkle with higher-degree nodes is that the through traffic has multiple possible path directions. For example, in the figure, traffic entering from the East could be directed to the West or to the South; the path is set by interconnecting a transponder on the East optical terminal to a transponder on the West or the South optical terminal, respectively. In many real-world implementations, the transponders are interconnected using a manual patch panel. Modifying the through path of a connection requires that a technician manually rearrange the patch panel, a process that is not conducive to rapid reconfiguration and is subject to operator error.

The reconfiguration process can be automated through the addition of an optical switch, as shown in Fig. 2.6. (The traffic patterns shown in Fig. 2.5 and Fig. 2.6 are not the same.) Each transponder at the node feeds into a switch port, and the switch is configured as needed to interconnect two transponders to create a through path. Additionally, the add/drop signals are fed into ports on the switch, so that they can be directed to/from transponders on any of the optical terminals. Furthermore, the switch allows any transponder to be flexibly used for either add/drop or through traffic, depending on how the switch is configured. Note that the degree-two architecture of Fig. 2.4 could benefit from a switch as well with respect to these latter

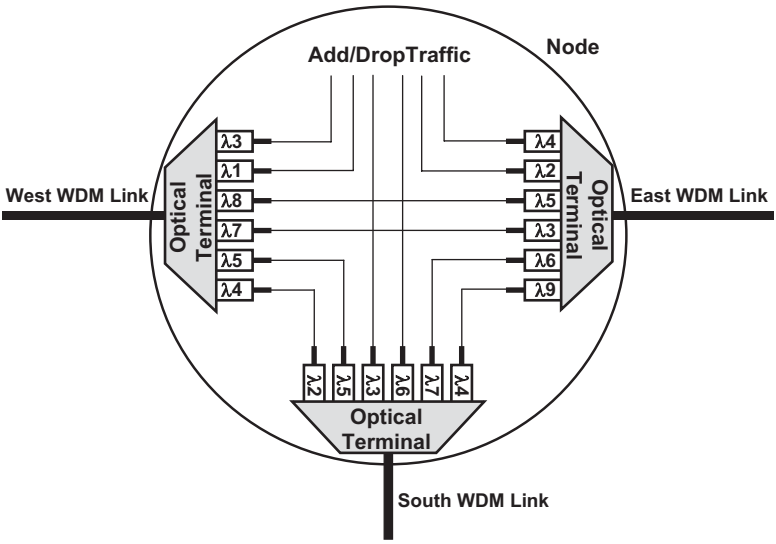


Fig. 2.5 O-E-O architecture at a degree-three node (without automated reconfigurability). There are three possible directions through the node. The path of a transiting connection is set by interconnecting a pair of transponders on the associated optical terminals

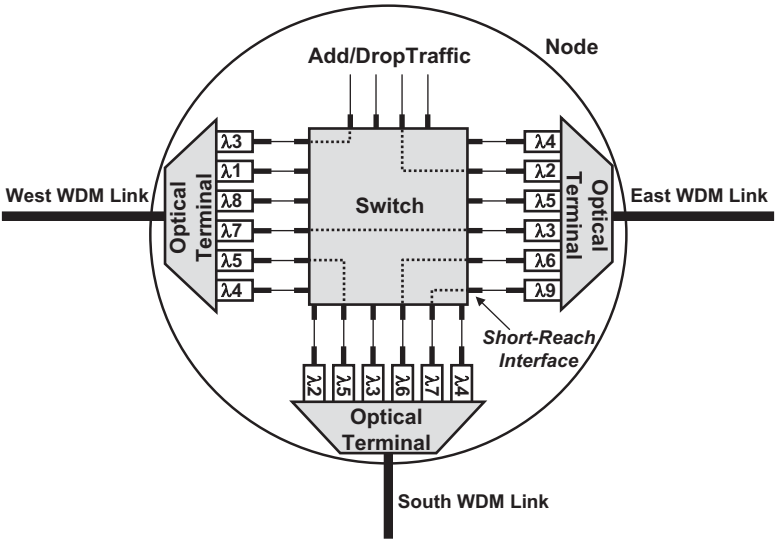


Fig. 2.6 A switch is used to automate node reconfigurability. The particular switch shown has an electronic switch fabric and is equipped with short-reach interfaces on all of its ports

two applications, i.e., directing add/drop traffic to any optical terminal and flexibly using a transponder for either add/drop or through traffic.

While deploying a switch enhances the network flexibility and reduces operational complexity due to less required manual intervention, the downside is the additional equipment cost. There are generally two types of optical switches that are used in such applications. Most commonly, a switch with an electronic switching fabric is used, where each port is equipped with a short-reach interface to convert the 1,310-nm optical signal from the transponder to an electrical signal (this is the option shown in Fig. 2.6). A second option is to use a switch with an optical switch fabric such as a MEMS-based switch. This technology can directly switch an optical signal (which in this case is a 1,310-nm signal), thereby obviating the need for short-reach interfaces on the switch ports (although it may require that the transponders be equipped with a special interface that is tolerable to the optical loss through the switch). As MEMS technology comes down in price, this type of switch may be the more cost-effective option. Electronic switches and MEMS switches are revisited in Sects. 2.10.1 and 2.10.2, respectively.

2.4.3 Advantages of the O-E-O Architecture

The fact that the O-E-O architecture processes all traffic entering the node in the electrical domain does offer some advantages. First, converting the signal to the electrical domain and back to the optical domain "cleans up" the signal. Optical signals undergo degradation as they are transmitted along a fiber. The O-E-O process reamplifies, reshapes, and retimes the signal, a process that is known as *3R-regeneration*.

Second, it readily allows for performance monitoring of the signal. For example, if Synchronous Optical Network (SONET) framing is being used, then a SONET-compatible transponder can examine the overhead bytes using electronic processing to determine if there are errors in the signal. Because this process can be done at every node, it is typically straightforward to determine the location of a failure.

Third, the O-E-O architecture is very amenable to a multivendor architecture because all communication within a node is via a standard 1,310-nm optical signal. Whereas the WDM transmission characteristics on a link may be proprietary to a vendor, the intra-nodal communication adheres to a well-defined standard. This allows the transmission systems on the links entering a node to be supplied by different vendors. Furthermore, the vendors of the switch (if present) and the transmission system can be different as well.

The O-E-O architecture also affords flexibility when assigning wavelengths to the traffic that passes through a node. Here, the term wavelength is used to indicate a particular frequency in the WDM spectrum. As noted above, the through traffic enters the node on one transponder and exits the node on a second transponder. These two transponders communicate via the 1,310 nm optical signal; there is no requirement that the WDM-compatible wavelengths of these two transponders be the same. This is illustrated in Figs. 2.4–2.6, where the wavelengths of two interconnected transponders are not necessarily the same. This process accomplishes what

is known as *wavelength conversion,* where the signal enters and exits the node on two different wavelengths. There is complete freedom in selecting the wavelengths of the two transponders, subject to the constraint that the same wavelength cannot be used more than once on any given fiber. The most important implication is that the choice of wavelength is local to a particular link; the wavelength assignment on one link does not affect that on any other link.

2.4.4 Disadvantages of the O-E-O Architecture

While the O-E-O architecture does have advantages, terminating every wavelength entering every node on a transponder poses many challenges in scaling this architecture to larger networks. For example, with 100 wavelengths per fiber, this architecture potentially requires several hundred transponders at a node. The first barrier is the cost of all this equipment, although transponder costs do continue to decrease (see Sect. 2.15). Second, there are concerns regarding the physical space at the sites that house the equipment. More transponders translate to more shelves of equipment, and space is already at a premium in many carrier offices. Furthermore, providing the power for all of this electronics and dissipating the heat created by them is another operational challenge that will only worsen as networks continue to grow.

Another barrier to network evolution is that electronics are often tied to a specific technology. For example, a short-reach interface that supports a 10-Gb/s signal typically does not also support a 40-Gb/s signal. Thus, if a carrier upgrades its network from a 10-Gb/s line rate to a 40-Gb/s line rate, a great deal of equipment, including transponders and any electronic switches, needs to be replaced.

Provisioning a connection can be cumbersome in the O-E-O architecture. A technician may need to visit every node along the path of a connection to install the required transponders. Even if transponders are pre-deployed in a node, a visit to the node may be required, for example, to manually interconnect two transponders via a patch panel. As noted above, manual intervention can be avoided through the use of a switch; however, as the node size increases, the switch size must grow accordingly. A switch, especially one based on electronics, will have its own scalability issues related to cost, size, power, and heat dissipation.

Finally, all of the equipment that must be deployed along a given connection is potentially a reliability issue. The connection can fail, for example, if any of the transponders along its path fails.

2.5 Optical Bypass

The scalability challenge of the O-E-O architecture was a major impetus to develop alternative technology where much of electronics could be eliminated. Given that through traffic is converted from a WDM-compatible signal to a 1,310 nm signal

only to be immediately converted back again to a WDM-compatible signal, removing the need for transponders for the through traffic was a natural avenue to pursue.

This gave rise to network elements, e.g., the OADM, that allow the through traffic to remain in the optical domain as it transits the node; transponders are needed only for the add/drop traffic. The through traffic is said to *optically bypass* the node. Studies have shown that in typical carrier networks, on average, over 50% of the traffic entering a node is through traffic; thus, a significant amount of transponders can be eliminated with optical bypass.

Networks equipped with elements that support optical bypass are referred to here as *optical-bypass-enabled* networks. Other terms used in the literature to describe this type of network are "all-optical" or "transparent." However, given that O-E-O-based regenerators are typically not entirely eliminated from all end-to-end paths (see Sect. 2.12) and that networks supporting protocol-and-format transparency have not materialized in a major way (see Sect. 1.3), these terms are not completely accurate. Another term that is sometimes used to indicate a network that supports optical bypass while still requiring some electronic regeneration is "translucent" [RFDH99; ShTu07].

2.5.1 Advantages of Optical Bypass

The advantages and disadvantages of the optical-bypass-enabled architecture are diametrically opposite to those that were discussed for the O-E-O architecture. We begin with the advantages. First, as the network traffic level increases, optical-bypass technology is potentially more scalable in cost, space, power, and heat dissipation because much of the electronics is eliminated. Second, optics is more agnostic to the wavelength line rate as compared to electronics. For example, a network element that provides optical bypass typically can support line rates over the range of 2.5–100 Gb/s (assuming the system wavelength spacing and signal spectrum are compatible with the element). Third, provisioning a connection is operationally simpler. In many scenarios, a new connection requires that a technician visit just the source and destination nodes to install the add/drop transponders. Fourth, the elimination of much of the electronics also improves the overall reliability. Even if the optical-bypass equipment has a higher failure rate than optical terminals, the removal of most of the transponders in the signal path typically leads to an overall lower failure rate for the connection [MaLe03].

2.5.2 Disadvantages of Optical Bypass

While removing many of the transponders from a connection path provides numerous cost and operational advantages, it does eliminate the functions provided by those transponders. First, the optical signal of the through traffic is not regenerated, thereby requiring extended optical reach, as described in Sect. 2.12. Second, remov-

ing the transponders from some or all of the intermediate nodes of a path also eliminates the node-by-node error-checking functionality they provided. In the absence of electronic performance monitoring, optical monitoring techniques are needed, as discussed in Chap. 7. Third, it may be challenging to support a multivendor environment in an optical-bypass-enabled network because not all intra-nodal traffic is converted to a standard optical signal as it is in an O-E-O network. Standards for extended-reach WDM transmission have not been defined, such that interoperability between multiple vendors is not guaranteed. For example, the transmission system of one vendor may not be compatible with the optical-bypass-enabled network elements of another vendor. Furthermore, without standard performance guidelines, it may be difficult to isolate which vendor's equipment is malfunctioning under a failure condition. Thus, in optical-bypass-enabled networks, it is common for a single vendor to provide both the transmission system and the optical networking elements. (The notion of "islands of transparency," where designated vendors operate within non-overlapping subsets of the network, is discussed in Chap. 4.)

Finally, and most importantly, a transiting connection enters and exits the node on the same wavelength; there is not the same opportunity for wavelength conversion as with the O-E-O architecture. Given that two signals on the same fiber cannot be assigned the same wavelength, this implies that the assignment of wavelengths on one link potentially affects the assignment of wavelengths on other links in the network. This *wavelength continuity constraint* is the major reason why advanced algorithms are required to efficiently operate a network based on optical-bypass technology. Such algorithms are covered in Chaps. 3, 4, and 5.

Note that early work on optical-bypass-enabled networks considered the possibility of changing the wavelength of a connection while in the optical domain, thereby eliminating the wavelength continuity constraint. However, even after much research, all-optical wavelength converters are largely impractical, due to their high cost and their compatibility with only simple, spectrally inefficient transmission formats. Thus, the wavelength continuity constraint is likely to remain a relevant factor in optical-bypass-enabled networks.

2.6 OADMs/ROADMs

The first commercial network element to support optical bypass was the degree-two OADM, shown in Fig. 2.7. In comparison to its O-E-O analog that required transponders for all traffic entering/exiting the node (Fig. 2.4), the OADM requires transponders only for the add/drop traffic. Traffic that is transiting the node can remain in the optical domain between the East and West links.

OADMs have been commercially available since the mid-1990s, although significant deployment did not start until after 2000. The name of the element derives from a SONET/Synchronous Digital Hierarchy (SDH) add/drop multiplexer (ADM), which is capable of adding/dropping lower-rate SONET/SDH signals to/from a higher-rate signal without terminating the entire higher-rate signal. Similar-

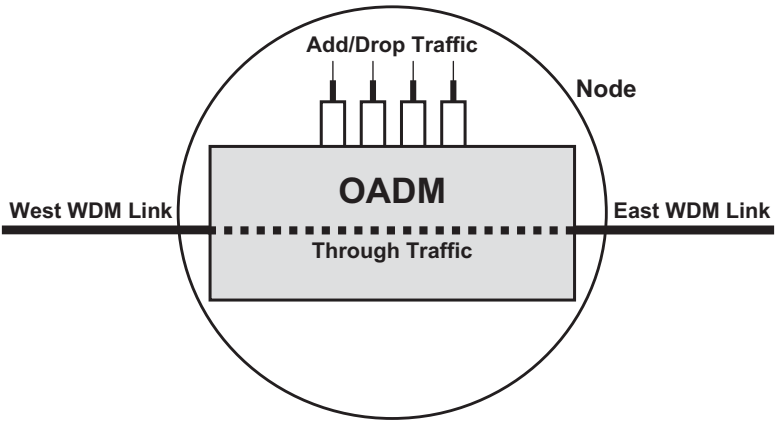


Fig. 2.7 Optical add/drop multiplexer **(***OADM***)** at a degree-two node. Transponders are required only for the add/drop traffic. The through traffic remains in the optical domain as it transits the node. (Adapted from Simmons [Simm05], © 2005 IEEE)

ly, the OADM adds/drops wavelengths to/from a fiber without having to electronically terminate all of the wavelengths comprising the WDM signal.

While the OADM itself costs more than two optical terminals, the reduction in transponders results in an overall lower nodal cost, assuming the level of traffic is high enough. The economics of optical bypass are explored further in Chap. 10.

Note that this textbook adopts the convention that the transponders are *not* considered to be part of the OADM or optical terminal; i.e., the transponders are inserted into the OADM, as opposed to being part of the network element itself. This viewpoint is common but not universal; some use the term OADM to include the transponders as well.

2.6.1 OADM Reconfigurability

One of the most important properties of an OADM is its degree of reconfigurability. The earliest commercial OADMs were not configurable. Carriers needed to specify up front which particular wavelengths would be added/dropped at a particular node, with all remaining wavelengths transiting the node. Once installed, the OADM was fixed in that configuration. Clearly, this rigidity limits the ability of the network to adapt to changing traffic patterns. Such OADMs are sometimes referred to as fixed OADMs, or FOADMs.

Today, however, most OADMs are configurable. This implies that any wavelength can be added/dropped at any node, and that the choice of add/drop wavelengths can be readily changed without impacting any of the other connections terminating at or transiting the node. Furthermore, it is highly desirable that the OADM be remotely configurable through software as opposed to requiring manual intervention. A fully configurable OADM is typically called a reconfigurable OADM, or ROADM

(pronounced rōd′-um). Due to the popularity of the term "ROADM" in the telecommunications industry, it will be used in the remainder of the book, unless the discussion is specifically addressing a non-configurable OADM.

One limitation of some ROADMs is that while they may be fully configurable, the amount of add/drop cannot exceed a given threshold [NYHS12]. A typical threshold in such ROADMs is a maximum of 50 % of the wavelengths supportable on a fiber can be added/dropped (e.g., a maximum of 40 add/drops from a fiber that can support 80 wavelengths). For many nodes in a network, this is sufficient flexibility. However, there is typically a small subset of nodes that would ideally add/drop a higher percentage than this. Such nodes could either route some of their traffic over alternative (less optimal) paths to avoid a fiber that has reached its add/drop limit or they could employ the traditional O-E-O architecture.

It may not seem worthwhile to build ROADMs that are capable of more than 50 % drop because the amount of transponders that are eliminated by taking advantage of optical bypass is relatively small. However, even if deploying a ROADM with more than 50 % drop does not save cost as compared to the O-E-O architecture, it is still preferable because a ROADM provides more agility than two optical terminals and a patch panel. Moreover, ROADMs that support up to 100 % add/drop allow the flexibility of using a ROADM at any node without having to estimate the maximum percentage drop that will ever occur at the node.

2.7 Multi-degree ROADMs

The ROADM network element was conceived to provide optical bypass at nodes of degree two. While all nodes in a ring architecture have degree two, a significant number of nodes in interconnected-ring and mesh topologies have a degree greater than two. Table 2.1 shows the percentage of nodes of a given degree averaged over several typical US mesh backbone networks. Tables 2.2 and 2.3 show the nodal-degree percentage averaged over several US metro-core networks for interconnected-ring and mesh topologies, respectively.

This raises the question of what type of equipment to deploy at nodes of degree three or higher. One thought is to continue using the optical-terminal-based O-E-O architecture at these nodes, while using ROADMs at degree-two nodes. This is sometimes referred to as the "O-E-O-at-the-hubs" architecture, where nodes of degree three or more are considered hubs. As all traffic must be regenerated at the hubs, it implies that electronic performance monitoring can be performed at the junction sites of the network, which may be advantageous for localizing faults. However, it also implies that the scalability issues imposed by O-E-O technology will still exist at a large percentage of the nodes.

Another option is to deploy degree-two ROADMs, possibly in conjunction with an optical terminal, at the hubs. Figure 2.8a depicts a degree-three node equipped with one ROADM and one optical terminal. Optical bypass is possible only for traffic transiting between the East and West links. Traffic transiting between the

Table 2.1 Nodal degree percentage averaged over several typical US backbone networks

Nodal degree	Percentage (%)
2	55
3	35
4	7
5	2
6	1

Table 2.2 Nodal degree percentage averaged over several typical US metro-core networks with interconnected-ring topologies

Nodal degree	Percentage (%)
2	85
4	9
6	4
8	2

Table 2.3 Nodal degree percentage averaged over several typical US metro-core networks with mesh topologies

Nodal degree	Percentage (%)	Nodal degree	Percentage (%)
2	30	6	10
3	25	7	3
4	15	8	2
5	15		

South and East links or the South and West links must undergo O-E-O conversion via transponders, as shown in the figure. Figure 2.8b depicts a degree-four node equipped with two ROADMs. Optical bypass is supported between the East and South links and between the North and West links, but not between any other link pairs. The design strategy with these quasi-optical-bypass architectures is to deploy the ROADM(s) in the direction where the most transiting traffic is expected. However, if the actual traffic turns out to be very different from the forecast, then there may be an unexpectedly large amount of required transponders; i.e., the architectures of Fig. 2.8 are not "forecast-tolerant."

With either of the above strategies, there is potentially still a large amount of electronics needed for transiting traffic. Another alternative is to deploy the network element known as the *multi-degree ROADM* (ROADM-MD), which extends the functionality of a ROADM to higher-degree nodes. A degree-three ROADM-MD is shown in Fig. 2.9. With a ROADM-MD, optical bypass is supported in *all* directions through the node to maximize the amount of transponders that can be eliminated. As with degree-two ROADMs, transponders are needed only for the add/drop traffic.

With the combination of the ROADM and the ROADM-MD, optical bypass can be provided in a network of arbitrary topology (subject to the maximum degree of the ROADM-MD). For example, in Fig. 2.10a, a degree-six ROADM-MD deployed in the node at the junction of the three rings allows traffic to pass all-optically from one ring to another; the remainder of the nodes have a ROADM. In the arbitrary mesh of Fig. 2.10b, a combination of ROADMs, degree-three ROADM-MDs, and degree-four ROADM-MDs is deployed to provide optical bypass in any direction through any node.

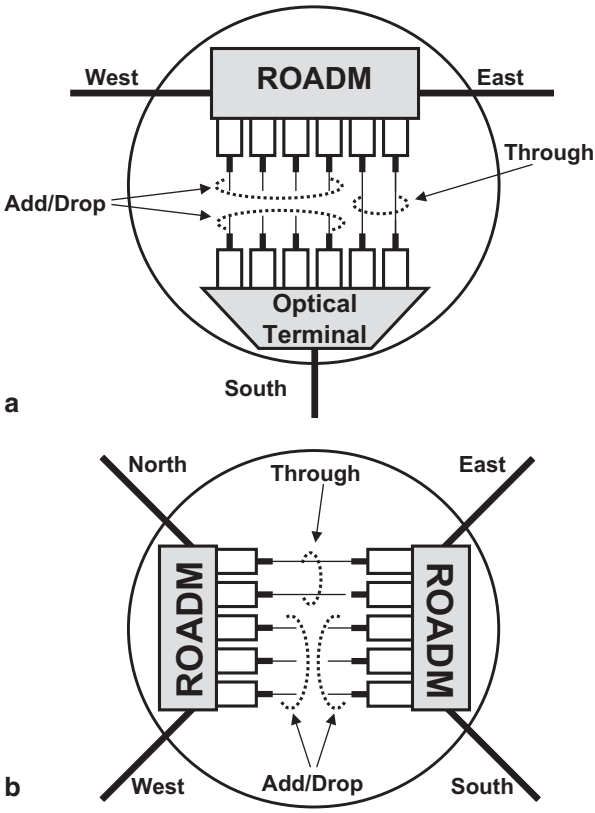


Fig. 2.8 **a** Degree-three node with one reconfigurable optical add/drop multiplexer (*ROADM*) and one optical terminal. **b** Degree-four node with two ROADMs. In these quasi-optical-bypass architectures, some of the transponders are used for transiting traffic that crosses two different network elements at the node

2.7.1 Optical Terminal to ROADM to ROADM-MD Upgrade Path

ROADM-MDs can be part of a graceful network growth scenario. Carriers may choose to roll out optical-bypass technology in stages in the network, where initially they deploy optical-bypass elements on just a few links to gradually grow their network. Consider deploying optical-bypass technology in just the small portion of the network shown in Fig. 2.11a. A ROADM would be deployed at Node B to allow bypass, with optical terminals deployed at Nodes A and C. As a next step, assume that the carrier wishes to extend optical bypass to the ring shown in Fig. 2.11b. Ideally, the optical terminals at Nodes A and C are in-service upgradeable (i.e., existing traffic is not affected) to ROADMs, so that all five nodes in the ring have a ROADM. In the next phase, shown in Fig. 2.11c, a link is added between Nodes A and E to enhance network connectivity. It is desirable that the ROADMs at Nodes A and E be in-service upgradeable to a degree-three ROADM-MD. This element upgrade path, from optical terminal to ROADM to ROADM-MD, is desirable for network growth, and is supported by most commercial offerings. Furthermore, upgrading to higher-degree ROADM-MDs is typically possible, up to the limit of the technology.

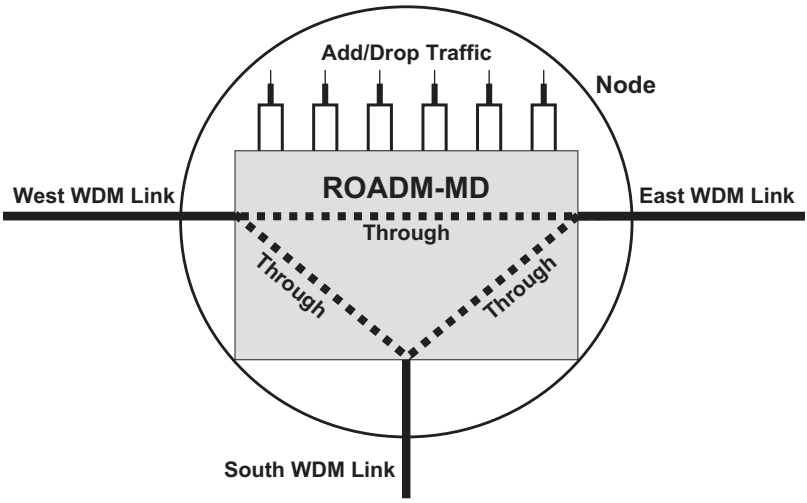


Fig. 2.9 Degree-three multi-degree reconfigurable optical add/drop multiplexer (*ROADM-MD*). Optical bypass is possible in all three directions through the node. Transponders are needed only for the add/drop traffic. (Adapted from Simmons [Simm05], © 2005 IEEE)

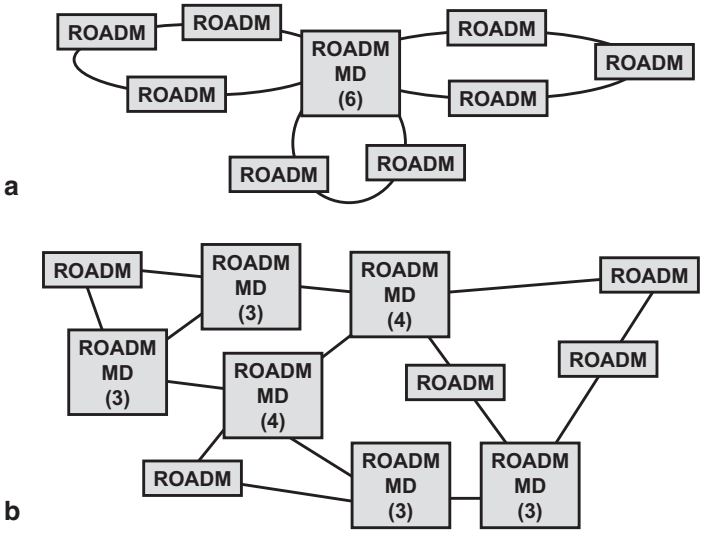


Fig. 2.10 **a** A degree-six multi-degree reconfigurable optical add/drop multiplexer (*ROADM-MD*) is deployed at the junction site of three rings, allowing traffic to transit all-optically between rings. The remaining nodes have reconfigurable optical add/drop multiplexers (*ROADMs*). **b** In this arbitrary mesh topology, a combination of ROADMs, degree-three ROADM-MDs, and degree-four ROADM-MDs is deployed according to the nodal degree. Optical bypass is supported in all directions through any node

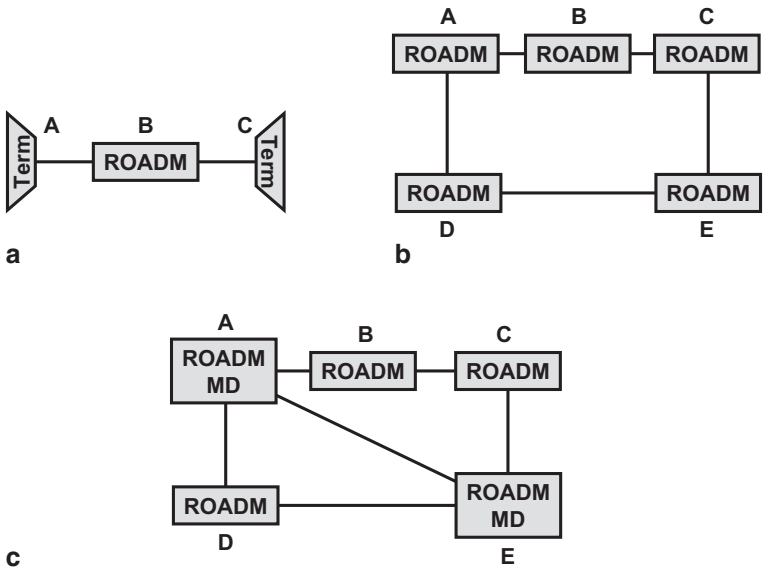


Fig. 2.11 In this network evolution, Node A is equipped with an optical terminal in (**a**), a reconfigurable optical add/drop multiplexer (*ROADM*) in (**b**), and a degree-three multi-degree reconfigurable optical add/drop multiplexer (*ROADM-MD*) in (**c**). Ideally, these upgrades are performed in-service, without affecting any existing traffic at the node

2.8 ROADM Architectures

In this section, three common ROADM architectures are presented and compared at a high level: *broadcast-and-select, route-and-select,* and *wavelength-selective*. The three architectures can be used for either a ROADM or a ROADM-MD. To simplify the text, however, they are referred to as ROADM architectures. For purposes of illustrating the architectures, degree-three ROADM-MDs are shown. It should be readily apparent how to modify the architectures for nodes with fewer or greater degrees.

There are numerous variations of the broadcast-and-select and route-and-select architectures, several of which are presented in Sect. 2.9 in order to illustrate various ROADM/ROADM-MD attributes.

An overview of the underlying ROADM technology can be found in Colbourne and Collings [CoCo11].

2.8.1 Broadcast-and-Select Architecture

Broadcast-and-select is a prevalent ROADM architecture as it is suitable for optically bypassing several consecutive nodes [BSAL02]. One common broadcast-and-select implementation, based on a $1 \times N$ WSS, is shown in Fig. 2.12, for a degree-

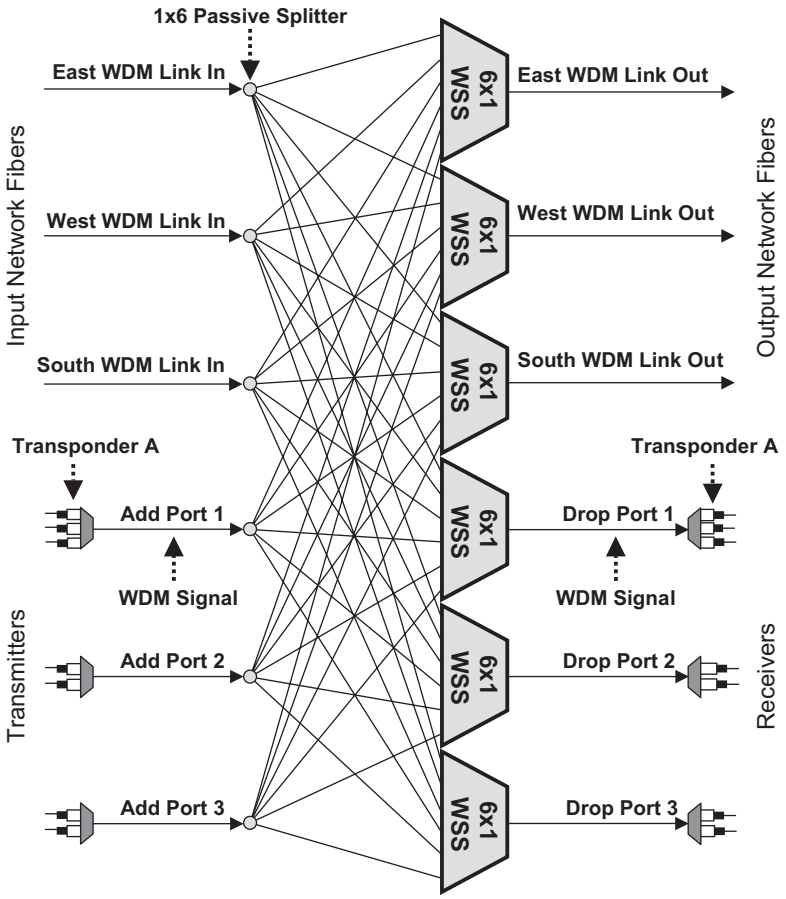


Fig. 2.12 One example of a broadcast-and-select ROADM architecture, with the number of add/drop ports equal to the number of network fibers. The input and output fibers are explicitly shown. The two transponders labeled *Transponder A* are the network transmit side and network receive side of the same transponder

three node [MMMT03]. Due to the architectural asymmetry in the input and output directions, the two directions are explicitly shown in the figure; i.e., input fibers are on the left and output fibers are on the right. Furthermore, the two directions of the transponders are explicitly shown; i.e., the network-side transmitters are on the left and the network-side receivers are on the right. The two transponders labeled *Transponder A* are actually two halves of the same transponder.

Additionally, the figure distinguishes between fibers corresponding to links at the node (i.e., input and output network fibers) and fibers used for add/drop traffic at the node (i.e., add and drop ports). Furthermore, in this figure, the number of add/drop ports is equal to the number of network fibers; however, this is not always the case as explored in Sect. 2.9.5.

Consider an incoming wavelength on a network fiber or an add port. The signal enters a 1×6 passive splitter, which directs the signal to each of the six 6×1 WSSs on the right-hand side. (Note that the WSS is the same component that has been discussed in Sect. 2.3.1 with regard to colorless optical terminals, although typically with fewer ports.) If, for example, λ_1 is present on each of the three input network fibers and on each of the three add ports, then each WSS receives six copies of λ_1, one on each of the WSS input ports. Each WSS is configured to allow at most one of the λ_1 wavelengths to pass through to its output port, thereby avoiding collisions on the output fiber or drop port.

For example, if it is desired that λ_1 optically bypass the node from the East fiber to the South fiber, then the WSS corresponding to the South fiber is configured to direct λ_1 from its first input port to its output port. If, instead, it is desired that λ_1 be transmitted by Transponder A to the South fiber, then the WSS corresponding to the South fiber is configured to direct λ_1 from its fourth input port to its output port. Finally, if it is desired that λ_1 drop from the East fiber to Drop Port 3, then the WSS corresponding to Drop Port 3 is configured to direct λ_1 from its first input port to its output port.

Clearly, the passive splitter is accomplishing the "broadcast" whereas the WSS is accomplishing the "select." While the passive splitter broadcasts each incoming signal, it does not typically split the power evenly among the network ports and the drop ports. Given that it is more important to maintain the signal integrity of the through traffic to allow the traffic to continue to the next node in its path, only a small portion of the incoming signal power, say 10%, is directed to each drop port, with the remainder directed to the network fibers. Additionally, there is typically amplification in a node, to help mitigate the splitting loss.

The broadcast-and-select architecture shown in Fig. 2.12 readily supports multicast in the optical domain, with multiple configurations possible. First, a signal entering the node from an input network fiber can be sent to multiple output network fibers; e.g., a signal can be multicast from the East link to both the West and South links. Second, a signal entering the node from an input network fiber can be sent to both a drop port and one or more output network fibers; e.g., a signal can be multicast from the East link to both the West link and a transponder on Drop Port 1. Such a function is known as *drop-and-continue*. Third, a signal from an add port can be directed to multiple output network fibers; e.g., a signal can be multicast from a transponder on Add Port 1 to both the East and South links. This capability can be very useful for providing protection against failures (see Chap. 7). Finally, a signal on an input network fiber can be directed to multiple transponders. (Depending on the technology used for the drop ports, the transponders receiving the multicast signal may need to be located on different drop ports; see Exercise 2.5.) For example, a signal on the East link may be sent to a transponder on Drop Port 1 and a transponder on Drop Port 3.

The architecture shown in the figure allows for connectivity that may not be required. For example, a signal on the East input fiber can be directed to the East output fiber. Or, a signal may be launched on an add port and be immediately directed to a drop port. While this connectivity may not appear to be beneficial, these

types of loopback may be useful for testing purposes. Additionally, there may be instances where traffic is routed along a path that loops back on itself due to equipment limitations or a link failure. If such loopback is not desired, then the six 6×1 WSSs can be replaced by six 5×1 WSSs.

Note that each add/drop port terminates in a "trapezoid." This represents optical-terminal-like equipment that is used to house the transponders and multiplex/demultiplex the add/drop wavelengths. The add/drop ports in this architecture are *multiwavelength* ports; i.e., *the add/drop fibers carry a WDM signal.* The ramifications of this are further discussed in Sect. 2.9.5.

2.8.2 Route-and-Select Architecture

The *route-and-select* architecture, illustrated in Fig. 2.13, is similar to that of broadcast-and-select. The chief difference is that the passive splitters at the inputs of Fig. 2.12 are replaced by WSSs in Fig. 2.13. This alternative architecture is motivated by the desire to eliminate the splitter loss, and reduce noise and crosstalk, to enable optical bypass of more nodes (see Sect. 2.9.1). The disadvantage is that the cost of the network element almost doubles. Additionally, multicast is not supported by the architecture of Fig. 2.13 (as indicated in Sect. 2.2, WSSs typically are not capable of multicasting a signal).

In a *hybrid* route-and-select/broadcast-and-select architecture, the 1×6 WSSs on the *add ports* of Fig. 2.13 are replaced by 1×6 passive splitters, similar to the add-port design of Fig. 2.12. This is a lower cost design than the pure route-and-select architecture of Fig. 2.13. In comparison with Fig. 2.12, it incurs the penalty of passing through a splitter just once, at the source, as opposed to at every optically bypassed node in the path. Furthermore, it supports limited multicast; a signal from a transponder on an add port can be sent to multiple output network fibers, which is advantageous for protection.

2.8.3 Wavelength-Selective Architecture

A third ROADM architecture is the *wavelength-selective* architecture, an implementation of which is shown for a degree-three node in Fig. 2.14. The core of this architecture is the optical switch, where the fabric of this switch must be optical, so that there is no need for O-E-O conversion for traffic transiting the node. The input and output of this architecture are symmetric; thus, we return to the convention of using a single line to represent both incoming and outgoing fibers.

Tracing an incoming signal from the East, the WDM signal from the East network fiber is demultiplexed into its constituent wavelengths, each of which is fed into a port on the optical switch. The optical switch is configured to direct the drop wavelengths to transponders. The remaining wavelengths (i.e., the through traffic) are directed to the multiplexers on either the West or South sides. The wavelengths

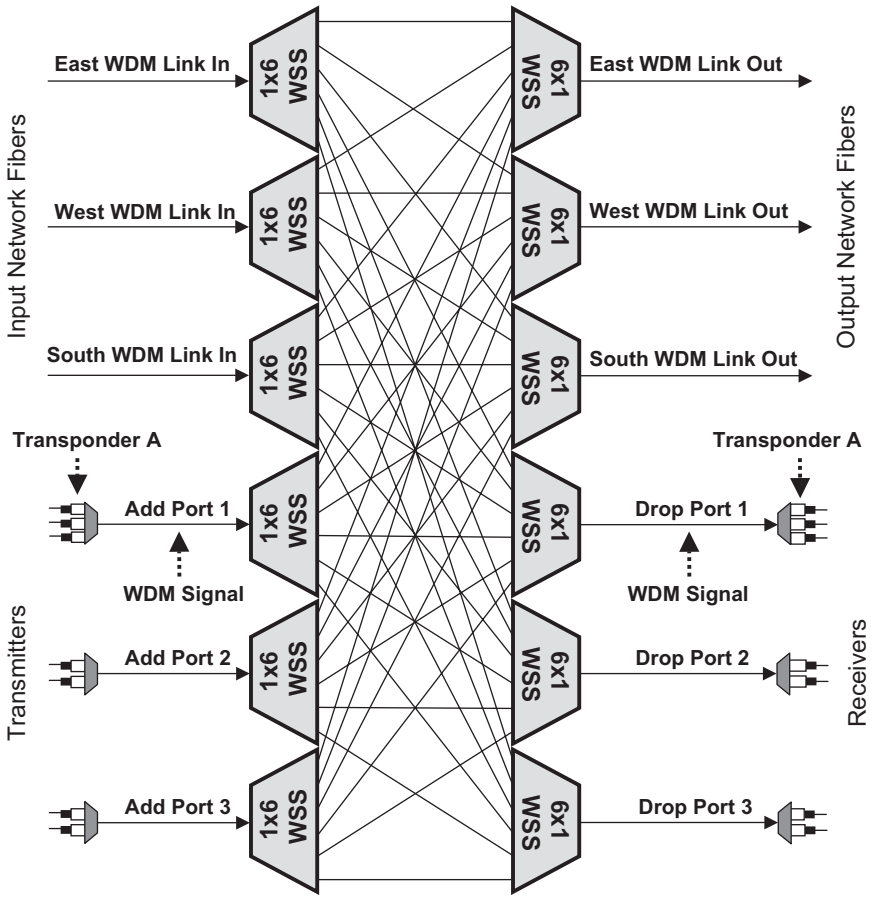


Fig. 2.13 One example of a pure route-and-select ROADM architecture, with the number of add/drop ports equal to the number of network fibers. In a hybrid route-and-select/broadcast-and-select architecture, the wavelength-selective switches (*WSSs*) on the add ports are replaced by passive splitters. This reduces the cost and supports some degree of multicast, while not appreciably degrading the cascadability of the ROADM

are multiplexed into a WDM signal, along with any add traffic, and sent out on the corresponding network fiber.

MEMS technology is likely to be used for the optical switch. As noted in Sect. 2.2, the MEMS elements themselves are not wavelength-selective; however, when combined with multiplexers and demultiplexers as shown, the combination is a wavelength-selective architecture.

The trapezoids in Fig. 2.14 represent any mux/demux technology; however, AWGs are sufficient for this architecture as opposed to more costly WSSs. With AWGs, the traffic that passes through the node must enter and exit on the same numbered port. For example, assume that λ_4 optically bypasses the node from East

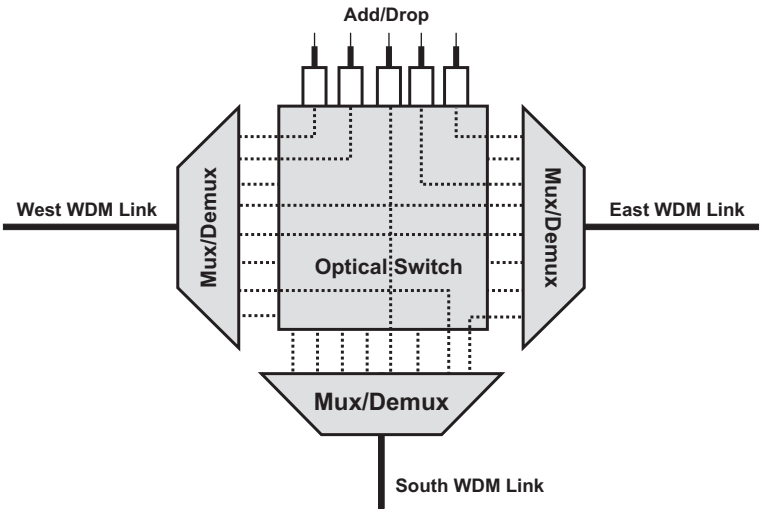


Fig. 2.14 Wavelength-selective ROADM architecture

to West; this wavelength enters the node on East AWG port 4 and must exit the West AWG on port 4 as well (because AWGs are not colorless). Similarly, if traffic is added at the node on λ_2, then it must be directed to port 2 on the associated AWG. Thus, only certain configurations of the optical switch are useful with respect to a given nodal traffic pattern. Paradoxically, this particular wavelength-selective ROADM architecture does not make use of WSSs, whereas the broadcast-and-select and route-and-select architectures do.

The biggest drawback of the wavelength-selective architecture is its scalability. The optical switch must be large enough to accommodate all of the wavelengths on the nodal fibers as well as all of the add/drop wavelengths. Consider a degree-four node, with 80 wavelengths on a fiber and a 50% drop requirement from each network fiber. The broadcast-and-select architecture can accommodate this configuration by using eight 8×1 WSSs (assuming the number of add/drop ports is four). In the wavelength-selective architecture, the optical switch must be of size 480×480. The largest commercially available MEMS switch in the 2015 time frame is on the order of 320×320; furthermore, this maximum size has not increased for several years. (However, see Exercise 2.8 for a wavelength-selective ROADM architecture utilizing WSSs for the mux/demux, instead of AWGs, that requires a smaller-sized MEMS switch.)

In terms of cost, the broadcast-and-select and route-and-select architectures are likely to be more cost effective as well, especially if a redundant optical switch fabric is required in the wavelength-selective architecture. Further comparisons among the three architectures are made in the next section with respect to various ROADM properties.

2.9 ROADM Properties

There are numerous important properties that characterize the various types of ROADMs and ROADM-MDs. These attributes, which can greatly affect network cost, operations, and functionality, are discussed in the following sections. Again, to simplify the text, the term ROADM is used in this section to imply both degree-two ROADMs and higher-degree ROADM-MDs.

2.9.1 Cascadability

A fundamental property of ROADMs is their cascadability, or the number of ROADMs a signal can be routed through before degradation of the signal requires that it be regenerated. In a backbone network, it is desirable to be capable of optically bypassing on the order of 5–10 consecutive nodes. In a metro-core environment, where nodes are more closely spaced, it is desirable to optically bypass on the order of 10–20 consecutive nodes.

One major factor in determining ROADM cascadability is the amount of loss in the ROADM through path. While optical amplifiers are used to boost the optical signal level to counteract the power loss suffered along the through path, any noise component of the signal is amplified as well. In addition, optical amplifiers add noise of their own. Thus, there is an overall degradation of the optical signal every time it optically bypasses a node.

A second significant factor that may limit the cascadability of a ROADM is the quality of its filters. As described earlier, a WDM signal enters a ROADM, with individual wavelengths in that signal being routed to different output ports. An internal filtering function is required to separate the WDM signal into its constituent wavelengths. Figure 2.15 depicts a generic filter bank that is designed for an eight-wavelength system. Each of the eight filters ideally passes just one wavelength while rejecting all of the others. However, note that the tops of the filters are not perfectly flat, such that the signal is distorted when it passes through. Furthermore, the bandwidth of the signal narrows when it passes through multiple ROADMs. The situation is exacerbated when the filters of the cascaded ROADMs are not aligned precisely. Additionally, the filters do not do a perfect job of rejecting adjacent wavelengths, thereby allowing some amount of crosstalk among the wavelengths. All of these impairments (as well as others covered in Chap. 4) limit the cascadability [HSLD12].

The cascadability of the three classes of ROADM architectures can vary widely depending on the quality of the components and, in the case of broadcast-and-select, on the degree of the ROADM. At a high level, the broadcast-and-select architecture should demonstrate better cascadability than the wavelength-selective architecture, in part due to less loss on the through path [TzZT03]. However, as the degree of the ROADM grows, the broadcast-and-select splitter loss increases, which ultimately restricts its cascadability. The route-and-select architecture was

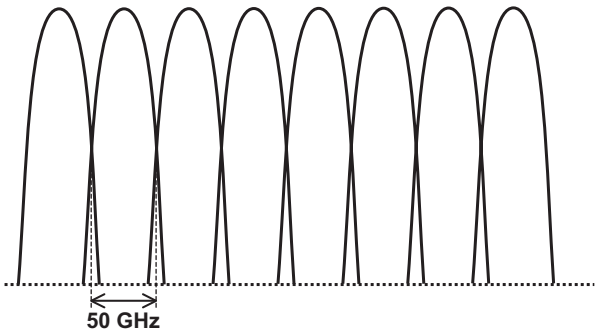


Fig. 2.15 A bank of eight filters, each designed to pass a signal with a bandwidth of less than 50 GHz. The filters are not ideal due to their rounded tops and the overlap with adjacent filters

purposely conceived to partially mitigate this limitation. It is expected to exhibit good cascadability even with high-degree ROADMs; however, this is predicated on having high-quality filters in the WSSs. (A signal passes through two WSSs per node in the route-and-select architecture, as opposed to just one in the broadcast-and-select architecture.)

2.9.2 Automatic Power Equalization

The wavelengths comprising a WDM signal may have originated at different nodes, such that the power levels entering any given node are unequal. Unbalanced power levels also result from uneven amplifier gain across the WDM spectrum. To maintain good system performance, it is necessary that these power levels be periodically balanced. It is desirable that the ROADM handle this functionality automatically, without requiring manual adjustment of power levels by a technician. Components such as WSSs are typically capable of automatic power equalization, such that this feature is available in most ROADMs.

2.9.3 Colorless

The *colorless* property for a ROADM directly parallels that for an optical terminal. It refers to the ability to plug a transponder of any wavelength into any (transponder) slot of the ROADM. As described in Sect. 2.3.1 with regard to the optical terminal, the colorless capability simplifies operations, enhances the value of using tunable transponders, and is highly desirable for a network with dynamic traffic. The colorless property is even more advantageous in a ROADM. Because of the wavelength continuity constraint in optical-bypass-enabled systems, it is often important to use a specific wavelength to carry a given connection. It is desirable to be able to insert the corresponding transponder in any add/drop slot of the ROADM, or to retune a tunable transponder to the corresponding wavelength without needing to move the transponder to a different slot.

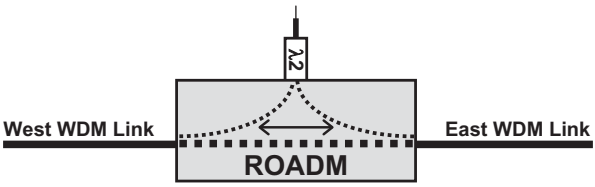


Fig. 2.16 In a directionless reconfigurable optical add/drop multiplexer (*ROADM*), a transponder can access any of the network links. In some implementations, one transponder can access multiple links simultaneously to provide optical multicast

The add/drop ports of broadcast-and-select and route-and-select ROADMs are similar to optical terminals (note the trapezoids on the add/drop ports in Fig. 2.12 and Fig. 2.13). They provide a muxing/demuxing function for the add/drop wavelengths, internal to the ROADM. Any of the technologies discussed in Sect. 2.3.1 for the optical terminal can be utilized for the ROADM add/drop ports as well. As a reminder, the add/drop architectures based on passive splitters/couplers and/or WSSs are colorless; the architecture based on AWGs is not.

The wavelength-selective architecture of Fig. 2.14 has a very different add/drop configuration. All of the muxing/demuxing occurs as the network fibers exit/enter the ROADM, such that the central optical switch operates on individual wavelengths. Each add/drop transponder is plugged directly into a *single-wavelength* port on the optical switch. The optical switch operates on whatever optical signal is present on a port; it is agnostic to the particular frequency of the signal. Thus, the wavelength-selective architecture of Fig. 2.14 is inherently colorless.

2.9.4 Directionless

Directionless refers to the ability of an add/drop transponder to access any of the network links that exit/enter a ROADM. A directionless ROADM is functionally illustrated in Fig. 2.16, where the transponder may access either the East link or the West link (or perhaps both simultaneously if multicast is supported).

This flexibility is especially useful in a dynamic environment, where connections are continually set up and torn down. It allows any connection associated with a particular transponder to be routed via any of the links at the node. It is also useful for protection in the optical layer, where at the time of failure, a connection can be sent out on an alternative path using the same transponder. Fewer transponders typically need to be pre-deployed at a directionless ROADM because the transponders are not tied to a particular network link (see Exercise 2.10).

There has been much debate surrounding the term “directionless,” as vendors have been reluctant to describe their products with this otherwise pejorative term. Other terms used for this same property are: steerable, edge configurable (used in the first edition of this textbook), and direction-independent. However, directionless has become the accepted term and is used here.

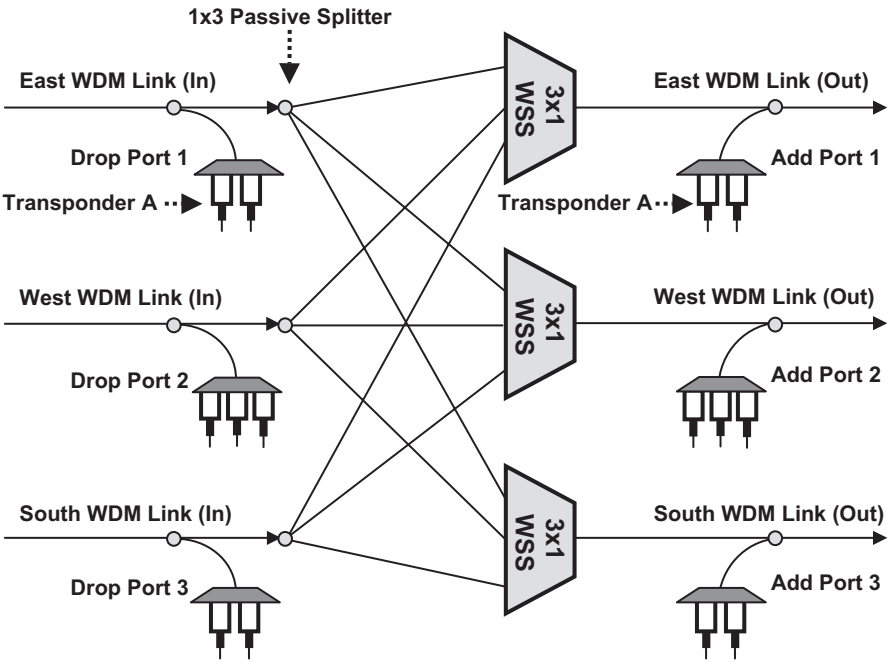


Fig. 2.17 One example of a non-directionless broadcast-and-select ROADM architecture. *Transponder A* can add/drop only to/from the East link. A non-directionless route-and-select architecture would look similar, with the 1×3 passive splitters at the input replaced by 1×3 WSSs

All three of the ROADM architectures illustrated in Fig. 2.12–2.14 are directionless. In the wavelength-selective architecture, the internal optical switch can be arbitrarily configured, allowing a transponder to be connected to any of the network links. In the broadcast-and-select and route-and-select architectures, each add/drop port has connectivity with each of the network links. However, in these latter two architectures, a *non-directionless* variation is also common. A degree-three non-directionless broadcast-and-select architecture is shown in Fig. 2.17. (Note that the drop ports are on the left and the add ports are on the right.) Each drop port is simply tapped off of its corresponding input network fiber, with say a 90:10 power split ratio between the through and drop paths. Similarly, each add port is coupled in to its corresponding output network fiber, also with an uneven power split ratio that favors the through path. Clearly, the transponders in this design can access just one network link; e.g., Transponder A is tied to the East link. This limits the flexibility of the ROADM, but removes roughly half of the cost; i.e., both fewer and smaller WSSs are required. In the broadcast-and-select architecture, it also reduces the loss of the through path.

The route-and-select architecture can be similarly modified, with add/drop ports directly tapped from the network fibers. Another non-directionless route-and-select option is shown in Fig. 2.18, where the incoming and outgoing WSSs also serve as the add/drop ports. (Note that this architecture provides colorless add/drop.) If the

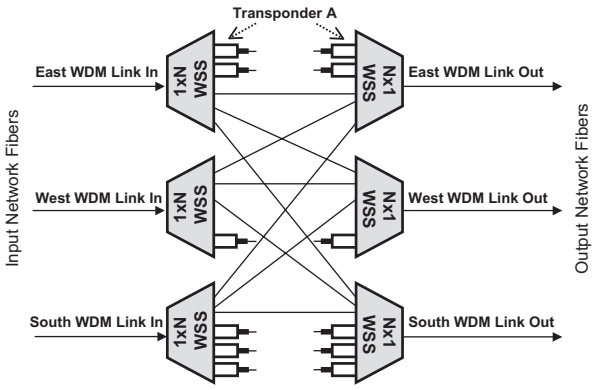


Fig. 2.18 A second option for a non-directionless route-and-select ROADM architecture, where some of the wavelength-selective switch (*WSS*) ports are used for the add/drop transponders

degree of the node is D and the WSSs are of size $1 \times N$, then the number of add/drop transponders that can be supported on each network fiber is limited to $N–D$. This configuration may be more appropriate for a node of low degree with a small amount of add/drop traffic.

2.9.4.1 Adding Configurability to a Non-Directionless ROADM

While a directionless ROADM is desirable, there are alternative means of achieving this same flexibility. One option is to deploy an optical switch in conjunction with a non-directionless ROADM, as shown in two different configurations in Fig. 2.19. The added switch is referred to as an "*edge switch.*" The traffic from the client layer (e.g., IP router) is fed through the edge switch so that it can be directed to the desired network link.

With the configuration of Fig. 2.19a, the client 1,310 nm optical signal is passed through the edge switch. The edge switch can have either an electronic or optical switch fabric; the latter is shown in the figure. The transponders are tied to a particular port, and the edge switch directs the client signal to the desired transponder. In Fig. 2.19b, the edge switch operates on the WDM-compatible signal and, thus, *must* have an optical switch fabric. In this configuration, the transponders can access any port. Thus, fewer transponders need to be pre-deployed with this configuration, assuming there is a single client (e.g., IP router) as shown. If there were multiple clients feeding the ROADM, then Fig. 2.19a may result in fewer transponders because it would allow the transponders to be shared among all of the clients. This is explored further in Exercise 2.12. There are some applications where the configuration of Fig. 2.19b as opposed to Fig. 2.19a must be used; e.g., see Sect. 2.13 and Sect. 4.7.2.

Because the edge switch is only for the add/drop traffic, and not the through traffic, the size is smaller as compared to a core switch that operates on all of the nodal wavelengths. For example, for a degree-three node with 80 wavelengths per fiber and 50 % add/drop from each fiber, the edge switch in either configuration of Fig. 2.19 needs to be of size 240×240, as opposed to 360×360 for a core switch.

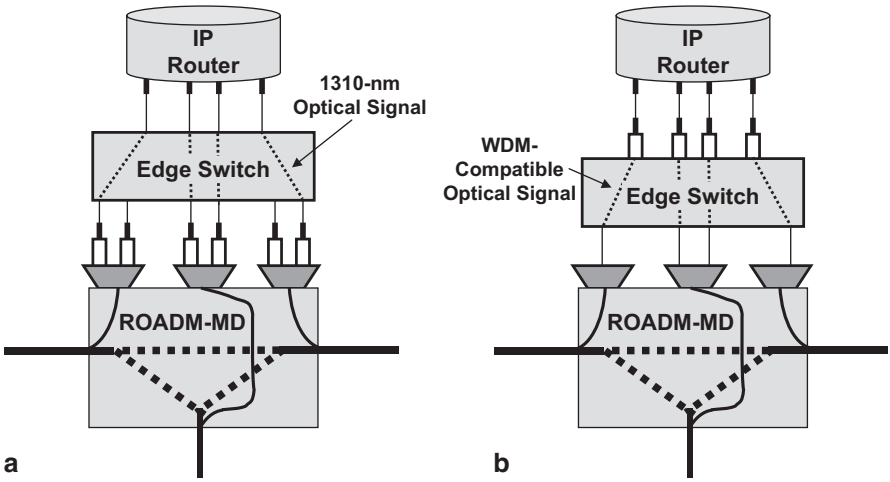


Fig. 2.19 An edge switch used in conjunction with a non-directionless multi-degree reconfigurable optical add/drop multiplexer (*ROADM-MD*) in order to add configurability. As shown, the edge switches operate as fiber cross-connects (*FXCs*). In **a**, the 1,310 nm optical signal is switched; in **b**, the wavelength-division multiplexing (*WDM*)-compatible optical signal is switched. The architecture that results in fewer required transponders depends upon the number of clients (e.g., Internet Protocol (*IP*) routers) and the traffic patterns

It is also possible to use a modular design for the edge switch, where multiple smaller switches are used in place of a single larger switch [KPWB12]. The clients (e.g., IP routers) would connect to multiple (or all) edge switches, and each edge switch would be capable of providing connectivity to any of the network links. Though some flexibility is lost as compared to deploying one large edge switch, the modular approach provides some protection against an edge-switch failure, allows smaller, more feasibly sized switches to be used, and is more compatible with a pay-as-you-grow strategy.

The architecture of Fig. 2.19b can also be used to provide a colorless capability if the ROADM itself is not colorless. Assume that the mux/demuxes (i.e., the trapezoids in the figure) are AWGs, which are not colorless. By placing the edge switch between the transponders and the AWGs, each transponder can be directed to the appropriate "colored" port on the AWG. However, the edge switch may need to be fairly large to provide this colorless feature (i.e., larger than if providing just the directionless feature), as investigated in Exercise 2.16.

Deploying an edge switch at a node can be useful for other purposes, e.g., reducing ROADM contention (Sect. 2.9.5.2) and protection (Chap. 7). The edge switches shown in Fig. 2.19 may also be called *fiber cross-connects* (FXCs) because they are essentially serving the same function as a fiber patch panel.

Another option for a non-directionless ROADM is to use the configurability of the IP router itself, rather than adding an edge switch. If the IP router has enough ports connected to each of the ROADM add/drop ports, then the router can establish connections on whatever network link is desired. However, due to the high cost of IP router ports, this may be more costly than adding an edge switch.

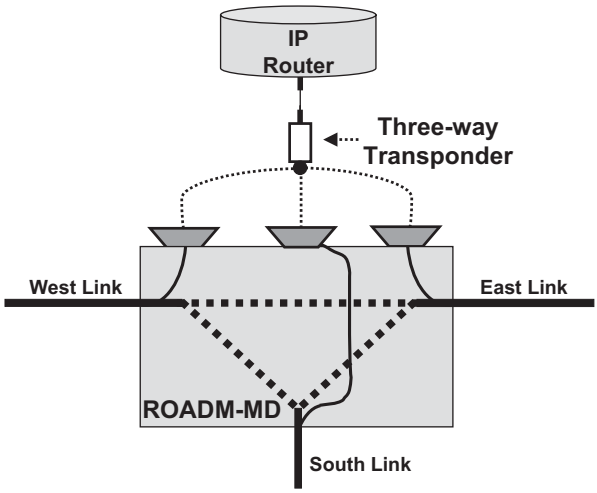


Fig. 2.20 A three-way transponder is able to access any of the network links at the degree-three node. It is desirable to use an optical backplane to simplify the cabling

A third option is to use *flexible WDM transponders* in conjunction with a non-directionless ROADM, where the output of the transponder is split into multiple paths and each path feeds into a different add port on the ROADM. In the reverse direction, the transponder is equipped with a switch to select a signal from one of the drop ports. A three-way flexible transponder that can access any of the three network links at the node is illustrated in Fig. 2.20. The flexible transponder option is discussed more fully in Simmons and Saleh [SiSa07].

2.9.5 Contentionless

In a *contentionless* ROADM, a new connection cannot be blocked solely due to contention within the ROADM. For a ROADM that is *not* contentionless, the contention typically arises in the form of wavelength conflicts (e.g., a wavelength is available on a network fiber, but is not available on an add/drop port). Contention may also arise if the ROADM includes components that have internal blocking (e.g., a switch that is not strictly non-blocking). Here, we assume that the components are internally non-blocking and focus on wavelength-contention scenarios within the ROADM.

The wavelength-selective architecture of Fig. 2.14 is contentionless. No multiplexing of wavelengths occurs within the ROADM other than the multiplexing that occurs as the signal is exiting onto a network fiber. Thus, the only constraint is that the wavelength must be available on the network fiber; the ROADM itself does not add any further wavelength constraints.

The broadcast-and-select and route-and-select architectures of Fig. 2.12 and Fig. 2.13, respectively, are *not* contentionless. The scenarios that cause contention are somewhat subtle; three specific scenarios are covered next. The scenarios arise because the add/drop ports are multiwavelength combined with the fact that the

WDM signal on an add/drop port does not have a one-to-one correspondence with a WDM signal on a network fiber. For illustration purposes, a degree-three broadcast-and-select architecture is shown in the figures, but the scenarios apply to any degree ROADM, and to the route-and-select architecture as well. Simulation studies of various forms of contention can be found in Feuer et al. [FWPW11].

It is important to note that with good routing and wavelength assignment algorithms (see Chaps. 3 and 5), contention in a ROADM can be significantly reduced even if the architecture is not contentionless. However, algorithms cannot completely eliminate the possibility of ROADM contention occurring, although it typically would occur only under high load conditions.

2.9.5.1 ROADM Contention Scenario 1

In Fig. 2.12, the number of ROADM add/drop ports equals the number of ROADM network links. If the number of add/drop wavelengths at a node is relatively small, then one may consider deploying fewer add/drop ports in order to reduce the ROADM cost and size. A ROADM with three network links but only two add/drop ports is shown in Fig. 2.21.

Assume that it is desired to establish three new connections at the node, one routed on each of the network links. Assume that each of the network links has only one available wavelength, and it happens to be λ_1 for all three links. There are only two add/drop ports; thus, at least two of the connections would need to be established on the same add/drop port. However, the add/drop wavelengths for a particular port are multiplexed together into a WDM signal (note the mux/demux trapezoids on the add/drop ports). Thus, two connections on the same wavelength cannot be established on the same add/drop port. One of the three desired connections would be blocked due to ROADM wavelength contention.

2.9.5.2 ROADM Contention Scenario 2

It may be tempting to assume that if the number of ROADM add/drop ports equals the number of ROADM network links, then contention cannot occur. However, this is not the case. Consider the example of Fig. 2.22, where the IP router is connected only to Add/Drop Port 1 (the same router is shown on both the add and drop sides). Assume that the router wants to establish two new connections, one on the East link and one on the West link. Additionally, assume that λ_1 is the only available wavelength on these two links. One of the connections will be blocked due to wavelength contention on Add/Drop Port 1.

It is possible to alleviate this particular wavelength contention scenario by feeding the outputs of the client (e.g., IP router) into an edge switch (e.g., an FXC), such that the client can access a transponder on *any* of the add/drop ports, and hence any of the network links. This is the same architecture that was illustrated in Fig. 2.19a for adding configurability to a non-directionless ROADM. Alternatively, the edge

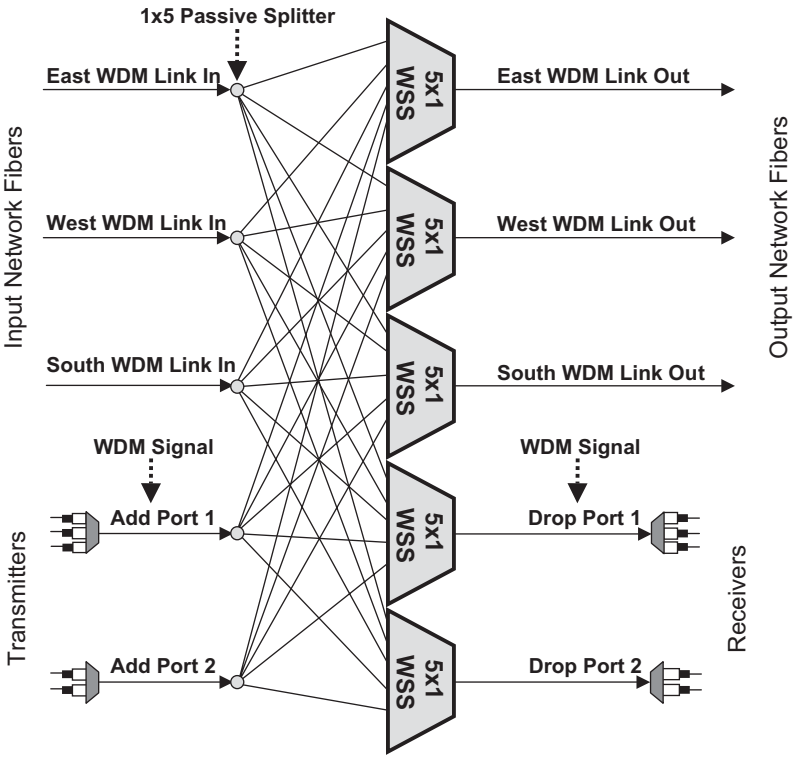


Fig. 2.21 A ROADM with three network links but only two add/drop ports. If it is desired to establish three connections, one per network link, each using the same wavelength, then wavelength contention on an add/drop port will block one of the connections

switch can be placed between the transponders and the mux/demuxes, as is illustrated in Fig. 2.19b.

Adding the edge switch addresses the contention of Fig. 2.22 because if λ_1 is available on two output network fibers, then it must be available on two different add ports (assuming that there are no λ_1 loopbacks from an add port to a drop port, and no "pre-lit" but unused λ_1 transponders on an add port). However, this is not necessarily true for the reverse direction if the ROADM is multicast-capable. It is possible, for example, that λ_1 is available on two input network fibers, but only available on one of the drop ports, due to this wavelength having been multicast to two different drop ports. Thus, the addition of the edge switch (in either of the Fig. 2.19 configurations) does not solve all contention issues of this type.

2.9.5.3 ROADM Contention Scenario 3

The third contention scenario is depicted in Fig. 2.23. Two transponders are deployed on each of the three add/drop ports. Both of the transponders on Add/Drop

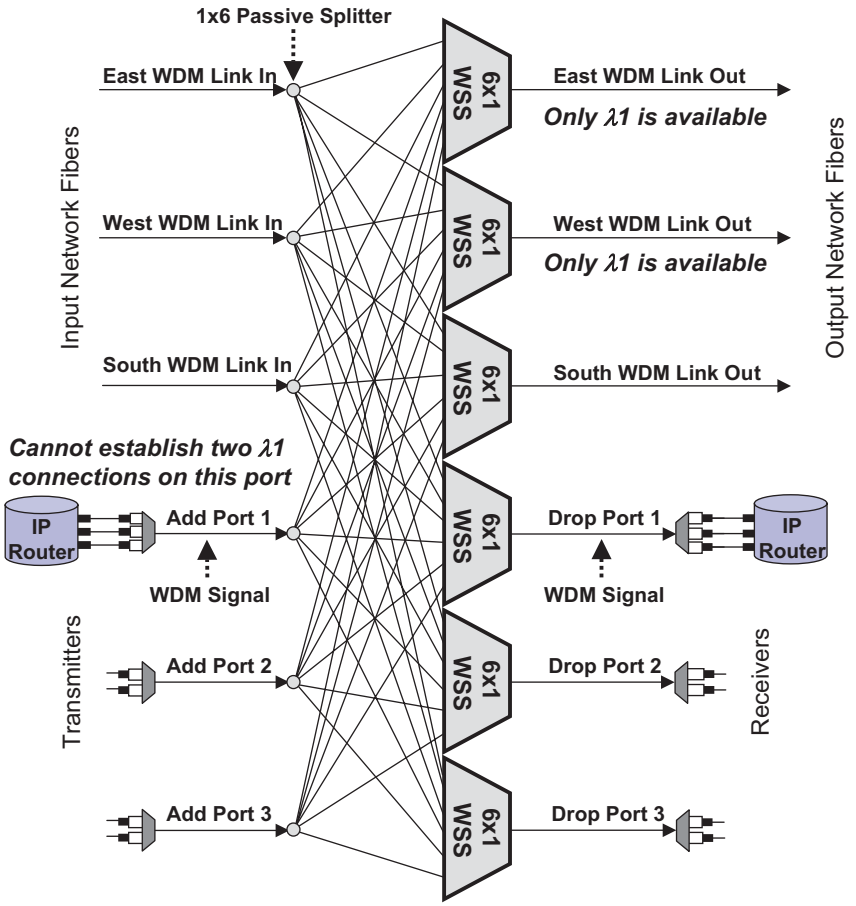


Fig. 2.22 The Internet Protocol (*IP*) router is connected only to Add/Drop Port 1. If it is desired that the router establish two connections, one on the East link and one on the West link, both of which have only λ_1 available, then one of the connections will be blocked due to wavelength contention on the add/drop port

Port 3 are currently being used for active connections. One of the transponders on both Add/Drop Ports 1 and 2 is currently being used for active connections, and both of these connections are carried on λ_1. The other transponder on these two ports is available.

Assume that a new connection request arrives, where the connection is to be routed on the South link. Further assume that the only available wavelength on the South link is λ_1. Then, once again, this connection will be blocked due to wavelength contention in the ROADM. The only available transponders are on add/drop ports where λ_1 is already in use, such that another connection on λ_1 cannot be established.

Note that adding an edge switch between the clients and the transponders (i.e., Fig. 2.19a) does not eliminate this third contention scenario. However, placing the edge switch between the transponders and the mux/demuxes (i.e., Fig. 2.19b) does

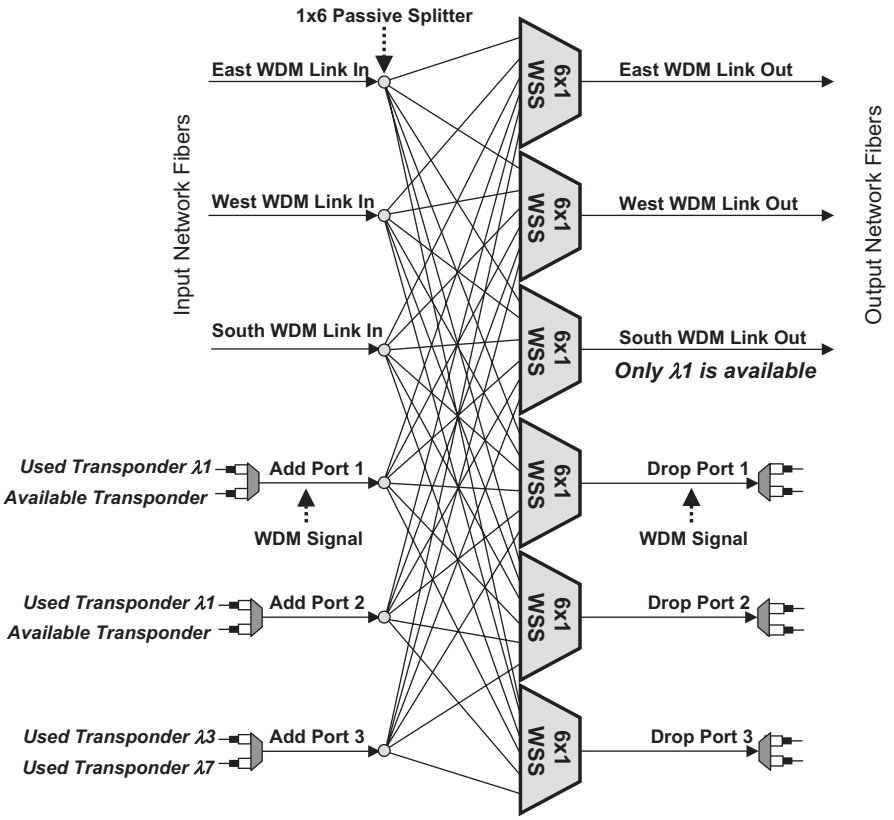


Fig. 2.23 A new connection on the South link is desired; only λ_1 is available on this link. The only available transponders are deployed on Add/Drop Ports 1 and 2. However, active connections are already established on these ports using λ_1. This prevents the new connection from being established

remove the contention. Alternatively, one could overprovision the number of transponders on each add/drop port, to ensure that there is always an available transponder; however, this could be costly.

Another solution, which addresses all three of the contention scenarios, in addition to being colorless and directionless, is presented next.

2.9.5.4 Contentionless Broadcast-and-Select and Route-and-Select ROADM Architectures

Two broadcast-and-select contentionless architectures are presented here. These architectures address all three contention scenarios discussed above. Similar approaches can be used for the route-and-select ROADM as well.

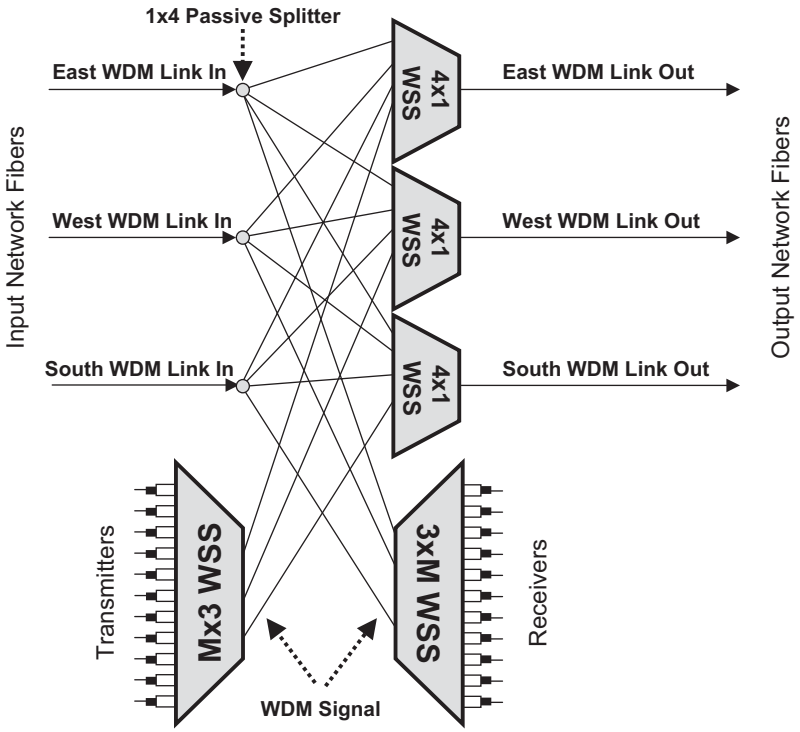


Fig. 2.24 A contentionless ROADM that uses an $M \times N$ wavelength-selective switch (*WSS*) for the add/drop traffic. In the figure, there are three wavelength-division multiplexing (*WDM*) add fibers and three WDM drop fibers. There is a one-to-one correspondence between the add/drop fibers and the network output/input fibers. Thus, if there is no wavelength contention on the network fibers, then there is no wavelength contention on the add/drop fibers. In fact, this architecture is colorless, directionless, and contentionless (*CDC*)

In the first architecture, illustrated in Fig. 2.24, an $M \times N$ WSS is used for the add/drop traffic as opposed to multiple $1 \times N$ WSSs [Coll09]. (M indicates the number of transponders that can be supported; N indicates the number of network links. It operates as an $M \times N$ WSS on the add port and an $N \times M$ WSS on the drop port.) This solution is analogous to placing an FXC between all of the transponders and all of the mux/demuxes, as in Fig. 2.19b. While the add/drop fibers do carry a WDM signal, there is a one-to-one correspondence between the add/drop fibers and the network fibers. Thus, if there is no wavelength contention on the network fibers, there will not be wavelength contention on the add/drop ports (assuming that the $M \times N$ WSS does not have any internal blocking).

This ROADM architecture is also colorless, because the $M \times N$ WSS can direct any wavelength to any transponder. Furthermore, this architecture is directionless, as any transponder can access any network link. A ROADM with all three properties (colorless, directionless, and contentionless) is referred to as a *CDC ROADM*.

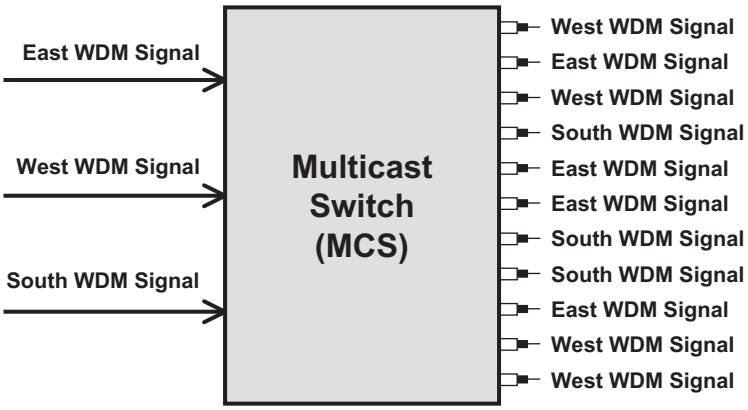


Fig. 2.25 A functional diagram of a $3 \times M$ multicast switch (*MCS*), for an arbitrary configuration. An incoming wavelength-division multiplexing (*WDM*) signal is multicast to the transponders corresponding to connections contained in that WDM signal. Only the drop direction is shown

The drawback of using an $M \times N$ WSS is that the device is likely to be costly, as well as large in physical size. An alternative is to use an $M \times N$ *multicast switch* (MCS) instead of an $M \times N$ WSS [Way12]. The MCS is *not* a wavelength-selective device; i.e., if there are WDM signals on the inputs, then there are WDM signals on the outputs. In the add direction, the $M \times N$ MCS operates similarly to the $M \times N$ WSS; the outputs of the individual transponders are combined into WDM signals corresponding to each of the N network output fibers. However, the operation of the $N \times M$ MCS on the drop side is somewhat different from that of the $N \times M$ WSS, due to the lack of wavelength selectivity. A $3 \times M$ MCS is functionally illustrated in Fig. 2.25 for an arbitrary configuration. The MCS multicasts the WDM signal from a given network input fiber to each of the M ports where there are transponders corresponding to connections in this WDM signal. Because a WDM signal is delivered to the transponders, it is necessary that the transponder receivers be equipped with a filter (or some other frequency-selective technology, such as coherent detection; see Sect. 4.2.3) to select the desired wavelength from the WDM signal. (If a filter is used, ideally it is tunable.)

Assuming that the MCS is internally contentionless, then substituting the $M \times N$ MCS in place of the $M \times N$ WSS in Fig. 2.24 still yields a CDC ROADM.

2.9.6 Gridless

As described in Sect. 2.9.1, a ROADM is usually equipped with filters that separate the WDM signals into their constituent wavelengths. The filters are typically of a fixed bandwidth, and align with wavelengths on a fixed grid. For example, in backbone networks, it is common to support 80 wavelengths on a fiber, with a wavelength centered at every 50 GHz in the C-band region of the spectrum. A por-

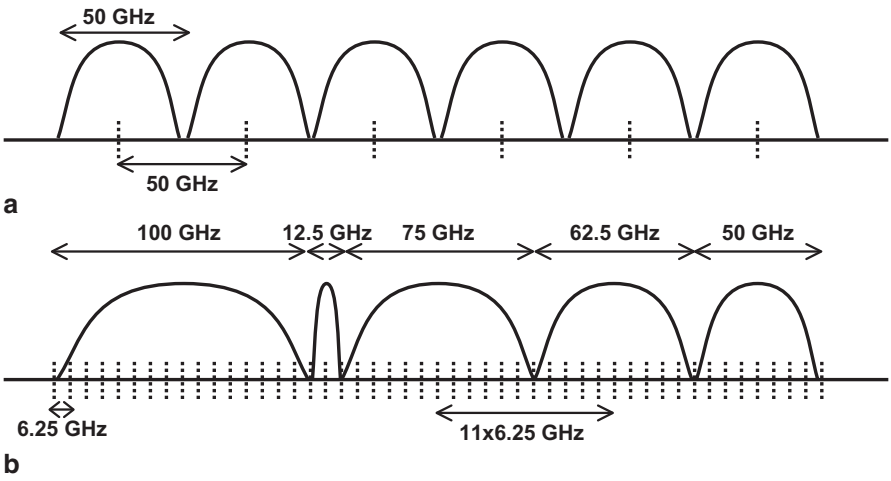


Fig. 2.26 **a** Wavelengths aligned on a 50-GHz grid. Each wavelength requires 50 GHz of bandwidth. **b** Wavelengths aligned on a 6.25-GHz grid. As specified by the International Telecommunication Union (*ITU*), each wavelength requires $N \times 12.5$ GHz of bandwidth, where N is an integer

tion of such a fixed spectral grid is illustrated in Fig. 2.26a. The center frequency of each wavelength in the figure is aligned on a 50-GHz grid, and each wavelength requires approximately 50 GHz of bandwidth. The corresponding ROADM filters would also be aligned on the same 50-GHz grid, with each filter having a passband bandwidth of about 50 GHz.

Such a fixed-filter arrangement has been sufficient for ROADM operation over several years. While the bit-rate of an individual wavelength has increased from 2.5 to 10 to 40 to 100 Gb/s, the required bandwidth of the signal in backbone networks has remained constant at 50 GHz.[2] (This implies that the transmission scheme has become 40 times more spectrally efficient as the line rate has increased. This is discussed further in Sect. 4.2.3.)

However, the convention of aligning wavelengths on a 50-GHz grid is likely to change. First, the line rate of a wavelength is expected to increase to 400 Gb/s (or higher). While it is theoretically possible that a 400-Gb/s signal may require a bandwidth of only 50 GHz in some scenarios [ZhNe12], it is more likely that a wider bandwidth, e.g., perhaps 62.5 GHz or 75 GHz, will be required. To efficiently accommodate such wider bandwidths, the WDM grid plan approved by the ITU in 2002 supports a spectral granularity as fine as 12.5 GHz [ITU02]. Thus, a 400-Gb/s wavelength will likely occupy five or six 12.5-GHz slots. (This is more efficient than allocating two 50-GHz slots.) The ITU recommendation also allows for uneven channel spacings on one fiber, such that some wavelengths may be spaced every 50 GHz, while others could be spaced every 62.5 GHz. Furthermore, the ITU modified its grid-plan recommendation in 2012 to include a "flexible grid" op-

[2] Early generations of 2.5-Gb/s technology actually required more than 50 GHz of bandwidth.

tion. This option supports any mix of wavelength spacings on one fiber, as long as each wavelength aligns with a 6.25-GHz grid and the bandwidth assigned to each wavelength is a multiple of 12.5 GHz [ITU12b]. This is illustrated in Fig. 2.26b, where wavelengths of different bandwidths are supported on one fiber. There have also been research proposals that go beyond this, to allow almost arbitrarily fine wavelength granularity [Jinn08]. (These flexible network approaches will be discussed extensively in Chap. 9.)

In order to remain compatible with this more flexible transmission approach, ROADMs must be designed with greater flexibility as well. Specifically, it is desirable that ROADMs support an additional type of configurability, where the filter shapes can be tailored, ideally through software control, to match the particular wavelength spacings, bandwidths, and transmission formats that are in use.

ROADMs that are not limited to a fixed transmission grid are aptly called *gridless*. The degree of flexibility in ROADMs classified as gridless can vary. For example, initial gridless products could be configured to support either 50-GHz or 100-GHz wavelength spacing. However, more recent products meet the flexibility specified by the ITU [GJLY12; Fini13].

The technology typically used for gridless ROADMs is *liquid crystal on silicon* (LCoS). For more details on gridless ROADM technology, see Baxter [BFAZ06], Fontaine [FoRN12], and Marom and Sinefeld [MaSi12]. Also see Sect. 9.7.1.

2.9.7 Wavelength Versus Waveband Granularity

Thus far, the discussion has implicitly assumed that ROADMs operate on a per-wavelength basis, where the choice of add/drop versus bypass can be made independently for each wavelength in the WDM signal. Alternatively, ROADMs can be configurable on the basis of a *waveband*. A waveband is a set of wavelengths that is treated as a single unit. Either the whole waveband is added/dropped or the whole waveband transits the node. Wavebands are usually composed of wavelengths that are contiguous in the spectrum. In most implementations, the wavebands are of equal size; however, nonuniform waveband sizes may be more efficient depending on the traffic [IGKV03].

Waveband granularity is clearly not as flexible as wavelength granularity. The chief motivation for using a waveband-based ROADM is the potential for reduced cost and complexity. For example, assuming that a waveband is composed of eight contiguous wavelengths, then the bank of eight 50-GHz per-wavelength filters shown in Fig. 2.15 can be replaced by a single 400-GHz filter. Waveband technology may be more suitable for metro networks, where sensitivity to cost is greater.

Wavebands are most effective when many connections are routed over the same paths in the network. Otherwise, inefficiencies may arise due to some of the bandwidth being "stranded" in partially filled wavebands. Studies have shown that under reasonable traffic conditions, and through the use of intelligent algorithms, the inefficiencies resulting from wavebands are small [BuWW03]. Nevertheless, some car-

riers are averse to using waveband technology because of the diminished flexibility. Additionally, hesitation with respect to employing wavebands arises out of concern that guardbands may be required to isolate adjacent wavebands. A guardband represents unused spectrum and hence lost capacity. However, waveband systems without guardbands have been deployed successfully [Busi00].

Any of the ROADM architectures discussed above are capable of waveband operation, including the wavelength-selective architecture [KZJP04]. Multigranular ROADMs that combine both waveband and wavelength granularities are discussed in Sect. 2.11.

2.9.8 Multicast

Some ROADMs support optical multicast, where a given signal is sent to multiple destinations rather than just one, and the signal replication occurs in the optical domain as opposed to in the electrical domain. This was discussed in Sect. 2.8.1, with respect to the broadcast-and-select architecture. As noted in that section, there are four types of multicast to consider: (1) one input network fiber to multiple output network fibers; (2) one input network fiber to both an output network fiber and a transponder on a drop port (i.e., drop and continue); (3) one transponder on an add port to multiple output network fibers; and (4) one input network fiber to multiple transponders on drop ports.

The directionless broadcast-and-select architecture of Fig. 2.12 is capable of all four multicast configurations. The non-directionless broadcast-and-select architecture of Fig. 2.17 is capable of the first two multicast configurations but not the third; it may be capable of the fourth, depending on the drop-port technology. The route-and-select architecture of Fig. 2.13 and the wavelength-selective architecture of Fig. 2.14 are not capable of multicast. The broadcast-and-select ROADM with the $M \times N$ WSS (Fig. 2.24) is capable of the first two multicast configurations, but not the third and fourth. (This assumes that WSSs are not capable of multicast, although in principle they could be. Also see Exercise 2.19.)

2.9.9 East/West Separability (Failure/Repair Modes)

East/West separability relates to the failure and repair modes of the ROADM. For a degree-three ROADM-MD, this extends to East/West/South separability, etc. It is desirable that a ROADM be designed such that a component failure brings down just one "direction" (unless there is a catastrophic failure). For example, if the WSS associated with the East direction of the ROADM fails, the remaining ROADM directions should remain operational. This provides an opportunity to restore the East add/drop traffic on another link. It would be undesirable to have an architecture where, for example, a failure of the ROADM East direction causes traffic to/from the West link to be shut down as well. Furthermore, it would be undesirable

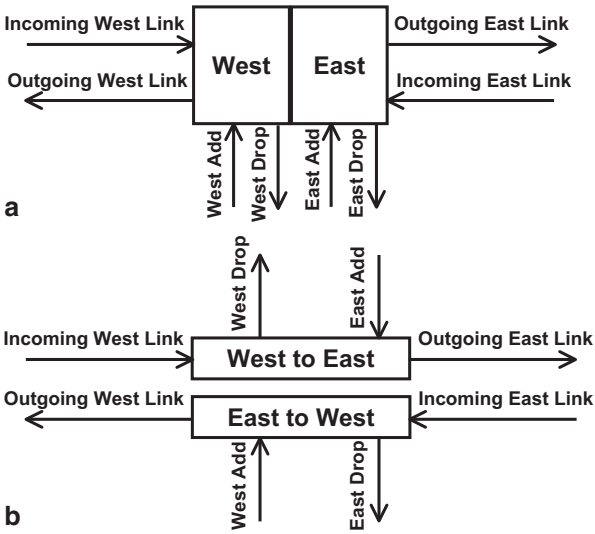


Fig. 2.27 **a** The desired East/West separability for failure and repair modes is shown, for a non-directionless ROADM. **b** An alternative partitioning that does not readily extend to ROADM-MDs and that leads to bidirectional traffic being routed on different links in the two directions under failure conditions

if the process of repairing a failure on the East side also requires that the West side be taken down.

The desired separability is shown in Fig. 2.27a for a non-directionless ROADM. An alternative ROADM partitioning that has been considered in some standards bodies is shown in Fig. 2.27b. It is not readily apparent how this latter partitioning would extend to ROADM-MDs. Additionally, a failure of either the top or bottom partition would result in bidirectional connections at the node being routed over two different paths. For example, if the bottom portion fails, then incoming connections would enter the node on the West link, but outgoing connections would exit the node on the East link.

Some large enterprises that deploy their own network equipment have been reluctant to take advantage of optical-bypass technology because of their concern over ROADM failures. They assume that adding/dropping all traffic on individual optical terminals (i.e., the O-E-O paradigm) results in a more reliable node, as a failure is less likely to bring down the entire node. However, with East/West separability, the typical ROADM failure is no worse than if an optical terminal fails. In fact, the ROADM, if directionless, potentially provides more automated recovery, as it can redirect add/drop traffic from the failed direction to another link at the node. Furthermore, it is possible to provide redundancy for the key ROADM components. For example, an extra WSS can be added to the ROADM to provide protection from a single WSS failure (see Exercise 2.9).

Thus, while a catastrophic ROADM failure does bring down all add/drop traffic at the node, such failures should not be a common event.

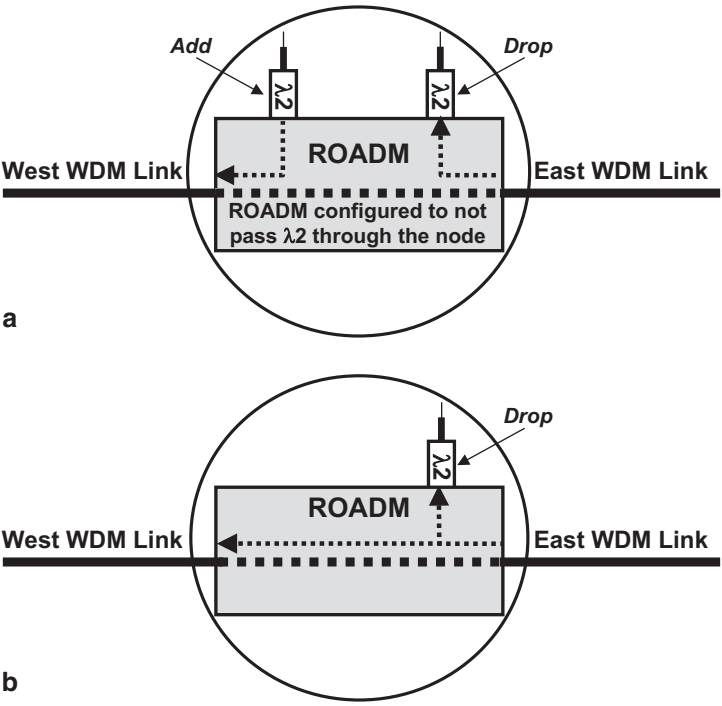


Fig. 2.28 **a** A reconfigurable optical add/drop multiplexer (*ROADM*) with wavelength reuse. A wavelength (λ_2) is dropped from the East link. The ROADM is configured to not pass this wavelength through the node, so that the same wavelength can be added on the West link. **b** A ROADM without wavelength reuse. The wavelength that is dropped from the East link continues to be routed through the node such that the same wavelength cannot be added on the West link

2.9.10 Wavelength Reuse

A ROADM property that potentially affects network efficiency is whether or not the ROADM supports *wavelength reuse*. With wavelength reuse, if a particular wavelength is dropped at a node from one of the network fibers, then the ROADM is capable of preventing that same wavelength from being carried on the through path. This is illustrated in Fig. 2.28a. In the figure, a particular wavelength (λ_2) enters the node on the East fiber and is dropped. After being dropped, the wavelength does not continue to be routed through the ROADM to the West fiber. This allows traffic sourced at the node to be added to the West fiber on that same wavelength, as is shown in the figure; i.e., the add traffic is "reusing" the same wavelength that was dropped. If the dropped wavelength had continued through the ROADM, then traffic could not have been added to the West fiber on this wavelength because the two signals would interfere.

Figure 2.28b illustrates a ROADM that does not have wavelength reuse. The wavelength entering from the East fiber is dropped but also continues through the ROADM to the West fiber. This wastes bandwidth, as this particular wavelength cannot be used to carry useful traffic on the West fiber (i.e., it is carrying traffic that has already reached its destination). If there are multiple consecutive ROADMs without reuse, the signal will continue to be routed through each one of them, preventing the wavelength from being reused to carry useful traffic on all of the intermediary links. (Note that the nodes on a ring cannot all be populated by no-reuse ROADMs, as the optical signals will continue to wrap around.)

Wavelength reuse is a desirable trait in ROADMs to maximize the useful capacity of the network, and most ROADMs do support this property. For example, in the broadcast-and-select architecture illustrated in Fig. 2.12, it is the WSS that prevents a dropped wavelength from being routed onto an output network fiber. Nevertheless, ROADMs without reuse can be useful network elements. First, no-reuse ROADMs are lower cost than ROADMs with reuse, making them a cost-effective option for nodes that drop a small amount of traffic. ROADMs without reuse typically allow just a small percentage of the wavelengths on a fiber to be added/dropped; e.g., a maximum of 8 add/drops from a fiber with 80 wavelengths. Additionally, the element is ideally deployed at nodes located on lightly loaded links, so that the wasted bandwidth is inconsequential.

In many implementations, a no-reuse ROADM is little more than an optical amplifier equipped with a coupler and splitter to add and drop traffic. Because of this design, it is often possible to upgrade from an optical amplifier to a no-reuse ROADM without affecting the traffic already passing through the amplifier. When a system is first installed, a network site that is not generating any traffic may be equipped with just an optical amplifier. As the network grows, the site may need to source/terminate traffic; when this occurs, the optical amplifier can be upgraded to the no-reuse ROADM. (Upgrading from an optical amplifier to a ROADM *with* reuse is generally not possible without a major overhaul of the equipment.)

2.10 Optical Switch Types

A variety of *optical switch* types were included in the architectures of the previous sections. As discussed in Sect. 2.2, the term optical switch indicates that the switch ports operate on the granularity of a wavelength (or a waveband). It does not imply that the switch supports optical bypass, nor does it imply that the switch fabric is optical. Optical switches are also referred to as optical cross-connects (OXCs).

There is often confusion surrounding the various types of optical switches. This section presents an optical-switch taxonomy that classifies switches primarily based on their functionality and fabric type. For more details of the underlying optical switch technologies, see Al-Salameh [Sala02], Papadimitriou [PaPP03], and El-Bawab [ElBa06].

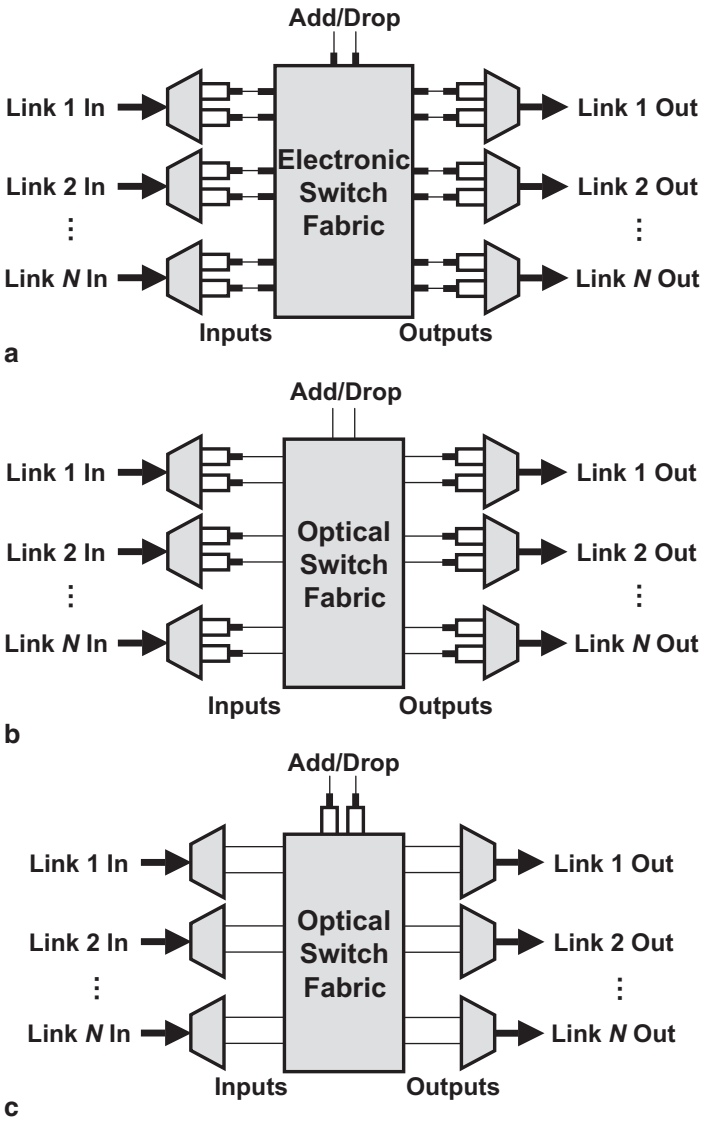


Fig. 2.29 Examples of optical-switch architectures. **a** O-E-O architecture with electronic switch fabric and electronic interfaces on all ports (Sect. 2.10.1). **b** Photonic switch that switches the 1,310-nm optical signal (Sect. 2.10.2). **c** A wavelength-selective all-optical switch (Sect. 2.10.3)

2.10.1 O-E-O Optical Switch

An optical switch based on O-E-O technology is shown in Fig. 2.29a (it also was shown as part of Fig. 2.6). The switch fabric is electronic, and each of the switch ports is equipped with a short-reach interface to convert the incoming 1,310-nm

optical signal to an electrical signal. These types of switches present scalability challenges in cost, power, and heat dissipation due to the amount of electronics. Consider using such a switch to provide configurability at a degree-four O-E-O node. Assume that each fiber carries 80 wavelengths and assume that the node needs to support 50% add/drop. There needs to be a port for each wavelength on the nodal fibers as well as for each add/drop wavelength. This requires a 480×480 switch, with each switch port having a short-reach interface.

2.10.2 Photonic Switch

The term *photonic switch* refers to an optical switch where the switch fabric is optical so that the incoming optical signal does not have to be converted to the electrical domain. MEMS technology is often used to build photonic switches. However, a photonic switch does not necessarily imply optical bypass through a node. For example, Fig. 2.29b illustrates a photonic switch that is being used to switch 1,310-nm optical signals. There are WDM transponders on both the input and output fibers and, thus, optical bypass is not supported. To more fully capture the switching architecture, Fig. 2.29b can be considered an OEO-O-OEO architecture, in contrast to the OEO-E-OEO architecture of Fig. 2.29a.

A carrier may choose to implement the configuration of Fig. 2.29b in order to isolate various technologies in the network. Because each wavelength is terminated on a WDM transponder in the node, the links are isolated from each other. This would allow, for example, the transmission technology used on each link to be supplied by different vendors. Furthermore, the switch vendor can be independent from the transmission vendor because the switch is operating on the standard 1,310-nm signal as opposed to a WDM-compatible signal.

This same vendor independence is provided by the O-E-O switch of Fig. 2.29a; however, the photonic switch configuration of Fig. 2.29b requires significantly less electronics. In terms of port count, however, the photonic switch is no smaller than the O-E-O switch. Using the same example of a degree-four node with 80 wavelengths per fiber and 50% add/drop, the required switch size, assuming wavelength granularity, is again 480×480. The advantage is that the switch fabric is not electronic and electronic interfaces are not required on the ports.

2.10.3 All-Optical Switch (ROADM)

All-optical switch is the term for a switch with an optical switch fabric that is being used to support optical bypass. This is illustrated in Fig. 2.29c (using the wavelength-selective architecture). The term *ROADM* is now more commonly used as a synonym for *all-optical switch*. (As a historic note, the term “all-optical switch” formerly implied a directionless ROADM; however, this distinction has been lost.) The overall architecture is also referred to as O-O-O to emphasize the ability to remain in the optical domain.

While the amount of electronics is significantly reduced compared to Fig. 2.29a and Fig. 2.29b, the number of ports on the switch fabric is no smaller. Assume that the all-optical switch of Fig. 2.29c has a core MEMS switch. Again, for a degree-four node supporting 50% add/drop, where each fiber carries 80 wavelengths, the switch fabric needs to be 480×480, assuming wavelength granularity. Scaling technologies such as MEMS to this size can be challenging. While the broadcast-and-select and route-and-select architectures effectively require the same-sized switch fabric, these architectures are composed of a collection of smaller components (e.g., WSSs) that internally operate on each separate wavelength, such that they are more scalable.

2.10.4 Fiber Cross-Connect

An FXC is an optical switch with an optical switch fabric that is essentially used to take the place of a fiber patch panel. An FXC was illustrated in Fig. 2.19, where it was used to provide configurability at a node with a non-directionless ROADM.

2.10.5 Grooming Switch

One special type of O-E-O switch is a grooming switch, which processes the sub-rate signals carried on a wavelength in order to better pack the wavelengths. While a switch with an optical switching fabric can conceivably perform grooming, this function is typically performed in the electrical domain. Grooming switches are discussed further in Chap. 6.

2.11 Hierarchical or Multigranular Switches

One means of fabricating scalable optical switches is through the use of a hierarchical architecture where multiple switch granularities are supported [HSKO99; SaSi99; SaHa09; WaCa12]. Coarse switching is provided to alleviate requirements for switches with very large port counts; a limited amount of finer granularity switching is also provided. A functional illustration of a three-level hierarchical switch is shown in Fig. 2.30, where switching can be performed on a per-fiber basis, a per-waveband basis, or a per-wavelength basis. While the three levels are shown as distinct switches (or cross-connects), it is possible to build such a multigranular switch with a single switching fabric, e.g., with MEMS technology [LiVe02].

If the traffic at a node is such that all of the traffic entering from one fiber is directed to another fiber, then that traffic is passed through the fiber-level switch only. The switch provides configurability if the traffic pattern changes, while providing optical bypass for all traffic on the fiber. If fiber-level bypass is not appropriate for a network, then that level of the switching hierarchy can be removed.

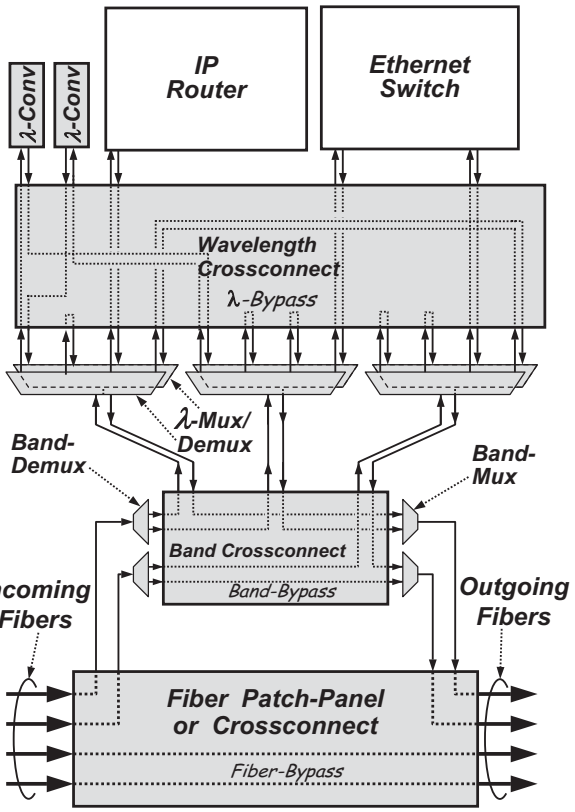


Fig. 2.30 A three-level hierarchical switch, allowing fiber-bypass, band-bypass, and wavelength-bypass. In the figure, each fiber contains two wavebands, and each waveband contains four wavelengths. Wavelength conversion (e.g., via back-to-back transponders) is used to groom the wavebands. (Adapted from Saleh and Simmons [SaSi99], © 1999 IEEE)

For traffic that needs to be processed on a finer granularity than a fiber, the band-level switch demultiplexes the WDM signal into its constituent wavebands. Some of the wavebands are switched without any further demultiplexing, providing band-level bypass, whereas some of the wavebands have to be further demultiplexed into their constituent wavelengths so that individual wavelengths can be dropped or switched. When equipped with wavelength converters (which are typically just back-to-back WDM transponders), the wavelength-level switch can also be used to better pack the wavebands (i.e., waveband grooming). Changing the frequency of a wavelength (i.e., a light channel) allows the wavelength to be shifted from one waveband to another.

The hierarchical approach addresses the port count issue while providing more flexibility than a single-layer switch of a coarse granularity. Studies have shown a significant number of ports can be saved through the use of a multigranularity switch, with the percentage of ports saved increasing with the level of traffic carried in the network [NoVD01; CaAQ04; YaHS08].

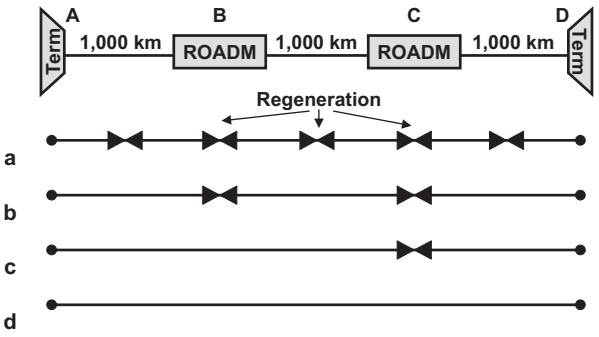


Fig. 2.31 Connection between Nodes A and D with: **a** 500-km optical reach; **b** 1,000-km optical reach; **c** 2,000-km optical reach; and **d** 3,000-km optical reach

2.12 Optical Reach

Having network elements capable of allowing transiting traffic to remain in the optical domain, such as the ROADM and ROADM-MD, is one requirement for optical bypass. In addition, the underlying transmission system must be compatible with a signal remaining in the optical domain as it traverses one or more nodes. An important property of a transmission system is the *optical reach,* which is the maximum distance an optical signal can be transmitted before it degrades to a level that requires the signal be regenerated. Regeneration typically occurs in the electrical domain. (All-optical regeneration, while feasible, has not been widely implemented. This is discussed further in Chap. 4.)

Consider the four-node linear network in Fig. 2.31. Nodes A and D are equipped with optical terminals, whereas Nodes B and C are equipped with ROADMs. The distance of each of the three links is 1,000 km. Assume that the connection of interest is between Node A and Node D.

In Fig. 2.31a, the optical reach is assumed to be 500 km. With this reach, not only must the connection be regenerated at Nodes B and C, but it must also be regenerated at intermediate dedicated regeneration sites along the links. These sites would otherwise be equipped with just an optical amplifier, but due to the limited reach need to regenerate all traffic that passes through them. The ROADMs at Nodes B and C, while capable of optical bypass, cannot be used for that purpose in this scenario because of the limited optical reach. In Fig. 2.31b, the optical reach is 1,000 km; this removes the regeneration at intermediate sites along the links, but still is not sufficient to allow the signal to optically bypass either Node B or C. With an optical reach of 2,000 km, as shown in Fig. 2.31c, the connection can make use of the ROADM at Node B to optically bypass that node, but it still must be regenerated at Node C (equivalently, it could be regenerated at Node B and optically bypass Node C). With an optical reach of 3,000 km, the connection is able to remain in the optical domain over the whole path, as shown in Fig. 2.31d.

As this example illustrates, the optical reach is critical in determining how much optical bypass is achieved in a network. In legacy transmission systems based on EDFA technology, the optical reach is on the order of 500–600 km; with newer

EDFA systems, the maximum reach is on the order of 1,500–2,500 km. To obtain significantly longer reach, Raman amplification is used (see Chap. 4). Raman systems generally have an optical reach in the range of 2,500–4,000 km, depending on the wavelength line rate and the equipment vendor. Such technology is sometimes referred to as "ultra-long-haul technology" to emphasize the extended optical reach. In typical backbone networks, the combination of optical-bypass elements and extended optical reach may eliminate on the order of 90% of the regenerations as compared to a system with no optical bypass and 500-km reach.

Many factors go into developing a transmission system that supports extended reach, some of which are touched on here; the subject is revisited in Chap. 4. As noted above, Raman amplification, sometimes in conjunction with EDFA amplification, is generally used in such systems. It is also necessary to deal with a host of optical impairments, such as chromatic dispersion, polarization-mode dispersion, linear crosstalk, four-wave mixing, and cross-phase modulation, all of which can degrade an optical signal (these are discussed in Chap. 4). Some of these impairments can be mitigated with special compensating equipment or with advanced receiver technology, while some of the problems from these impairments can be avoided by transmitting signals at a low enough power level (as long as the power level is still sufficiently larger than the noise level).

High-quality transmitters and receivers, as well as robust transmission schemes, are also important for attaining extended optical reach. Furthermore, the optical network elements themselves must be compatible with extended reach. For example, they must be of low loss, and not cause excessive distortion of the signal. Another key system component is advanced forward error correction (FEC). FEC allows errors picked up in transmission to be corrected at the destination. The stronger the FEC coding, the more errors can be corrected. This allows the signal to degrade further before it needs to be regenerated.

This list is not meant to be a comprehensive discussion of what goes into achieving extended optical reach. Suffice it to say it is quite an engineering accomplishment.

The desirable optical reach for a network depends on the geographic tier of where the system is deployed. In the metro-core, the longest connection paths are on the order of a few hundred kilometers. Thus, an optical reach of 500 km is sufficient to remove most, if not all, of the regeneration in the network. In regional networks, the longest connection paths are typically in the range of 1,000–1,500 km, requiring an optical reach of that order to eliminate regeneration. In backbone networks, the longest connection paths can be several thousand kilometers. For example, in the continental USA, the longest backbone connections, when including protection paths, are on the order of 8,000 km. A system with 4,000-km optical reach will eliminate much of the regeneration, but not all of it.

Note that the key factor in determining whether extended optical reach is useful in a particular network is the distribution of connection distances. A common mistake is to focus on the link distances instead, where it is erroneously assumed that extended optical reach provides no benefit unless the link distances are very long. In fact, optical bypass can be more effective in networks with high nodal density and relatively short links because a particular connection is likely to transit several

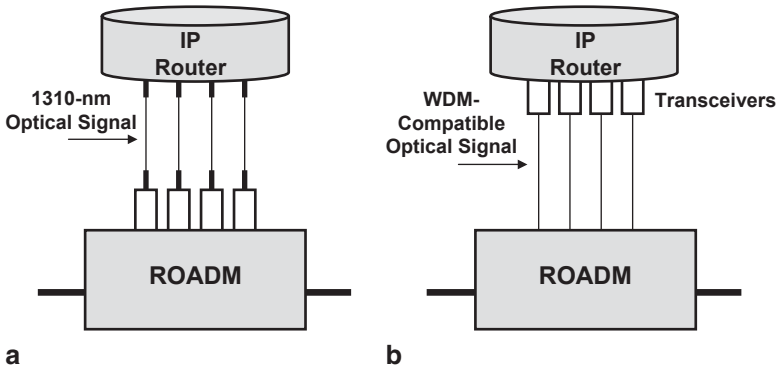


Fig. 2.32 In **a**, the Internet Protocol (*IP*) router communicates with the reconfigurable optical add/drop multiplexer (*ROADM*) via a standard 1,310-nm optical signal. In **b**, the electrical interfaces can be removed as wavelength-division multiplexing (*WDM*) transceivers are deployed directly on the IP router

intermediate nodes. Chapter 10 investigates the optimal optical reach of a network from a cost perspective.

Regarding terminology, it is perhaps not clear whether traffic that is regenerated at a node should be considered through traffic or add/drop traffic. As regeneration is typically accomplished via O-E-O means, it is usually considered add/drop traffic because it is dropping from the optical domain. This is the convention that is adopted here.

Chapter 4 discusses various regeneration architectures and strategies.

2.13 Integrating WDM Transceivers in the Client Layer

Much of the discussion so far has focused on removing the transponders for the traffic transiting a node, by implementing optical bypass. However, the add/drop traffic also provides an opportunity to remove some of the electronics from the node. The electronic higher layers and the optical layer typically communicate via a standard 1,310-nm optical signal. In Fig. 2.32a, the IP router is equipped with an interface to generate the 1,310-nm signal, and the WDM transponder plugged into the ROADM converts the 1,310-nm signal to a WDM-compatible signal.

The equipment can be simplified by having a WDM-compatible transceiver deployed directly on the IP router, as shown in Fig. 2.32b. This is referred to as *integrating* the transceivers with the IP router. (The term transceiver is used rather than transponder because the input to the transceiver is an electrical signal from the IP router rather than a 1,310-nm optical signal; i.e., there is no transponding of an optical signal.) The IP router and ROADM would then communicate via a WDM-compatible signal, which would be added to the WDM signal exiting the ROADM. This eliminates the electronic interfaces between the router and the ROADM. Clearly,

the transceiver output must meet the specifications of the WDM transmission system for this configuration to work.

Transceivers can be integrated with other electronic switching elements. For example, referring back to Fig. 2.29a, the combination of a WDM transponder and an electronic interface on each switch port of the O-E-O switch can be replaced by a transceiver that is integrated with the O-E-O switch, thereby eliminating a lot of electronics. Note, however, this integrated architecture negates one of the advantages of O-E-O switches, namely the independence of the switch and transmission-system vendors. With integrated transceivers, either one vendor supplies both the switch and the transmission system, or separate vendors collaborate to ensure compatibility. (Also see Sect. 5.10.)

Consider combining this integrated-transceiver architecture with a *non-directionless* ROADM. Assume that an edge switch is deployed at the node as discussed in Sect. 2.9.4.1 in order to achieve greater flexibility. The edge switch would be deployed between the transceivers and the ROADM, and would need to be capable of switching WDM-compatible signals. This is an example of where the architecture of Fig. 2.19b must be used as opposed to that of Fig. 2.19a.

2.14 Packet-Optical Transport

The packet-optical transport architecture takes integration much further, where a single platform supports switches in Layers 0–2 [Elby09b]. By taking advantage of current powerful processing capabilities, a single hybrid switch fabric performs both packet and circuit switching. For example, in a metro-core network, a packet-optical transport platform (P-OTP) (also known as a packet-optical transport system, P-OTS) might support a ROADM at Layer 0, a SONET/SDH switch at Layer 1, and an Ethernet switch at Layer 2. In a backbone network, the combination might be a ROADM, an OTN switch, and an MPLS router.

One motivation for a multilayer platform is to eliminate the transponders/transceivers that are necessary for communication between layer-specific platforms. A second driver is to streamline control across layers, to produce more efficient networks. With unified multilayer control, more of the switching can be pushed to lower layers, where it is more cost effective and consumes significantly less power. For example, the power consumption of a ROADM is approximately three orders of magnitude lower than that of an Ethernet switch or an IP router, on an energy-per-bit basis [Tuck11b]. Additionally, a single hybrid switch fabric can apportion the switch resources according to the mix of packet and circuit services, which is more efficient than deploying a dedicated switch platform per service.

We revisit the topics of multilayer traffic grooming and unified multilayer control in Chaps. 6 and 8, respectively.

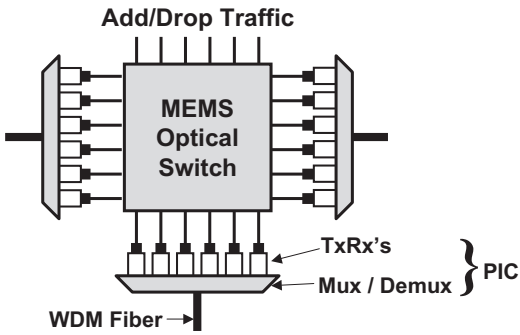


Fig. 2.33 A scalable OEO-O-OEO network node architecture, which combines photonic integrated circuit (*PIC*) and micro-electro-mechanical-system (*MEMS*) technologies. (Adapted from Saleh and Simmons [SaSi12], © 2012 IEEE)

2.15 Photonic Integrated Circuits

One of the motivations for removing electronics from a node is to reduce cost, power, and physical space requirements, and improve reliability. This led to the development of optical-bypass technology, which is now deployed in most major telecommunications carrier networks. An alternative approach is PIC technology used in combination with the more traditional O-E-O architecture [MDLP05, Welc06, Kish11]. The idea is not to eliminate transponders but to make them lower cost and smaller through integration.

Optical systems traditionally have been assembled using discrete components, e.g., each transponder may be a separate "pizza-box" sized card that is plugged into a chassis (e.g., Fig. 2.2). This contributes to the cost, power, space, and reliability issues associated with O-E-O technology. However, with PIC technology, numerous WDM components are monolithically integrated on a chip. For example, ten lasers, a multiplexer, and several control components may be integrated on a single PIC transmitter chip. Through integration, the cost, power, space, and reliability burdens are significantly reduced. A variety of materials can be used for PIC chips, including indium phosphide (InP).

Because the architecture remains O-E-O with this approach, there are no wavelength continuity constraints, thereby simplifying the algorithms needed to run the network. It also allows for performance monitoring at every node.

However, one bottleneck that the PIC-based O-E-O architecture has thus far not completely addressed is switching, particularly at large network nodes. Core switching remains in the electrical domain [Kish11; ELST12]. Thus, the scalability issues of core electronic switches are not eliminated. A more scalable option might be to combine MEMS technology with PIC technology [SaSi12], as shown in Fig. 2.33 (a *small* electronic edge switch could be added for grooming of low-rate traffic). Ideally, these two technologies can be combined, or integrated, in a compact, cost-effective way. The MEMS switch fabric is optical, thereby taking advantage of the scalability of optics. Note that this is an example of the OEO-O-OEO architecture of Fig. 2.29b.

While it is possible to achieve extended optical reach with PIC technology, thus far, commercial offerings have not supported optical bypass [Kish11]. In addi-

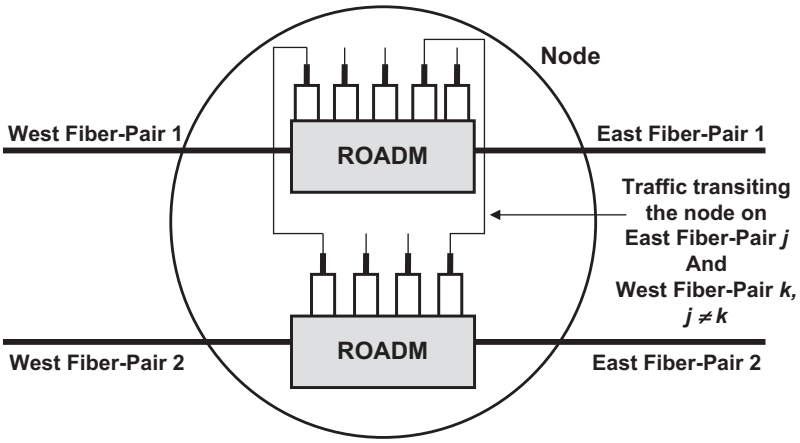


Fig. 2.34 Two reconfigurable optical add/drop multiplexers (*ROADMs*) deployed in parallel at a degree-two node with two fiber pairs per link

tion, commercial PIC transmitters and receivers can tune over just a limited range [Kish11]. Nevertheless, PIC technology is considered a key enabler of future flexible spectrum schemes. A typical approach in these schemes is to take advantage of *multi-carrier* technology, where multiple lower-rate subcarriers are combined to create a higher-rate signal. More specifically, a number of closely spaced (in frequency) subcarriers are generated to form a single "superchannel." By varying the number of subcarriers, the bandwidth of the superchannel can be adjusted, with relatively fine granularity. PICs are well suited to generating the comb of tightly spaced subcarriers needed in this methodology. Such technology is discussed in more detail in Chap. 9.

2.16 Multi-Fiber-Pair Systems

Thus far, the discussion has implicitly assumed that each network link is populated by one fiber pair. In this section, the multi-fiber-pair scenario is considered.

The O-E-O architecture does not change with multiple fiber pairs per link. There is an optical terminal for every fiber-pair incident on a node. The signal from every incoming fiber is fully demultiplexed and terminated on WDM transponders. In order to reduce the amount of required electronics, it may be possible to have an incoming fiber from one link be directly connected to an outgoing fiber on another link, so that fiber-bypass is achieved, though this arrangement is not readily reconfigurable.

For a network with optical bypass, there are a few options for architecting the node. First, assume that every link has N fiber pairs. One option is to deploy N copies of a network element at a node. For example, Fig. 2.34 shows a degree-two node with two fiber pairs per link. The node is equipped with two ROADMs. Optical

bypass is supported for traffic that bypasses the node using fiber pair 1 or fiber pair 2. However, traffic routed on fiber pair 1 of one link and fiber pair 2 of the other link requires O-E-O conversion.

Alternatively, if optical bypass is desired regardless of how the traffic is routed, then a degree-four ROADM-MD could be used. This may provide more flexibility than is required, however, as it allows a signal to be all-optically routed between fiber pair 1 of a link and fiber pair 2 of that *same* link. In the broadcast-and-select and route-and-select architectures, the ROADM-MD designs can be simplified to allow optical bypass in just the desired directions. For example, the non-directionless ROADM-MD architecture of Fig. 2.17 can be modified for a "quasi-degree-four" node as shown in Fig. 2.35. Support is provided for loopback on the same fiber pair, but not for loopback on different fiber pairs on the same link. 3×1 WSSs are used instead of the 4×1 WSSs that would be used in a fully functional degree-four non-directionless ROADM-MD. A similar type of simplification is possible with the directionless architecture of Fig. 2.12. (The route-and-select versions of these architectures can be similarly modified as well.)

If the number of fiber pairs on the links is unequal, due to some links being more heavily utilized, then either one large ROADM-MD could be deployed at a node to provide full optical bypass, or some combination of optical bypass and O-E-O could be used. For example, if a degree-two node has one link with two fiber pairs and one link with one fiber pair, a degree-three ROADM-MD would provide optical bypass in all directions. Again, this is more flexibility than is typically needed as it provides an all-optical path between the two fiber pairs that are on the same link; the ROADM-MD can be simplified, analogous to what is shown in Fig. 2.35. Another option is to deploy a ROADM in combination with an optical terminal, in which case the network planning process should favor routing the through traffic on the fiber pairs interconnected by the ROADM.

When considering the maximum desired size of a ROADM-MD, it is important to take into account the possibility of multiple fiber pairs. The maximum desired sizes of the various network elements may be larger than what is indicated by simply looking at the nodal-degree distribution. For example, some carriers require that ROADM-MDs be potentially scalable to a degree of ten, despite not having any nodes in their network with more than five incident links. This allows full optical bypass if, for example, two fiber pairs are eventually deployed on every link.

2.17 Exercises

2.1 The input/output configuration of a network device can be described by a table, where the (ith, jth) table entry corresponds to the output port to which the jth wavelength is directed when input on the ith input port. Note that if a wavelength is multicast by the device, then multiple output ports will be included in a single table entry. (a) Which of the tables shown below represent valid configurations for a typical 4×4 WSS? (b) The configuration table of an

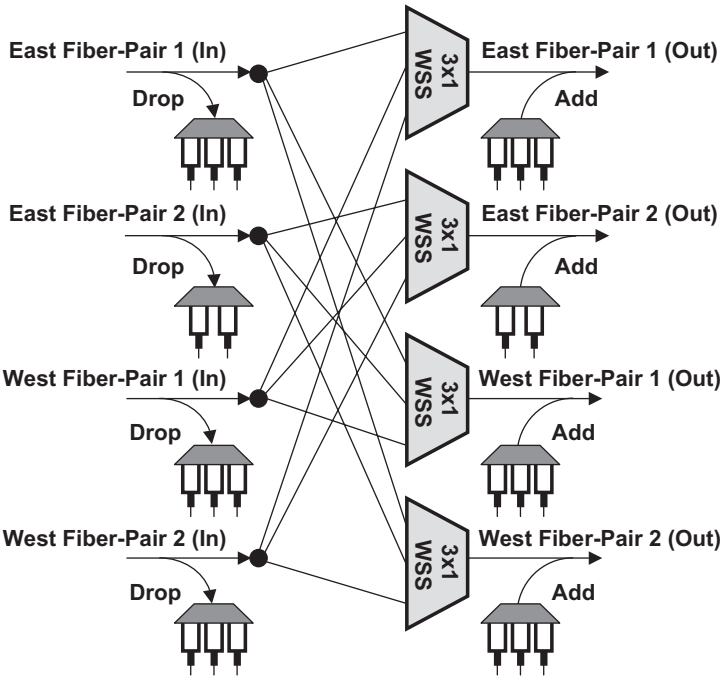


Fig. 2.35 A simplified non-directionless ROADM-MD at a "quasi-degree-four" node where routing between fiber pairs on the same link is not required. As compared to a full degree-four ROADM-MD, each through path is split three ways rather than four, and there are 3×1 (wavelength-selective switches) WSSs rather than 4×1 WSSs

$N \times N$ AWG, with N wavelengths, must satisfy the properties of a Latin Square; i.e., each output port must appear once and only once in each row and in each column of the table (and each table entry contains just one output port). Create a valid configuration table for a 4×4 AWG with 4 wavelengths; start with [1 2 3 4] as the first row and the first column. (c) Show the configuration table for a 1×4 passive splitter with three wavelengths.

Table 1		Wavelength #			
		1	2	3	4
Input Port	1	3	2	1	4
	2	2	3	4	1
	3	1	2	3	4
	4	4	1	2	3

Table 2		Wavelength #			
		1	2	3	4
Input Port	1	4	1	2	1
	2	3	4	1	2
	3	2	3	4	3
	4	1	2	3	4

Table 3		Wavelength #			
		1	2	3	4
Input Port	1	2,3	1	4	1
	2	4	1,4	2	3
	3	1	4	2,3	3
	4	1	2	3	1,2,4

2.2 With many of the components discussed in this chapter, an individual port may function as an input port or an output port depending upon the direction of light propagation (e.g., consider passive combiners and passive splitters). If such a device allows multiple input ports to direct the same wavelength to an output port, then what feature does it also support?

2.3 Define the *nodal drop ratio* as

$$\frac{Number\ of\ Wavelengths\ that\ Drop\ At\ Node}{Number\ of\ Wavelengths\ that\ Enter\ Node}.$$

In the following networks, assume that there is one wavelength of traffic in both directions between every pair of nodes. Assume that all nodes support optical bypass and that no regeneration is required. Assume shortest-path routing. What is the nodal drop ratio for: (a) Each node of an N-node ring, with N odd? (b) The center node of a linear chain of N nodes, with N odd? (c) The center node of an $N \times N$ grid, with N odd? In the $N \times N$ grid, assume that routing is along a row and then a column; note that bidirectional traffic may follow different paths in the two directions.

2.4 Consider a 3×5 grid network where every link is 1,000 km in length. Assume that all nodes support optical bypass and that the optical reach is 3,000 km. Assume that there is one wavelength of traffic in both directions between every pair of nodes. What is the average nodal drop ratio in this network, assuming all connections are routed over the shortest distance path? The average nodal drop ratio is defined as

$$\frac{\sum_i Number\ of\ Wavelengths\ that\ Drop\ At\ Node\ i}{\sum_i Number\ of\ Wavelengths\ that\ Enter\ Node\ i}.$$

2.5 Of the drop-port technologies shown in Fig. 2.3, which ones support multicasting a wavelength to more than one transponder on a given drop port? (Note that each of the diagrams in Fig. 2.3 represents a single drop port.)

2.6 Consider the optical-terminal architecture shown in Fig. 2.3c, with one passive splitter and multiple $1 \times N$ WSSs. An alternative architecture can be constructed with one $1 \times N$ WSS followed by multiple passive splitters. For example, if $4N$ transponders need to be supported on the terminal, the $1 \times N$ WSS directs 4 wavelengths to each output port, and the passive splitters are of size 1×4. Compare this architecture to that of Fig. 2.3c on criteria such as high-level cost, loss, multicasting capability, etc.

2.7 In the broadcast-and-select architecture of Fig. 2.12, $1 \times N$ passive splitters are on the input side, and $N \times 1$ WSSs are on the output side. What is potentially wrong with an architecture where there are $1 \times N$ WSSs on the input side and $N \times 1$ passive couplers on the output side? (Hint: Consider that when a WSS switches a wavelength from port i to port j (assume $i<j$) a small portion of the power may leak out on ports $i+1$ to $j-1$ during the switching process, depending on the WSS design.)

2.8 (a) Why is the wavelength-selective ROADM architecture of Fig. 2.14 colorless even though it uses AWGs for the multiplexers/demultiplexers rather than WSSs? (b) Is there any advantage to using WSSs for the multiplexers/demultiplexers instead of AWGs? Assume that $1 \times N$ WSSs with large N are possible. Hint: A $1 \times N$ WSS can direct multiple wavelengths from a WDM signal on its

input port to one of its output ports, whereas a typical AWG can direct only one wavelength from a WDM signal to each of its output ports. (c) If a ROADM at a degree-four node requires that up to 50% add/drop be supported *to/from each fiber,* what sized MEMS switch is required if WSSs are used as the mux/demux rather than AWGs? Assume that a fiber supports 80 wavelengths. (d) How about if up to 50% add/drop *at the node* is required?

2.9 Consider the broadcast-and-select architecture of Fig. 2.12 at a degree-two node with two add/drop ports. Modify the architecture in order to provide 1:4 protection for the WSSs (i.e., one spare WSS to protect the four primary WSSs). Discuss the trade-offs involved with this protected architecture. For example: How much does the loss of the through path increase (assume any 1:*N* splitters/couplers have an ideal loss of 1/*N*; any small switches have a loss of 1 dB)? How much does the availability of any one through path increase; assume that the WSS has an availability of 0.99999, any small switches have an availability of 0.9999996, and splitters/couplers have an availability of 1.0? (Availability is the probability of being in a working state at a given instant of time.)

2.10 Consider a dynamic network, where connection requests (each one requiring a full wavelength) arrive at a degree-two node equipped with a ROADM according to a Poisson process of 20 Erlangs. Assume that the connections are randomly routed on either of the two network links at the node, with equal probability. Assume that no regeneration occurs at the node. (a) First, assume that the ROADM is *not* directionless. How many transponders must be pre-deployed on the add/drop ports to yield a blocking probability (due to no available transponders) of less than 10^{-4}? (b) Second, assume that the ROADM is directionless. How many total transponder cards must be pre-deployed at the node to yield a blocking probability (due to no available transponders) of less than 10^{-4}?

2.11 Consider a degree-N node with *W* wavelengths per fiber, where up to a fraction *P* of the total wavelengths at the node may add/drop. Consider a core switch at the node that operates on all incoming/outgoing wavelengths and all add/drop wavelengths (e.g., the optical switch shown in Fig. 2.14), and an edge switch that operates only on the add/drop wavelengths (e.g., the edge switch shown in Fig. 2.19b). Both switches are assumed to have a per-wavelength granularity. What is the general formula for how large of a core switch is required at this node? What is the general formula for how large of an edge switch is required at this node? What is the ratio of the two switch sizes?

2.12 Consider the two architectures of Fig. 2.19, where an edge switch is used to add configurability to a non-directionless ROADM at a degree-three node. (a) First, assume that there is just one client (e.g., an IP router), as shown in the figure. Assume that connection requests (each one requiring a full wavelength) are generated by the client according to a Poisson process of 30 Erlangs, and assume that the connections are randomly routed on one of the three network links with equal probability (refer to this as traffic T). Calculate how many transponders must be deployed in the two architectures of Fig. 2.19 to yield a blocking probability (due to no available transponders) of less than 10^{-4}. (b)

Second, assume that there are four clients at the node, each of which (independently) generates traffic T. Again, calculate how many transponders must be deployed in the two architectures of Fig. 2.19 to yield a blocking probability (due to no available transponders) of less than 10^{-4}.

2.13 $M \times 3$ WSSs are used in Fig. 2.24 to provide the contentionless feature for the broadcast-and-select architecture. However, the feasible size of M may be small, thereby limiting the amount of add/drop transponders. Draw an architecture using $M \times 3$ WSSs that supports $3 \cdot M$ add/drop transponders at a node. Does the ROADM remain contentionless?

2.14 Consider a two-level waveband/wavelength hierarchical switch at a degree-four node. Assume that each fiber supports B wavebands. Assume that the switch is architected such that half of the wavebands on the input fibers (i.e., any 2B of the 4B wavebands) can be dropped from the waveband-level switch to the wavelength-level switch. Similarly, up to 2B wavebands can be added from the wavelength-level switch to the waveband-level switch. Client add/drop traffic enters/exits at the wavelength-level switch. Assume that a wavelength-selective architecture (e.g., MEMS-based) is used for both the waveband-level and wavelength-level switches. Draw an architecture for a switch that meets these specifications. Is this architecture colorless? Directionless? Contentionless?

2.15 Consider constructing an $M \times N$ WSS by cascading one $M \times 1$ WSS with one $1 \times N$ WSS. (a) Is this $M \times N$ WSS internally contentionless? (b) If not, can a contentionless $M \times N$ WSS be constructed from *multiple* $M \times 1$ WSSs and *multiple* $1 \times N$ WSSs? If so, how many of these WSSs are needed? (Note: For cost reasons, the $M \times N$ WSS is ideally constructed as an integrated component; the designs explored here are for investigating the functionality.)

2.16 Consider constructing a contentionless drop port for a CDC ROADM by combining AWG-based demultiplexers with a single fiber cross-connect (FXC). Draw this architecture for N network fibers, W wavelengths per fiber, and M transponders on the drop port. How large of an FXC would be required for a node with three network fibers, 80 wavelengths per fiber, and the ability to support a total of ten transponders of any wavelength? Is this an efficient architecture for a node with a small amount of drop traffic?

2.17 Draw an architecture for the $N \times M$ multicast switch (e.g., see Fig. 2.25), using small switches, passive splitters, and/or passive couplers. The architecture should allow for a signal on one of the N input fibers to be multicast to an arbitrary number of the M output ports. Assume that the receivers are equipped with a filter that can select any wavelength from a WDM signal.

2.18 Show the configuration table (see Exercise 2.1) corresponding to the 3×11 MCS shown in Fig. 2.25. Assume that there are four wavelengths. (Number the ports from top to bottom.)

2.19 Consider the contentionless broadcast-and-select ROADM with the $M \times N$ MCS. Which of the four multicast configurations listed in Sect. 2.9.8 can this architecture support?

2.20 The end of Sect. 2.9.5.2 describes a scenario with the ROADM of Fig. 2.12 where a particular wavelength is free on two of the incoming network fibers but is free on only one drop port, due to that wavelength having been multicast from one network fiber to two drop ports. The addition of an edge switch between the clients and the transponders does not address this potential contention scenario. Does having *more* add/drop ports than network links address this contention scenario, assuming the presence of an edge switch?

2.21 In an O-E-O network, there is an additional cost of one regeneration every time a connection is routed through a node. In a pure "all-optical" network (i.e., no regenerations), there is no additional cost incurred for routing a connection through a node. How might this affect the choice of network topology, e.g., in terms of fiber connectivity or network diameter, in the two architectures?

2.22 Research Suggestion: As described in Sect. 2.8.3, a wavelength-selective ROADM that utilizes AWGs for the mux/demuxes needs to support only a limited number of optical switch configurations. Investigate whether this limited connectivity allows more scalable optical switches to be used in the wavelength-selective architecture.

References

[BFAZ06] G. Baxter, S. Frisken, D. Abakoumov, H. Zhou, I. Clarke, A. Bartos, S. Poole, Highly programmable wavelength selective switch based on liquid crystal on silicon switching elements. Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'06), Anaheim, 5–10 Mar 2006, Paper OTuF2

[BSAL02] A. Boskovic, M. Sharma, N. Antoniades, M. Lee, Broadcast and select OADM nodes application and performance trade-offs, Proceedings, Optical Fiber Communication (OFC'02), Anaheim, 17–22 Mar 2002, Paper TuX2

[Busi00] Business Wire, Broadwing Communications to use Corvis technology for enhanced wavelength services, 10 May 2000. http://www.thefreelibrary.com/Business+Wire/2000/May/10-p53. Accessed 20 Mar 2014

[BuWW03] P. Bullock, C. Ward, Q. Wang, Optimizing wavelength grouping granularity for optical add-drop network architectures, Proceedings, Optical Fiber Communication (OFC'03), Atlanta, 23–28 Mar 2003, Paper WH2

[CaAQ04] X. Cao, V. Anand, C. Qiao, Multi-layer versus single-layer optical cross-connect architectures for waveband switching, Proc. IEEE INFOCOM 2004, **3**, 1830–1840 Hong Kong, 7–11 March 2004

[ChLH06] A. L. Chiu, G. Li, D.-M. Hwang, New problems on wavelength assignment in ULH networks, Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'06), Anaheim, 5–10 Mar 2006, Paper NThH2

[CoCo11] P. Colbourne, B. Collings, ROADM switching technologies, Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'11), Los Angeles, 6–10 Mar 2011, Paper OTuD1

[Coll09] B. C. Collings, Wavelength selectable switches and future photonic network applications, International Conference on Photonics in Switching, Pisa, 15–19 September 2009

[DoOk06] C. R. Doerr, K. Okamoto, Advances in silica planar lightwave circuits, J. Lightwave Technol. 24(12), 4763–4789 (December 2006)

[ElBa06] T. S. El-Bawab, (ed.), *Optical Switching*, (Springer, New York, 2006)

[Elby09b] S. Elby, The future Internet—a service provider's long term view, IEEE/LEOS Summer Topicals Meeting, Newport Beach, 20–22 Jul 2009, 137–138
[ELST12] V. Eramo, M. Listanti, R. Sabella, F. Testa, Definition and performance evaluation of a low-cost/high-capacity scalable integrated OTN/WDM switch, J. Optic. Commun. Netw. **4**(12), 1033–1045, (December 2012)
[Fini13] Finisar, Wavelength selective switches for ROADM applications, Product Data Sheet, August 2013
[FoRN12] N. K. Fontaine, R. Ryf, D. T. Neilson, N × M wavelength selective crossconnect with flexible passbands, Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'12), Los Angeles, 4–8 March 2012, Paper PDP5B.2
[FWPW11] M. D. Feuer, S. L. Woodward, P. Palacharla, X. Wang, I. Kim, D. Bihon, Intra-node contention in dynamic photonic networks, J. Lightwave Technol. **29**(4), 529–535, (15 February 2011)
[GJLY12] O. Gerstel, M. Jinno, A. Lord, S. J. B. Yoo, Elastic optical networking: A new dawn for the optical layer? IEEE Communications Magazine, **50**(2), S12–S20, (February 2012)
[HSKO99] K. Harada, K. Shimizu, T. Kudou, T. Ozeki, Hierarchical optical path cross-connect systems for large scale WDM networks, Proceedings, Optical Fiber Communication (OFC'99), San Diego, 21–26 February 1999, Paper WM55
[HSLD12] Y. Hsueh, A. Stark, C. Liu, T. Detwiler, S. Tibuleac, M. Filer, G. K. Chang, S. E. Ralph, Passband narrowing and crosstalk impairments in ROADM-enabled 100G DWDM networks, J. Lightwave Technol. **30**(24), 3980–3986, (15 December 2012)
[IGKV03] R. Izmailov, S. Ganguly, V. Kleptsyn, A. C. Varsou, Non-uniform waveband hierarchy in hybrid optical networks, Proceedings, IEEE INFOCOM 2003, **2**, 1344–1354. San Francisco, 30 March – 3 April, 2003
[ITU02] International Telecommunication Union, Spectral grids for WDM applications: DWDM Frequency Grid, ITU-T Rec. G.694.1, Edition 1.0, June 2002
[ITU12b] International Telecommunication Union, Spectral grids for WDM applications: DWDM Frequency Grid, ITU-T Rec. G.694.1, Edition 2.0, February 2012
[Jinn08] M. Jinno, et al., Demonstration of novel spectrum-efficient elastic optical path network with per-channel variable capacity of 40 Gb/s to over 400 Gb/s, Proceedings, European Conference on Optical Communication (ECOC'08), Brussels, 21–25 September 2008, Paper Th.3.F.6
[Kish11] F. A. Kish, et al., Current status of large-scale InP photonic integrated circuits, IEEE J. Selected Topics in Quantum Electronics. **17**(6), 1470–1489, (November/December 2011)
[KPWB12] I. Kim, P. Palacharla, X. Wang, D. Bihon, M. D. Feuer, S. L. Woodward, Performance of colorless, non-directional ROADMs with modular client-side fiber cross-connects, Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'12), Los Angeles, 4–8 March 2012, Paper NM3F.7
[KZJP04] V. Kaman, X. Zheng, O. Jerphagnon, C. Pusarla, R. J. Helkey, J. E. Bowers, A cyclic MUX–DMUX photonic cross-connect architecture for transparent waveband optical networks, IEEE Photonics Technol. Lett. **16**(2), 638–640, (February 2004)
[LiVe02] R. Lingampalli, P. Vengalam, Effect of wavelength and waveband grooming on all-optical networks with single layer photonic switching, Proceedings, Optical Fiber Communication (OFC'02), Anaheim, 17–22 March 2002, Paper ThP4
[MaLe03] B. Manseur, J. Leung, Comparative analysis of network reliability and optical reach, National Fiber Optic Engineers Conference (NFOEC'03), Orlando, 7–11 September 2003
[Maro05] D. M. Marom, et al., Wavelength-selective 1 x K switches using free-space optics and MEMS micromirrors: Theory, design, and implementation, J. Lightwave Technol. **23**(4), 1620–1630, (April 2005)
[MaSi12] D. M. Marom, D. Sinefeld, Beyond wavelength-selective channel switches: Trends in support of flexible/elastic optical networks, Proceedings, International Conference on Transparent Optical Networks (ICTON'12), United Kingdom, 2–5 July 2012, Paper Mo.B1.4
[MDLP05] S. Melle, R. Dodd, C. Liou, D. Perkins, M. Sosa, M. Yin, Network planning and economic analysis of an innovative new optical transport architecture: The digital optical network,

National Fiber Optic Engineers Conference (NFOEC'05), Anaheim, 6–11 March 2005, Paper NTuA1

[MMMT03] S. Mechels, L. Muller, G. D. Morley, D. Tillett, 1D MEMS-based wavelength switching subsystem, IEEE Commun. Magazine, **41**(3), 88–94, (March 2003)

[NoVD01] L. Noirie, M. Vigoureux, E. Dotaro, Impact of intermediate traffic grouping on the dimensioning of multi-granularity optical networks, Proceedings, Optical Fiber Communication (OFC'01), Anaheim, 17–22 March 2001, Paper TuG3

[NYHS12] F. Naruse, Y. Yamada, H. Hasegawa, K. Sato, Evaluations of OXC hardware scale and network resource requirements of different optical path add/drop ratio restriction schemes, J. Optic. Commun. Netw. **4**(11), B26-B34, (November 2012)

[Okam98] K. Okamoto, Tutorial: Fundamentals, technology and applications of AWG's, Proceedings, European Conference on Optical Communication (ECOC'98), Madrid, 20–24 September 1998, pp. 35–37

[PaPP03] G. I. Papadimitriou, C. Papazoglou, A. S. Pomportsis, Optical switching: Switch fabrics, techniques, and architectures, J. Lightwave Technol. **21**(2), 384–405, (February 2003)

[RaSS09] R. Ramaswami, K. N. Sivarajan, G. Sasaki, *Optical networks: A practical perspective*, 3rd edn. (Morgan Kaufmann Publishers, San Francisco, 2009)

[RFDH99] B. Ramamurthy, H. Feng, D. Datta, J. P. Heritage, B. Mukherjee, Transparent vs. opaque vs. translucent wavelength-routed optical networks, Proceedings, Optical Fiber Communication (OFC'99), **1,** 59–61 San Diego, 21–26 February 1999, Paper TuF2

[SaHa09] K. Sato, H. Hasegawa, Optical networking technologies that will create future bandwidth-abundant networks, J. Optic. Commun. Netw. **1**(2), A81-A93, (July 2009)

[Sala02] D. Y. Al-Salameh, Optical switching in transport networks: Applications, requirements, architectures, technologies and solutions, in *Optical Fiber Telecommunications IV A*, ed. by I. Kaminow, T. Li, (Academic Press, San Diego, 2002), pp. 295–373

[SaSi99] A. A. M. Saleh, J. M. Simmons, Architectural principles of optical regional and metropolitan access networks, J. Lightwave Technol., **17**(12), 2431–2448, (December 1999)

[SaSi12] A. A. M. Saleh, J. M. Simmons, All-optical networking—evolution, benefits, challenges, and future vision, Proc. IEEE, **100**(5), 1105–1117, (May 2012)

[ShTu07] G. Shen, R. Tucker, Translucent optical networks: The way forward, IEEE Commun. Magazine. **45**(2), 48–54, (February 2007)

[Simm05] J. M. Simmons, On determining the optimal optical reach for a long-haul network, J. Lightwave Technol. **23**(3), 1039–1048, (March 2005)

[SiSa07] J. M. Simmons, A. A. M Saleh, Network agility through flexible transponders, IEEE Photonics Technol. Lett. **19**(5), 309–311, (1 March 2007)

[StWa10] T. A. Strasser, J. L. Wagener, Wavelength-selective switches for ROADM applications, IEEE J. Selected Topics in Quantum Electronics, **16**(5), 1150–1157, (September/October 2010)

[Telc05a] Telcordia Technologies, *NEBS™ requirements: Physical protection*, GR-63-CORE, Issue 4, Apr. 2005.

[Tuck11b] R. S. Tucker, Green optical communications—Part II: Energy limitations in networks, IEEE Journal of Selected Topics in Quantum Electronics, **17**(2), 261–274, (March/April 2011)

[TzZT03] A. Tzanakaki, I. Zacharopoulos, I. Tomkos, Optical add/drop multiplexers and optical cross-connects for wavelength routed networks, Proceedings, International Conference on Transparent Optical Networks (ICTON'03), Warsaw, 29 Jun–3 Jul, 2003, pp. 41–46

[WaCa12] Y. Wang, X. Cao, Multi-granular optical switching: A classified overview for the past and future, IEEE Communications Surveys & Tutorials, **14**(3), 698–713, Third Quarter, (2012)

[Way12] W. I. Way, Optimum architecture for M × N multicast switch-based colorless, directionless, contentionless, and flexible-grid ROADM, Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'12), Los Angeles, 4–8 March 2012, Paper NW3F.5

[Welc06] D. F. Welch, et al., The realization of large-scale photonic integrated circuits and the associated impact on fiber-optic communication systems, J. Lightwave Technol. **24**(12), 4674–4683, (December 2006)

[WFJA10] S. L. Woodward, M. D. Feuer, J. L. Jackel, A. Agarwal, Massively-scaleable highly-dynamic optical node design, Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'10), San Diego, 21–25 March 2010, Paper JThA18

[WuSF06] M. C. Wu, O. Solgaard, J. E. Ford, Optical MEMS for lightwave communication, J. Lightwave Technol. **24**(12), 4433–4454, (December 2006)

[YaHS08] I. Yagyu, H. Hasegawa, K. Sato, An efficient hierarchical optical path network design algorithm based on a traffic demand expression in a Cartesian product space, IEEE Journal on Selected Areas in Communications, **26**(6), 22–31, (August 2008)

[ZhNe12] X. Zhou, L. E. Nelson, 400G WDM transmission on the 50 GHz grid for future optical networks, J. Lightwave Technol. **30**(24), 3779–3792, (15 December 2012)

Chapter 3
Routing Algorithms

3.1 Introduction

Telecommunications networks are generally so large and complex that manually designing a network in a reasonable amount of time is prohibitively difficult. Network designers primarily rely on automated algorithms to determine, for example, how to route traffic through the network, how to protect the traffic, and how to bundle the traffic into wavelengths. In systems with optical bypass, additional algorithms are needed to handle regeneration and to ensure that wavelength contention issues are minimal. The fact that networks have reached a stage where algorithms are essential in producing cost-effective and efficient network designs can be daunting. The good news is that extensive research has been done in this area and much expertise has been gained from actual network deployments, resulting in the development of relatively straightforward algorithms that produce very effective network designs.

When designing network algorithms, it is important to consider the size of the problem in terms of the number of network nodes, the amount of traffic carried in the network, and the system specifications. Metro-core networks have tens of nodes whereas backbone networks may have as many as 100 nodes or more. The size of the demand set depends on whether the traffic requires grooming or not. First, consider a backbone network. If all of the traffic is at the wavelength line rate (no grooming needed), then there are typically a few hundred to a couple of thousand demands in the network. If all of the traffic is subrate, such that grooming is needed, then there could be tens of thousands of demands. The number of demands in a metro-core network is significantly lower than this. Another key system parameter is the number of wavelengths per fiber. Metro-core WDM networks generally have no more than 40 wavelengths per fiber, whereas backbone networks typically have 80 (or as many as 160) wavelengths per fiber. Any algorithms used in the network planning process must be scalable under these conditions.

The run time of the network planning algorithms is very important. In a dynamic real-time environment, a new connection may need to be established in less than 1 s. The process of planning the route of the connection and determining which network resources should be allocated to it may need to be completed in less than 100 ms to

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4_3,

allow time for the network to be configured appropriately to carry the connection. Furthermore, real-time design may be implemented in a distributed manner at the network nodes, where processing and memory capabilities may be limited.

In long-term network planning, design time is not as critical, although it is still important. In some long-term planning exercises, a large number of demands (e.g., thousands of subrate demands) may need to be processed at once. Furthermore, if working with a "greenfield" network (i.e., a completely new network), numerous design scenarios typically are considered. For example, the process may include comparing different network topologies, different line rates, or different protection strategies. Due to the number of designs that need to be run, and the often short time allocated to the network design process (especially when the design exercise is performed by a system vendor in response to a carrier request), it is desirable that the planning process for each scenario take no more than a couple of minutes.

Producing a network design that is *optimal* relative to a set of metrics is typically computationally complex and would require an inordinate amount of time for realistic networks. Thus, many steps in the planning process rely on heuristics, which experience has shown produce very good, though not always optimal, results and which run in a reasonable amount of time.

A brief overview of the network planning process can be found in Simmons [Simm06]. This chapter focuses on the routing component of the algorithms, whereas Chaps. 4 and 5 discuss regeneration and wavelength assignment, respectively. The initial emphasis is on treating these three components as sequential steps in the planning process; however, performing routing, regeneration, and wavelength assignment in a single step is covered in Chap. 5.

Routing is the process of selecting a path through the network for a traffic demand, where there are typically many possible paths to get from the demand source to the demand destination. It is important to take into account several factors when selecting a route. First, cost is a key consideration. The selected route should require adding minimal cost to the network when possible. The path distance and number of links in the path may also be relevant, as these are indicators of the bandwidth occupied by the path; these factors also may affect the reliability of the connection. However, this does not imply that demands should always be routed over the shortest possible path or the path with the fewest links. In fact, such a strategy may lead to inefficient designs. It is also necessary to consider the total potential capacity of the network, where a particular path may be chosen because it leaves the network in a better state to accommodate future traffic.

Several routing strategies are discussed in this chapter, with a focus on relatively straightforward methodologies that are effective in network planning for practical optical networks. The chapter is not meant to be a review of all known routing algorithms and strategies. Much of the chapter is equally applicable to a network with optical bypass as to a network based on optical-electrical-optical (O-E-O) technology.

The first few sections of the chapter specifically examine routing a demand with one source and one destination over a single path. Sect. 3.2 introduces shortest-path

routing algorithms, and Sect. 3.3 covers how such algorithms are used depending on the underlying technology of the network. Sects. 3.4 and 3.5 describe some effective routing strategies that take into account network cost and network utilization. Sect. 3.6 considers the more detailed network modeling that may be needed when routing demands in real time, where only the equipment that is already deployed in the network can be used. Other aspects of real-time routing, such as distributed path computation, contention due to concurrent demand requests, and routing with stale network state information, are covered in Chap. 8 on dynamic networking.

Routing with protection is covered in Sect. 3.7, where two or more diverse paths are required for a demand to enable recovery from a failure. Specific protection schemes, however, are not covered until Chap. 7.

Section 3.8 is relevant to planning scenarios where multiple demand requests are processed at one time. Assuming that the routing policy is adaptive, where the selected path is dependent on the state of the network, the order in which the demands are routed is important. Various effective ordering strategies are presented in this section.

Section 3.9 provides an overview of flow-based routing, which relies on linear programming techniques for tractability. This is still an active area of research.

Multicast routing, from one source to multiple destinations, is covered in Sect. 3.10. This includes a discussion on a multicast variant known as "manycast," where one source communicates with *any N* of the *M* possible destination nodes, for some specified *N* and *M*. Finally, Sect. 3.11 covers multipath routing, where a demand is split into multiple lower-rate streams and routed over more than one path. The key challenge is finding a path set where the difference in delay among the paths is below an acceptable threshold. Several methodologies for finding a feasible path set are presented.

3.2 Shortest-Path Algorithms

Most routing strategies incorporate some type of shortest-path algorithm to determine which path minimizes a particular metric. A general discussion of shortest-path algorithms can be found in Cormen et al. [CLRS09]. In the shortest-path algorithms discussed here, it is assumed that the metric for an end-to-end path is the sum of the metrics of the links comprising the path. Any such additive metric can be used, depending on the goal of the routing process. For example, to find the path with the shortest geographic distance, each link is assigned a metric equal to its own distance. As another example, assume that each link is assigned a metric of unity. The shortest-path algorithm then finds the path that traverses the fewest hops. (The term *hop* is often used to refer to each link in a path.) As a third example, assume that each network link has a certain probability of being available (i.e., not failed), and assume that each link in the network fails independently such that the availability of a path is the product of the availabilities of each link in the path (ignoring node failures). The link metric can be chosen to be the negative of the logarithm

of the link availability, where the logarithm function is used in general to convert a multiplicative metric to an additive metric. Higher link availability corresponds to a smaller link metric; thus, running the shortest-path algorithm with this metric produces the path with the highest availability. As this last example demonstrates, the metric may be unrelated to distance; thus, the term "shortest path" is in general a misnomer. Nevertheless, this term will be used here to represent the path that minimizes the desired metric.

One well-known shortest-path algorithm is the Dijkstra algorithm, where the inputs to the algorithm are the network topology, the source, and the destination. This is a "greedy" algorithm that is guaranteed to find the shortest path from source to destination, assuming a path exists. Greedy algorithms proceed by choosing the optimal option at each step without considering future steps. In the case of the Dijkstra algorithm, this strategy produces the optimal overall result. (In general, however, greedy algorithms do not always yield the optimal solution.)

Another shortest-path algorithm is the breadth-first-search (BFS) algorithm, which proceeds by considering all one-hop paths from the source, then all two-hop paths from the source, etc., until the shortest path is found from source to destination [Bhan99]. If there is a unique shortest path from source to destination, the BFS and Dijkstra algorithms produce the same result. However, if there are multiple paths that are tied for the shortest, the BFS algorithm finds the shortest path with the fewest number of hops. This can be helpful in network planning because fewer hops can potentially translate into lower cost or less wavelength contention, as is discussed in the next section. The Dijkstra algorithm does not in general have this same tie-breaking property. Furthermore, the BFS algorithm works with negative link metrics, as long as there are no cycles in the network where the sum of the link metrics is negative. This is relevant for one of the graph transformations that is commonly used as part of an algorithm to find two or more diverse routes, as discussed in Sect. 3.7. The Dijkstra algorithm needs a small modification to be used with negative link metrics. Overall, then, the BFS shortest-path algorithm is somewhat better suited for network planning; code for this algorithm is provided in Chap. 11.

The Dijkstra and BFS algorithms can be applied whether or not the links in the network are bidirectional. A link is bidirectional if traffic can be routed in either direction over the link. Furthermore, the algorithms work if different metrics are assigned to the two directions of a bidirectional link. However, if the network is bidirectionally symmetric, such that the traffic flow is always two-way and such that the metrics are the same for the two directions, then a shortest path from source to destination also represents, in reverse, a shortest path from destination to source. (This is also called an *undirected* network.) In this scenario, which is typical of telecommunications networks, it does not matter which endpoint is designated as the source and which is designated as the destination.

The shortest-path algorithm can be incorporated as part of a larger procedure to find the *K*-shortest paths (KSP). KSP algorithms find the shortest path between the source and destination, the second shortest path, etc., until the *K*th shortest path is found or until no more paths exist. Note that the paths that are found are not necessarily completely disjoint from each other, i.e., the paths may have links

and/or nodes in common. Many KSP algorithms exist, e.g., Yen [Yen71], Eppstein [Epps94], where the ones that find only simple paths (i.e., paths without loops) are the most relevant for network design. The code for one such KSP algorithm is provided in Chap. 11 (the code follows the procedure described in Hershberger et al. [HeMS03]).

A variation of the shortest-path problem arises when one or more constraints are placed on the desired path; this is known as the *constrained shortest-path* (CSP) problem. Some constraints are straightforward to handle. For example, if one is searching for the shortest-path subject to all links of the path having at least N wavelengths free, then prior to running a shortest-path algorithm, all links with fewer than N free wavelengths are removed from the topology. As another example, the intermediate steps of the BFS shortest-path algorithm can be readily used to determine the shortest-path subject to the number of path hops being less than H, for any $H>0$ (similar to Guerin and Orda [GuOr02]). However, more generally, the CSP problem can be difficult to solve, e.g., determining the shortest path, subject to the availability of the path being greater than some threshold, where the availability is based on factors other than distance. Various heuristics have been proposed to address the CSP problem, e.g., Korkmaz and Krunz [KoKr01], Liu and Ramakrishnan [LiRa01] (the latter reference addresses the constrained KSP problem). Some heuristics have been proposed to specifically address the scenario where there is just a single constraint; this is known as the *restricted shortest-path* (RSP) problem. Additionally, a simpler version of the multi-constraint problem arises when *any* path satisfying all of the constraints is desired, not necessarily the shortest path; this is known as the *multi-constrained path* (MCP) problem. An overview, including a performance comparison, of various heuristics that address the RSP and MCP problems can be found in Kuipers et al. [KKKV04].

3.3 Routing Metrics

As discussed in the previous section, a variety of metrics can be used with a shortest-path algorithm. Two common strategies are find the path with the fewest hops and find the path with the shortest distance. With respect to minimizing network cost, the optimal routing strategy to use is dependent on the underlying system technology, as discussed next. (In this section, issues such as grooming and shared protection that also may have an impact on cost are not considered; these issues are discussed in Chaps. 6 and 7, respectively.)

3.3.1 Minimum-Hop Path Versus Shortest-Distance Path

In a pure O-E-O network, a connection is electronically terminated (i.e., regenerated) at every intermediate node along its path, where the electronic terminating equipment is a major component of the path cost, and is typically a major source of

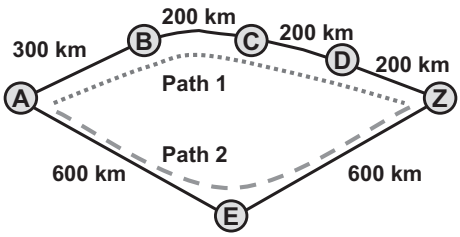


Fig. 3.1 *Path 1, A-B-C-D-Z,* is the shortest-distance path between Nodes *A* and *Z,* but *Path 2, A-E-Z,* is the fewest-hops path. In an O-E-O network, where the signal is regenerated at every intermediate node, *Path 2* is typically the lower-cost path

failures along a path as well. Thus, searching for the path from source to destination with the fewest hops is generally favored as it minimizes the amount of required regeneration. This is illustrated in Fig. 3.1 for a connection between Nodes A and Z. Path 1 is the shortest-distance path at 900 km, but includes four hops. Path 2, though it has a distance of 1,200 km, is typically lower cost because it has only two hops and thus requires fewer regenerations.

In networks with optical bypass, regeneration is determined by the system optical reach, which is typically based on distance. For example, an optical reach of 2,000 km indicates the connection can travel no further than 2,000 km before it needs to be regenerated. (In reality, the optical reach is determined by many factors as is discussed in Chap. 4, but for simplicity, it is usually specified in terms of a distance.) This favors searching for the shortest-distance path between the source and destination. However, with optical-bypass systems, there is a wavelength continuity constraint, such that the connection must remain on the same wavelength (i.e., lambda) as it optically bypasses nodes. Finding a wavelength that is free to carry the connection is potentially more difficult as the number of links in the path increases. This implies that the number of path hops should be considered as well.

Overall, a good strategy for optical-bypass systems is to search for a route based on distance, but of the paths that meet the minimal regeneration, favor the one with the fewest hops. This is illustrated by the three paths between Nodes A and Z shown in Fig. 3.2, where the optical reach is assumed to be 2,000 km. Path 1, with

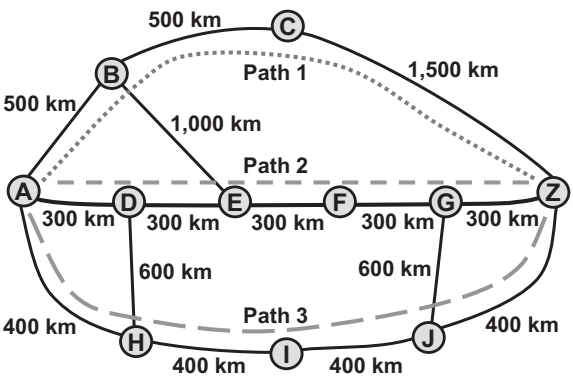


Fig. 3.2 Assume that this is an optical-bypass-enabled network with an optical reach of 2,000 km. *Path 1, A-B-C-Z,* has the fewest hops but requires one regeneration. *Path 2, A-D-E-F-G-Z,* and *Path 3, A-H-I-J-Z,* require no regeneration. Of these two lowest-cost paths, *Path 3* is preferred because it has fewer hops

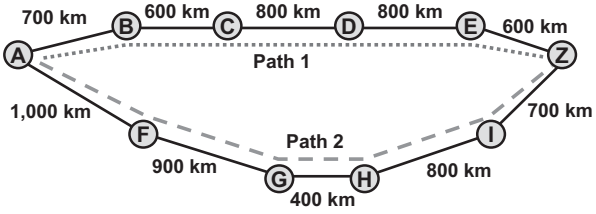


Fig. 3.3 Assume that this is an optical-bypass-enabled network with an optical reach of 2,000 km. *Path 1* is 3,500 km, but requires two regenerations. *Path 2* is longer at 3,800 km but requires only one regeneration. (Adapted from Simmons [Simm06]. © 2006 IEEE)

a distance of 2,500 km, has the fewest hops but requires one regeneration. Path 2, with a distance of 1,500 km, and Path 3, with a distance of 1,600 km, do not require any regeneration and are thus lower-cost paths. Of these two paths, Path 3 is generally more favorable (all other factors, e.g., link load, being equal), even though it is somewhat longer than Path 2, because it has only four hops compared to the five hops of Path 2.

3.3.2 Shortest-Distance Path Versus Minimum-Regeneration Path

While path distance is clearly related to the amount of regeneration in an optical-bypass-enabled network, it is important to note that the path with shortest physical distance is not necessarily the path with minimum regeneration. In addition to considering the distance over which a signal has traveled in determining where to regenerate, carriers generally require that any regeneration occur in a network node (i.e., the add/drop and switching locations of a network), as opposed to at an arbitrary site along a link. First, sites along a link, e.g., amplifier huts,[1] may not be large enough to house regeneration equipment. Second, from a maintenance perspective, it is beneficial to limit the number of locations where regeneration equipment is deployed. (If the optical reach of the system is shorter than the distance between two nodes, then there is no choice but to deploy a dedicated regeneration site along the link where all transiting traffic is regenerated. In this scenario, the dedicated regeneration site can be considered a network node.)

This additional design constraint can lead to more regeneration than would be predicted by the path length. Consider the example shown in Fig. 3.3, with a connection between Node A and Node Z, and assume that the optical reach is 2,000 km. Path 1 is the shortest path for this connection, with a length of 3,500 km. Based on the path length and the optical reach, it is expected that one regeneration would be required. However, because regeneration occurs only at

[1] Typically small buildings that house the optical amplifiers.

Table 3.1 Percentage of shortest-distance paths that required an extra regeneration

Network	# of nodes	Optical reach			
		1,500 km	2,000 km	2,500 km	3,000 km
1	75	7%	1.5%	0.9%	0.4%
2	60	2%	0.8%	0.5%	0.5%
3	30	1.4%	3%	1.1%	0.2%

nodes, two regenerations are actually required (e.g., the regenerators could be at Nodes C and E). Path 2, while longer, with a length of 3,800 km, requires just one regeneration (at Node G).

To get an idea of how often this phenomenon occurs, we examined the three reference backbone networks of Sect. 1.10. For each possible source/destination pair, we compared the minimum number of required regenerations over all paths to the number of required regenerations in the shortest path. A range of optical reach assumptions were considered. Table 3.1 indicates the percentage of source/destination pairs where the shortest path required one additional regeneration as compared to the path that required the minimum number of regenerations. (No shortest path required more than one extra regeneration.) In the worst case, 7% of the shortest paths in Network 1 required one extra regeneration, assuming 1,500 km optical reach.

Another factor that affects the amount of regeneration is whether the network fully supports optical bypass in all directions at all nodes. There may be some nodes not fully equipped with optical-bypass equipment. For example, a degree-three node may be equipped with a reconfigurable optical add/drop multiplexer (ROADM) and an optical terminal (see Sect. 2.7); any transiting traffic entering via the optical terminal needs to be regenerated regardless of the distance over which it has been transmitted.

As the examples of this section illustrate, simply running a shortest-path algorithm may not produce the most desirable route. Rather, it is often advantageous to generate a *set* of candidate paths, and then select a particular path to use based on other factors. For example, a particular route may be selected because of its lower cost or because it avoids a "bottleneck" link that is already heavily loaded.

The next section discusses how to generate a good set of candidate paths, and Sect. 3.5 discusses how the path set is used in the routing process.

3.4 Generating a Set of Candidate Paths

Two different strategies are presented for generating a set of candidate paths. The first is a more formal strategy that uses a KSP algorithm to find minimum-cost paths. The second is a somewhat ad hoc methodology to find a set of paths with good link diversity.

3.4.1 K-Shortest Paths Strategy

The marginal cost of adding a new demand to the network is largely a function of the amount of electronic terminating equipment needed to support the end-to-end connection. This, in turn, is dependent upon the number of regenerations that are required. Most of the other network costs, e.g., amplification, are incurred when the network is first installed and are amortized over the whole demand set.

Thus, roughly speaking, in an O-E-O network, paths that have the same number of hops have an equivalent cost because the number of required regenerations is the same. To find a set of lowest-cost paths, one can run a KSP algorithm with the metric set to unity for all links. The first N of the paths returned by the algorithm will satisfy the minimum-hop criterion, for some N, where $1 \leq N \leq$ K. Setting K to around 10 will typically ensure that all, or almost all, lowest-cost paths are found. In the three reference networks of Sect. 1.10, setting K equal to 10 finds *all* minimum-hop paths for 99.5 % of the source/destination pairs in Network 1, 99.9 % of the pairs in Network 2, and 100 % of the pairs in Network 3.

Similarly, in an optical-bypass-enabled network, paths with the same amount of regeneration can be considered equivalent-cost paths. First, consider networks where any loopless path is shorter than the optical reach, as may be the case in a metro-core network. All paths can be considered lowest-cost because there is no need for regeneration in the network. In this scenario, one can run a KSP algorithm with a link metric of unity to generate a set of lowest-cost paths that have the fewest, or close to the fewest, number of hops.

In more general optical-bypass-enabled networks where there is regeneration, it is not quite as straightforward to find a set of paths that meet the minimum regeneration. A KSP algorithm can be run with distance as the link metric. However, as was illustrated by the example of Fig. 3.3, each of the returned paths must be examined to determine the actual number of required regenerations. There is no guarantee that running the KSP algorithm with some fixed value of K finds a path with the fewest possible number of required regenerations. (In Chap. 4, an alternative link metric that is more tied to the underlying optical system is presented that may be a better predictor of regeneration.) However, if the minimum number of regenerations found in any path returned by the KSP algorithm is R, and at least one of the paths found by the KSP algorithm has a distance greater than $(R+1)$*[Optical Reach], then the set of minimum-regeneration paths must have been found (see Exercise 3.3). Practically speaking, setting K to about 10 in the KSP algorithm is sufficient to generate a good set of least-cost paths. In the three reference backbone networks, setting K equal to 10 finds at least one minimum-regeneration path for virtually all source/destination pairs (over the range of optical-reach settings shown in Table 3.1). Note that searching for *all* minimum-regeneration paths in an optical-bypass-enabled network may be undesirable, as the number of such paths could be in the thousands for a large network.

Alternatively, one can use a more complex topology transformation (i.e., the reachability graph), as described in Sect. 3.6.2, to ensure a minimum-regeneration path is found. However, the simpler method described above is generally sufficient.

Note: One could use the intermediate steps of the BFS algorithm to find the shortest-distance path subject to a maximum number of hops to assist in finding the minimum-regeneration path with the minimum number of hops. Again, because distance does not directly translate to regeneration, this strategy does not necessarily always succeed. Furthermore, with K large enough, such paths are generally found by the KSP process anyway.

3.4.2 Bottleneck-Avoidance Strategy

While the KSP technique can generate a set of lowest-cost paths, the paths that are found may not exhibit good link diversity. For example, a link that is expected to be heavily loaded may appear in every lowest-cost path found between a particular source and destination. If the link diversity is not sufficient for a particular source/destination pair, then an alternative strategy can be used to generate a candidate path set, as described here.

The first step is to determine the links in the network that are likely to be highly loaded (i.e., the "hot spots"). One methodology for estimating load is to perform a preliminary routing where each demand in the forecasted traffic set is routed over its minimum-hop path (O-E-O networks) or its shortest-distance path (optical-bypass-enabled networks). While a traffic forecast may not accurately predict the traffic that will actually be supported in the network, it can be used as a reasonably good estimator of which links are likely to be heavily loaded. (Alternatively, one can combine the traffic forecast with the maximum-flow method of [KaKL00] to determine the critical links.)

It is important to consider not just single links that are likely to be bottlenecks, but also sequences of consecutive links that may be heavily loaded, and find routes that avoid the whole sequence of bad links. This is illustrated in Fig. 3.4, where Links BC, CD, and DZ are assumed to be likely bottlenecks. The shortest-distance path between Nodes A and Z is Path 1, which is routed over all three of the problem links. If one were to look for a path between these nodes that avoids just Link BC, then Path 2 is the remaining shortest-distance path. This is not satisfactory as the path still traverses Links CD and DZ. It would be better to simultaneously avoid all three bottleneck links and find Path 3.

After identifying the top 10–20 "hot spots" in the network, the next step is to run the shortest-path algorithm multiple times, where in each run, one bad link or one bad sequence of links is removed from the topology. (If a particular "hot spot" does not appear in the lowest-cost paths for a source/destination pair, then that hot spot can be skipped in this process. Furthermore, it is not desirable to remove *all* potentially bad links at once when running the shortest-path algorithm, as the resulting set of paths may be circuitous and may shift the hot spots to other locations in the network.) This process finds paths that avoid particular bottleneck areas, if possible.

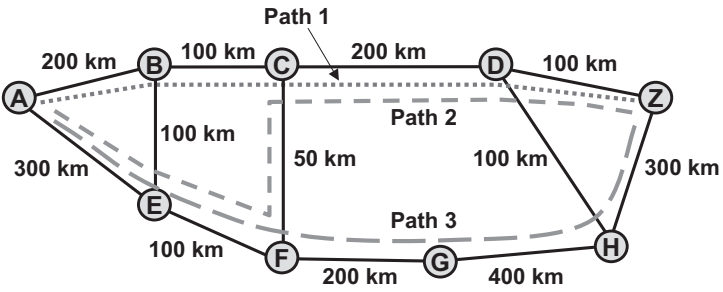


Fig. 3.4 Links *BC, CD,* and *DZ* are assumed to be bottleneck links. *Path 1, A-B-C-D-Z,* crosses all three of these links. If Link *BC* is eliminated from the topology, the resulting shortest path, *Path 2, A-E-F-C-D-Z,* still crosses two of the bottleneck links. All three bottleneck links must be simultaneously eliminated to yield *Path 3, A-E-F-G-H-Z*

The bottleneck-avoidance strategy can be combined with the KSP method such that, overall, the candidate path set includes lowest-cost paths and paths that are diverse with respect to the expected hot spots. There are clearly other methods one can devise to generate the candidate paths; however, the strategy described above is simple and is effective in producing designs for practical optical networks with relatively low cost and good load balancing.

3.5 Routing Strategies

The previous section discussed methods of producing a good set of candidate paths for all relevant source/destination pairs. This section considers strategies for selecting one of the paths to use for a given demand. These strategies hold for scenarios where demand requests enter the network one at a time, as well as scenarios where there is a whole set of demands that need to be routed but it is assumed that the demands have been ordered so that they are considered one at a time. (Ordering the demand set is covered in Sect. 3.8.)

In either real-time or long-term planning, a particular candidate path may not be feasible as the network evolves because there is no free bandwidth on one or more of the path links. Furthermore, with real-time operation, there is the additional constraint that all necessary equipment to support the connection must already be deployed. Thus, if a particular path requires regeneration at a node, and the node does not have the requisite available equipment, the path is considered infeasible.

In optical-bypass-enabled networks, selecting a wavelength for the route is an important step of the planning process. This section focuses on selecting a route independent of wavelength assignment, where wavelength assignment is performed as a separate step later in the process. Chapter 5 considers treating both of these aspects of the planning process in a single step.

3.5.1 Fixed-Path Routing

In the strategy known as fixed-path routing, the set of candidate paths is generated prior to any demands being added to the network. For each source/destination pair, one path is chosen from the associated candidate path set, and that path is used to route *all* demand requests for that source/destination pair (the other candidate paths are never used). Ideally, the path is a lowest-cost path, although load balancing, based on the traffic forecast, can also be a consideration for selecting a particular path.

This is clearly a very simple strategy, with any calculations performed up front, prior to any traffic being added. However, the performance of this strategy can be very poor, as it usually results in certain areas of the network becoming unnecessarily congested. The same path is always used for a given source/destination pair, providing no opportunity to adapt to the current network state. This often results in premature blocking of a demand even though feasible paths do exist for it.

Nevertheless, many telecommunications carriers continue to use fixed-path routing, where the shortest path, in terms of distance, is used for all demand requests. The main motivation for using shortest-path routing is minimization of the end-to-end path latency. For applications such as electronic trading in the financial markets, latency may be of critical importance, where excess delays of even a few microseconds can be unacceptable [Bach11]. However, most latency requirements are not that stringent, such that requiring shortest-path routing is overly restrictive.

Not only can fixed shortest-path routing lead to excess blocking but it can also lead to extra regeneration, as was shown in Table 3.1 in Sect. 3.3.2. In many of the scenarios of Table 3.1 where the shortest path resulted in an extra regeneration, it is possible to find a minimum-regeneration path with a length within 200 km of the shortest-path length. The extra path distance would add less than 1 ms to the latency, which is likely acceptable for most applications.

Overall, fixed-path routing, including fixed shortest-path routing, is not recommended, except for niche applications with very stringent latency requirements.

3.5.2 Alternative-Path Routing

In alternative-path routing, the set of candidate paths is also generated prior to any demands being added to the network. However, in this strategy, the candidate set is narrowed down to M paths (as opposed to one path in fixed-path routing), for some small number M, for each source/destination pair. When a demand request arrives for a given source/destination, one of the M paths is selected to be used for that particular demand. This allows some degree of state-dependent routing. In practice, selecting about three paths per each source/destination pair is a good strategy, although in real-time planning, where utilizing equipment that is already deployed is an issue, it may be desirable to select somewhat more paths.

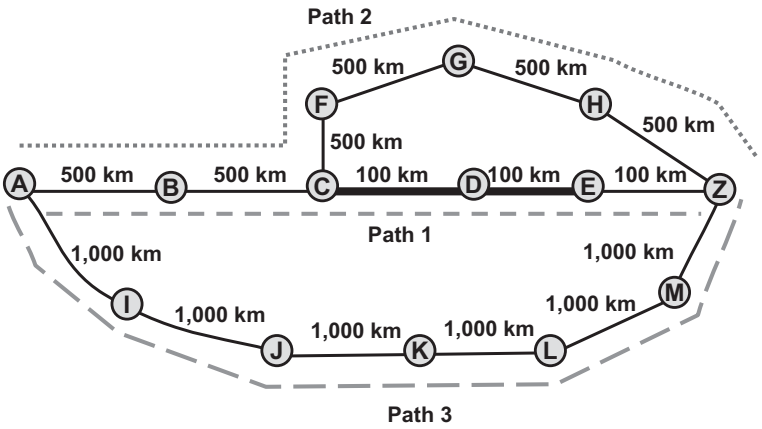


Fig. 3.5 Two alternative paths between Nodes *A* and *Z* are desired for load balancing, with Links *CD* and *DE* assumed to be the bottleneck links. With bottleneck diversity, *Paths 1* and *2* are selected. If total diversity is required, then *Path 3* must be included, which is a significantly longer path

There have been numerous studies that have shown alternative-path routing results in lower blocking than fixed-path routing, e.g., Karasan and Ayanoglu [KaAy98], Chu et al. [ChLZ03]. One of the more recent studies demonstrated one to two orders of magnitude lower blocking probability using alternative-path routing, even when limiting the length of the alternative paths for purposes of latency [Stra12].

3.5.2.1 Selecting the Set of Alternative Paths

In some research regarding alternative-path routing, it is assumed that the set of *M* alternative paths must have no links in common; however, this is unnecessarily restrictive. (Selecting paths with no links in common, however, is important for protection, as is covered in Sect. 3.7.) The goal is to select the *M* paths such that the same expected "hot spots" do not appear in all of the paths. Furthermore, it is not necessary to pick the *M* paths such that no "hot spot" appears in any of the paths. This would have the effect of simply shifting the heavy load to other links as opposed to balancing the load across the network.

By not requiring total diversity, the *M* paths are potentially shorter, resulting in lower latency, cost, and capacity utilization, and likely lower failure rate. This is illustrated in Fig. 3.5, where it is desired to find two alternative (unprotected) paths between Nodes A and Z, and where it is assumed that Links CD and DE are the bottleneck links. If diversity with respect to the bottleneck links is the only requirement, then Paths 1 and 2 are selected for the path set. If *total* diversity is required, then the path set would need to include Path 3 and either Path 1 or 2. Path 3 is significantly longer, has more hops, and more regeneration (e.g., assuming an optical

reach of 2,500 km) than either of Paths 1 and 2. Utilizing Path 3 for purposes of load balancing is undesirable.

The methodology for choosing the *M* paths was analyzed in a study using Reference Network 2 with realistic traffic that was dynamically established and torn down [Simm10]. Alternative-path routing with *M* equal to three was employed. The optical reach was assumed to be 2,500 km. Requiring only *bottleneck* diversity as opposed to *total* diversity for each set of *M* paths reduced both the average routed path length and the path hops by 5 % and reduced the amount of regeneration in the network by 15 %. The blocking probability was reduced by 55 %. This demonstrates the benefits of requiring only bottleneck diversity.

Ideally, the set of *M* paths selected for alternative-path routing are lowest-cost paths. However, in order to get enough "hot-spot" diversity among the paths, it may be necessary to include a path that does not meet the lowest cost, e.g., one of the *M* paths may have an additional regeneration. While selecting a path with a small amount of extra cost is not ideal, it is typically preferable to blocking a demand request due to poor load balancing.

All other factors being equal (e.g., cost, expected load), paths with shorter distance and fewer hops should be favored for inclusion in the set of *M* paths, to minimize delay and potentially improve reliability.

3.5.2.2 Selecting a Path for a Demand Request

When a demand request arrives for a particular source/destination, any of the *M* paths can be potentially used to carry the connection, assuming the path is feasible (i.e., it has the necessary available bandwidth and equipment). Typically, the current state of the network is used in determining which of the *M* paths to use. A common strategy is to select the feasible path that will leave the network in the "least-loaded" state. Assume that the most heavily loaded link in the *i*th path has W_i wavelengths already routed on it. Then the selected path is the one with the minimum W_i. If multiple paths are tied for the lowest W_i, then the load on the second most heavily loaded link in these paths is compared, and so on. (This is also known as *least congested path* routing [ChYu94].) If multiple paths continue to be tied with respect to load, or if the tie is broken only when comparing links with load much less than the maximum, then one can consider other factors, e.g., in an optical-bypass-enabled network, the path with fewest hops can be used to break the tie. Furthermore, if one of the *M* paths requires more regeneration than the other paths, then this path should not be selected unless its W_i is significantly lower than that of the other paths, or unless it is the only feasible path.

In another strategy for selecting which of the *M* paths to use, congestion and hops are jointly considered. For example, when choosing between two candidate paths, a path with *H* more hops is selected only if its most heavily loaded link has *L* fewer wavelengths routed on it, where the parameters *H* and *L* can be tuned as desired.

In real-time routing, it may also be beneficial to consider the available equipment at the nodes when selecting one of the *M* paths. If a particular path requires a regeneration at a node and there is very little free regeneration equipment at the node, then that path may not be favored, especially if there are other paths that have similar link loading and greater equipment availability.

The alternative-path routing strategy works very well in practice. It uses the traffic forecast to assist in generating the initial candidate path set, whereas it uses current network conditions to select one of the *M* paths for a particular demand request. One can add a larger dynamic component to the algorithm by allowing the set of *M* paths to be updated periodically as the network evolves. With a good choice of paths, the network is generally fairly well loaded before all *M* paths for a particular source/destination pair are infeasible. When this occurs, one can revert to dynamically searching for a path, as is covered in the next section.

A more restrictive form of alternative-path routing is known as *fixed-alternate routing,* where the *M* candidate paths are considered in a fixed order, and the first such path that has available capacity is selected. This is simpler than more general alternative-path routing because it only needs to track whether a path is available or not; it does not need to track the load on every link. However, this method is not as effective at load balancing and typically leads to higher blocking [ChLZ03].

3.5.3 Dynamic-Path Routing

In dynamic-path routing (also called *adaptive unconstrained routing* [MoAz98]), there is no predetermination of which paths to use for a particular source/destination combination. The path calculation is performed at the time of each demand request, based on the current state of the network. The first step is to determine if there are any links in the network with insufficient available bandwidth to carry the new demand. Any such links should be temporarily eliminated from the network topology. In addition, in real-time design with an O-E-O network, any node that does not have available regeneration equipment should be temporarily removed from the topology, because regeneration would be required at any intermediate node in the path.

After the topology has been pruned based on the current network state (more advanced topology transformations are discussed in Sect. 3.6), the procedure for generating a candidate set of paths can be followed, i.e., the KSP algorithm can be run and/or the bottleneck-avoidance strategy can be used where the current hot spots in the network are systematically eliminated from the already-pruned topology. (The links with *no* available capacity were already eliminated up front; thus, the hot spots that are eliminated at this point are the relatively heavily loaded links that still have available capacity.) This process takes tens of milliseconds to complete, so it is possible to generate a candidate set of paths every time a new demand request is received, assuming the total provisioning time is on the order of 1 s or more. One can use a smaller *K* in the KSP algorithm or consider fewer hot spots

in the bottleneck-avoidance methodology in order to reduce the processing time further. Many dynamic implementations simply look for a *single* shortest path in the pruned topology.

A variety of metrics can be used for dynamic routing. Typically, the metric reflects number of hops, distance, or current congestion, e.g., Bhide et al. [BhSF01]. One method suggested in Zhang et al. [ZTTD02] for optical-bypass-enabled networks uses a metric based on the number of wavelengths that are free on consecutive links, as this is an indicator of the likelihood of being able to assign wavelengths to the links. This was shown to provide better performance than simply considering link congestion; however, it does involve a graph transformation in order to capture the relationship between adjacent links in the shortest-paths algorithm.

After the candidate paths are generated, one path is selected for the new demand based on the current network state. (In implementations where only a single candidate path is generated, clearly this step of choosing one of the candidate paths is not needed.) Note that in real-time planning, some of the candidate paths may be infeasible due to a lack of resources, even though some amount of topology pruning occurred up front. For example, in an optical-bypass-enabled network, a candidate path may require that regeneration occur at a node that does not have available regeneration equipment. With optical bypass, it is typically not known ahead of time whether an intermediate node in a path will require regeneration; thus, the node is not pruned from the topology during the preprocessing step. After eliminating any of the candidate paths that are infeasible, a technique such as selecting the least-loaded path, as described in Sect. 3.5.2.2, can be used to pick among the remaining paths.

The dynamic path selection methodology provides the greatest adaptability to network conditions. While this may appear to be desirable, studies have shown that routing strategies that consider a more global design based on traffic forecasts can be advantageous [EMSW03]. Thus, a strictly adaptive algorithm is not necessarily ideal. Furthermore, the dynamic methodology may result in many different routes for the demands between a given source and destination. This has the effect of decreasing the network "interference length," which can potentially lead to more contention in the wavelength assignment process for optical-bypass-enabled networks. The interference length is the average number of hops shared by two paths that have at least one hop in common, as defined in Barry and Humblet [BaHu96]. In addition, if the network makes use of wavebands, where groups of wavelengths are treated as a single unit, then the diversity of paths produced by a purely dynamic strategy can be detrimental from the viewpoint of efficiently packing the wavebands. Finally, the dynamic methodology is the slowest of the three routing strategies discussed and involves the most computation. It may not be suitable, for example, if the demand request must be provisioned in a sub-second time frame.

Given the good results that are produced by the simpler alternative-path routing strategy of Sect. 3.5.2, it is often favored over a purely dynamic routing strategy. When the network is so full that none of the alternative paths are feasible, the dynamic strategy can be used instead.

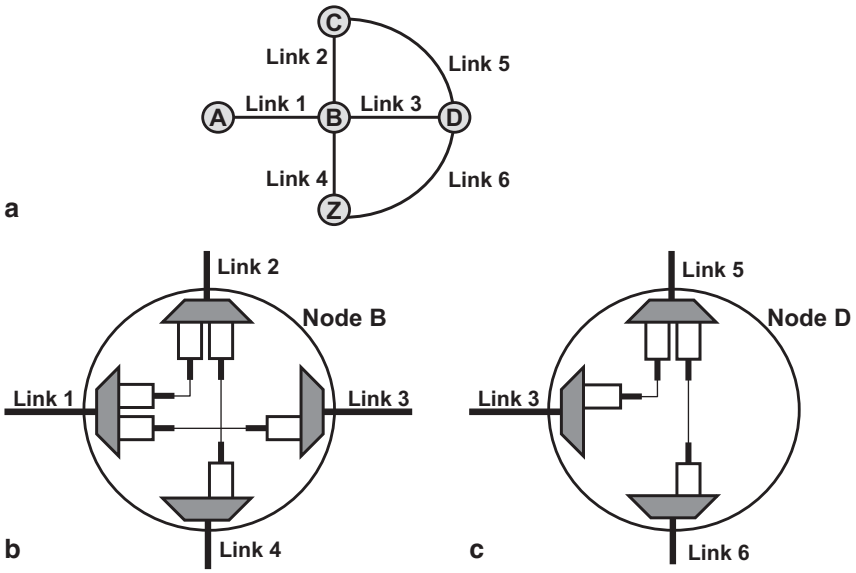


Fig. 3.6 **a** Network topology, where a new demand is requested from Node *A* to Node *Z*. **b** The available equipment at *Node B*. **c** The available equipment at *Node D*

3.6 Capturing the Available Equipment in the Network Model

In real-time planning, some, or even all, of the candidate paths may be infeasible due to a lack of available equipment in particular nodes. As described above, the dynamic routing process first prunes out the links and nodes that would clearly be infeasible for a new demand. In a scenario where there is little available equipment, it may be necessary to perform more involved topology transformations to model the available equipment in more detail. In the examples below, it is assumed that there is available equipment at the source and destination nodes; otherwise, the demand is rejected without further analysis.

3.6.1 O-E-O Network

First, consider an O-E-O network with the topology shown in Fig. 3.6a and assume that a new demand request arrives where the source is Node A and the destination is Node Z. Node B and Node D of the network are illustrated in more detail in Fig. 3.6b, c, respectively. In this example, it is assumed that pairs of transponders are interconnected via patch cables rather than through a flexible switch (i.e., the nodal architecture is that of Fig. 2.5, not Fig. 2.6). Thus, in order for a new demand

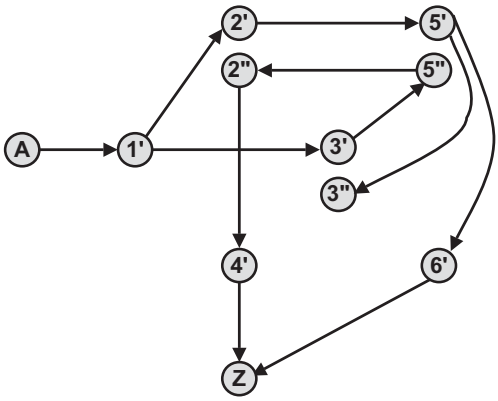


Fig. 3.7 Graph transformation to represent the available equipment in the network of Fig. 3.6. The numbered nodes correspond to the links in Fig. 3.6 with the same number, with the *prime* and *double prime* representing the two directions of the link. Nodes *A* and *Z* are the demand endpoints

to transit Node B from Link j to Link k, ($1 \leq j, k \leq 4$), there must be an available transponder on the optical terminal for Link j, an available transponder on the optical terminal for Link k, and the two transponders must be interconnected. Assuming that Fig. 3.6b depicts all of the available equipment at Node B, then the only possible paths through the node for a new demand are between Links 1 and 2, Links 1 and 3, and Links 2 and 4. Similarly, the only possible paths through Node D are between Links 3 and 5 and between Links 5 and 6. It is assumed that the remaining nodes in the network have sufficient available equipment to support any path, i.e., Node A has an available transponder on Link 1, Node Z has available transponders on both Links 4 and 6, and there are available transponders at Node C to support a path between Link 2 and Link 5.

To capture the path restrictions imposed by the limited amount of available equipment at Nodes B and D, one can perform a graph transformation where each link in the original topology becomes a node in the new topology. To be more precise, because each link shown in Fig. 3.6a actually represents bidirectional communication, each direction of a link becomes a node. These nodes are interconnected in the new topology only if there is equipment available in the real network to allow a new path to be routed between them. Nodes also have to be added to represent the source and destination of the new demand, i.e., Node A and Node Z, respectively.

The resulting transformed graph is illustrated in Fig. 3.7, where the node numbers in this graph correspond to the link numbers of Fig. 3.6. The single-prime nodes in the transformed graph represent the links in the direction from the (alphabetically) lower letter to the higher letter, and the double-prime nodes represent the reverse link direction. Thus, Node 2′ represents Link 2 in the original graph in the direction from Node B to Node C; Node 2″ represents Link 2 in the direction from Node C to Node B. There is no need to add a node representing Link 1″ because this link enters the demand source; similarly, there is no need to add a node representing Link 4″ or Link 6″ because these links exit the demand destination. Note that the node representing Link 1′ is connected to the nodes representing Link 2′ and Link 3′, but not Link 4′ due to the lack of a transponder pair connecting these links (in Node B). Similarly, there is no link connecting the node representing Link 3′ and the node representing Link 6′.

A shortest-path algorithm is run on the transformed topology to find a feasible path, using unity as the link metric to minimize the number of hops, and hence minimize the number of regenerations. The desired path from Node A to Node Z in the transformed graph is A-1′-2′-5′-6′-Z, which corresponds to path A-B-C-D-Z in the original graph.

Routing constraints such as those imposed by the available transponder pairs, where only certain directions through a node are possible, are known as *turn constraints* (this term arises from vehicular routing, where only certain turns are permissible). The graph transformation described above is one means of solving such problems, which allows a standard shortest-path algorithm to be run. An alternative is to use the original graph but modify the Dijkstra algorithm or the BFS algorithm to take the turn constraints into account when building out the path from source to destination [BoUh98, SoPe02]. With this methodology, an explicit graph transformation is not required.

Note that if the O-E-O nodes are equipped with switches, as in Fig. 2.6, then turn constraints do not arise, because the switch can interconnect any two transponders in the node. Additionally, if there are many available transponders at each node, then this level of modeling is not needed as typically any path through the node can be supported.

3.6.2 Optical-Bypass-Enabled Network

In an optical-bypass-enabled network, a different graph transformation can be used in real-time planning, similar to Gerstel and Raza [GeRa04]. The methodology is described first, followed by an example. It is assumed that the network is equipped with directionless ROADMs, or non-directionless ROADMs in combination with an edge switch. (If this flexibility is not present, then there will be turn constraints based on which pairs of transponders are pre-connected, similar to the O-E-O network.) A new topology is created, composed of only those nodes that have available regeneration equipment, along with the source and destination of the new demand. Any two of these nodes are interconnected by a link in the new topology if a regeneration-free path with available bandwidth exists between the two nodes in the real topology. Even if there are multiple regeneration-free paths between a pair of nodes, only one link is added in the new topology. The resulting graph is often referred to as the *reachability graph*. A shortest-path algorithm is run on this transformed topology to find a feasible path. A link metric such as [*LARGE*+*NumHops*] can be used, where *LARGE* is some number greater than any possible path hop count, and *NumHops* is the number of hops in the minimum-hop regeneration-free path (in the real topology) between the nodes interconnected by the link. With this metric, the shortest-path algorithm finds a minimum-regeneration feasible path with the minimum number of hops, assuming one exists.

An example of such a graph transformation is shown in Fig. 3.8. The full network is shown in Fig. 3.8a. Assume that this is an optical-bypass-enabled

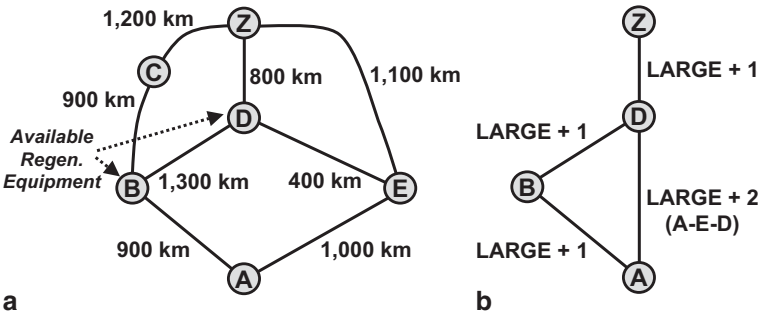


Fig. 3.8 **a** Nodes *A* and *Z* are assumed to be the endpoints of a new demand request, and Nodes *B* and *D* are assumed to be the only nodes with available regeneration equipment. The optical reach is 2,000 km. **b** The resulting reachability graph, where *LARGE* represents a number larger than any possible path hop count

network with directionless ROADMs and an optical reach of 2,000 km. Furthermore, assume that Nodes A and Z are the demand endpoints, and that only Nodes B and D are equipped with available regeneration equipment. Figure 3.8b illustrates the associated reachability graph; Nodes C and E do not appear in this graph because they do not have available regeneration equipment. Each of the links in the reachability graph represents a regeneration-free path in the real network. For example, Link AD in the reachability graph represents the two-hop path A-E-D in the real network. Note that there is no link connecting Nodes B and Z because all paths between these two nodes are longer than the optical reach. Running a shortest-path algorithm on the reachability graph yields the path A-D-Z, corresponding to A-E-D-Z in the true network.

Such graph transformations as described above for O-E-O and optical-bypass-enabled networks would need to be performed every time there is a new demand request to ensure that the current state of the network is taken into account, which adds to the complexity of the routing process. Ideally, sufficient equipment is pre-deployed in the network and the candidate paths are sufficiently diverse that one can avoid these transformations, at least until the network is heavily loaded. Strategies for determining how much equipment to pre-deploy are covered in Chap. 8.

3.7 Diverse Routing for Protection

The previous sections focused on finding a single path between a source and a destination. If any of the equipment supporting a connection fails, or if the fiber over which the connection is routed is cut, the demand is brought down. Thus, it is often desirable to provide protection for a demand to improve its *availability,* where availability is defined as the probability of the demand being in a working state at a given instant of time. Numerous possible protection schemes are covered in Chap. 7. Here, it is simply assumed that two paths are required from source to

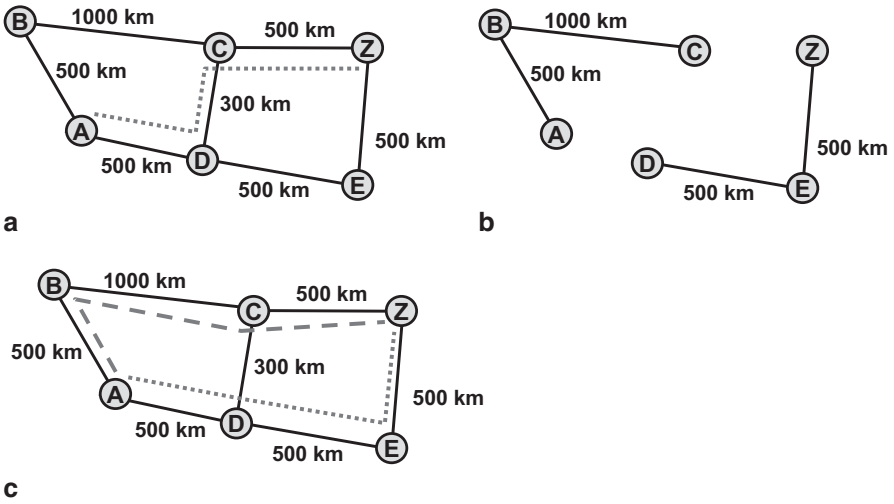


Fig. 3.9 The shortest pair of disjoint paths, with distance as the metric, is desired between Nodes *A* and *Z*. **a** The first call to the shortest-path algorithm returns the path shown by the *dotted line*. **b** The network topology after pruning the links comprising the shortest path. The second call to the shortest-path algorithm fails as no path exists between Nodes *A* and *Z* in this pruned topology. **c** The shortest pair of disjoint paths between Nodes *A* and *Z,* shown by the *dotted* and *dashed lines*

destination, where the two paths should be disjoint to ensure that a single failure does not bring down both paths. If one is concerned only with link failures, then the two paths can be simply link-disjoint, where nodes can be common to both paths. If one is concerned with both link and node failures, then the two paths should be both link-and-node disjoint (except for the source and destination nodes). The question is how to find the desired disjoint paths in a network.

Network designers sometimes resort to the simple strategy of first searching for a single path using a shortest-path algorithm. The links in the returned path are then pruned from the topology and the shortest-path algorithm is invoked a second time. (If node-disjointness is required, then the intermediate nodes from the first found path are also pruned from the topology.) If a path is found with this second invocation, then it is guaranteed to be disjoint from the first path.

While a simple strategy, it unfortunately fails in some circumstances. Consider the network topology shown in Fig. 3.9a, and assume that two *link-diverse* paths are required from Node A to Node Z. The first invocation of the shortest-path algorithm returns the path shown by the dotted line. Removing the links of this path from the topology yields the topology of Fig. 3.9b. It is not possible to find a path between Nodes A and Z on this pruned topology, causing the strategy to fail. In fact, two diverse paths can be found in the original topology as shown in Fig. 3.9c. This type of scenario is called a "trap topology," where two sequential calls to the shortest-path algorithm fail to find disjoint paths even though they do exist.

Even if the simple two-call strategy succeeds in finding two disjoint paths, the paths may not be optimal. In the network of Fig. 3.10a, it is assumed that O-E-O technology is used such that minimizing the number of hops is desirable. The mini-

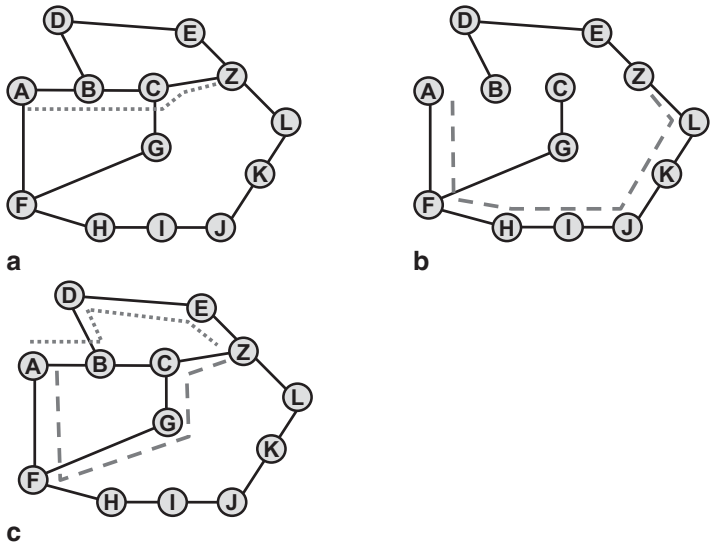


Fig. 3.10 The shortest pair of disjoint paths, with hops as the metric, is desired between Nodes *A* and *Z.* **a** The first call to the shortest-path algorithm returns the path shown by the *dotted line.* **b** The network topology after pruning the links comprising the shortest path. The second call to the shortest-path algorithm finds the path indicated by the *dashed line.* The total number of hops in the two paths is ten. **c** The shortest pair of disjoint paths between Nodes *A* and *Z,* shown by the *dotted* and *dashed lines*; the total number of hops in the two paths is only eight

mum-hop path from Node A to Node Z is shown by the dotted line. The links of this path are pruned from the topology resulting in the topology shown in Fig. 3.10b. The second call to the shortest-path algorithm returns the path shown by the dashed line in Fig. 3.10b. While indeed link-disjoint, the two paths of Fig. 3.10a, b cover a total of ten hops. However, the lowest-cost pair of disjoint paths has a total of only eight hops, as shown in Fig. 3.10c.

As these examples illustrate, the two-call strategy may not be desirable. It is preferable to use an algorithm specifically designed to find the shortest pair of disjoint paths, as described next. In this context, *shortest* is defined as the pair of paths where the sum of the metrics on the two paths is minimized.

3.7.1 Shortest Pair of Disjoint Paths

The two best-known shortest pair of disjoint paths (SPDP) algorithms are the Suurballe algorithm [Suur74, SuTa84] and the Bhandari algorithm [Bhan99]. Both algorithms involve calls to a regular shortest-path algorithm; however, they require different graph transformations (e.g., removing links, changing the link metrics) to ensure that the shortest pair of disjoint paths is found. The graph transformations of the Bhandari algorithm may generate links with negative metrics, which is why

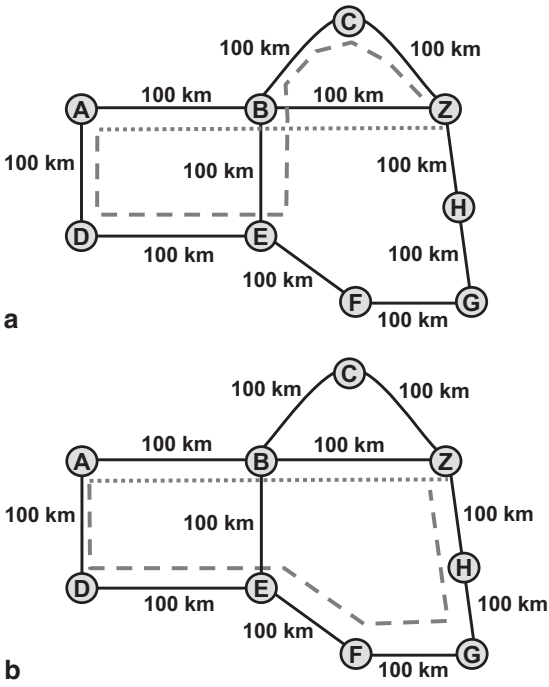


Fig. 3.11 **a** The shortest link-disjoint pair of paths is shown by the *dotted* and *dashed lines*. **b** The shortest link-and-node-disjoint pair of paths is shown by the *dotted* and *dashed lines*

it requires an associated shortest-path algorithm such as BFS, which can handle graphs with negative link metrics.

Both the Suurballe and the Bhandari algorithms are guaranteed to find the pair of disjoint paths between a source and destination where the sum of the metrics on the two paths is minimized, assuming that at least one pair of disjoint paths exists. As illustrated by the examples of Fig. 3.9, 3.10, the shortest single path may not be a part of the shortest-disjoint-paths solution. The run times of the Suurballe and Bhandari algorithms are about the same; however, the latter may be more readily extensible to other applications [Bhan99]. Chapter 11 provides the code for the Bhandari algorithm.

The SPDP algorithms can be used to find either the shortest pair of link-disjoint paths or the shortest pair of link-and-node-disjoint paths. To illustrate the difference, Fig. 3.11a shows the shortest pair of link-disjoint paths between Nodes A and Z, where the paths have Node B in common; together, the paths cover seven links and 700 km. Figure 3.11b shows the shortest link-and-node-disjoint paths between A and Z for the same topology; these paths cover eight links and 800 km.

Furthermore, the SPDP algorithms can be modified to find the shortest *maximally* link-disjoint (and optionally node-disjoint) paths when *totally* disjoint paths do not exist. Consider the topology shown in Fig. 3.12, where a protected connection is required between Nodes A and Z. A pair of completely disjoint paths does not exist between these two nodes. However, the maximally disjoint pair of paths, with one

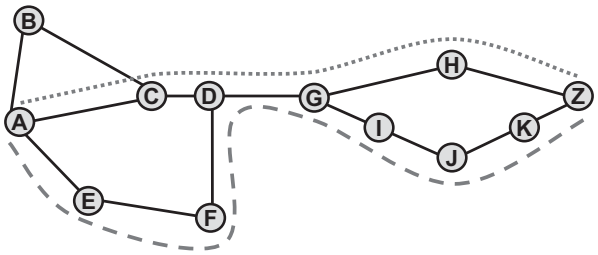


Fig. 3.12 There is no completely disjoint pair of paths between Nodes *A* and *Z.* The set of paths shown by the *dotted* and *dashed lines* represents the shortest maximally disjoint pair of paths. The paths have Nodes *D* and *G,* and the link between them, in common

common link (Link DG) and two common nodes (Nodes D and G), is shown by the dotted and dashed lines in the figure. This pair of paths minimizes the number of single points of failures for the connection.

If a demand is very susceptible to failure, or the availability requirements are very stringent, then it may be desirable to establish more than two disjoint paths for the demand. The SPDP algorithms can be extended to search for the *N* shortest disjoint paths between two nodes, for any *N,* where the *N* paths are mutually disjoint. In most optical networks, there are rarely more than just a small number of disjoint paths between a given source and destination, especially in a backbone network. For example, in the three reference backbone networks of Sect. 1.10, there are never more than five disjoint paths between any two nodes; for almost all node pairs, there are just two to four disjoint paths. (Detailed connectivity statistics for these networks are provided in Sect. 7.6.3.1.) If *N* is larger than this, the SPDP algorithms can be used to return the *N* shortest *maximally* disjoint paths.

3.7.2 Minimum-Regeneration Pair of Disjoint Paths

As noted in Sect. 3.3.2, the paths of minimum distance in an optical-bypass-enabled network do not necessarily correspond to the paths of minimum regeneration. With single paths, this phenomenon arises because regeneration typically must occur in network nodes as opposed to at arbitrary sites along the links (and because some nodes may not be equipped with network elements that support optical bypass in all directions). Limiting regeneration to node sites also may result in extra regenerations when dealing with a pair of disjoint paths. Furthermore, the fact that regeneration is determined independently on the disjoint paths may also cause the minimum-distance pair of disjoint paths to require extra regeneration as compared to the minimum-regeneration pair of disjoint paths, as shown in the next example.

In Fig. 3.13, assume that a pair of link-and-node-disjoint paths is required between Nodes A and Z, and assume that the network supports optical bypass with

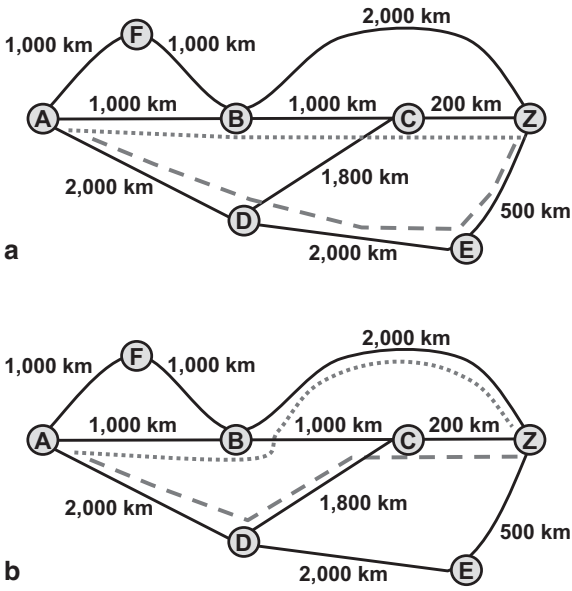


Fig. 3.13 Assume that the optical reach is 2,000 km. **a** For a protected connection from Node *A* to Node *Z,* the combination of diverse paths *A-B-C-Z* and *A-D-E-Z* is the shortest (6,700 km), but requires three regenerations. **b** The combination of diverse paths *A-B-Z* and *A-D-C-Z* is longer (7,000 km), but requires two regenerations. (Adapted from Simmons [Simm06]. © 2006 IEEE)

an optical reach of 2,000 km. The shortest-distance pair of disjoint paths, shown in Fig. 3.13a, is A-B-C-Z and A-D-E-Z. These two paths have a combined distance of 6,700 km, and require a total of three regenerations (at Nodes C, D, and E). However, the minimum-regeneration pair of disjoint paths, shown in Fig. 3.13b, is A–B–Z and A-D-C-Z, which covers a total of 7,000 km, but requires a total of only two regenerations (at Nodes B and D). As this example illustrates, each of the candidate pairs of disjoint paths must be explicitly examined to determine the number of required regenerations.

The likelihood of the minimum-distance pair of disjoint paths requiring more than the minimum amount of regeneration is affected by the network topology and the optical reach. This was investigated using the three reference backbone networks of Sect. 1.10. Table 3.2 indicates the percentage of source/destination pairs where the number of regenerations required in the shortest-distance pair of link-and-node-disjoint paths is greater than the number of regenerations required in the minimum-regeneration pair of link-and-node-disjoint paths, for different optical reach values. In most cases, the difference is just one regeneration; for a small number of cases in Network 1, the difference is two. As with the results for unprotected routing (i.e., Table 3.1), the worst case is Network 1 with 1,500 km optical reach, with 14 % of the shortest-distance pairs of disjoint paths requiring at least one extra regeneration over the minimum.

A heuristic that increases the probability of finding a minimum-regeneration pair of disjoint paths was presented in Beshir et al. [BKOV12]. The heuristic requires a graph transformation similar to what was performed in Sect. 3.6.2, where a reachability graph is created with links added between any pair of nodes that have a regeneration-free path between them. (In a large network such as Reference

Table 3.2 Percentage of shortest-distance pairs of link-and-node-disjoint paths that required extra regeneration

Network	# of nodes	Optical reach			
		1,500 km	2,000 km	2,500 km	3,000 km
1	75	14%	7%	5%	5%
2	60	6%	3%	3%	3%
3	30	0.5%	0.2%	0%	0%

Network 1, and with 3,000-km optical reach, the degree of most of the nodes in the reachability graph is about 50.) A process similar to that of the SPDP algorithms is then run on the reachability graph (with some additional transformations needed). The results for the sample networks in Beshir et al. [BKOV12] indicate that although this heuristic is not guaranteed to find a minimum-regeneration pair of disjoint paths, it often does.

Another effective (and simple) heuristic is to take the shortest-distance dual paths produced by the SPDP algorithm, and one at a time, eliminate a link included in these paths. The SPDP algorithm is rerun on the slightly pruned topology. This is often sufficient to find a pair of disjoint paths requiring minimum regeneration. Using this strategy on the three reference networks finds a minimum-regeneration pair of disjoint paths for all source/destination pairs in Network 3, virtually all source/destination pairs in Network 2, and 97.5–99.5% of the source/destination pairs in Network 1, depending on the reach. Furthermore, the paths that are found using this strategy tend to have a *relatively* short distance. For example, in the cases where this strategy found a minimum-regeneration pair of disjoint paths that had fewer regenerations than the minimum-distance pair of disjoint paths, the increase in the total distance of the two paths averaged about 200–500 km. The increase in the distance of just the primary path (i.e., the shorter of the two paths) averaged about 20–100 km. Thus, the latency impact of using minimum-regeneration disjoint paths as opposed to minimum-distance disjoint paths should be small. In fact, for many source/destination pairs, the primary path in the minimum-regeneration solution that was found was actually shorter than the primary path in the minimum-distance solution.

To *guarantee* that a minimum-regeneration pair of disjoint paths is found for each source/destination pair in a network, one strategy is to look at all disjoint-path combinations where the distances of the paths potentially could result in fewer regenerations than is required in the shortest-distance pair of disjoint paths. This can be performed with multiple calls to a KSP routine, using outer and inner loops that look for the primary and secondary paths, respectively (see Exercise 3.4). For a network the size of Network 1, this algorithm requires several seconds of run time. This strategy has the additional benefit of favoring finding disjoint paths with a relatively short primary path.

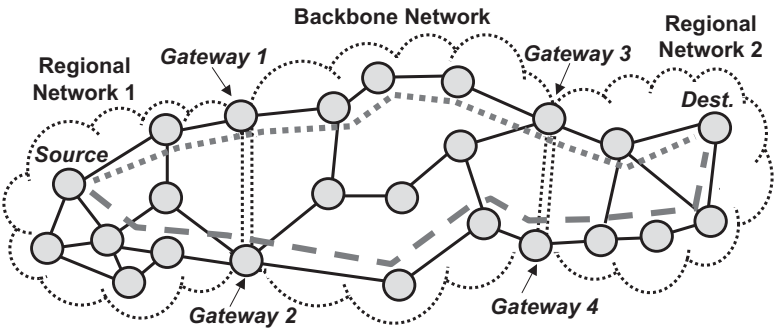


Fig. 3.14 A protected connection between the source and destination is routed from *Regional Network 1,* through the backbone network, to *Regional Network 2.* The gateways are the nodes at the boundaries between the regional networks and the backbone network. It is desirable to have diverse paths from the source to *Gateways 1* and *2,* diverse paths between *Gateways 1* and *2* and *Gateways 3* and *4,* and diverse paths from *Gateways 3* and *4* to the destination

3.7.3 Shortest Pair of Disjoint Paths: Dual Sources/Dual Destinations

Another interesting twist to the problem of finding the shortest pair of disjoint paths arises when there are two sources and/or two destinations [Bhan99]. There are several scenarios where this type of problem arises.

First, consider the scenario shown in Fig. 3.14, where the source is in one regional network, the destination is in another regional network, and the two regional networks are interconnected by a backbone network. As shown in the figure, there are two nodes that serve as the gateways between each regional network and the backbone network. Each network is a separate domain, where routing occurs separately within each domain. It is desired that a protected connection be established between the source and destination. Thus, the paths in Regional Network 1 between the source and Gateways 1 and 2 should be diverse. Similarly, the paths in Regional Network 2 between Gateways 3 and 4 and the destination should be diverse. Additionally, the backbone paths between Gateways 1 and 2 and Gateways 3 and 4 should be diverse. Thus, moving from left to right in the figure, this particular application requires finding the shortest pair of disjoint paths between one source and two destinations in the first regional network, between two sources and two destinations in the backbone network, and between two sources and one destination in the second regional network. (In this context, the sources and destinations are with respect to a particular domain; they are not necessarily the ultimate source and destination of the end-to-end connection.)

A second application arises in backhauling traffic to diverse sites. (Backhauling refers to the general process of transporting traffic from a minor site to a major site for further distribution.) As covered in Sect. 6.6, backhauling is often used when low-rate traffic from a small node needs to be groomed (i.e., packed into a wavelength) at a node that is equipped with a grooming switch. For reliability purposes,

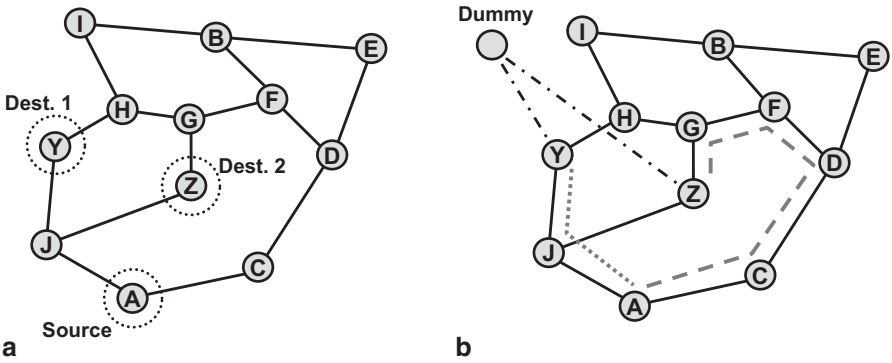


Fig. 3.15 **a** A disjoint path is desired between one source (Node *A*) and two destinations (Nodes *Y* and *Z*). **b** A *dummy* node is added to the topology and connected to the two destinations via links that are assigned a metric of zero. An SPDP algorithm is run between Node *A* and the *dummy* node to implicitly generate the desired disjoint paths, as shown by the *dotted line* and the *dashed line*

the traffic from the small node is typically backhauled to two such grooming nodes, where the paths to the two nodes should be diverse. This is another example of where it is desirable to find the shortest diverse set of paths between one source and two destinations (or, in the reverse direction, two sources and one destination).

A third application arises with cloud computing. With cloud services, the data and computing resources for a particular application are typically replicated at some set of *M* data centers distributed throughout the network. The cloud user (e.g., an enterprise) requires connectivity to any one of the data centers at a given time. If the path to the data center fails, it is often more bandwidth-efficient to roll the user over to a different data center rather than using an alternative path to the original data center [DBSJ11]. If the protection resources are preplanned, then it is necessary to find diverse paths between the user and two of the data centers. This is an instance of diverse paths between one source and two destinations, but where the two destinations come from a set of *M* possible destinations (i.e., *any* two of the *M* destinations are suitable). The routing technique described below can be extended to this scenario as well. Additionally, for mission-critical cloud applications, it may be desirable to preplan protection paths to more than two data centers; the technique below extends to the shortest set of diverse paths between one source and an arbitrary number of destinations.

The one source/two destination problem setup is illustrated in Fig. 3.15a where Node A is the source and Nodes Y and Z are the two destinations. (This figure is a general illustration; it is not related to Fig. 3.14.) Finding the shortest pair of disjoint paths is quite simple. A "dummy" destination node is added to the topology as shown in Fig. 3.15b; Nodes Y and Nodes Z are connected to the dummy node via links that are assigned a metric of zero. An SPDP algorithm is then run using Node A as the source and the dummy node as the destination. This implicitly finds the shortest pair of disjoint paths from Node A to Nodes Y and Z.

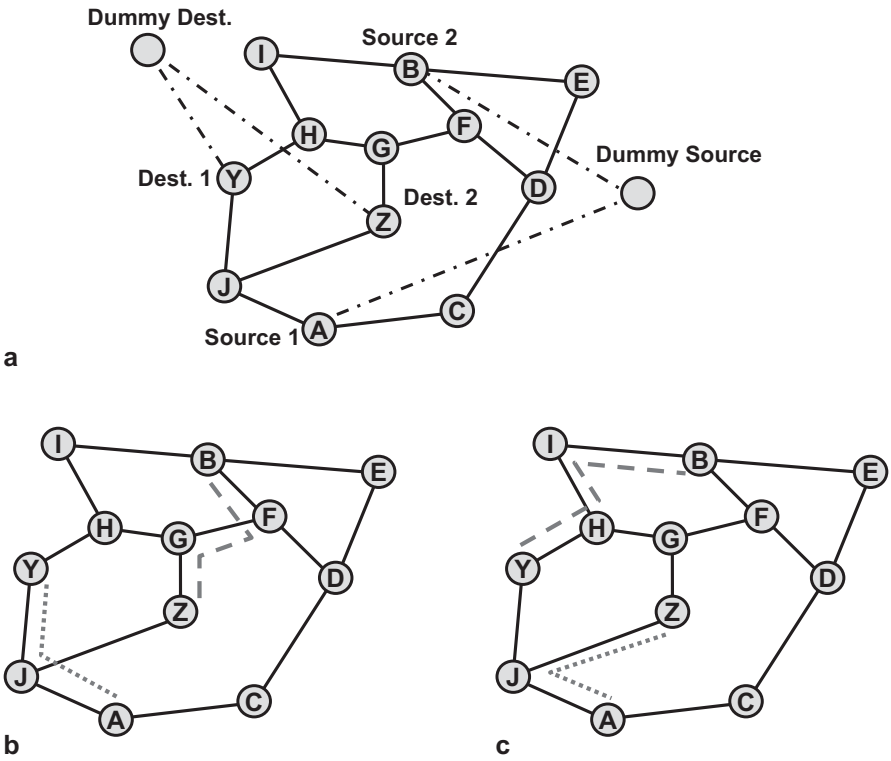


Fig. 3.16 Diverse paths are required from two sources (Nodes *A* and *B*) to two destinations (Nodes *Y* and *Z*). **a** One dummy node is connected to the two sources and one dummy node is connected to the two destinations via links that are assigned a metric of zero. **b** In this solution, one path is between Nodes *A* and *Y* and the other path is between Nodes *B* and *Z*. **c** In this solution, one path is between Nodes *A* and *Z* and the other path is between Nodes *B* and *Y*

For the cloud-computing scenario described above, the *M* data-center nodes would each be connected to the dummy destination node via links with a metric of zero. Running an SPDP algorithm (that looks for disjoint *dual* paths) between the source node and the dummy destination node then implicitly finds the shortest disjoint dual path to two of the *M* data-centers nodes. If greater protection is required, then, as noted in Sect. 3.7.1, the SPDP algorithm can be extended to find the *N* shortest disjoint paths between the source node and the dummy node, where $N \leq M$. This implicitly finds *N* mutually disjoint paths (if they exist) between the cloud user and *N* of the data centers.

If disjoint paths are desired between two sources and two destinations, then both a dummy source and a dummy destination are added, and the SPDP algorithm is run between the two dummy nodes. (As noted above, this arises in Fig. 3.14, where it is desired to find disjoint paths in the backbone network between the two sets of gateway nodes.) This procedure is illustrated in the network of Fig. 3.16a where it is assumed that a shortest pair of disjoint paths from Nodes A and B to Nodes Y and Z is

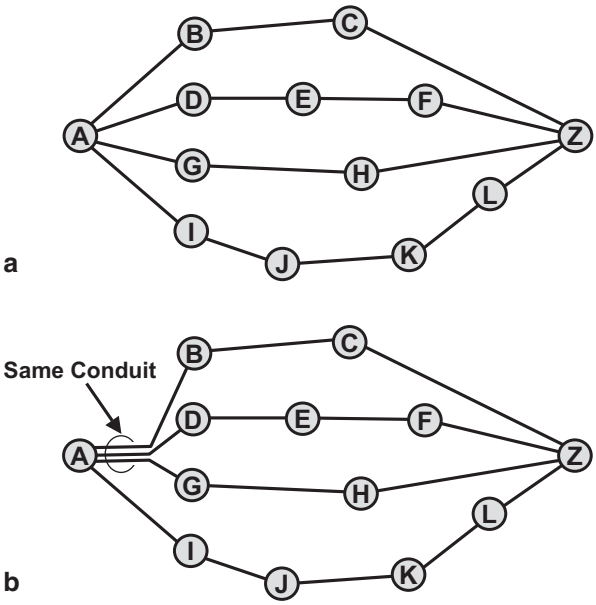


Fig. 3.17 **a** In the link-level view of the topology, Links *AB, AD,* and *AG* appear to be diverse. **b** In the fiber-level view, these three links lie in the same conduit exiting Node *A,* and thus are not diverse. A single cut to this section of conduit can cause all three links to fail

required. One dummy node is connected to Nodes A and B, and the other to Nodes Y and Z, via links with a metric of zero. After running the SPDP algorithm between the two dummy nodes, it is interesting that the two paths that are found may be between Nodes A and Y and between Nodes B and Z, as shown in Fig. 3.16b, or the two paths may be between Nodes A and Z and between Nodes B and Y, as shown in Fig. 3.16c. The SPDP algorithm does not allow control over which source/destination combinations will be produced by the algorithm.

3.7.4 Shared Risk Link Groups

When searching for disjoint paths for protection, it may be necessary to consider the underlying physical topology of the network in more detail. Two links that appear to be disjoint when looking at the network from the link level may actually overlap at the physical fiber level. For example, portions of the two links may lie in the same fiber conduit such that a failure along that conduit would simultaneously disrupt both links. Links that are part of the same failure group comprise what is known as a *shared risk link group* (SRLG). (There may be network resources other than links that fail as a group. Thus, the more general term is *shared risk group* (SRG).)

A common SRLG configuration, known as the *fork configuration,* occurs when multiple links lie in the same conduit as they exit/enter a node. This is illustrated in Fig. 3.17. The link-level view of the network is shown in Fig. 3.17a, where it appears Links AB, AD, and AG are mutually disjoint. The fiber-level view is shown in Fig. 3.17b, where it is clear that these three links lie in the same conduit

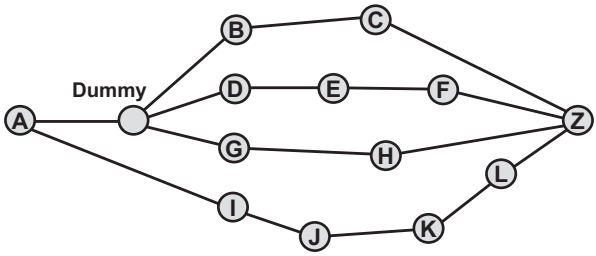


Fig. 3.18 Graph transformation performed on Fig. 3.17 to account for the SRLG extending from Node *A*. A *dummy* node is added, and each link belonging to the SRLG is modified to have this *dummy* node as its endpoint instead of Node *A*. A link is added between Node *A* and the *dummy* node, where this link is assigned a metric of zero

at Node A. Thus, it would not be desirable to have a protected demand where, for example, one path includes Link AB and the other path includes Link AG, because the common conduit is a single point of failure. To find paths that are truly diverse requires that the SPDP algorithm be modified to account for the SRLGs, as described next.

In SPDP algorithms such as the Bhandari algorithm, the first step is to find the single shortest path from source to destination, which is then used as a basis for a set of graph transformations. If the source or destination is part of an SRLG fork configuration, and one of the links included in the SRLG lies along the shortest path that is found in the first step of the SPDP, then an additional graph transformation such as the one shown in Fig. 3.18 for Node A is required [Bhan99]. A dummy node is temporarily added to the topology and the SRLG links with an endpoint of Node A are modified to have the dummy node as the endpoint instead (e.g., Link AB is modified to be between the dummy node and Node B). The link metrics are kept the same. Another link, with a metric of zero, is added between Node A and the dummy node. The SPDP algorithm then proceeds, with node-disjointness required (see Exercise 3.9). Because of the presence of the dummy node and the node-disjointness requirement, the shortest pair of disjoint paths will not include two links from the same SRLG fork configuration.

In addition to links with a common conduit resulting in SRLGs, certain graph transformations may produce the same effect. Consider the graph transformation for capturing available regeneration equipment in an optical-bypass-enabled network, which was presented in Sect. 3.6.2. In this transformation, links are added to a transformed graph where the links represent regeneration-free paths in the original graph. Links in the new graph may appear to be diverse that actually are not. This effect is illustrated in Fig. 3.19. The true network topology is shown in Fig. 3.19a. Assume that the optical reach is 2,000 km and assume that a *protected* path is desired between Nodes A and Z. Assume that the nodes with available regeneration equipment are Nodes B, D, H, J, K, L, and M. The transformed graph, i.e., the reachability graph, is shown in Fig. 3.19b. A link is added to the reachability graph

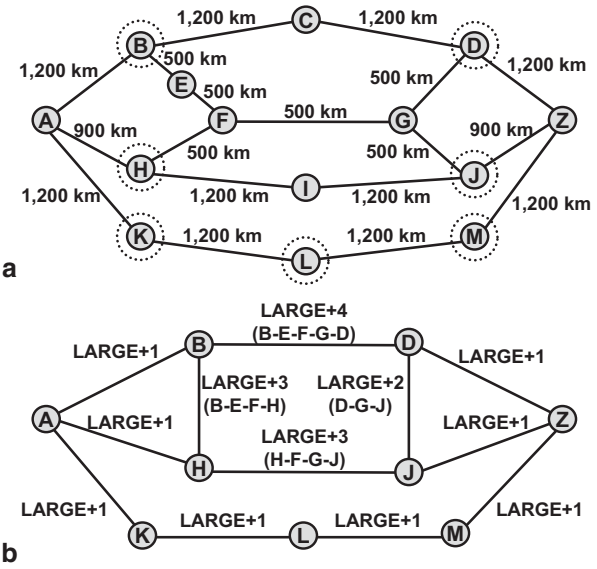


Fig. 3.19 **a** Example network, where the source and destination are *A* and *Z,* respectively, and the optical reach is 2,000 km. Only Nodes *B, D, H, J, K, L,* and *M* have available regeneration equipment (i.e., the *circled nodes*). **b** Resulting transformed graph (i.e., the reachability graph), where links represent regeneration-free paths. Links *BD* and *HJ* appear to be diverse but actually represent paths with a common link (*FG*)

if a regeneration-free path exists in the true network between the link endpoints and the two endpoints have available regeneration equipment. (The link metrics in the reachability graph are based on the number of hops in the regeneration-free path. *LARGE* is some number greater than any possible path hop count.)

The next step is to look for diverse paths between Nodes A and Z in the reachability graph. Note that Links BD and HJ appear to be diverse in this graph. However, Link BD corresponds to path B-E-F-G-D in the real network and Link HJ corresponds to path H-F-G-J in the real network. Thus, both links correspond to paths that contain the link FG in the real network, implying that Links BD and HJ in the reachability graph comprise an SRLG. This type of SRLG formation, where only the middle portions of the links overlap, is referred to as the *bridge configuration*. (This configuration occasionally arises in actual network topologies due to a portion of the conduits of multiple links being deployed along the same bridge in order to cross a body of water.) To handle this SRLG scenario, the SPDP algorithm is run multiple times, each time eliminating one of the links comprising the bridge SRLG, i.e., for the scenario of Fig. 3.19b, the SPDP algorithm must be run twice. The best result from any of the runs is taken as the solution. In this example, this yields A-H-J-Z and A-K-L-M-Z in the reachability graph, corresponding to A-H-F-G-J-Z and A-K-L-M-Z in the real network.

To find the optimal pair of diverse paths in a network with multiple bridge configurations, all possible combinations have to be considered where only one link from each bridge is present. The number of possibilities grows exponentially with the number of bridge configurations. Thus, this methodology becomes intractable if the number of bridge configurations is large. (In the reachability graph of a large network, there may be numerous such bridge configurations.) Furthermore, there

are other classes of SRLGs, though not very common, for which finding an optimal routing solution is difficult, as described in Bhandari [Bhan99]. In general, there are no computationally efficient algorithms that are guaranteed to find the optimal pair of disjoint paths in the presence of any type of SRLG; thus, heuristics are generally used when "difficult" SRLGs are present in a network.

One such heuristic to handle arbitrary SRLGs is proposed in Xu et al. [XXQL03]. The first step in this heuristic is to find the single shortest path from source to destination; call this Path 1. The links in Path 1 are temporarily pruned from the topology. Furthermore, any link that belongs to an SRLG to which at least one of the links in Path 1 also belongs is assigned a very large metric to discourage its use. The shortest-path algorithm is called again on the modified graph.

This second call may fail to find a diverse path. First, it may fail to find a path because of Path 1's links having been removed from the topology. In this case, the process restarts with a different Path 1 (one could use a KSP algorithm to generate the next path to use as Path 1). Second, it may fail because the path that is found, call it Path 2, includes a link that is a member of an SRLG that also includes a link in Path 1, in which case Path 1 and Path 2 are not diverse. (While the large metric discourages the use of such links in Path 2, it does not prevent it.) In this scenario, the most "risky" link in Path 1 is determined; this is the link that shares the most SRLGs with the links in Path 2. The links in Path 1 are restored to the topology, with the exception of the risky link, and the process restarts, i.e., a new Path 1 is found on this reduced topology. The process continues as above, until a diverse pair of paths can be found, or until no more paths between the endpoints remain in the topology (due to risky links being sequentially removed), in which case it fails. If the procedure does find a diverse pair of paths, there is no guarantee that they are the shortest such paths; however, Xu et al. [XXQL03] report achieving good results using this strategy, with reasonable run time.

3.7.5 Routing Strategies with Protected Demands

Sections 3.4 and 3.5 covered generating candidate paths and using various routing strategies for unprotected demands (i.e., demands with a single path between the source and destination). In this section, these same topics are revisited for protected demands. We focus on the bottleneck-avoidance strategy, combined with an SPDP algorithm, to generate a set of candidate diverse paths.

As in Sect. 3.4, the initial step is to determine the links that are expected to be heavily loaded. This can be done by performing a preliminary design with all demands in the traffic forecast routed over their shortest paths, where in general the traffic forecast will include both unprotected and protected demands. (An alternative means of estimating the critical links for protected traffic, based on maximum-flow theory, is detailed in Kar et al. [KaKL03]; this method can be combined with the traffic forecast to determine the links expected to be the most heavily loaded.) One can then generate a set of candidate paths for the protected demands by using

the bottleneck-avoidance strategy, where the most heavily loaded links or sequence of links are systematically removed from the topology, with the SPDP algorithm run on the pruned topology. The goal is to generate a set of lowest-cost (or close to lowest-cost) disjoint path pairs that do not all contain the same expected "bad" links. Note that a given source/destination pair may support both unprotected and protected demands. A candidate path set is independently generated for each source/destination/protection combination.

Any of the three routing strategies discussed in Sect. 3.5—fixed-path routing, alternative-path routing, and dynamic-path routing—are applicable to protected demands. (Variations of these strategies may be used for demands that share protection bandwidth, as covered in Chap. 7.) As with unprotected demands, alternative-path routing is an effective strategy for routing protected demands in practical optical networks.

3.8 Routing Order

In real-time network planning, demand requests generally are received and processed one at a time. In long-term network planning, however, there may be a set of demands to be processed at once. If the routing strategy used is adaptive, such that the network state affects the choice of route, then the order in which the demands in the set are processed will affect how they are routed. This in turn can affect the cost of the network design, the loading in the network, and the blocking probability. Thus, some attention should be paid to the order in which the demands are processed.

One common strategy is to order the demands based on the lengths of the shortest paths for the demands, where the demands with longer paths are processed first. The idea is that demands with longer paths are harder to accommodate and thus should be handled earlier to ensure that they are assigned to optimal paths. This criterion can be combined with whether or not the demand requires protection, where protected demands are routed earlier as they require more bandwidth and there is generally less flexibility in how they can be routed.

This scheme often yields better results when combined with a round-robin strategy. Within a given "round," the ordering is based on the required protection and the path length. However, at most one instance of a particular source/destination/protection combination is routed in each round. For example, if there are two protected demands between Node A and Node B, and three protected demands between Node A and Node C, where the AB path is longer than the AC path, plus one unprotected demand between Nodes A and D, then the routing order is: AB, AC, AD, AB, AC, AC.

Another strategy that is compatible with alternative-path routing is to order the demands based on the quality of the associated path set. If a particular source/destination/protection combination has few desirable path options, then the demands between this source and destination with this level of protection are routed

earlier. For example, a particular path set may have only one path that meets the minimum cost. It is advantageous to route the associated demands earlier to better ensure that the minimum-cost path can be utilized. In addition, the expected load on the links comprising the path set should be considered; the heavier the projected congestion for a particular source/destination/protection combination, the earlier it is routed.

Other ordering strategies are clearly possible. In general, no one strategy yields the best results in all network planning exercises. However, when using alternative-path routing, the routing process is so fast that multiple routing orders can be tested to determine which yields the best results. For example, routing thousands of demands with three different ordering strategies takes on the order of a couple of seconds. This is acceptable for long-term network planning.

Another possibility is to make use of meta-heuristics that take an initial solution (in this case, a particular route order), and then explore the neighborhood around that solution (e.g., exchange the routing order of two demands) to determine if the solution can be improved. Such meta-heuristics typically include mechanisms to avoid getting stuck in local minima. For example, one such meta-heuristic, *simulated annealing* [VaAa87], is used in Bogliolo et al. [BoCM07] to improve the order in which demands are routed and assigned wavelengths when various physical-layer impairments are taken into account. Simulated annealing is also used in Christodoulopoulos et al. [ChMV11] (to improve the routing order when dealing with multiple line rates on one fiber). Another meta-heuristic, *tabu search* [GlLa97], is used in Charbonneau and Vokkarane [ChVo10] (to test various orderings for routing of manycast trees). If the number of demands is very large, however, such techniques may not be able to explore very much of the possible solution space to be effective.

3.9 Flow-Based Routing Techniques

With global optimization techniques such as *integer linear programming* (ILP), the routing order of the demands is not relevant. ILP implicitly considers the whole solution space to find the optimal solution. However, ILP methodologies often have a very long run time and are impractical except for very small networks with little traffic. A more practical approach is to use efficient *linear programming* (LP) techniques (e.g., the Simplex algorithm), combined with strategies that drive the solution to integer values.

For example, routing a set of traffic demands can be formulated as a *multicommodity flow* (MCF) problem, where each source/destination pair in the traffic set can be considered a different commodity that needs to be carried by the network [BaMu96, OzBe03, ChMV08]. While integer solutions to the MCF problem are usually desired in an optical network, corresponding to routing each demand over a single path, the integrality constraints are relaxed in the LP approach, to make the problem more tractable. Despite not enforcing integer solutions, the LP can be

combined with various clever techniques to improve the likelihood that such a solution is found.

In a typical MCF formulation, a flow cost function that monotonically increases with W_l, the number of wavelengths utilized on each link l, is used to reduce the number of "wavelength-hops" in the solution. This is equivalent to reducing the number of regenerations in an O-E-O network. (Modified cost functions may be desirable for networks with optical bypass.) Using a cost function that is convex, such as W_l^2, favors solutions with good load balancing, as the incremental cost increases as a link becomes more heavily loaded. A nonlinear cost function should be input as a piecewise linear function into the LP formulation, where the breakpoints occur on integers to encourage integer solutions [OzBe03].

The results of Ozdaglar and Bertsekas [OzBe03] and Christodoulopoulos et al. [ChMV08] indicate that by employing cost functions with integer breakpoints and using random perturbation techniques (where the slope of the cost function is changed very slightly on each link to reduce the likelihood of two paths looking equally good [ChMV08]), integer solutions are produced in most of the routing instances tested. If integer solutions are not produced by the LP, then rounding (or alternative techniques) can be used, although the results may be suboptimal. In some of the proposed strategies, the LP jointly solves both the routing and wavelength assignment problems rather than treating the routing as a separate step; this is discussed in Chap. 5.

Even with LP relaxation techniques, run times grow rapidly with the number of connections, as reported in Banerjee and Mukherjee [BaMu96]. The run times reported in Christodoulopoulos et al. [ChMV08] are reasonable; however, the sample network is small.

More research is needed to determine how practical these approaches are to realistic network design instances, and how extensible they are to various grooming and protection schemes. Designs on actual carrier networks often need to account for numerous real-world limitations and conditions. These would need to be incorporated as additional constraints in the LP problem. Furthermore, while LP (or ILP) techniques may be able to minimize an objective function, the solution may not be robust. Small changes to the assumptions regarding the network may result in a markedly different solution. It is typically preferable to find a solution that, while perhaps not yielding the absolute minimum cost, has a broad "trough," such that it is relatively immune to changes in network conditions. In some scenarios, heuristics may be better at achieving such solutions, as the heuristics may incorporate "engineering judgment" as opposed to attacking the design problem from a more mathematical perspective.

3.10 Multicast Routing

Multicast traffic involves one source communicating with multiple destinations, where the communication is one-way. Multicast is also referred to as point-to-multipoint communication, in contrast to a point-to-point connection between a

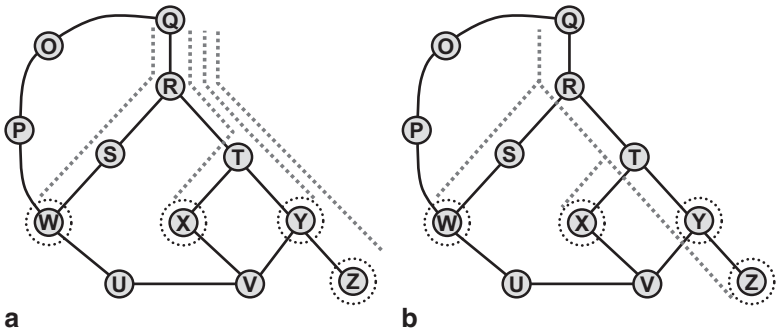


Fig. 3.20 **a** Four unicast connections are established between the source, *Q,* and the destinations, *W, X, Y,* and *Z.* **b** One multicast connection is established between Node *Q* and the four destinations

single source and single destination. (A tutorial on multicast routing can be found in Sahasrabuddhe and Mukherjee [SaMu00].) The need for multicast could arise, for example, if the optical network is being used to distribute video simultaneously to multiple cities. Rather than setting up a separate unicast connection between the source and each of the destinations, where multiple copies of the signal may be transmitted on a link, a multicast tree is constructed to reduce the amount of required capacity. The multicast tree connects the source to each of the destinations (without any loops), such that just one copy of the signal is sent on any link of the tree. This is illustrated in Fig. 3.20, where Node Q is the source and Nodes W, X, Y, and Z are the destinations. In Fig. 3.20a, four separate unicast connections are established between Node Q and each of the destinations. Note that four connections traverse the link between Nodes Q and R. In Fig. 3.20b, a single multicast tree is established, as shown by the dotted line, which requires significantly less capacity. Nodes R and T are branching nodes of this multicast tree.

To investigate the capacity benefits of multicast in a realistic network, a study was performed on Reference Network 2 [SaSi11]. Five thousand multicast sets were generated, with one source node and D destination nodes, where D was uniformly distributed between 5 and 15. The destinations were chosen based on their traffic weightings. (A typical demand set for this network was used to obtain the weightings. However, the results were similar when the nodes were selected with equal likelihood.) The study compared routing D unicast connections versus one multicast connection. The results showed that multicast provided a factor of roughly three benefit in capacity as compared to multiple unicast connections, where capacity was measured as the average number of wavelengths required on a link. (Approximately the same capacity benefits were obtained when capacity was measured in terms of bandwidth-distance, or in terms of the number of wavelengths needed on the most heavily utilized link.) If the number of destination nodes was uniformly distributed between 2 and 6 (instead of 5 and 15), the capacity benefit was a factor of roughly 1.5.

A tree that interconnects the source and all of the destinations is known as a Steiner tree (where it is assumed that all links are bidirectionally symmetric, i.e.,

two-way links with the same metric in both directions). The weight of the tree is the sum of the metrics of all links that comprise the tree. Finding the Steiner tree of minimum weight is in general a difficult problem to solve (unless the source is broadcasting to every node in the network); however, several heuristics exist to find good approximate solutions to the problem [Voss92].

We present two of the heuristics below, which we refer to as *minimum spanning tree with enhancement* (MSTE) and *minimum paths* (MP) (various other names are used in the literature, e.g., [Voss92] refers to these two heuristics as *shortest distance graph* and *cheapest insertion,* respectively). MSTE and MP were tested on the three reference networks of Sect. 1.10 to compare their relative performance. Thousands of scenarios were run, with the number of multicast destinations varying between 5 and 15. The source and the set of destinations nodes were randomly selected.

When the objective was minimum-distance trees, the two heuristics produced the same (or close to the same) results in about 75 % of the tests with Networks 1 and 2 and 85 % of the tests with the smaller Network 3. In the scenarios where there was a significant difference in the results, MP outperformed MSTE roughly 85 % of the time for Networks 1 and 2, and about 70 % of the time for Network 3.

When the objective was minimum-hop trees, the two heuristics produced the same results in about 55 % of the Network 1 tests, 65 % of the Network 2 tests, and 80 % of the Network 3 tests. In the scenarios where the results differed, MP outperformed MSTE roughly 70 % of the time for Network 1, 60 % of the time for Network 2, and 55 % of the time for Network 3.

Overall, MP performed better in more scenarios, but not uniformly. The relative performance of MSTE tended to improve when there were more destinations in the test set. The details of the two heuristics are presented next; the code for both heuristics is included in Chap. 11.

3.10.1 Minimum Spanning Tree with Enhancement

The steps of the MSTE heuristic are presented first, followed by a small example. This heuristic follows the MST heuristic of Kou et al. [KoMB81] combined with the enhancement of Waxman [Waxm88].

The original network topology is referred to here as *A*. The first step is to create a new topology *B* composed of just the source and destination nodes. In this particular heuristic, the procedure is the same regardless of which of the nodes is the source. All of the nodes are fully interconnected in *B,* where the metric of the link connecting a pair of nodes equals the metric of the shortest path between the two nodes in *A*. A minimum spanning tree is found on *B* using an algorithm such as Prim's [CLRS09]. (A minimum spanning tree is a tree that touches all nodes in the topology where the sum of the metrics of the links comprising the tree is minimized. This is different from a minimum Steiner tree in that *all* nodes are included and is easier to solve.) Each link in the resulting minimum spanning tree, where a link connects

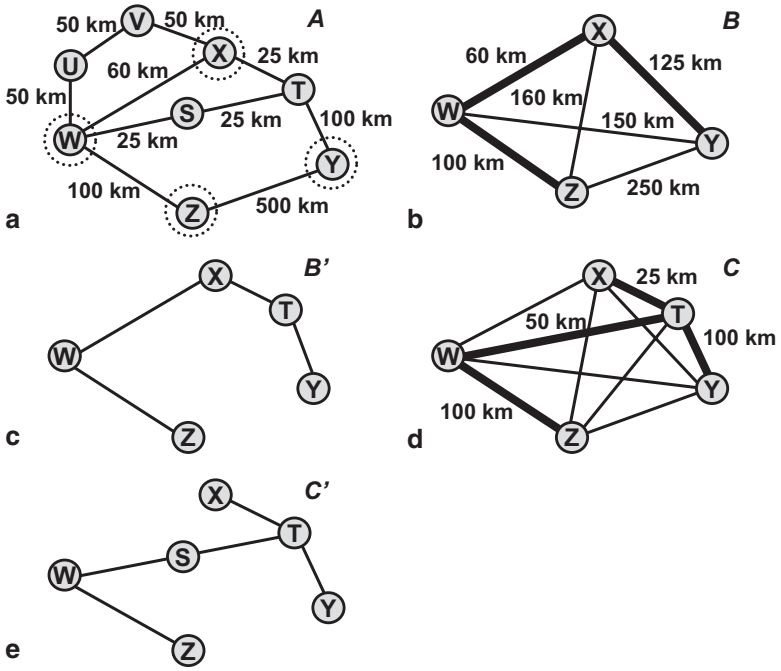


Fig. 3.21 **a** Original topology *A,* where the source and destination nodes are *circled.* **b** Topology *B* formed by interconnecting all source and destination nodes. **c** The minimum spanning tree on *B* expanded into paths, forming topology *B′*. **d** Topology *C* formed by interconnecting all nodes of *B′*. **e** The minimum spanning tree on *C* expanded into paths, forming topology *C′*

two nodes *i* and *j,* is expanded into the shortest path in the true topology between nodes *i* and *j* (e.g., if the shortest path in topology *A* between nodes *i* and *j* has three links, then the link between nodes *i* and *j* in topology *B* is replaced by the three links). Call the resulting topology *B′*. (MST without the enhancement terminates here.) Another new topology is then formed, *C,* composed of all of the nodes in *B′,* with all of the nodes fully interconnected. The operations performed on topology *B* are repeated for topology *C,* resulting in the topology *C′*. A minimum spanning tree is then found on *C′*. Any links in the resulting tree that are not needed to get from the source to the destinations, if any, are removed, leaving the approximation to the minimum Steiner tree.

This heuristic is illustrated with the small example of Fig. 3.21. Using the notation from above, the topology shown in Fig. 3.21a is the *A* topology. The source is Node W and the multicast destinations are Nodes X, Y, and Z. However, as noted above, the procedure is the same regardless of which of the four nodes is actually the source. A fully interconnected topology composed of these four nodes is shown in Fig. 3.21b. This is the *B* topology. The link metrics are assumed to be based on distance, e.g., the metric of Link WY, 150 km, is the shortest distance between Nodes W and Y in the *A* topology. The minimum spanning tree on *B* is shown by the

thick lines. Each link in this tree is expanded into its associated shortest path in the original topology. This produces the B' topology shown in Fig. 3.21c. For example, Link XY in B is expanded into X-T-Y in B'. A new fully interconnected topology is formed with the five nodes of B', as shown in Fig. 3.21d. This is the C topology, where the minimum spanning tree is shown by the thick lines. The links of the tree are expanded into their associated shortest paths to form the C' topology, shown in Fig. 3.21e. This topology is already a Steiner tree; thus, no further operations are needed, leaving C' as the final solution. In this example, the solution is optimal but that is not always the case.

3.10.2 Minimum Paths

The MP heuristic was first proposed in Takahashi and Matsuyama [TaMa80]. In contrast to the MSTE heuristic, the MP heuristic may produce different results depending upon which node is the source node. The first step in MP is to begin building a tree, where initially the tree is composed of just the source node. In each successive step, the destination node that is closest to *any* of the nodes already in the tree is added to the tree, along with the shortest path between the tree node and that destination node (i.e., all nodes and links of the shortest path that are not already in the tree are added to the tree). This continues until all destination nodes have been added to the tree, resulting in a multicast tree.

Running the MP heuristic on the example of Fig. 3.21, where Node W is assumed to be the source node, does *not* produce the optimal solution, as explored further in Exercise 3.15.

To improve the performance of MP, the algorithm can be run N+1 times, where N is the number of destination nodes; in each run a different node is designated as the "source" node. The best resulting tree is taken as the solution. Another enhancement is treating the tree produced by MP as topology B' in the MSTE heuristic; steps (d) and (e), as illustrated in the example of Fig. 3.21, are then applied to this topology.

3.10.3 Regeneration in a Multicast Tree

Heuristics such as the ones described above can be used to generate an approximate minimum-cost multicast tree, where the link metric is set to unity for O-E-O networks and is set to link distance for optical-bypass-enabled networks. If it is known in advance the possible sets of nodes that will be involved in multicast demand requests, then one can pre-calculate the associated multicast trees. Furthermore, alternative multicast trees can be generated for each set of nodes where certain bottleneck links are avoided in each tree. The best tree to use would be selected at the time the multicast demand request arrives. If the multicast groups are unknown in advance, then the tree can be generated dynamically as the requests arrive. If it is necessary to add nodes to an existing multicast group, then greedy-like heuristics can be used to determine how to grow the tree [Waxm88].

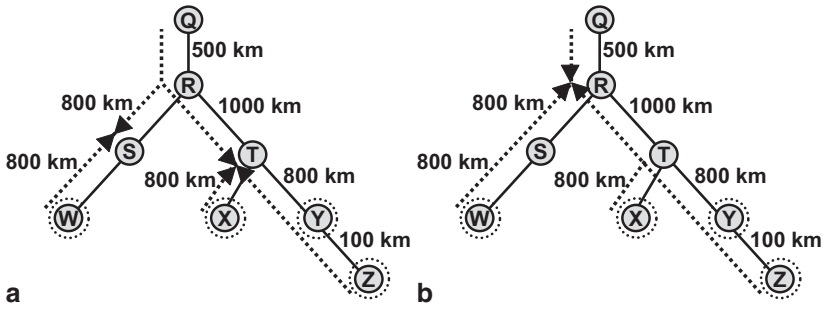


Fig. 3.22 Assume that the optical reach is 2,000 km. **a** Regeneration at both Nodes *S* and *T*. **b** Regeneration only at Node *R*

As with unicast connections, the actual amount of regeneration in a multicast tree with optical bypass will likely depend on factors other than distance. For example, the branching points in the tree may be favored for regeneration. Refer to Fig. 3.22, which shows just the links included in the multicast tree of example Fig. 3.20, where the source is Node Q and the multicast destinations are Nodes W, X, Y, and Z. Assume that the optical reach is 2,000 km. In Fig. 3.22a, the signal is regenerated at the furthest possible node from the source without violating the optical reach. This results in regenerations at Nodes S and T. If, however, the regeneration occurs at the branching point Node R as in Fig. 3.22b, then no other regeneration is needed.

Note that optical bypass does not permit the wavelength of the signal to be changed, except when the signal is regenerated. Thus, in Fig. 3.22b, the signal is carried on the same wavelength for all links in the tree below Node R; this wavelength can be different, however, from the wavelength used on Link QR. The wavelength continuity constraint in a multicast tree may make wavelength assignment difficult. It may be necessary to add in a small number of regenerations for the purpose of wavelength conversion.

A *directionless* ROADM *with multicast capability* (Sect. 2.9.8) is very useful in a multicast tree. At a branching node where regeneration does *not* occur, e.g., Node T in Fig. 3.22b, the ROADM can all-optically multicast the signal onto the outgoing branches (the ROADM includes amplification such that the outgoing power level is not cut in half). Second, at a node with regeneration, e.g., Node R in the same figure, the ROADM can multicast the regenerated signal onto both Links RS and RT. One transponder is used to receive the signal from Link QR, and one transponder is used to transmit the signal on both Links RS and RT. Finally, the drop-and-continue feature of the ROADM can be used at Node Y, where the signal drops and continues on to Node Z.

A *non-directionless* ROADM *with multicast capability* can perform the first and third of these functions, i.e., all-optical multicast from one input network fiber to multiple output network fibers, and drop-and-continue. However, to implement multicast after a regeneration with a non-directionless ROADM, multiple transponders are needed to transmit the signal on each of the outgoing links. For

example, in Fig. 3.22b, one transponder is needed to transmit the signal on Link RS and a second transponder is needed to transmit the signal on Link RT. (Another option is to use a flexible regenerator card, as is described in Sect. 4.7.2.)

If the ROADMs are not capable of any type of optical multicast (and assuming flexible transponders/regenerators are not used), then the signal must be duplicated *in the electrical domain* at all branching points. This provides further motivation to attempt to align the branching points with the required regeneration points, to minimize the number of transitions to the electrical domain.

3.10.4 Multicast Protection

Consider providing protection for a multicast connection such that the source remains connected to each of the destinations under a failure condition. One simplistic approach is to try to provision two disjoint multicast trees between the source and the destinations. However, because of the number of links involved in a multicast tree, it is often difficult, if not impossible, to find two completely disjoint trees.

Segment-based multicast protection has also been proposed, where the multicast tree is conceptually partitioned into segments and each segment is protected separately. In one scheme, the source node, each tree branch node, and each destination node are segment endpoints [WGPD08]. These nodes cannot serve as intermediate nodes along any other segment. A disjoint protection path is found for each individual segment. Assuming that there is never more than one failure in the tree, then the segment protection paths do not have to be diverse from one another. Furthermore, the protection paths can incorporate portions of the multicast tree that survive the failure.

Another segment-based multicast protection scheme is *level protection,* as proposed in [PaEA12]. The source node and each destination node of the multicast tree serve as segment endpoints; they cannot also be intermediate nodes of any segment. The source is assigned to level 0; the destination nodes are assigned a level that is determined by the number of segments that are traversed to get to that destination node. The first directed protection tree is found from the source to all of the destination nodes in level 1. In the next step, a second directed protection tree is found from *any* node in levels 0 or 1 to all of the destination nodes in level 2. This process continues, where the $(j+1)$st directed protection tree is found from *any* node in levels 0, 1,…, j to all of the destination nodes in level $j+1$. There are various criteria for disjointness that must be enforced between the protection trees and the arcs that connect the nodes in different levels of the tree (an arc is a directed link); however, the scheme does allow for a relatively high degree of capacity sharing as well (e.g., between protection trees) to improve its efficiency. The protection trees guarantee that under a single link failure, a path can still be found from the source to any destination node in any level.

Another protection approach is to arbitrarily order the multicast destinations, and then sequentially find a disjoint pair of paths between the source and each destination, using an SPDP algorithm [SiSM03]. When looking for a disjoint pair of paths between the source and the $(i+1)$st destination, the links that comprise the path pairs between the source and the first i destinations are favored. The resulting sub-graph, which may not have a tree structure, provides protection against any single failure.

Finally, a very different approach that has been applied to multicast protection is *network coding,* where the destinations receive independent, typically linear, combinations of signals over diverse paths [KoMe03, MeGa08, MDXA10]. The linear combinations are such that if one signal is lost due to a failure, it can be recovered (almost) immediately from the other signals that are received. By sending combinations of signals, rather than sending duplicate copies of each individual signal, network resources may be used more efficiently. The topic of protection based on network coding is covered in more detail in Chap. 7. Network coding may also reduce the capacity requirements in unprotected multicast scenarios, which was the original context in which this methodology was investigated [ACLY00].

3.10.5 Manycast

In the multicast variant known as *manycast,* one source communicates with *any* N of the destination nodes, for some specified N. This could be useful, for example, for distributing huge data sets that require processing. The processing centers may be distributed among M cities, but it is necessary to utilize only N of them (where $N<M$). There are $\binom{M}{N} = \frac{M!}{(M-N)!N!}$ possible ways to select the N processing centers. If this number is relatively small, then one can calculate the shortest tree for each possible set of N processing centers, e.g., using one of the Steiner tree heuristics presented earlier, and then select the best solution. If the number of possibilities is prohibitively large, then a heuristic can be used.

One such manycast heuristic, which is derived from the MP algorithm of Sect. 3.10.2, is presented in Charbonneau and Vokkarane [ChVo10]. Assume that the number of possible destination nodes is M, but that only N of them must be reached by the manycast tree. M candidate manycast trees are created by the heuristic. The first step in creating the ith tree is to add the shortest path between the source node and the ith destination node. In all successive steps, the destination node that is closest to *any* of the nodes already in the tree is added to the tree, along with the shortest path between the tree node and that destination node (i.e., all nodes and links of the shortest path that are not already in the tree are added to the tree). This continues until a total of N destination nodes have been added to the tree. The M trees created in this manner can then be compared on criteria such as

the total tree distance, number of tree hops, amenability to wavelength assignment, etc., where the best tree is chosen as the solution. (This algorithm is explored further in Exercise 3.16.)

If N equals one, the problem is referred to as *anycast*. Because the source communicates with just one destination at a given time, *anycast* is not generally considered a form of multicast. *Anycast* arises, for example, with cloud computing, where particular data resources and applications are replicated in a set of M data centers [DDDP12]. It is assumed that the cloud user (e.g., an enterprise) needs to access *any one* of the M data centers, ideally via a low-latency path.

Another variation of manycast occurs when the available resources are not distributed equally among the possible destination nodes. In this scenario, referred to as *multi-resource manycast,* the number of required destination nodes depends upon the set of destinations that are selected. Several heuristics are proposed for this problem in [ShJu07], all of which are derived from the MP tree-growing methodology of Sect. 3.10.2. These heuristics differ in the criterion used for selecting the order in which destination nodes are added to the tree. Of the criteria considered, the most effective one is to add the destination node that minimizes the quantity (*path cost of adding the node*)/(*available node resources*). Destination nodes are added to the tree until the sum of the added resources reaches the desired threshold.

3.11 Multipath Routing

Typically, a demand between two endpoints is routed over a single path in the optical layer. However, as discussed below, there are scenarios where it is beneficial to split the aggregate signal of the demand into lower-rate signals, with each lower-rate signal potentially being transmitted over a different path from source to destination. The process of splitting the aggregate signal into lower-rate signals is known as *inverse multiplexing*. The International Telecommunication Union (ITU) standard that supports inverse multiplexing in the optical layer (e.g., in SONET/SDH or OTN) is *Virtual Concatenation* (VCAT) [Choy02, BCRV06]. This allows, for example, a 40-Gb/s demand to be split into four 10-Gb/s signals. Furthermore, VCAT supports *multipath routing,* where the lower-rate streams can be routed over different paths.

It is the responsibility of the destination node to reconstruct the aggregate signal from the lower-rate streams. The lengths of the paths over which the demand has been split are unlikely to be equal, resulting in different fiber propagation delays. To account for the *differential delay* between the longest of these paths and the other paths, the destination must buffer the traffic until the data arrives on all of the paths. The speed of light in fiber is approximately 2×10^8 m/s; thus, a difference of 2,000 km in route distance corresponds to a 10-ms difference in end-to-end delay. Passing through nodal equipment can also contribute to delay; however, in the optical layer, this delay component is typically much smaller than the fiber propagation delay.

In order to maintain feasible buffer sizes, especially at high data rates, it is important to limit the differential delay. This motivates developing strategies for finding a set of paths where the differential delay between any two of the paths is less than some maximum. More precisely, one should consider the sum of the differential delays between the longest path and each of the other paths, because that determines the required buffer size [SrSr06]. However, for simplicity, the focus is typically on the differential delay between the longest and shortest paths. Concern over the differential delay is more relevant in a backbone network where path distances can be very long. VCAT specifies the maximum allowable differential delay for a particular connection rate [BCRV06]; however, in practice, the enforced maximum is typically much lower than what the standard allows.

3.11.1 Non-Disjoint Multipath Routing

Splitting traffic over multiple paths can be beneficial when adding a new demand to a congested network. While routing the demand over a single path may be desirable, it may not be possible to find a single path with enough bandwidth to support the new connection, or the only paths that do have enough bandwidth are very circuitous. Finding a single path with sufficient bandwidth is especially challenging when routing large demands composed of multiple wavelengths [CJDD09]. Rather than blocking the new demand request, partitioning the demand into lower-rate signals that can be routed over different paths may allow it to be accepted. (This strategy is especially relevant in the flexible bandwidth approach covered in Chap. 9.) Note that the set of paths over which the demand is routed are not required to be completely disjoint, i.e., there can be links that are common to more than one of the paths.

One strategy for finding a path set of sufficient capacity that meets a specified *differential delay constraint* (DDC) utilizes a "sliding window" approach [SAAG05]. A KSP algorithm generates a list of K paths between the source node and destination node, where a K of 20–25 is recommended in Srivastava et al. [SAAG05]. In order of increasing path distance, we label these paths $P_1, P_2, \ldots, P_k$. The selection process starts with path P_1. It assigns as much of the required demand bandwidth as possible to this path. It then considers P_2. If the differential delay between P_1 and P_2 is less than the DDC, then it assigns as much of the remaining demand bandwidth as possible to P_2. The process continues until either all of the required bandwidth has been assigned, in which case the algorithm terminates successfully, or a path P_j is reached where the differential delay between P_j and P_1 is greater than the DDC. In the latter scenario, all of the bandwidth is unassigned and the bandwidth assignment process restarts with P_2, where the differential delays are now relative to the delay of P_2. If the process fails again, it restarts with P_3, etc. This process continues until enough capacity is found to carry the new demand or the strategy fails.

Because the set of K paths may have links in common, assigning bandwidth to one path may preclude a path with a higher index from being used. Thus, the sliding

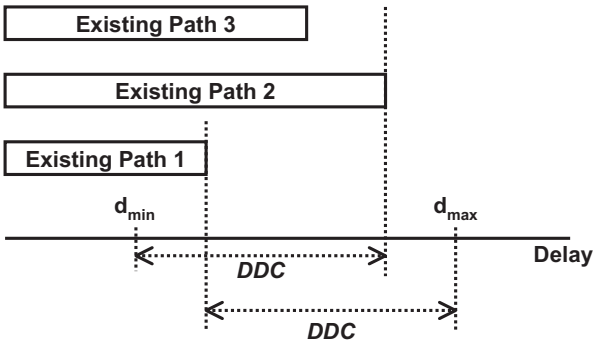


Fig. 3.23 Assume that a demand is carried over three existing paths, where the end-to-end delays of these existing paths are shown. A new path is desired that satisfies the differential delay constraint (*DDC*) relative to all of the existing paths. Thus, the delay of the new path must fall between $d_{\min}$ and $d_{\max}$

window strategy is not necessarily optimal, in terms of minimizing the bandwidth-links utilized or minimizing the number of paths over which the demand is split. However, because the strategy starts with P_1, the solution tends to favor routing over shorter paths, which is desirable from the point of view of *absolute* latency of the paths. Note that the absolute latency of the demand is determined by the delay of the *longest* path.

If the number of path hops is more important than absolute latency, an extra step can be implemented: in the jth iteration, where paths P_j , P_{j+1},…, P_{j+m} are the possible paths in the iteration (i.e., P_{j+m} is the longest of the K paths that still satisfies the DDC with respect to P_j), these $m+1$ paths are first sorted from fewest hops to most hops. As much of the demand bandwidth as possible is assigned to the path with the fewest hops; as much of the remaining bandwidth as possible is then assigned to the path with the second fewest hops, etc. This process continues until all of the required bandwidth is assigned or until the jth iteration fails. In case of the latter, the algorithm then moves to the $(j+1)$st iteration. This modified strategy favors routing over paths with fewer hops, which is desirable for reducing cost in an O-E-O network.

Next, consider a demand that has already been established over multiple divergent paths. Assume that the client associated with this demand requests more bandwidth and that increasing the bandwidth of one of the existing paths is not possible, i.e., a new path must be found to satisfy the increase in bandwidth. Additionally, assume that the existing paths that carry the demand cannot be rerouted. (The ITU standard that supports dynamic adjustment of service bandwidth is the *Link Capacity Adjustment Scheme* (LCAS) [ITU06].) The differential delay between the new path and each one of the existing paths must be less than the DDC.

This is illustrated in Fig. 3.23, where it is assumed that there are three existing paths for the demand; the figure indicates the end-to-end delay of each of the paths. (Note that the three existing paths are not necessarily the three shortest paths between the source and destination, such that the new path could be shorter than all of them.) To meet the DDC, it is necessary to specify both lower and upper bounds on the delay of the new path, i.e., $d_{\min}$ and $d_{\max}$, as shown in the figure. This is equivalent to specifying lower and upper bounds on the *distance* of the new path, which

we represent by D_{min} and D_{max}, respectively. The objective is to find a new path (not necessarily completely disjoint from the existing paths) with a distance that falls between D_{min} and D_{max}.

One option is to run a KSP algorithm (first removing any links that do not have enough capacity to support the additional bandwidth) until a path is found with the desired distance. However, if D_{min} is large, hundreds of iterations of the KSP algorithm may be needed. Another strategy, inspired by Lagrangian methods, modifies the link distances prior to running the KSP algorithm [AhKK06]. A link of length L is set to have a link metric of:

$$L\left(1/\left(D_{min}\cdot D_{max}\right)\right)+1/L.$$

The KSP algorithm is run with the new link metrics until a path of the desired distance is found. The purpose of the modified metric is to drive the KSP algorithm to a path of suitable distance in fewer iterations.

To get an idea of the efficacy of this methodology, it was tested using Reference Networks 1 and 2. (Reference Network 3 is relatively small, such that using a KSP algorithm with the *real* link lengths as the metric to find a path of a desired distance typically does not require a lot of iterations.) For each source/destination pair in these two networks, the shortest dual path was found; assume that the longer of the two paths has length D km. Then, the link-metric-modification scheme was run to find a path with distance falling between D and $D+1000$ km; these bounds were arbitrarily chosen for purposes of the test. For 90% of the source/destination pairs, using the modified link metrics required 15 or fewer KSP iterations to find a path satisfying the distance criteria. In contrast, for roughly 10% of these same pairs, running KSP with the *real* link lengths required more than 50 iterations to find a path of the desired length; many required more than 100 iterations. Interestingly, for all of the source/destination pairs where more than 100 iterations were required using the modified link metrics, running KSP with the real link length as the metric found a feasible path with 6 or fewer iterations. Thus, the two sets of link metrics are somewhat complementary in their efficiency, at least in this test.

3.11.2 Disjoint Multipath Routing

The next multipath application that we consider is where traffic protection is required. In order to reduce the amount of protection capacity that needs to be allocated, it is possible to employ schemes where the demand is split over multiple working paths and/or there are multiple backup paths for the demand. For example, in the protection scheme discussed in Huang et al. [HuMM11], there are multiple working paths, all of which are protected by a single backup path. Depending on how protection is implemented, it may be necessary that all of the working paths and the backup path be mutually link-disjoint. Furthermore, the differential delay

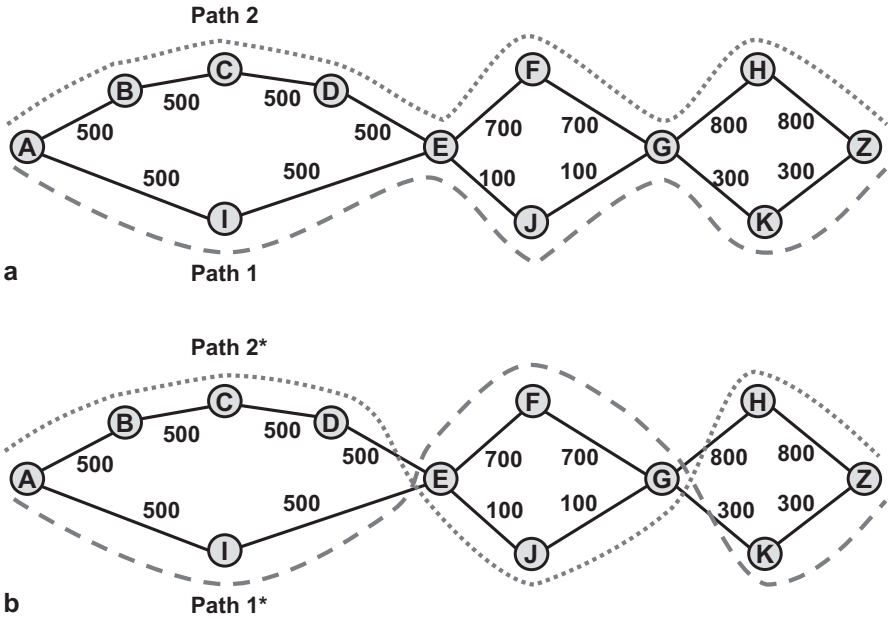


Fig. 3.24 Assume that the maximum allowable differential delay is 5 ms, corresponding to a maximum difference in path distance of 1,000 km. The link labels indicate link distance, in km. **a** The initial paths have distances of 1,800 km and 5,000 km; the difference, 3,200 km, exceeds the 1,000-km limit. **b** The subpaths that run between nodes *A, E, G,* and *Z* can be swapped to form two new paths of distance 3,000 km and 3,800 km; the difference, 800 km, satisfies the differential delay constraint

among the working paths must be less than some maximum. (For some protection schemes, it may be desirable to extend the differential delay limit to the backup path as well.) This corresponds to finding a set of *disjoint* paths that satisfy a DDC. Note that another benefit of partitioning a demand into disjointly routed lower-rate streams is that if a working path and the backup path both fail, or if there is no backup path, then there is still connectivity between the source and destination over the remaining working paths, albeit at a reduced rate.

The strategy proposed in Huang et al. [HuMM11] for finding a set of disjoint paths satisfying a DDC starts by enumerating the *N* shortest *link-disjoint* paths between the source and destination (e.g., using the Bhandari algorithm), where *N* is usually no more than two to four in a backbone network. Assuming that the DDC is not met by this path set, the process identifies the "merge nodes" that fall in two or more of the paths, where the source and destination are included as merge nodes. It then forms a list of new paths from different combinations of the subpaths that lie between the merge nodes. From the paths that are formed by this process, a set of link-disjoint paths that satisfies the DDC is selected, if possible.

This subpath swapping process is illustrated in Fig. 3.24 for a demand between Nodes A and Z. Assume that the DDC is 5 ms, corresponding to a maximum

difference in path distance of 1,000 km. The link-disjoint paths, Path 1 and Path 2, shown in Fig. 3.24a with distances of 1,800 km and 5,000 km, respectively, do not satisfy the DDC. The merge nodes are A, E, G, and Z. Four different link-disjoint path sets between A and Z can be formed from the subpaths that run between the merge nodes. The two link-disjoint paths with the minimum differential delay are Path 1* and Path 2* as shown in Fig. 3.24b, with distances of 3,000 km and 3,800 km, respectively. These two paths satisfy the DDC.

If this method fails to find a satisfactory solution, then one could create a larger list of possible subpaths by finding more circuitous paths between the merge nodes (but still maintaining disjointness). If no merge nodes exist for the path set, or if both link and node disjointness are desired, then this is equivalent to looking for more circuitous end-to-end paths in order to reduce the differential delay. The link-metric-modification strategy described above can be helpful in finding paths or subpaths in a desired distance range.

Ideally, the distance of the *longest* working path remains the same in this process, or is reduced due to subpath swapping, as that will prevent the *absolute latency* of the demand from increasing. Some, or all, of the remaining paths will increase in distance to reduce the differential delay. Essentially, *the network itself is being used as "storage,"* to reduce the required buffer size at the destination.

In addition to protection applications, splitting a demand over disjoint paths may be useful for security as well. For example, sensitive transmissions may be partitioned into multiple lower-rate signals that are sent over disjoint paths to reduce the probability of the data being intercepted by adversaries.

3.12 Exercises

3.1 Provide an example of where the Dijkstra shortest-path algorithm yields a different result from the Breadth-First-Search shortest-path algorithm. In your example, if the source and destination are swapped, do the two algorithms still yield different results?

3.2 Consider bidirectional and unidirectional rings of N nodes, with N odd. In the bidirectional ring architecture, assume that the traffic is routed over the fewest hops; the ring is equipped with two fibers, one for the clockwise traffic and the other for the counter-clockwise traffic. In the unidirectional ring architecture, assume that all traffic is routed in the clockwise direction, over just one fiber. (a) If every node in the ring sends one wavelength to each of the other nodes in the ring, how many wavelengths are *utilized* on the fibers in the two types of rings? (Note that it is asking for the utilization, not a wavelength assignment, i.e., optical bypass is not relevant in this exercise.) (b) Next, assume that the traffic is protected, where the protect path is routed over a disjoint path (i.e., in the bidirectional ring, this traffic is routed over the longer path; in the unidirectional ring, this traffic is routed counter-clockwise over a second fiber). How many wavelengths are *utilized* on the fibers in the two types of rings?

3.3 In Sect. 3.4.1, it is stated that if the minimum number of regenerations found is *R*, and at least one of the paths found by the KSP algorithm has a distance strictly greater than $(R+1)$*[Optical Reach], then the full set of minimum-regeneration paths must have been found. Prove that this statement is true.

3.4 Specify the pseudo-code for an algorithm that is guaranteed to find a minimum-regeneration pair of disjoint paths, without looking at every possible pair of paths. One suggestion is to have an outer loop use a KSP algorithm to search for the primary path and an inner loop use a KSP algorithm to search for a diverse secondary path. Consider judicious conditions that allow the loops to terminate, such that the run time is reasonable.

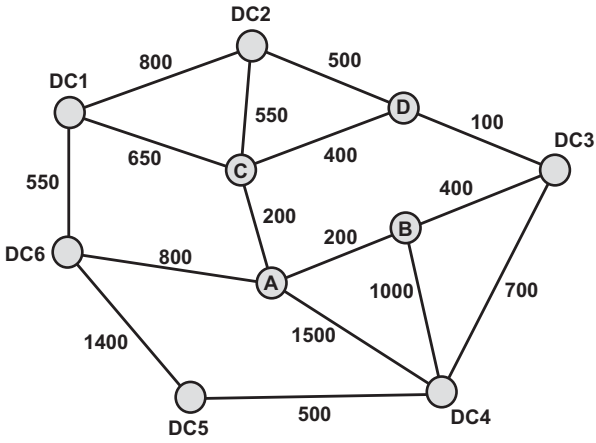
3.5 Assume that the optical reach in a network is *R* and that no links have a distance longer than *R*. We say that a path *P* requires the number of regenerations predicted based on its distance, D_p, if the number of required regenerations equals *floor* $[(D_p-\varepsilon)/R]$, for some arbitrarily small $\varepsilon>0$. (a) How many hops (i.e., links) must be in a path before the number of required regenerations could be greater than that predicted based on its distance? (b) Assume that a multi-hop path from Node A to Node B and a multi-hop path from Node B to Node C both require the number of regenerations predicted by their respective path distances. Does this imply that if the two paths are concatenated to form a path from A to C that this new path will also require the number of regenerations predicted by its distance?

3.6 (a) Assume that the minimum number of regenerations required for any path between a given source and destination is R*. What is the maximum number of regenerations that could be required for a *shortest* path between that same source and destination? (Assume that all nodes support optical bypass, regeneration occurs only in nodes, regeneration is solely determined by distance, and no link has a distance that is longer than the optical reach.) (b) Assume that the minimum number of regenerations required for any pair of diverse paths between a given source and destination is *Q**. What is the maximum number of regenerations that could be required for a *shortest* pair of diverse paths between that same source and destination? (*Shortest* is where the sum of the distances on the two paths is minimized.)

3.7 One metric that may be used for selecting which candidate path to use for a demand request in alternative-path routing is $\sqrt{H}/W$, where *H* is the number of hops in the path and *W* is the number of wavelengths that are free on the most heavily loaded link of the path. The path that minimizes the metric is selected. Discuss the merits of this metric.

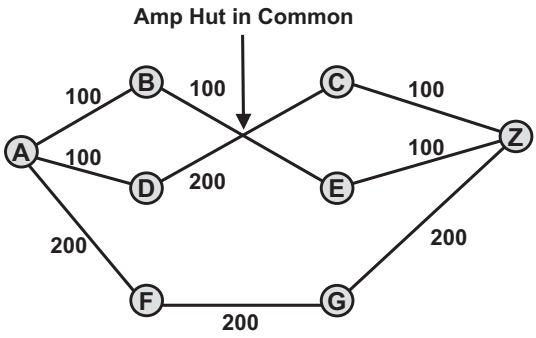
3.8 The network shown below has six data centers (DCs), labeled DC1 through DC6, where resources are replicated at all six DCs. The link labels indicate the link distance. Assume that a cloud-computing user is located at Node B. (a) Assume that the user is being served by DC3. Find the minimum-distance pair of link-diverse paths from Node B to DC3. Is it possible to find a shorter pair of paths if protection can be provided using link-diverse paths to DC3 and to a

second DC (any of the five other DCs)? *For the remainder of this exercise assume that DC3 is congested and cannot be assigned to the user. Traffic can still be routed through the node where this DC is located.* (b) Assume that the user is assigned to DC4 instead of DC3. Find the minimum-distance pair of link-diverse paths from Node B to DC4. Is it possible to find a shorter pair of paths if protection can be provided using link-diverse paths to DC4 and to a second DC (any of the other DCs, except for DC3)? (c) Assume that the user's application is mission critical. Find the shortest link-diverse paths from Node B to three *different* DCs (i.e., any three of the DCs, except for DC3). What is the total distance of the three paths? (d) Assume that three link-diverse paths are desired but only two-way diversity is required among the DCs, i.e., the three link-diverse paths can terminate on either two or three different DCs, whichever produces the shortest total path distance. How should the graph be transformed so that an SPDP algorithm (that looks for three diverse paths) can be used? What is the shortest set of three link-diverse paths that meets the requirements, and which DCs are used?

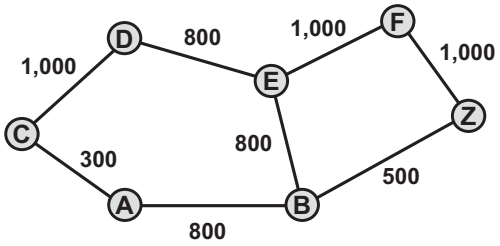


3.9 For the fork configuration SRLG shown in Fig. 3.17, it would be sufficient to look for *link-disjoint* paths in the transformed graph (i.e., Fig. 3.18) to avoid selecting two links in the same SRLG. Provide an example of a fork configuration SRLG, where both link-and-node-disjointness must be enforced on the transformed graph.

3.10 Consider an SRLG configuration consisting of two links, where the only point in common between the two links is an amplifier hut somewhere in the middle of the two links, i.e., the links cross (see figure below). How can such an SRLG be handled in an SPDP algorithm to ensure that the two crossing links are not treated as being diverse?

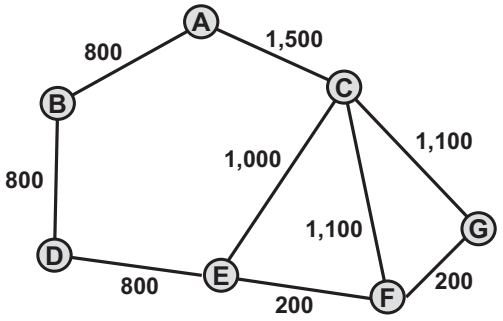


3.11 In the optical-bypass-enabled network shown below, assume that the optical reach is 2,000 km. The distance of each link, in km, is shown in the figure. Assume that Node A is equipped with a *non-directionless* ROADM that supports less than 100% add/drop to/from any link. Assume that the add/drop limit to/from Link AB has been reached in this ROADM. (a) Assume that a unidirectional connection is required from Node A to Node Z. What is the minimum-distance feasible path for this connection? (b) Next, assume that bidirectionally symmetric paths between Nodes A and Z are required, where the paths must utilize the same links and be assigned the same wavelength on a given link (in the reverse direction). What is the minimum-regeneration path that satisfies these requirements, and how many regenerations are required?



3.12 In the network shown below, assume that a multicast connection is desired between Node A and Nodes E, F, and G. Assume that: the optical reach is 1,500 km, back-to-back transponders are used for regeneration, and the nodes are equipped with directionless ROADMs/ROADM-MDs with full multicast capability. The distance of each link, in km, is shown in the figure. (a) What is the minimum-distance multicast tree? (b) How many transponders are required in this tree? (c) What is the minimum amount of transponders required by any multicast tree (i.e., not necessarily one of minimum distance)? (d) Of the multicast trees that require the minimum number of transponders, which one has the minimum distance? (e) Assume that the nodes are equipped with

ROADMs/ROADM-MDs that support all forms of multicast except drop-and-continue. What multicast tree requires the minimum number of transponders?



3.13 Consider a set of k multicast nodes, $N_1 \ldots N_k$. In the optimal minimum-distance Steiner tree, it does not matter which of the k nodes is designated as the source and which $k-1$ nodes are the destinations. (a) Does it matter which of the k nodes is the source node when finding the tree that requires the fewest number of transponders, assuming the network has an O-E-O architecture? (b) How about if the network is optical-bypass-enabled, with directionless ROADMs/ROADM-MDs with full multicast capability?

3.14 Multicast trees generally save capacity as compared to multiple unicast connections. Can you say anything about how the number of required transponders compares in the multicast versus multiple-unicast scenarios if: (a) all network nodes are equipped with directionless ROADMs/ROADM-MDs capable of full optical multicast? (b) all network nodes are equipped with ROADMs/ROADM-MDs that are *not* capable of any type of optical multicast? In both parts (a) and (b) assume that there is at least one branching point in the multicast structure, the path followed from the source to each destination is the same in all scenarios, the optical reach is the same in all scenarios, regenerations are not required for purposes of wavelength conversion, and flexible transponders/regenerators are not used.

3.15 (a) What multicast tree is produced when running the MP heuristic of Sect. 3.10.2 on the network of Fig. 3.21, with Node W treated as the source, and Nodes X, Y, and Z the destinations? (b) How does this compare with the solution found by the MSTE heuristic, shown in Fig. 3.21e? (c) Can a better solution be found by the MP heuristic if a different node (i.e., either X, Y, or Z) is treated as the source and W is treated as one of the destinations? (d) If the result from part (a), where W is the source, is treated as topology B' in the MSTE heuristic of Sect. 3.10.1, is the result improved (i.e., the remaining steps of the MSTE heuristic are performed on this B')?

3.16 Consider the network of Exercise 3.8. Assume that a grid user, located at Node B, requires access to *any three of the six* data centers. (a) Using the manycast routing heuristic presented in Sect. 3.10.5, with the metric being link distance, which three data centers are selected? (b) What is the total distance of the tree that is generated by the heuristic? (c) If the objective is to minimize the total resulting tree distance, is this the optimal choice of data centers? (d) Based on the results, can you suggest a modification to the manycast heuristic to further optimize it?

3.17 Consider the multipath example shown in Fig. 3.24, and assume that a demand will be split and routed on the two paths. (a) What effect does the subpath swapping shown in the figure have on the absolute latency of the demand? (b) Can anything be said regarding how this subpath swapping affects the utilized network capacity, in terms of utilized bandwidth-km?

3.18 Consider using multipath routing such that a demand requiring a total service rate of R is split *equally* among M *diverse* paths (of equal cost). Assume that if one of the M paths fails, it is required that a fraction P of the original total service rate still be achievable, where $0 \leq P \leq 1$. (a) How large must M be such that no explicit protection capacity is required for the demand (i.e., the working capacity of the $M-1$ surviving paths is sufficient)? (b) Assume that M is smaller than this threshold. If protection is provided by utilizing the $M-1$ surviving paths, where the working capacity of these paths is now supplemented by protection capacity, what is the *total* amount of protection capacity that is required? (c) Again, assume that supplemental protection capacity must be deployed. If *all* of this protection capacity is deployed along a diverse $(M+1)$st path, how much protection capacity is required?

3.19 Consider a four-node ring, with the nodes labeled sequentially 1–4. Assume that there is only one wavelength available per fiber. Assume that there is one bidirectional demand request between Nodes 1 and 3, and one bidirectional demand request between Nodes 2 and 4 (the two demand requests each require one wavelength of bandwidth). Requirement 1 is that the routing for each demand must be bidirectionally symmetric (i.e., the path from A to Z is followed in reverse for Z to A). Requirement 2 is that a demand cannot be split into lower-rate streams that are routed over different paths. (a) If both requirements are enforced, is it possible to route both demands? (b) How about if just requirement 2 is enforced? (c) How about if just requirement 1 is enforced?

3.20 *Research Suggestion:* Sect. 3.11.1 outlined a process for finding a path with a distance that falls between a lower and upper bound. The process involves adjusting the link metrics prior to running a KSP algorithm. In the tests that were run, the modified link metrics worked well in most cases to find a path of desired length (i.e., KSP required few iterations). In the cases where it did not work well (i.e., KSP required many iterations), running a KSP with the real link lengths as the metric did work well. Investigate this phenomenon further. Is it possible to develop a single link metric that works well in all cases?

References

[ACLY00] R. Ahlswede, N. Cai, S.-Y. R. Li, R. W. Yeung, Network information flow. IEEE Trans. Inform. Theory. **46**(4), 1204–1216 (Jul. 2000)

[AhKK06] S. Ahuja, M. Krunz, T. Korkmaz, Optimal path selection for minimizing the differential delay in Ethernet over SONET. Comput. Netw. **50**(13), 2349–2363 (15 Sept. 2006)

[Bach11] A. Bach, The financial industry's race to zero latency and terabit networking, in *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'11)*, Los Angeles, CA, Service Provider Summit Keynote Address, 6–10 Mar. 2011

[BaHu96] R. A. Barry, P. A. Humblet, Models of blocking probability in all-optical networks with and without wavelength changers. IEEE J. Sel. Areas Commun. **14**(5), 858–867 (Jun. 1996)

[BaMu96] D. Banerjee, B. Mukherjee, A practical approach for routing and wavelength assignment in large wavelength-routed optical networks. IEEE J. Sel. Areas Commun. **14**(5), 903–908 (Jun. 1996)

[BCRV06] G. Bernstein, D. Caviglia, R. Rabbat, H. Van Helvoort, VCAT-LCAS in a clamshell. IEEE Commun. Mag. **44**(5), 34–36 (May 2006)

[Bhan99] R. Bhandari, in *Survivable Networks: Algorithms for Diverse Routing*, (Kluwer, Boston, 1999)

[BhSF01] N. M. Bhide, K. M. Sivalingam, T. Fabry-Asztalos, Routing mechanisms employing adaptive weight functions for shortest path routing in optical WDM networks. Photonic Netw. Commun. **3**(3), 227–236 (Jul. 2001)

[BKOV12] A. Beshir, F. Kuipers, A. Orda, P. Van Mieghem, Survivable routing and regenerator placement in optical networks, in *4th International Workshop on Reliable Networks Design and Modeling (RNDM 2012)*, St. Petersburg, Russia, 3–5 Oct. 2012

[BoCM07] G. Bogliolo, V. Curri, M. Mellia, Considering transmission impairments in RWA problem: Greedy and metaheuristic solutions, *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'07)*, Anaheim, CA, 25–29 Mar. 2007, Paper JWA69

[BoUh98] A. Boroujerdi, J. Uhlmann, An efficient algorithm for computing least cost paths with turn constraints. Inform. Process. Lett. **67**, 317–321 (1998)

[ChLZ03] X. Chu, B. Li, Z. Zhang, A dynamic RWA algorithm in a wavelength-routed all-optical network with wavelength converters, in *Proceedings, IEEE INFOCOM 2003*, San Francisco, CA, 30 Mar.–3 Apr. 2003, vol. 3, pp. 1795–1804

[ChMV08] K. Christodoulopoulos, K. Manousakis, E. Varvarigos. Comparison of routing and wavelength assignment algorithms in WDM networks, in *Proceedings, IEEE Global Communications Conference (GLOBECOM'08)*, New Orleans, LA, 30 Nov.–4 Dec. 2008

[ChMV11] K. Christodoulopoulos, K. Manousakis, E. Varvarigos, Reach adapting algorithms for mixed line rate WDM transport networks. J. Lightwave Technol. **29**(21), 3350–3363 (1 Nov. 2011)

[Choy02] L. Choy, Virtual concatenation tutorial: Enhancing SONET/SDH networks for data transport. J. Opt. Netw. **1**(1), 18–29 (Jan. 2002)

[ChVo10] N. Charbonneau, V. M. Vokkarane, Tabu search meta-heuristic for static manycast routing and wavelength assignment over wavelength-routed optical WDM networks, in *Proceedings, IEEE International Conference on Communications (ICC'10)*, Cape Town, South Africa, 23–27 May 2010

[ChYu94] K. Chan, T. P. Yum, Analysis of least congested path routing in WDM lightwave networks, in *Proceedings, IEEE INFOCOM 1994*, Toronto, Ontario, 12–16 Jun. 1994, vol. 2, pp. 962–969

[CJDD09] X. Chen, A. Jukan, A. C. Drummond, N. L. S. da Fonseca, A multipath routing mechanism in optical networks with extremely high bandwidth requests, in *Proceedings, IEEE Global Communications Conference* (*GLOBECOM'09)*, Honolulu, HI, 30 Nov.–4 Dec. 2009

[CLRS09] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, in *Introduction to Algorithms*, 3rd edn. (MIT Press, Cambridge, 2009)

[DBSJ11] C. Develder, J. Buysse, A. Shaikh, B. Jaumard, M. De Leenheer, B. Dhoedt, Survivable optical grid dimensioning: Anycast routing with server and network failure protection, in *IEEE International Conference on Communications (ICC'11)*, Kyoto, Japan, 5–9 Jun. 2011

[DDDP12] C. Develder, M. De Leenheer, B. Dhoedt, M. Pickavet, D. Colle, F. De Turck, P. Demeester, Optical networks for grid and cloud computing applications. Proc. IEEE. **100**(5), 1149–1167 (May 2012)

[EMSW03] A. Elwalid, D. Mitra, I. Saniee, I. Widjaja, Routing and protection in GMPLS networks: From shortest paths to optimized designs. J. Lightwave Technol. **21**(11), 2828–2838 (Nov. 2003)

[Epps94] D. Eppstein, Finding the *k* shortest paths, in *Proceedings, 35th Annual Symposium on Foundations of Computer Science*, Santa Fe, NM, 20–22 Nov. 1994, pp. 154–165

[GeRa04] O. Gerstel, H. Raza, Predeployment of resources in agile photonic networks. J. Lightwave Technol. **22**(10), 2236–2244 (Oct. 2004)

[GlLa97] F. W. Glover, M. Laguna, in *Tabu Search*, (Kluwer, Boston, 1997)

[GuOr02] R. Guerin, A. Orda, Computing shortest paths for any number of hops. IEEE/ACM Trans. Netw. **10**(5), 613–620 (Oct. 2002)

[HeMS03] J. Hershberger, M. Maxel, S. Suri, Finding the k shortest simple paths: A new algorithm and its implementation, in *Proceedings, Fifth Workshop on Algorithm Engineering and Experiments*, Baltimore, MD, 11 Jan. 2003, pp. 26–36

[HuMM11] S. Huang, C. U. Martel, B. Mukherjee, Survivable multipath provisioning with differential delay constraint in telecom mesh networks. IEEE/ACM Trans. Netw. **19**(3), 657–669 (Jun. 2011)

[ITU06] International Telecommunication Union, Link Capacity Adjustment Scheme (LCAS) for Virtual Concatenated Signals," ITU-T Rec. G. 7042/Y.1305, Mar. 2006

[KaAy98] E. Karasan, E. Ayanoglu, Effects of wavelength routing and selection algorithms on wavelength conversion gain in WDM optical networks. IEEE/ACM Trans. Netw. **6**(2), 186–196 (Apr. 1998)

[KaKL00] K. Kar, M. Kodialam, T. V. Lakshman, Minimum interference routing of bandwidth guaranteed tunnels with MPLS traffic engineering applications. IEEE J. Sel. Areas Commun. **18**(12), 2566–2579 (Dec. 2000)

[KaKL03] K. Kar, M. Kodialam, T. V. Lakshman, Routing restorable bandwidth guaranteed connections using maximum 2-route flows. IEEE/ACM Trans. Netw. **11**(5), 772–781 (Oct. 2003)

[KKKV04] F. Kuipers, T. Korkmaz, M. Krunz, P. Van Mieghem, Performance evaluation of constraint-based path selection algorithms. IEEE Netw. **18**(5), 16–23 (Sep./Oct. 2004)

[KoKr01] T. Korkmaz, M. Krunz, Multi-constrained optimal path selection, in *Proceedings, IEEE INFOCOM 2001*, Anchorage, AK, 22–26 Apr. 2001, vol. 2, pp. 834–843

[KoMB81] L. Kou, G. Markowsky, L. Berman, A fast algorithm for Steiner trees. Acta Inform. **15**(2), 141–145 (Jun. 1981)

[KoMe03] R. Koetter, M. Médard, An algebraic approach to network coding. IEEE/ACM Trans. Netw. **11**(5), 782–795 (Oct. 2003)

[LiRa01] G. Liu, K. G. Ramakrishnan, A*Prune: An algorithm for finding K shortest paths subject to multiple constraints, in *Proceedings, IEEE INFOCOM 2001*, Anchorage, AK, 22–26 Apr. 2001, vol. 2, pp. 743–749

[MDXA10] E. D. Manley, J. S. Deogun, L. Xu, D. R. Alexander, All-optical network coding. J. Opt. Commun. Netw. **2**(4), 175–191 (Apr. 2010)

[MeGa08] R. C. Menendez, J. W. Gannett, Efficient, fault-tolerant all-optical multicast networks via network coding, in *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'08)*, San Diego, CA, 24–28 Feb 2008, Paper JThA82

[MoAz98] A. Mokhtar, M. Azizoglu, Adaptive wavelength routing in all-optical networks. IEEE/ACM Trans. Netw. **6**(2), 197–206 (Apr. 1998)

[OzBe03] A. E. Ozdaglar, D. P. Bertsekas, Routing and wavelength assignment in optical networks. IEEE/ACM Trans. Netw. **11**(2), 259–272 (Apr. 2003)

[PaEA12] T. Panayiotou, G. Ellinas, N. Antoniades, Segment-based protection of multicast connections in metropolitan area optical networks with quality-of-transmission considerations. J. Opt. Commun. Netw. **4**(9), 692–702 (Sep. 2012)

[SAAG05] A. Srivastava, S. Acharya, M. Alicherry, B. Gupta, P. Risbood, Differential delay aware routing for Ethernet over SONET/SDH, in *Proceedings, IEEE INFOCOM 2005*, Miami, FL, 13–17 Mar. 2005, vol. 2, pp. 1117–1127

[SaMu00] L. H. Sahasrabuddhe, B. Mukherjee, Multicast routing algorithms and protocols: A tutorial. IEEE Netw. **14**(1), 90–102 (Jan./Feb. 2000)

[SaSi11] A. A. M. Saleh, J. M. Simmons, Technology and architecture to enable the explosive growth of the Internet. IEEE Commun. Mag. **49**(1), 126–132 (Jan. 2011)

[ShJu07] Q. She, J. P. Jue, Min-cost tree for multi-resource manycast in mesh networks, in *Proceedings, First International Symposium on Advanced Networks and Telecommunication Systems*, Mumbai, India, 17–18 Dec. 2007

[Simm06] J. M. Simmons, Network design in realistic 'all-optical' backbone networks. IEEE Commun. Mag. **44**(11), 88–94 (Nov. 2006)

[Simm10] J. M. Simmons, Diversity requirements for selecting candidate paths for alternative-path routing, *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'10)*, San Diego, CA, 21–25 Mar 2010, Paper NThA4

[SiSM03] N. K. Singhal, L. H. Sahasrabuddhe, B. Mukherjee, Provisioning of survivable multicast sessions against single link failures in optical WDM mesh networks. J. Lightwave Technol. **21**(11), 2587–2594 (Nov. 2003)

[SoPe02] H. Soliman, C. Peyton, An efficient routing algorithm for all-optical networks with turn constraints, *IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS'02)*, Fort Worth, TX, 12–16 Oct. 2002, pp. 161–166

[SrSr06] A. Srivastava, A. Srivastava, Flow aware differential delay routing for next-generation Ethernet over SONET/SDH, in *Proceedings, IEEE International Conference on Communications (ICC'06)*, Istanbul, Turkey, 11–15 Jun. 2006, vol. 1, pp. 140–145

[Stra12] J. L. Strand, Integrated route selection, transponder placement, wavelength assignment, and restoration in an advanced ROADM architecture. J. Opt. Commun. Netw. **4**(3), 282–288 (Mar. 2012)

[SuTa84] J. W. Suurballe, R. E. Tarjan, A quick method for finding shortest pairs of disjoint paths. Networks **14**, 325–336 (1984)

[Suur74] J. W. Suurballe, Disjoint paths in a network. Networks. **4**, 125–145 (1974)

[TaMa80] H. Takahashi, A. Matsuyama, An approximate solution for the Steiner problem in graphs. Math. Japon. **24**(6), 573–577 (1980)

[VaAa87] P. J. M. van Laarhoven, E. H. L. Aarts, in *Simulated Annealing: Theory and Applications*, (D. Reidel Publishing Co., Boston, 1987)

[Voss92] S. Voss, Steiner's problem in graphs: Heuristic methods. Discrete Appl. Math. **40**(1), 45–72 (1992)

[Waxm88] B. M. Waxman, Routing of multipoint connections. IEEE J. Sel. Areas Commun. **6**(9), 1617–1622 (Dec. 1988)

[WGPD08] X. Wang, L. Guo, L. Pang, J. Du, F. Jin, Segment protection algorithm with load balancing for multicasting WDM mesh networks, in *10th International Conference on Advanced Communication Technology (ICACT)*, Pyeongchang, Korea, 17–20 Feb 2008, vol. 3, pp. 2013–2016

[XXQL03] D. Xu, Y. Xiong, C. Qiao, G. Li, Trap avoidance and protection schemes in networks with shared risk link groups. J. Lightwave Technol. **21**(11), 2683–2693 (Nov. 2003)

[Yen71] J. Y. Yen, Finding the K shortest loopless paths in a network. Manage. Sci. **17**(11), 712–716 (Jul. 1971)

[ZTTD02] Y. Zhang, K. Taira, H. Takagi, S. K. Das, An efficient heuristic for routing and wavelength assignment in optical WDM networks, in *Proceedings, IEEE International Conference on Communications (ICC'02)*, New York, NY, 28 Apr.–2 May 2002, pp. 2734–2739

Chapter 4
Regeneration

4.1 Introduction

Using the routing techniques of Chap. 3, a path is selected for each demand request entering the network. The next step in the planning process is selecting the regeneration sites for the demand, if any. (Chapter 5 considers techniques for accomplishing routing, regeneration, and wavelength assignment in a single step.) Regeneration "cleans up" the optical signal, typically by reamplifying, reshaping, and retiming it; this is referred to as "3R" regeneration. As discussed in Chap. 3, paths are usually selected to minimize the amount of required regeneration, as regeneration adds to the network cost and typically reduces the reliability of a path.

If the network is based on optical-electrical-optical (O-E-O) technology, then determining the regeneration locations for a selected path is straightforward because the connection is regenerated at every intermediate node along the path. For an optical-bypass-enabled network, where it is possible to transit an intermediate node in the optical domain depending on the quality of the optical signal, determining the regeneration locations for a path may be more challenging. Numerous factors affect when an optical signal must be regenerated, including the underlying transmission technology of the system, the properties of the network elements, and the characteristics of the fiber plant on which the system is deployed. When these factors are considered together, one can estimate the nominal distance over which an optical signal can travel before requiring regeneration (i.e., the optical reach). However, while quoting the optical reach in terms of physical distance is expedient for benchmarking the system performance at a high level, it is not sufficient for determining the necessary regeneration locations in an actual network. Many factors other than distance need to be considered.

Section 4.2 presents a high-level discussion of some of the optical impairments and system properties that have an impact on when a signal must be regenerated. Ultra-long-reach technology has succeeded because it is possible to mitigate many of these impairments or design a system such that their effects are negligible. In Sect. 4.3, the discussion focuses on one of the more important impairments, namely noise. This leads to a link metric that can be used in the routing process instead of distance, such that the routing algorithm is more likely

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4_4, © Springer International Publishing Switzerland 2014

to produce paths that minimize the amount of required regeneration. Section 4.4 discusses capturing impairments other than noise in the routing and regeneration processes.

Given that regeneration, and thus implicitly network cost, are dependent on the physical-layer design, it may be beneficial to integrate algorithms for the physical-network design (e.g., algorithms to optimize the configuration of the optical amplifiers) with the architectural design and planning tool. Some of the benefits of this approach are discussed in Sect. 4.5.

In Sect. 4.6, the discussion moves from the physical-layer aspects of regeneration to the architectural facets. Several regeneration strategies are presented, where the trade-off is between operational simplicity and cost. The strategy employed will likely affect how much regeneration is required and where a connection must be regenerated; thus, it is an important aspect of the network design.

Finally, in Sect. 4.7, different options for actually implementing regeneration in a node are presented. Again, there is a trade-off of operational flexibility versus cost. While most of this chapter is relevant only to optical-bypass-enabled networks, much of this final section applies to O-E-O networks as well.

There are a few key points to be emphasized in this chapter. First, while having to encompass physical-layer phenomena in a network planning tool may seem imposing, there are known methodologies for tackling this problem that have been successfully implemented in live networks. Furthermore, when optical-bypass technology is developed by system vendors, there needs to be close collaboration between the systems engineers and the network architects. If the requirements of the physical layer are so exacting that the ensuing complexities in the network planning tool are unmanageable (e.g., if the wavelengths have to be assigned in a precise order on all links), then it may be necessary to modify the technological approach. Finally, while occasional regeneration may appear to be undesirable because of its cost, it does provide an opportunity to change the wavelength on which an optical signal is carried. This can be advantageous in the wavelength assignment step of the planning process, which is covered in Chap. 5.

4.2 Factors That Affect Regeneration

4.2.1 Optical Impairments

One of the major impairments that an optical signal encounters is accumulated noise. The principal sources of noise are the spontaneous emissions of optical amplifiers, which are amplified along with the signal as they propagate together through the network. This is suitably referred to as *amplified spontaneous emissions (ASE) noise*. The strength of the signal compared to the level of the noise is captured by the signal's *optical signal-to-noise ratio* (OSNR), where signals with lower OSNR are more difficult to receive without errors.

Many other optical impairments arise from the physical properties of light propagating in a fiber. For example, the propagation speed of light within a fiber depends on the optical frequency. This causes the optical signal pulses, which have a finite spectral width, to be distorted as they propagate along a fiber (on most fiber types, the pulses spread out in time). This phenomenon is known as *chromatic dispersion,* or simply *dispersion*. Dispersion accumulates as a linear function of the propagation distance. Higher-bit-rate signals, where pulses are closer together, are more susceptible to errors due to this effect; the amount of tolerable dispersion decreases with the square of the bit rate.

A different type of dispersion, known as *polarization-mode dispersion* (PMD), stems from different light polarizations propagating in the fiber at different speeds. In simplistic terms, an optical signal can be locally decomposed into two orthogonal principal states of polarization, each of which propagates along the fiber at a different speed, resulting in distortion. PMD accumulates as a function of the square root of the propagation distance. As with chromatic dispersion, PMD is a larger problem as the bit rate of the signal increases; the PMD-limited reach decreases with the square of the bit rate.

Several nonlinear optical effects arise as a result of the fiber refractive index being dependent on the optical intensity. (The refractive index governs the speed of light propagation in a fiber.) As the optical signal power is increased, these nonlinearities become more prominent. One such nonlinearity is *self-phase modulation* (SPM), where the intensity of the light causes the phase of the optical signal to vary with time. This potentially interacts with the system dispersion to cause significant pulse distortion. *Cross-phase modulation* (XPM) is a similar effect, except that it arises from the interaction of two signals, which is more likely to occur when signals are closely packed together in the spectrum. Another nonlinear effect is *four-wave mixing* (FWM). This arises when signals carried on three particularly spaced optical frequencies interact to yield a stray signal at a fourth frequency, or two frequencies interact to generate two stray signals. These stray signals can potentially interfere with desired signals at or near these frequencies.

There are various other fiber nonlinearities that can distort the signal. For detailed coverage of optical impairments, see Forghieri et al. [FoTC97], Gnauck and Jopson [GnJo97], Poole and Nagel [PoNa97], and Bayvel and Killey [BaKi02].

4.2.2 Network Element Effects

In addition to impairments that accumulate due to propagation in a fiber, there are a number of deleterious effects that a signal may suffer when transiting an optical-bypass-enabled network element. For example, a network element may utilize optical filters to internally demultiplex the wavelengths entering from a wavelength-division multiplexing (WDM) network port. Each time a signal passes through such a filter, the bandwidth of the channel through which the signal propagates "narrows" to some degree, distorting the signal. Another source of signal degradation

is the crosstalk caused by "leakage" within a switching element. This occurs when a small portion of the input signal power appears at outputs other than the desired output. Additionally, the optical loss of a network element may depend on the state of polarization of the signal; this is known as *polarization dependent loss* (PDL). Since the signal polarization may vary with time, the loss may also vary over time, which is undesirable. Furthermore, the network element may contribute to dispersion, and the dispersion level may not be flat across the transmission band, making compensation more difficult.

Factors such as filter narrowing, crosstalk, PDL, and dispersion contribute to a limit on the number of network elements that can be optically bypassed before a signal needs to be regenerated; i.e., they limit the *cascadability,* as discussed in Sect. 2.9.1. However, as optical-bypass technology has matured, the performance of the network elements has significantly improved. Many commercial systems support optical bypass of up to 10 (backbone) or 20 (metro-core) network elements prior to requiring regeneration. With this capability, the number of network elements bypassed is not usually the limiting factor in determining where regeneration must occur, especially in backbone networks where the distances between nodes may be very long; i.e., other limiting effects "kick-in" prior to ten nodes being traversed. (However, the network elements do have an impact on the OSNR, as discussed in Sect. 4.3.1.)

4.2.3 Transmission System Design

The characteristics of the transmission system clearly influence the optical reach of the system. One of the most important system design choices is the type of amplification to employ, where Raman technology is often used to attain an extended optical reach. Distributed Raman amplification uses the fiber itself to amplify the optical signal, so that the rate of OSNR degradation is less steep compared to amplification using erbium-doped fiber amplifiers (EDFAs). This trend is illustrated in Fig. 4.1, which depicts the OSNR level as a function of transmission distance, in a hypothetical system, for both distributed Raman and lumped EDFA amplification (lumped indicates the amplification occurs only at the amplifier sites).

The acceptable OSNR level at the receiver depends on the receiver sensitivity and the desired system margin. (The receiver sensitivity is the minimum average optical power necessary to achieve a specified bit error rate (BER) [RaSS09]. A network is typically designed to perform better than the minimum acceptable level to account for degradation as the system ages.) As shown in Fig. 4.1, for a desired level of OSNR at the receiver and for a given amplifier spacing, distributed Raman amplification supports a significantly longer transmission distance as compared to EDFA amplification. See Islam [Isla02] and Rottwitt and Stentz [RoSt02] for more details on Raman amplifiers.

Two other important system properties that affect optical reach include the spacing between wavelengths, where closer spacing reduces the reach, and the initial

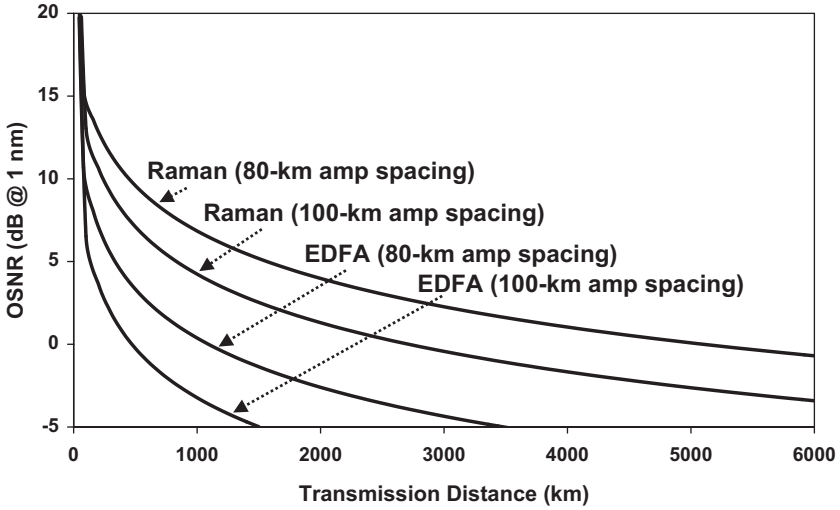


Fig. 4.1 Optical signal-to-noise ratio (*OSNR*) as a function of transmission distance for distributed Raman amplification and for lumped amplification using erbium-doped fiber amplifiers (*EDFAs*), for both 80-km and 100-km amplifier spacings. The *OSNR* degrades more slowly with Raman amplification. For a given amplification type, the *OSNR* degrades more slowly with amplifiers spaced closer together. Amplifier spacing is typically on the order of 80 km; however, some networks have a spacing closer to 100 km

launched powers of the optical signals. Increasing the launched power increases the optical reach, up to a point. However, if the signal power is too high, the nonlinear optical impairments will have an appreciable negative impact on the reach, implying that there is an optimum power level, which is system dependent.

Other important design choices are the signal modulation format, which is the format used for coding the data on a lightstream, and the detection method at the receiver. With bit rates of 10 Gb/s and below, the modulation format is typically *on–off keying* (OOK), where the presence of light indicates a "1" and the relative absence of light indicates a "0." The corresponding receiver makes use of *direct detection,* where the determination of 1s and 0s is based on measurements of the signal energy only. At 40 Gb/s, two common modulation formats are *differential phase-shift keying* (DPSK) and *differential quadrature phase-shift keying* (DQPSK); the receiver uses direct detection with differential demodulation.

At 100 Gb/s, the modulation format recommended by the Optical Internetworking Forum (OIF) is *dual-polarization quadrature phase-shift keying* (DP-QPSK). As the name suggests, data are transmitted using two polarizations, with both in-phase and quadrature (i.e., 90° offset) components. This is combined with *coherent detection* at the receiver, where the receiver bases its decisions on the recovery of the full electric field, which contains both amplitude and phase information [ILBK08]. Digital coherent technology is enabled by very fast analog-to-digital converters and advanced signal processing in the electrical domain. Due to several advantageous

properties of such coherent systems [VSAJ09], this technology is now favored for new 40-Gb/s deployments as well [Hans12].

One of the advantages of coherent detection is its improved compatibility with advanced modulation formats, as compared to direct detection [Savo07]. This enables the use of modulation formats with increased *spectral efficiency,* such as QPSK. Spectral efficiency is defined as the ratio of the information bit rate to the total bandwidth consumed. Increased spectral efficiency translates to greater capacity on a fiber. Support for advanced modulation formats will be even more important in the future, as spectrally efficient 400-Gb/s and 1-Tb/s line rates will likely require modulation formats of even greater complexity, e.g., *multilevel* amplitude and/or phase modulation [Conr02, Winz12].

Another benefit of coherent detection is that it can more easily support polarization multiplexing [Robe11]. By using two polarizations, each with the dual phase components of QPSK, DP-QPSK encodes four bits per symbol. Thus, the symbol rate, or baud rate, is reduced by a factor of four in comparison to the line rate. This was critical in developing a 100-Gb/s format that required approximately the same bandwidth as legacy 10-Gb/s and 40-Gb/s wavelengths. In addition to providing increased spectral efficiency, the lower baud rate effectively reduces the speed of the required electronics by a factor of four. This has enabled 100-Gb/s bit rates to be achieved using currently available electronic components.

As noted in Sect. 4.2.1, impairments such as chromatic dispersion and PMD become more problematic as the bit rate increases, initially prompting concern that certain fibers would not be capable of supporting 100-Gb/s rates. Instead, receiver technology based on coherent detection has improved the situation. Recovery of both amplitude and phase information at the receiver allows it to compensate for linear impairments such as dispersion and PMD. The compensation is performed electronically in the digital signal processor of the receiver [IpKa10]. Furthermore, coherent systems are more tolerant to the bandwidth-narrowing effects induced by optically bypassing several reconfigurable optical add/drop multiplexers (ROADMs).

Coherent detection also provides frequency selectivity [OSul08]. A coherent receiver is capable of picking out a particular frequency from a WDM signal without requiring a filter. This feature is desirable, for example, for operation with the multicast switch (MCS)-based contentionless ROADM architecture described in Sect. 2.9.5.4. However, depending on the coherent receiver design, the number of wavelengths in the received WDM signal may need to be limited, to avoid receiver overload and noise penalties [GBSE10]. Thus, the MCS architecture may need to be modified, such that each slot of the MCS receives no more than some subset of the wavelengths in the spectrum [Way12].

Current 100-Gb/s systems achieve a reasonable optical reach (e.g., 2,000–2,500 km) in spite of the high bit rate. This is in part due to the ability of coherent detection to mitigate linear impairments. Additionally, coherent detection improves the receiver sensitivity as compared to direct or differential detection, which is also beneficial in extending the reach. Another major factor in attaining extended reach has been the use of advanced forward error correction (FEC). The stronger the

FEC code, the greater its ability to detect and correct errors, which allows a longer optical reach (i.e., with stronger FEC, a lower signal-to-noise ratio (SNR) produces the same error rate). State-of-the-art FEC (in the 2015 time frame) provides a net coding gain of roughly 10–11 dB at a post-FEC BER of 10^{-15}, with an overhead of about 20 % [ChOM10, MSMK12].

4.2.4 Fiber Plant Specifications

The characteristics of the physical fiber plant also have a large impact on system regeneration. For example, if the amplifier huts are spaced further apart, then the OSNR degrades more quickly leading to more frequent regeneration. This is illustrated in Fig. 4.1, where the OSNR decreases more sharply for 100-km amplifier spacing as compared to 80-km spacing, for a given amplification type.

The type of fiber in the network may have an impact on optical reach as well, where the most commonly used fiber types are classified as either *non dispersion-shifted fiber* (NDSF) or *non-zero dispersion-shifted fiber* (NZ-DSF; SMF-28® and AllWave® are examples of NDSF fiber; LEAF® and TrueWave® REACH are examples of NZ-DSF fiber[1]). Older fiber plants may include *dispersion-shifted fiber* (DSF, e.g., SMF/DS™). As the names imply, these classes of fiber differ in the dispersion level in the portion of the spectrum occupied by WDM systems, thereby requiring different levels of dispersion compensation and having different mitigating effects on the nonlinear optical impairments (dispersion can be helpful in combatting these impairments). For example, SMF-28 fiber has roughly four times the level of dispersion as LEAF fiber in the WDM region of interest. SMF/DS fiber has zero, or near-zero, dispersion in this region (which generally makes it unfavorable for use with WDM systems). In addition to dispersion differences, fiber types may differ in their Raman gain efficiency, where higher efficiency may lead to longer reach.

With the advent of optical-bypass technology, carriers have paid greater attention to the types of fiber deployed in their networks due to the impact it can have on system performance. Most new long-haul system deployments have employed NDSF fiber due to its relatively large dispersion, which, as noted above, helps combat nonlinear optical impairments. Some larger carriers have taken a long-term view and deployed fiber sheaths containing multiple fiber types, to enable them to light up whichever type may be best suited for future systems.

Smaller carriers are more likely to have a heterogeneous fiber plant, where disparate fiber types are deployed in different regions of the network. This scenario often stems from purchasing existing fiber from a collection of other carriers, rather than deploying new fiber throughout the network themselves. This heterogeneity needs to be accounted for in the routing, regeneration, and wavelength assignment processes.

[1] SMF-28 and LEAF are registered trademarks of Corning Incorporated; AllWave and TrueWave are registered trademarks of OFS FITEL, LLC.

4.2.5 Mitigation of Optical Impairments

Section 4.2.1 outlined a host of impairments that can be detrimental to the optical signal. However, there are well-known techniques for mitigating some of these optical impairments. For example, dispersion compensation is typically used to combat chromatic dispersion, where the level of compensation needed depends on the amount of dispersion in the transmission fiber and the dispersion tolerance of the system. Note that it is not desirable to reduce the dispersion along a fiber to zero, as its presence helps reduce some of the harmful nonlinear effects [Kurt93, TkCh94, TCFG95].

In early optical-bypass-enabled system deployments, dispersion compensation was accomplished by installing dispersion compensating fiber (DCF) having inverse dispersion relative to the transmission fiber, at various sites along each link. For example, if the link fiber has positive dispersion, which is the typical case, then DCF with negative dispersion is utilized. (With positive-dispersion fiber, the signal pulses spread out; with negative dispersion, the pulses are compressed.) DCF is generally expensive, increases the loss, provides only a static means of compensation, and adds a small amount of latency. Further problems with DCF may result from the dispersion level of the transmission fiber not being constant across the transmission band; typically, the dispersion level of the fiber has a particular slope across the band. The DCF may not have precisely the same inverse dispersion slope, leading to different levels of residual dispersion depending on the transmission wavelength.

As networks evolved, electronic dispersion compensation (EDC) was used as an enhancement to, or as a replacement of, the DCF strategy [KaSG04]. EDC can be deployed on a per-wavelength basis (e.g., as part of the WDM transponder), and can be dynamically tuned over a range of dispersion levels to better match the compensation requirements of a given connection. For example, receivers based on maximum likelihood sequence estimation (MLSE) are one means of combating chromatic dispersion, as well as possibly other impairments [CaCH04, ChGn06]. (MLSE operates on a sequence of bits rather than a single bit at a time, and selects the data sequence that is statistically most likely to have generated the detected signal.) In another EDC strategy, pre-compensation is used at the transmitter, based on feedback from the receiver [MORC05]. Alternatively, post-compensation can be implemented at the receiver, which may be more suitable for dynamic networking.

PMD compensation is more challenging because the level of PMD may vary as a function of time, necessitating adjustable PMD compensators, which can be costly. PMD is more of a problem on older fibers, as newer fiber types tend to have low PMD.

It was initially anticipated that chromatic dispersion and PMD would be more problematic as line rates increased to 100 Gb/s and higher. However, as noted in Sect. 4.2.3, digital coherent technology allows for the compensation of linear impairments such as chromatic dispersion and PMD in the receiver, via signal processing. Implementing additional techniques to further mitigate these impairments is typically not required.

In addition to the linear effects, there are numerous nonlinear impairments. Many of the problems from nonlinear effects can be avoided by maintaining the signal power at a low enough level, but still sufficiently higher than the noise level. In addition, as mentioned above, maintaining a small amount of residual system dispersion can be effective in reducing some of the nonlinear effects. There is also ongoing research into using coherent detection to counteract some of the nonlinear effects [Tayl10].

4.2.6 Mixed Line-Rate Systems

The rollouts of 10-Gb/s, 40-Gb/s, and 100-Gb/s backbone systems have generally come at 5–7-year intervals. Most large carriers light up another fiber when deploying a system with a new line rate, such that there are not heterogeneous line rates on one fiber. However, smaller carriers, whose network traffic grows more slowly, may reach a point where they require more network capacity but not enough to justify lighting up new fibers. An alternative strategy to increase capacity is to populate the remainder of the current fiber using wavelengths of a higher line rate, resulting in a mix of line rates on one fiber. As the system line rates have increased, more spectrally efficient transmission schemes have been employed, such that 10-Gb/s, 40-Gb/s, and 100-Gb/s line rates are all compatible with 50-GHz channel spacing. However, these line rates entail different modulation formats and demonstrate different susceptibility to impairments, which can pose problems when they co-propagate on the same fiber.

Numerous studies have investigated the compatibility of the various line rates and modulation formats, e.g., 100-Gb/s and 40-Gb/s coherent DP-QPSK wavelengths co-propagating with 10-Gb/s OOK wavelengths, or 40-Gb/s DPSK wavelengths co-propagating with 10-Gb/s OOK wavelengths [CRBT09, BBSB09, BRCM12]. These tests showed a significant performance penalty for DP-QPSK wavelengths due to XPM induced by adjacent legacy 10-Gb/s wavelengths. The penalty was more severe at 40-Gb/s line rate than at 100 Gb/s. Reducing the performance penalty to an acceptable level required employing fairly large guardbands between the different wavelength formats. Guardbands represent wasted capacity; clearly, their usage should be minimized. This type of effect needs to be considered in the wavelength assignment process, such that these different wavelength types are segregated as much as possible. For example, the 10-Gb/s wavelengths could be assigned to one end of the spectrum and the coherent-system wavelengths to the other end. This topic is revisited in Chap. 5, on wavelength assignment.

One of the more important factors in multi-rate systems is managing dispersion. As discussed earlier, coherent systems require little to no exogenous dispersion compensation because the coherent receiver itself is capable of compensation, whereas legacy 10-Gb/s systems typically have spools of dispersion compensating fiber strategically deployed along a link. Furthermore, the presence of dispersion can be helpful in mitigating some of the deleterious nonlinear interchannel effects.

A dispersion map that meets the requirements of 10-Gb/s wavelengths without excessively penalizing co-propagating coherent-system wavelengths is suggested in Anderson et al. [ADZS12]. The general philosophy is to allow dispersion to build up along a link, but reduce the accumulated dispersion to near zero at the link endpoints.

4.2.7 System Regeneration Rules

As the above discussion indicates, there are numerous factors to take into account when determining where an optical signal needs to be regenerated. Every vendor has a different approach to the problem, such that there is no uniform set of rules for regeneration across systems. It is generally up to the individual vendors to analyze their own particular implementation and develop a set of rules that governs routing and regeneration in a network. While there are many potential aspects to consider, in some systems it is possible to come up with a minimal set of rules that is sufficient for operating a network in real time. For example, the system rules may be along the lines of: regenerate a connection if the OSNR is below N; regenerate (or add an effective OSNR penalty) if the accumulated dispersion is above D, or if the accumulated PMD is above P, or if the number of network elements optically bypassed is greater than E (where N, D, P, and E depend on the system). When determining these rules, vendors usually factor in a system margin to account for aging of the components, splicing losses (i.e., optical losses that arise when fiber cuts are repaired), and other effects. Furthermore, it is important that the rules be immune to the dynamics of the traffic in the network. Connections are constantly brought up and down in a network, either due to changing traffic patterns or due to the occurrence of, or recovery from, failures; thus, it is important that the regeneration rules be independent of the number of active channels on a fiber.

While this may appear to be quite challenging, it is important to point out that there are optical-bypass-enabled networks with optical reaches over 3,000 km that have operated over a number of years using a set of relatively simple rules. It may be necessary to sacrifice a small amount of optical reach in order to come up with straightforward rules. However, as analyzed in Chap. 10, the marginal benefits of increasing the optical reach beyond a certain point are small anyway.

As pointed out in Sect. 4.2.4, the performance of a system depends on the characteristics of the fiber plant in the carrier's network. In the early stages of evaluating a system for a given carrier, prior to any equipment deployment, the exact specifications of the fiber plant may not be known. Thus, in the beginning stages of network design, one may have to rely on selecting regeneration locations based on path distance, where the optical reach in terms of a nominal distance is used. For purposes of system evaluation and cost estimation, this is generally acceptable. After the fiber spans have been fully characterized, the network planning tools can implement the more precise system regeneration rules.

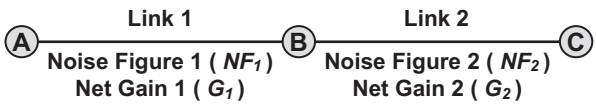


Fig. 4.2 Two consecutive links, each with their respective NFs and net gains

4.3 Routing with Noise Figure as the Link Metric

In this section, we consider a system where the signal power levels are low enough, or dispersion levels are high enough, that nonlinear optical impairments can be neglected. Section 4.4 considers routing and regeneration when this assumption does not hold. Assume that the fiber PMD is low, such that the most important factors to consider with respect to regeneration are the OSNR, chromatic dispersion, and the number of network elements optically bypassed. The focus first is on OSNR; the other two factors are considered in Sect. 4.3.3 to ensure a cohesive system design.

As an optical signal propagates down a fiber link, its OSNR degrades. The *noise figure* (NF) of a link is defined as the ratio of the OSNR at the start of the link to the OSNR at the end of a link, i.e.,

$$\mathrm{NF}_{\mathrm{Link}} = \frac{\mathrm{OSNR}_{\mathrm{Link\,Start}}}{\mathrm{OSNR}_{\mathrm{Link\,End}}}. \tag{4.1}$$

The NF is always greater than or equal to unity. Low NF is desirable as it indicates less signal degradation. NF is a quantity that can be measured in the field for each link once the amplifiers have been deployed.

Numerous factors affect the NF of a link. For example, the type of amplification is very important, where Raman amplification generally produces a lower NF than amplification using EDFAs. Large fiber attenuation and large splicing losses can contribute to a higher NF. Longer span distances (i.e., the distances between amplifier huts) also generally increase the NF. The NF is also affected by the fiber type of the link.

OSNR is important in determining when an optical signal needs to be regenerated. Transponder receivers generally have a minimum acceptable OSNR threshold, below which the signal cannot be detected properly. Thus, the evolution of OSNR as the signal traverses an optical path is critical. Consider the two consecutive links shown in Fig. 4.2, where each link i has an associated NF, NF_i, and a net gain, G_i. Net gain refers to the total amplification on the link compared to the total loss. The cumulative NF for an optical signal traversing Link 1 followed by Link 2 is given by Haus [Haus00]:

$$\mathrm{NF}_{\mathrm{Total}} = \mathrm{NF}_1 + \frac{(\mathrm{NF}_2 - 1)}{G_1}. \tag{4.2}$$

In most systems, the net gain on a link is unity (in linear units), because the total amplification is designed to exactly cancel the total loss. In addition, typical values

for the link NF are in the hundreds (in linear units), such that the "1" term is negligible. The formula then simplifies to

$$NF_{Total} \approx NF_1 + NF_2. \tag{4.3}$$

Extending this formula to multiple links, the NF of an end-to-end path is the sum of the NFs on each link of the path, where it is desirable to minimize this sum. Link NF is thus a suitable additive link metric that can be used in the shortest-path algorithms of Chap. 3. Using this as a metric yields the path with least NF or, equivalently, the path with the highest overall OSNR (assuming other impairments are properly managed). NF is typically a better metric than distance in finding paths that minimize the number of required regenerations. However, it is still true that the minimum-noise-figure path may not be the minimum-regeneration path, e.g., due to regenerations occurring only at network nodes. Thus, when generating candidate paths, as described in Chap. 3, each path must be evaluated to determine the actual regeneration locations, to ensure minimum (or close to minimum) regeneration paths are chosen.

When working with the formulas for NF, it is important to use the correct units. The NF of a link is typically quoted in decibels (dB). However, Equations 4.2 and 4.3 require that the NF be in linear units. The following formula is used to convert from dBs to linear units:

$$\text{Linear Units} = 10^{\text{Decibel Units}/10}. \tag{4.4}$$

One of the benefits of using NF as a metric is that it is a field-measurable quantity that implicitly subsumes the heterogeneity of a real network, e.g., different fiber types, different amplifier types, different span distances, etc. If there is not an opportunity to measure the NF, then it is also possible to estimate the NF through formulas that incorporate these network characteristics [Desu94]. Experience has shown such calculations to be fairly accurate.

To determine where regeneration is required based on accumulated noise, one adds up the NF on a link-by-link basis. Regeneration must occur before the total NF grows to a certain threshold, i.e., before the OSNR decreases below a threshold. There are other factors that may force regeneration to be required earlier; however, in a well-designed system, ideally, there should be a consistent confluence of factors, so that regeneration can be minimized (see Sect. 4.3.3).

It should be noted that the noise-figure/OSNR-based approach was successfully used for routing and regeneration in some of the earliest optical-bypass-enabled commercial deployments. Pertinent factors other than noise were translated into an effective OSNR penalty, so that regeneration could be determined solely from the effective OSNR level. Despite transmission systems becoming more complex and challenging due to higher line rates, there are vendors that have advocated a similar methodology for routing/regeneration in more recent systems [ShSS11].

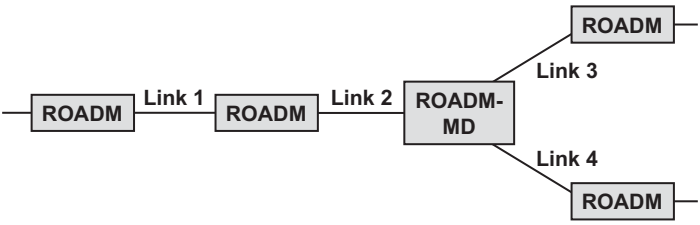


Fig. 4.3 The NF of each link needs to be adjusted to account for the NF of the network elements at either end of the link. For example, the NF of *Link 2* is incremented by half of the NF of a reconfigurable optical add/drop multiplexer (*ROADM*) and half of the NF of a multi-degree ROADM (*ROADM-MD*)

4.3.1 Network Element Noise Figure

In addition to each link having a measurable NF, the nodal network elements contribute to OSNR degradation as well. Thus, each network element, such as a ROADM, has an associated NF. To account for this effect in the routing process, the link metric should be adjusted based on the type of equipment deployed at either end of the link. (This is simpler than modeling the network elements as additional "links" in the topology.) One strategy is to add half of the NF of the elements at the endpoints to the link NF. (This adjustment is used only for routing purposes; it does not imply that the NF of the element add/drop path is half that of the through path.)

Consider Link 2 shown in Fig. 4.3, which is equipped with a ROADM at one end and a multi-degree ROADM (ROADM-MD) at the other end. Assume that the NF of the ROADM is about 16 dB and the NF of the ROADM-MD is about 17 dB. Halving these values yields roughly 13 and 14 dB, respectively (subtracting 3 dB is roughly equivalent to dividing by two in linear terms). These amounts should be added to the NF of Link 2, where the additions must be done using linear units. For example, if Link 2 has a NF of 25 dB, then it should be assigned a link metric of $10^{25/10} + 10^{13/10} + 10^{14/10} = 361.3$. By adding half of the element NF to each link connected to the element, the full NF of the element is accounted for regardless of the direction in which the element is traversed. (If, instead, the *full* amount of the element NF were added to each link connected to the element, then the element NF would be double-counted along a path. If the full amount of the element NF were added to just one of the links entering the element, then some paths may not count the element penalty at all. For example, if the penalty of the ROADM-MD were added to Link 2 only, then a path from Link 3 to Link 4 would not include any ROADM-MD penalty.)

4.3.2 Impact of the ROADM without Wavelength Reuse

There is one network element that warrants special consideration with respect to noise and regeneration: the ROADM that does not have wavelength reuse (see

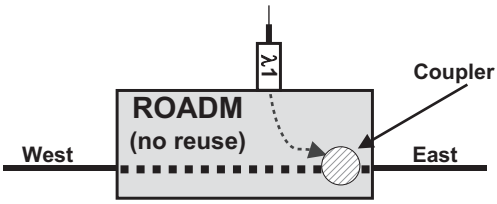


Fig. 4.4 A connection carried on λ_1 is added at a node equipped with a reconfigurable optical add/drop multiplexer (*ROADM*) without wavelength reuse. The added signal is combined with the WDM signal entering the node from the *West* link. The noise in the λ_1 region of the spectrum from the *West* link is added to the new connection's signal

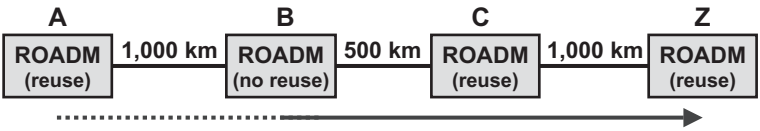


Fig. 4.5 Assume that the optical reach is 2,000 km. The reconfigurable optical add/drop multiplexer (*ROADM*) at Node *B* does not have wavelength reuse. A signal from Node *B* to Node *Z* has a level of noise as if it originated at Node *A,* and thus, needs to be regenerated at Node *C*

Sect. 2.9.10). Such elements are typically little more than an optical amplifier with a coupler/splitter for adding/dropping traffic at the node. A simplified illustration of such a ROADM is shown in Fig. 4.4. An optical signal added at this node is coupled to the light passing through the node from the network links. Assume that the added signal in the figure is carried on λ_1 and assume that it is sent out on the East link. While there is no signal at λ_1 entering the ROADM from the West link, the noise in the λ_1 region of the spectrum has propagated down the fiber and undergone amplification along with the rest of the spectrum. This noise will be coupled with the signal being added on λ_1 at the node. From an OSNR perspective, it appears as if the added signal has effectively been transmitted over some distance, thus affecting where it needs to be regenerated.

The effect on regeneration is illustrated more clearly in Fig. 4.5. Node B in the figure is equipped with a ROADM without wavelength reuse, whereas the remaining nodes have a ROADM with reuse. Assume that the nominal optical reach of the system is 2,000 km (for simplicity, we will still discuss reach in terms of a distance), and assume that a connection is established from Node B to Node Z. As the path distance from B to Z is 1,500 km, it would appear that no regeneration is required. However, the connection added at Node B has accumulated noise as if it originated at Node A. Thus, from an OSNR perspective, it is equivalent to a connection from Node A to Node Z, which covers 2,500 km. Therefore, a regeneration is required at Node C in order to clean up the signal.

This same effect does not occur with elements that have wavelength reuse because such elements are equipped with means of blocking any wavelength on an

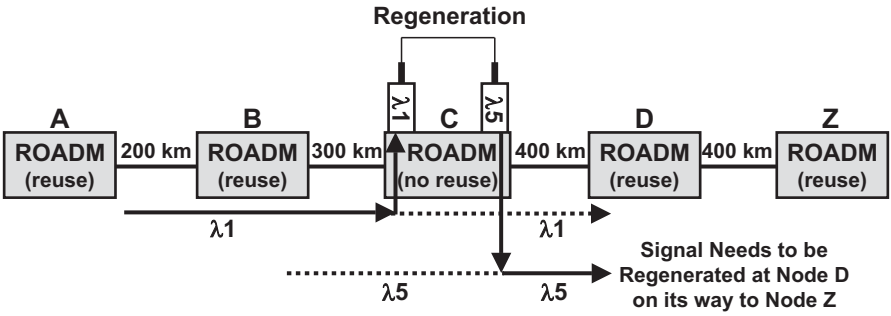


Fig. 4.6 The desired connection is between Nodes *A* and *Z*. Assume that the optical reach is 1,000 km. If the connection is regenerated at Node *C,* which has a no-reuse reconfigurable optical add/drop multiplexer (*ROADM*), it will need to be regenerated again at Node *D*

input fiber along with the noise in the region of the spectrum around it. Thus, the noise is not combined with a wavelength being added at the node.

Note that regenerating a connection at a node with a no-reuse ROADM is not desirable. Consider the setup shown in Fig. 4.6, where Node C has a no-reuse ROADM and the remaining nodes have ROADMs with reuse. Assume that the optical reach is 1,000 km, and assume that a connection between Nodes A and Z is launched from Node A using wavelength λ_1. If this connection is dropped at Node C for regeneration, λ_1 will continue on through the node because of the inability of the ROADM at Node C to block the wavelength. Thus, after the signal is regenerated at Node C, it must be relaunched on a different wavelength, say λ_5. This one connection will then "burn" two wavelengths on the link between Nodes C and D (λ_1 and λ_5). Furthermore, because of noise being combined with the added signal, as described above, the signal on λ_5 will have a noise level as if it had been added at Node B. The added noise will cause the signal to require regeneration at Node D. Thus, the regeneration at Node C is not effective because one regeneration at Node D would have been sufficient for the whole end-to-end path. Additionally, the amount of add/drop supported at a no-reuse ROADM is usually very limited. Using it for regeneration may prevent the node from sourcing/terminating additional traffic in the future. For these reasons, regeneration is generally not recommended at a node with a no-reuse ROADM.

4.3.3 Cohesive System Design

The focus thus far has been on OSNR, where routing is performed using NF as the link metric. Next, the chromatic dispersion and the cascadability of the network elements are considered to ensure that there is a cohesive system design. This type of unified approach to system design is also appropriate when nonlinear impairments are a factor, as in Sect. 4.4.

Assume that based solely on the OSNR analysis, it is anticipated that the optical reach of a system is on the order of L km. It is beneficial to coordinate the dispersion management and the network element performance to align with this number. For example, if the OSNR forces a path to be regenerated after roughly L km, then there is no need to add dispersion compensation suitable for transmission well beyond L km.

To be more concrete, assume that, based on the OSNR analysis, the optical reach is nominally on the order of 2,000 km. In addition, assume that the dispersion tolerance of the system is 10,000 ps/nm; i.e., after accumulating this much dispersion, a signal needs to be regenerated. Assume that the fiber has a dispersion level of 15 ps/(nm · km), and assume that the network elements have negligible dispersion. Assuming DCF is used for dispersion compensation, enough DCF should be added such that after 2,000 km, the 10,000 ps/nm limit is not exceeded. The required level of dispersion compensation can be determined by solving for C in the following equation:

$$2{,}000\,\text{km} \cdot (15\,\text{ps}/(\text{nm} \cdot \text{km}) - C) = 10{,}000\,\text{ps}/\text{nm}. \tag{4.5}$$

This yields a C of 10 ps/(nm · km), which is a guideline as to how much average dispersion compensation is needed per km of fiber. (Note that the DCF would have negative dispersion to counteract the positive dispersion of the fiber.)

It should be noted that the advent of coherent detection has simplified this aspect of system design. As discussed in Sect. 4.2.3, coherent technology allows for electronic dispersion compensation in the receiver. Coherent systems typically have a dispersion tolerance of more than 50,000 ps/nm [GrBX12]. Given that commonly deployed fibers have a dispersion level on the order of 17 ps/(nm · km) or lower, dispersion management does not have to be explicitly designed into the system (assuming the optical reach is less than 3,000 km). This applies to PMD as well. Coherent systems generally have a differential-group-delay (DGD) tolerance of more than 100 ps (DGD is a measure of the PMD phenomenon). Most fibers have a PMD coefficient below $1\ ps/\sqrt{km}$ (PMD coefficients on newer fibers are typically below $0.1\ ps/\sqrt{km}$), such that PMD compensation does not have to be included in the design process.

Next, consider any limitations due to optically bypassing consecutive nodes. Assume that nodes in a backbone network are spaced such that the typical distance between nodes is 250 km. If the optical reach of the system based on OSNR is 2,000 km, then it is not likely that more than seven or so consecutive network elements will be optically bypassed. This can serve as a guideline in designing the cascadability properties of the network elements. Greater cascadability is likely desirable for metro-core systems, where the node density is typically much higher.

As this example illustrates, by coordinating the different aspects of system design, the overall system can be made more cost effective.

4.4 Impairment-Based Routing Metrics Other Than Noise Figure

Section 4.3 primarily focused on using NF as the routing metric, with dispersion and network element cascadability also taken into consideration in determining where regeneration is required. In this section, we consider systems that are susceptible to a wider range of impairments such that NF alone may not be the best predictor of regeneration. This is likely to be the case, for example, if the power levels are relatively high such that nonlinear impairments play a larger role. There has been a large research effort in the area of *impairment-aware routing and wavelength assignment* (IA-RWA), leading to a number of proposed link metrics and design methodologies that account for various impairments during the routing and regeneration processes. Survey papers that strive to classify the myriad approaches can be found in Azodolmolky et al. [AKMC09] and Rahbar [Rahb12].

One methodology focuses on the performance measure known as the *Q-factor,* where higher Q correlates to lower BER [Pers73]. In Kulkarni et al. [KTMT05] and Markidis et al. [MSTT07], a Q penalty is calculated for each link, where this penalty can take into account noise, crosstalk, dispersion, and FWM, among other impairments. Using the Q penalty as the link metric in a "shortest path" routing algorithm then favors finding paths with relatively good performance. (However, in general, the Q-factor of an end-to-end path cannot be calculated directly by the Q-factors of the links that comprise the path.) A related methodology in Morea et al. [MBLA08] defines a *quality of transmission* function that captures various impairments via an OSNR penalty, where the penalties are determined based on experimental results.

Rather than attempt to capture all impairments in a single link metric, Manousakis et al. [MKCV10] propose using a multicost approach, where each link is assigned a cost *vector*. The set of costs includes various additive metrics, e.g., the link length, the noise variance of a "1" signal, the noise variance of a "0" signal, the wavelength utilization, etc. (The noise variance includes components due to ASE noise, crosstalk, FWM, and XPM. Using noise variance as a link metric was first proposed in He et al. [HBPS07].) A path between two nodes that has all of its metrics higher (i.e., worse) than those of another path between the same two nodes is considered "dominated." A modified version of Dijkstra's algorithm is run to find paths from source to destination, where multiple non-dominated paths from source to intermediate nodes are tracked. Any dominated paths are eliminated from further consideration. After generating a set of non-dominated paths from source to destination, a function is applied to the cost vector to generate a scalar value, where the function is monotonic with respect to the various metrics. The path producing the best result is selected. This methodology can be used to insert regenerations along the path, where the number of regenerations is one of the metrics for the path. This produces multiple "versions" of a given path, each with its own set of metrics, depending on where the regenerations are placed. The cost-vector approach can be used to incorporate wavelength assignment as well; see Chap. 5.

There have also been proposals to run a modified shortest-path algorithm where each of the running metrics in the cost vector is compared to an associated threshold. Any partial path that exceeds one or more thresholds is excluded from further consideration in the algorithm. However, for some impairments, there is not an absolute tolerance. For example, a system may nominally tolerate a chromatic dispersion of *D*. Rather than eliminating a path with dispersion higher than *D,* a penalty that captures the excess dispersion could be added to the overall performance metric. If the path performs well with respect to all other metrics, it may still be a viable path. Thus, setting strict thresholds for some of the impairments may be too conservative of a strategy.

As will be discussed in Sect. 4.6.2, some networks may restrict regeneration to just a subset of the nodes; optical bypass must occur in the remaining nodes. An IA-RWA algorithm designed for such an architecture was proposed in Zhao et al. [ZhSB12]. Prior to adding any traffic, numerous paths are calculated for each source/destination pair, e.g., using a K-shortest-paths algorithm, with *K* on the order of 40. The all-optical segments of a path are those portions of the path that *must* be all-optical due to a lack of regeneration capability at any of the intermediate nodes. If a particular path contains an all-optical segment that violates the optical reach, taking into account just static impairments, then that path is eliminated from further consideration (static impairments, e.g., noise, are those that do not depend on what connections are present on the fiber). The remaining paths are sorted from shortest to longest length. As demand requests arrive for a particular source/destination, the paths are checked one by one for viability, where both wavelength continuity and optical reach must be satisfied on each all-optical segment of the path. The optical reach calculation in this stage takes into account the dynamic impairments as well, i.e., those that are based on the state of the network (e.g., XPM and FWM due to neighboring wavelengths). Furthermore, each path is checked to ensure that its addition would not cause an existing connection to violate its desired quality of transmission (QoT). The first path in the candidate set that is viable is selected for the new demand. A dynamic programming strategy is then used to minimize the number of regenerations for the path, subject to the restriction that regeneration can only occur at designated nodes.

Some of the proposed IA-RWA methods are quite complex, especially when compared to the relatively simple approach of shortest-path routing using NF as the link metric. Clearly, it is desirable to simplify the system-engineering rules as much as possible while still producing accurate results.

4.5 Link Engineering

Link engineering (or span engineering) is the process of designing the physical infrastructure for a particular network. For example, this may include determining the amplifier type for each site (e.g., pure Raman, pure EDFA, or hybrid Raman/EDFA), the gain setting of each amplifier, the amount of dispersion compensation,

and the locations of the dispersion compensation. For a given set of amplifiers, proper setting of the gain and proper distribution of dispersion compensation can provide extra system margin (say on the order of 0.25–0.5 dB). Algorithms specifically designed to optimize the system performance can be very helpful in this process. For example, *simulated annealing* [VaAa87] proved to be a good technique for selecting the locations of the dispersion compensation modules.

Clearly, the design choices in the physical layer have an impact on the network design at higher layers. It can be very advantageous to have a unified network design tool that incorporates the physical-layer design. For example, consider a system that has two types of amplifiers, where the Type A amplifier provides greater gain than the Type B amplifier, but it costs more. Installing the Type A amplifier as opposed to the Type B amplifier in certain sites reduces the NF of the corresponding links. However, the effect on the overall amount of regeneration in the network may be small. By integrating physical-layer design with network planning, one can evaluate whether the extra cost of the Type A amplifier is justified by the expected reduction in regeneration. This allows better performance and cost optimization of the network as a whole.

4.6 Regeneration Strategies

The previous sections addressed some of the physical-layer factors that affect where regeneration is required. In this section, some of the architectural issues related to regeneration are considered.

There are several approaches to managing regeneration in a network. Three strategies are presented here, where the three differ with respect to flexibility, operational complexity, and cost. Although the above discussion has elucidated several factors other than distance that affect optical reach, the illustrative examples here will continue to refer to optical reach in terms of distance, for simplicity.

4.6.1 Islands of Transparency

In the architectural strategy known as "islands of transparency" [Sale98a, Sale00], a network is partitioned into multiple "islands." The geographic extents of the islands are such that any intra-island (loopless) path can be established without requiring regeneration. However, a regeneration is required whenever a path crosses an island boundary, regardless of the path distance.

An example of a network partitioned into three islands is shown in Fig. 4.7. Island 1 is composed of Nodes A through H; regeneration is not required for a connection between any two of these nodes. Node A is also a member of Island 2, and serves as the regeneration point for any connection routed between Island 1 and Island 2. It is assumed that Node A is equipped with two ROADMs. One ROADM

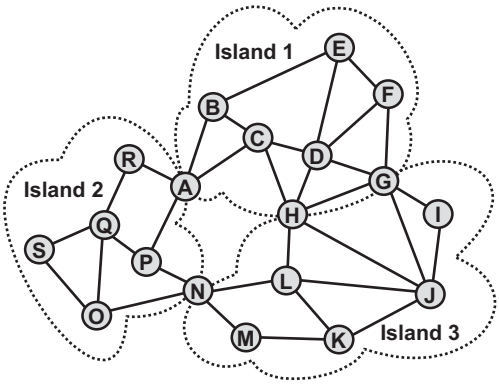


Fig. 4.7 This network is partitioned into three islands of transparency. Any traffic within an island does not need regeneration. Any traffic between islands does require regeneration

is oriented to allow traffic routed between Links AB and AC to optically bypass the node; the other ROADM is oriented to allow optical bypass between Links AR and AP. Traffic between Islands 1 and 2 is dropped from one ROADM and added to the other ROADM, with an O-E-O regeneration occurring in between. (Refer to Fig. 2.8(b) for an illustration of a degree-four node equipped with two ROADMs.) The other nodes that fall on the boundaries between islands, i.e., Nodes H, G, and N, have similar types of configurations.

There are several operational advantages to this architecture. First, it completely removes the need to calculate where to regenerate when establishing a new connection. Regeneration is strictly determined by the island boundaries that are crossed, if any. Second, the islands are isolated from each other, so that a carrier could potentially deploy the equipment of a different vendor in each island without having to be concerned with interoperability in the optical domain. For example, each vendor's system could support a different number of wavelengths on a fiber. This isolation is advantageous even in a single-vendor system, as it allows the carrier to upgrade the equipment of the different islands on independent time scales.

The chief disadvantage of the "islands" architecture is that it results in extra regeneration. In the figure, assume that a connection is required between Nodes B and P, along path B-A-P, and assume that the path between them is shorter than the optical reach. Based on distance, no regeneration is required; however, because of the island topology, the connection is regenerated at Node A. Thus, the simplicity and flexibility of the system comes with an additional capital cost.

A methodology for partitioning a network into transparent islands was presented in Karasan and Arisoylu [KaAr04] and Shen et al. [ShST09]. The first step is to enumerate all of the faces of the network graph (assuming the graph is planar), where a face is a region bounded by links such that no other links are included in the region. For example, in Fig. 4.7, links BC, CD, DE, and EB form a face. The following procedure can be used to enumerate the faces. Start with any link. Form a counterclockwise cycle starting at that link, where at each node the outgoing link of the cycle is the one that is immediately clockwise to the incoming link. The traversed links in the cycle constitute a face. (It is assumed that all nodes have a degree of at least

two.) Repeat these steps, where in each iteration the cycle-forming process starts by selecting a "non-traversed" link. Continue until all links in the network graph have been traversed at least once (some links are traversed in two different directions), at which point all faces have been enumerated.

The next step is to select one face, and create a new island with it. (It is assumed that the geographic extent of any face is small enough that regeneration-free paths exist between each pair of nodes in the face.) Next, a neighboring face, i.e., one with a common link, is temporarily added to this island. If there is a regeneration-free path between all nodes in this new island, then this face is permanently added to the island; if not, it is removed. This process continues until all neighboring faces of the island have been checked and no more can be added to the island (as the island continues to grow, there are new neighboring faces to be checked; additional criteria can be used in determining which neighboring face to add first [KaAr04]). If faces still exist that are not in any island, then one such face is selected and the island-growing process is repeated. This process does a reasonably good job of minimizing the number of islands formed, which is desirable from the point of view of minimizing the number of O-E-O nodes at island boundaries.

Rather than creating islands with nodes being at the intersection of neighboring islands, it is possible to partition the network such that links are the common entity between neighboring islands; see Exercise 4.15.

Note that the isolation provided by the island paradigm typically exists with respect to different geographic tiers of a network. For example, it is not common for traffic to be routed all-optically from a metro network into a backbone (or regional) network. The metro WDM system usually has coarser wavelength spacing and lower-cost components. The tolerances of the metro optics may not be stringent enough to be compatible with the system of the backbone network. Thus, the metro traffic typically undergoes O-E-O conversion prior to being carried on a backbone network regardless of the connection distance. (However, the metro and backbone tiers do not necessarily represent islands according to the definition above because regeneration may also be required *within* these tiers.)

4.6.2 Designated Regeneration Sites

A second architectural strategy designates a subset of the nodes as regeneration sites, and allows regeneration to occur only at those sites. If an end-to-end connection is too long to be carried solely in the optical domain, then it must be routed through one or more of the regeneration sites. Either the designated regeneration sites are equipped with O-E-O equipment such that all traffic that transits them needs to be regenerated, or they have optical-bypass equipment so that regeneration occurs only when needed.

The routing process must be modified to take into account the limited number of sites that can provide regeneration. To ensure that a valid path is found that transits the necessary regeneration sites, one can use a graph transformation similar

to the one discussed in Sect. 3.6.2, i.e., the "reachability graph" is created. (Other strategies are presented in Carpenter et al. [CMSG04] and Yang and Ramamurthy [YaRa05a].) In Sect. 3.6.2, the reachability graph was discussed in the context of real-time network design, where only certain nodes may have available regeneration equipment. Restricting regeneration to certain nodes in a network is an equivalent problem. Only those nodes that are designated as regeneration sites (plus the source and destination of the demand to be routed) appear in the reachability graph. A link between two of these nodes is added only if a regeneration-free path exists between them in the true topology. Finding a path in the reachability graph guarantees regeneration feasibility.

4.6.2.1 Strategies for Determining the Set of Regeneration Sites

Several strategies have been proposed for selecting the set of nodes at which regeneration can occur. They vary with respect to complexity and design objective.

One simple strategy for selecting the regeneration sites is to first route all of the forecasted traffic over its shortest path. A greedy type of strategy can then be employed where nodes are sequentially picked to be regeneration sites based on the number of paths that become feasible with regeneration allowed at that site [CSGJ03, YeKa03]. (More than one regeneration site may need to be added in a single step.) Enough nodes are picked until all paths are feasible.

Another strategy for selecting the regeneration sites, which utilizes the network topology rather than a traffic forecast, is based on *connected dominating sets* (CDSs) [CSGJ03]. A dominating set of a graph is a subset of the nodes, S, such that all nodes not in S are directly connected to at least one of the nodes in S. The dominating set is connected if there is a path between any two nodes in S that does not pass through a node not in S. The first step of the CDS methodology is to create the reachability graph. All network nodes are included in the graph, and two nodes are connected by a link only if there is a regeneration-free path between them in the true topology. A minimal CDS is found for the reachability graph, using heuristics such as those described in Guha and Khuller [GuKh98]. (The minimal CDS is the CDS with the fewest number of nodes.) The nodes in the minimal CDS are designated as the regeneration sites. By the definition of a CDS and by virtue of how links are added to the reachability graph, any node not in the CDS is able to reach a node in the CDS without requiring intermediate regeneration. This guarantees that for any source/destination combination, a path exists that is feasible from a regeneration standpoint.

The goal of the CDS strategy is to minimize the number of nodes selected as regeneration sites. However, this may result in demands needing to follow circuitous paths or in paths that require excess regeneration. The scheme proposed in Bathula et al. [BSCF13] addresses both of these issues. Assume that the goal is to select a set of regeneration nodes that minimizes the amount of required regeneration. We represent this set by R. As in the CDS strategy, the first step is to create the reachability graph containing all nodes in the network. Note that a minimum-hop path in

this network corresponds to a minimum-regeneration path. For each source/destination pair, if all minimum-hop paths in the reachability graph for that pair contain a particular node, or group of nodes, then those nodes are added to *R*. After this step is completed for each source/destination pair, a check is performed to see if there are any pairs such that a minimum-hop path cannot be found (in the reachability graph) where all intermediate nodes of the path belong to *R*; let *P* represent these source/destination pairs. If *P* is nonempty, then more nodes need to be added to *R*. A variety of strategies can be used. For example, a greedy algorithm can be used, where, in each iteration, the node that belongs to a minimum-hop path of the most source/destination pairs still in *P* is added to *R*. Enough iterations of the greedy algorithm are run until *P* is empty. A final post-processing is performed to see if any of the regeneration nodes added via the greedy strategy can be removed without causing any source/destination pairs to be moved back into *P*.

This scheme can be modified to minimize cost, where cost takes into account both regenerations and wavelength-km of bandwidth. Assume that the cost of a regeneration is equivalent to the cost of *D* km of bandwidth. The links in the reachability graph are assigned their corresponding distance in the true network topology, plus *D* to account for regeneration cost. The same process as described above for selecting *R* to minimize regeneration can then be used to minimize cost, except minimum-distance paths are considered in the reachability graph instead of minimum-hop paths.

4.6.2.2 Advantages and Disadvantages of Designated Regeneration Sites

As noted above, the designated-regeneration-site architecture may result in extra regenerations depending on the strategy used to select the sites. This is illustrated in Fig. 4.8. Assume that the optical reach is 1,000 km, and assume that regeneration is permitted only at Nodes B, D, and E. Assume that the nodes are equipped with optical-bypass elements, including the regeneration-capable sites. A connection between Nodes A and Z is ideally regenerated at just Node C. However, because regeneration is not permitted at Node C, the connection is regenerated at both Nodes B and E (or at Nodes B and D), resulting in an extra regeneration.

Moreover, if the designated regeneration sites are equipped with O-E-O network elements (e.g., back-to-back optical terminals), then the amount of excess regeneration is likely to be significantly higher, as any connection crossing such a node needs to be regenerated. The extra regeneration cost may be partially offset by somewhat lower network element costs. Consider two adjacent degree-two network nodes where the two nodes are close enough together such that any required regeneration could equivalently occur in either node. Assume that all of the transiting traffic needs to be regenerated in one of the two nodes. In Scenario 1, assume that both nodes are equipped with ROADMs and assume that both regenerate 50 % of the traffic that traverses them. In Scenario 2, one of the nodes has a ROADM and regenerates none of the traffic, and the other node has two optical terminals and regenerates 100 % of the transiting traffic. The total amount of regeneration is the

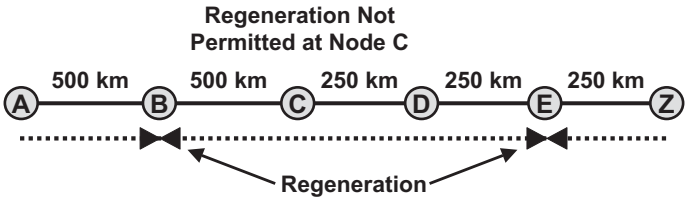


Fig. 4.8 Assume that the optical reach is 1,000 km. A connection between Nodes *A* and *Z* is ideally regenerated in just Node *C.* However, because it is assumed that regeneration is not permitted at this site, the connection is regenerated at both Nodes *B* and *E* instead

same in either scenario. However, because two optical terminals cost less than a ROADM, Scenario 2 is overall less costly. (This is an extreme example because it assumed 100 % of the transiting traffic needed to be regenerated.) Setting up O-E-O-dedicated regeneration sites in a network may have this same effect to a degree, but the overall cost is still likely to be higher because of the extra regeneration.

A possible benefit of designating only certain nodes as regeneration sites is more streamlined equipment pre-deployment. With regeneration occurring at a limited number of sites, the process of pre-deploying equipment is more economical; i.e., with fewer pools of regeneration equipment, it is likely that less regeneration equipment needs to be pre-deployed across the network to reduce the blocking probability below a given threshold (see Exercise 4.16). It may also simplify network operations.

Overall, given that regeneration can be accomplished without any special equipment (i.e., the same transponders used for traffic add/drop can be used for regeneration), it is not clear if it is necessary to severely limit the number of nodes at which regeneration is supported. The biggest disadvantage is it makes the routing process more challenging and less flexible, and it may cost more depending on the amount of extra regenerations it produces. Even if the cost is not greater, the congestion on particular links may be higher, due to the greater constraints placed on routing. While it may make sense to eliminate some nodes as possible regeneration sites (e.g., because the nodal offices are not large enough to house a lot of terminating equipment), in general, designating only a subset of the nodes for regeneration may not be optimal.

4.6.3 Selective Regeneration

The third strategy is *selective regeneration,* which allows any (or almost any) node to perform regeneration; a connection is regenerated only when needed. The decision as to whether regeneration is needed, and if so, where to implement it, is made on a per-connection basis. This strategy is the one most commonly used in actual network deployments. Given the freedom in selecting regeneration locations, this approach yields the fewest regenerations (assuming enough available regeneration equipment is deployed at the nodes) and also allows the most flexibility when routing.

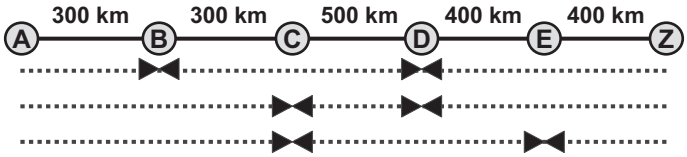


Fig. 4.9 Assume that the optical reach is 1,000 km, and assume that regeneration is permitted at any node. A connection between Nodes *A* and *Z* can be regenerated at Nodes *B* and *D,* at Nodes *C* and *D,* or at Nodes *C* and *E*

Often, there will be several options as to where the regeneration can occur for a given connection. Consider a connection between Nodes A and Z in Fig. 4.9. Assume that the optical reach is 1,000 km, and assume that regeneration is possible in any of the nodes. The minimum number of regenerations for the connection is two. As shown in the figure, there are three possible scenarios that yield two regenerations: regenerate at Nodes B and D, regenerate at Nodes C and D, or regenerate at Nodes C and E.

There are several factors that should be considered when selecting one of these regeneration scenarios for the AZ connection. In real-time planning, the amount of available equipment at each node should be considered, where regeneration is favored in the nodes that have more free equipment. Additionally, if the nodes are equipped with network elements that have a limit on the total add/drop (see Sect. 2.6.1), then it is important to favor regeneration at the nodes that are not close to reaching this limit. (Reaching the maximum amount of add/drop at a node could severely impact future growth, as that node will not be able to source/terminate more traffic.)

One needs to also consider the subconnections that will result from a particular regeneration scenario. The term *subconnection* is used here to refer to the portions of the connection that fall between two regeneration points or between an endpoint and a regeneration point. For example, if the connection in the figure is regenerated at Nodes C and D, the resulting subconnections are A-C, C-D, and D-Z. By aligning the newly formed subconnections with those that already exist in the network (i.e., producing subconnections with the same endpoints, on the same links), the wavelength assignment process may encounter less contention. Furthermore, in a waveband system where bands of wavelengths are treated as a single unit, creating subconnections with similar endpoints yields better packing of the wavebands.

One could also consider the system margin of the resulting subconnections when selecting where to regenerate. If the connection in Fig. 4.9 is regenerated at Nodes C and E, then one of the resulting subconnections, C-E, has a length of 900 km. With the other two regeneration options, no subconnection is longer than 800 km, such that there is somewhat greater system margin. Of course, if the optical reach is specified as 1,000 km, then any of these options should work. Thus, while taking the resulting distances of each subconnection into account may produce *extra* margin, adding this consideration to the algorithms should not be required, assuming the system works as specified.

Another factor is that the optical reach of a small number of wavelengths may be significantly shorter than that of the other wavelengths, e.g., due to regions of very low fiber dispersion. It may be beneficial to select regeneration points such that a short subconnection is produced that can make use of one of the "impaired" wavelengths.

Note that some of the aforementioned factors may be at odds with one another. The desire to align the subconnections favors continuing to regenerate at the same nodes, but the add/drop limits of the equipment may require regeneration to be more dispersed. Given that reaching the add/drop limit at a node could restrict network growth, the limits of the network elements, if any, should dominate as the network becomes more full.

4.7 Regeneration Architectures

This chapter has thus far covered many of the physical effects, as well as the architectural strategies, that affect where a given connection must be regenerated. All of the relevant factors must be incorporated in the network planning tool to ensure that regeneration sites are selected as needed for each demand. This section examines how regeneration is actually implemented within a node.

4.7.1 Back-to-Back WDM Transponders

As has been pointed out several times already, one means of regenerating a signal is to have it exit the optical domain on one WDM transponder and reenter the optical domain on a second WDM transponder. Figure 4.10 illustrates this architecture where the pairs of transponders used for regeneration are connected via a patch cable. The process of O-E-O conversion typically achieves full 3R regeneration, where the signal is reamplified, reshaped, and retimed. It usually provides an opportunity to change the wavelength of the optical signal as well.

The flexibility of the back-to-back transponder architecture largely depends on the capabilities of the equipment. Consider regeneration of a connection entering from the East link and exiting on the West link. If the corresponding transponders connected to the East and West links are fully wavelength tunable, then each transponder can be independently tuned such that any (East/West) input/output wavelength combination is supported. If the transponders are not tunable, then the possible input/output wavelength combinations depend on the wavelengths of the transponders that are cabled together.

If the ROADM-MD is *directionless,* as illustrated in Fig. 4.10a, then any pair of back-to-back transponders can access any two of the network links. For example, on the left-hand side of Fig. 4.10a, there are two regenerations shown, between the East and South links and the South and West links. On the right-hand side,

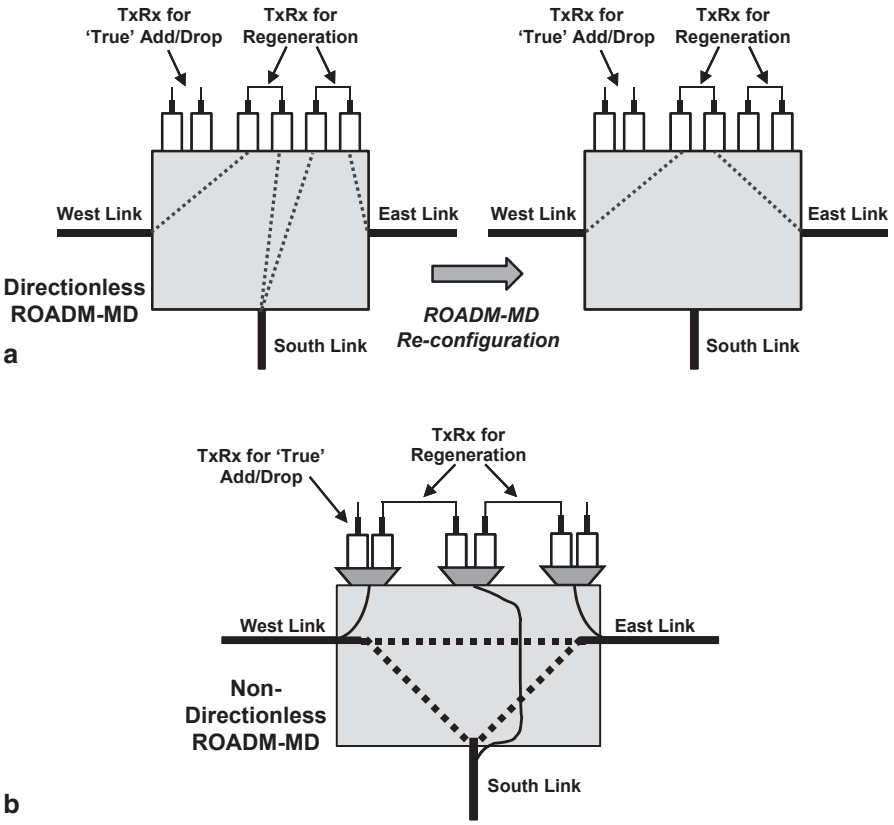


Fig. 4.10 Regeneration via back-to-back transponders (*TxRx*'s) that are interconnected by a patch cable. **a** The directionless multi-degree reconfigurable optical add/drop multiplexer (*ROADM-MD*) allows any transponder pair to access any two network links. The *left-hand side* shows regeneration between the East/South links and the South/West links. After the *ROADM-MD* is reconfigured, the *right-hand side* shows regeneration between the East/West links. **b** In the non-directionless *ROADM-MD,* the transponders are tied to a particular network link; thus, in this example, regeneration is possible only between the East/South and South/West links. (Adapted from Simmons [Simm05]. © 2005 IEEE)

after the ROADM-MD has been reconfigured, there is one regeneration between the East and West links. Contrast this with Fig. 4.10b, where the ROADM-MD is *non-directionless,* such that the transponders are tied to a network link. With this network element, the possible regeneration directions are determined by which pairs of transponders are interconnected. Thus, in Fig. 4.10b, regeneration is possible between the East/South links and between the South/West links, but not between the East/West links. Manual intervention is required to rearrange the patch cables to allow for different configurations.

In Fig. 4.10, one limitation is that the transponders must be partitioned between the "true" add/drop function and the regeneration function. (As mentioned previ-

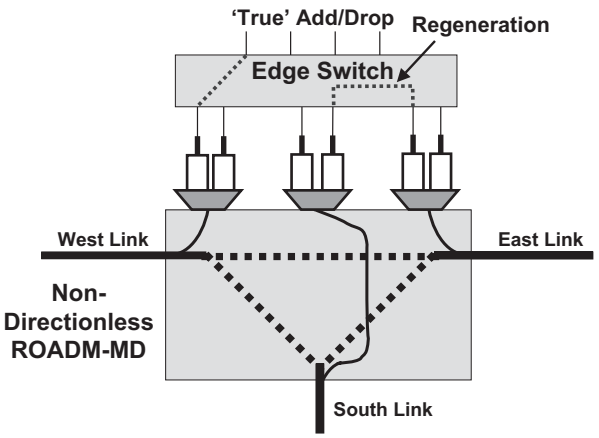


Fig. 4.11 An edge switch provides flexibility at a node with a non-directionless multi-degree reconfigurable optical add/drop multiplexer (*ROADM-MD*). Any transponder can be used for either "true" add/drop or regeneration; any regeneration direction through the node is supported. (Adapted from Simmons [Simm05]. © 2005 IEEE)

ously, a regeneration can be considered add/drop traffic because the signal drops from the optical layer. The term "*true* add/drop" is used here to distinguish those signals that actually originate from, or terminate at, the node.) Only those transponders that are cabled together can be used for regeneration, where manual intervention is required to adjust the apportionments.

To address this limitation, one can use an edge switch, e.g., an FXC, as shown in combination with a non-directionless ROADM-MD in Fig. 4.11. Adding the edge switch allows any transponder to be used for either "true" add/drop or regeneration, depending on the switch configuration. While this architecture incurs the cost of the edge switch, it reduces the number of transponders that have to be pre-deployed at a node and reduces the amount of manual intervention required to configure the node. The presence of the edge switch also provides edge configurability for the non-directionless ROADM-MD (as discussed in Sect. 2.9.4.1). Thus, it allows the transponders to be used to regenerate in any direction through the node, as could already be achieved with the directionless ROADM-MD.

An edge switch could also be used in combination with a directionless ROADM, so that any transponder can be used for "true" add/drop or regeneration.

4.7.2 Regenerator Cards

A WDM transponder converts an incoming WDM-compatible optical signal to a 1,310-nm optical signal. When two transponders are cabled together for regeneration, they communicate via the 1,310-nm signal. However, for regeneration purposes, this conversion is unnecessary. A simpler device capable of 3R regeneration, referred to as a *regenerator,* is shown in the architectures of Fig. 4.12. The received optical signal on one side of the regenerator is converted to an electrical signal that directly modulates the optical transmitter on the other side, eliminating the need for short-reach interfaces (i.e., the 1,310-nm interfaces). Note that the regenerator

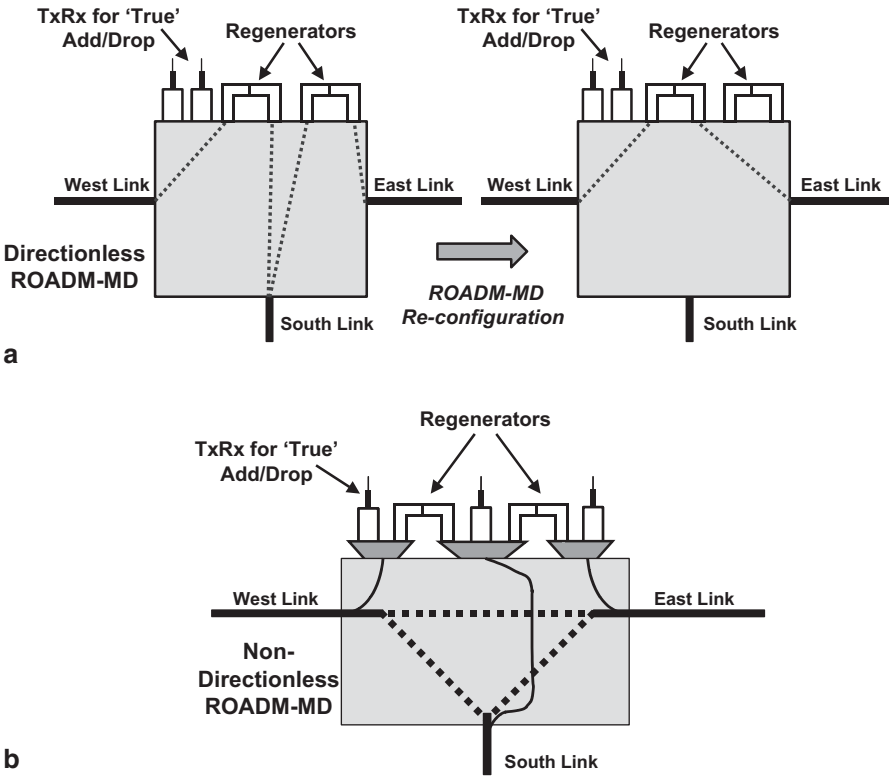


Fig. 4.12 **a** Regenerators used in conjunction with a directionless multi-degree reconfigurable optical add/drop multiplexer (*ROADM-MD*). By reconfiguring the *ROADM-MD,* a regenerator card can be used for regeneration between different combinations of links. **b** Regenerator used in conjunction with a non-directionless *ROADM-MD.* In the configuration shown, regeneration is supported only between the East/South links and the South/West links. (Adapted from Simmons [Simm05]. © 2005 IEEE)

is a bidirectional device; i.e., one regenerator supports both directions of a connection. The motivation for the regenerator is cost; the cost of one regenerator card is roughly 70–80 % of the cost of two transponder cards.

Regenerator cards cannot be used for "true" add/drop, as they lack the short-reach interface. Thus, there is a clear division between the add/drop equipment and the regeneration equipment. The attributes of the regenerator card can have a profound impact on network operations. Consider using a regenerator card for a connection routed on the East and West links. It is desirable for the regenerator to allow the incoming wavelength from the East link to be different from the outgoing wavelength on the West link (the same also applies for traffic in the reverse direction). Otherwise, wavelength conversion would be prohibited from occurring in concert with a regeneration, which could be a significant restriction. Ideally, the regenerator card is fully tunable, such that any combination of incoming and outgo-

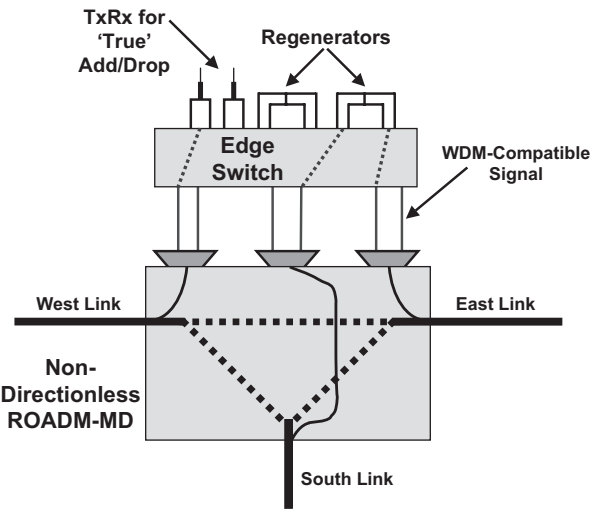


Fig. 4.13 With an edge switch used in combination with regenerator cards, four ports on the switch are utilized for each regeneration. Additionally, the edge switch must be capable of switching a wavelength-division multiplexing (*WDM*)-compatible signal (e.g., the switch could be a MEMS-based fiber cross-connect)

ing wavelengths can be accommodated with a single card. If the regenerator cards are not tunable, then inventory issues become problematic if every combination of input and output wavelengths is potentially desired (see Exercise 4.19). Storing thousands of different regenerator combinations would be impractical.

Regenerator cards can be used with a *directionless* ROADM-MD, as shown in Fig. 4.12a, or with a *non-directionless* ROADM-MD, as shown in Fig. 4.12b. As with the back-to-back transponder architecture, the directionless ROADM-MD allows a regenerator to be used for regeneration in any direction through the node. In the non-directionless ROADM-MD, the regenerator is tied to a particular regeneration direction (e.g., East/South and South/West in the figure).

Deploying an edge switch with a non-directionless ROADM-MD to gain flexibility with the regenerator card is inefficient with respect to switch port utilization. The configuration is shown in Fig. 4.13. The signals that need to be regenerated are directed by the edge switch to regenerator cards, which ideally are tunable. The signals that are truly dropping at the node are fed into transponders. The edge switch allows wavelengths from any two network ports to be fed into a particular regenerator, thereby providing flexibility in the regeneration direction. For example, in the figure, the edge switch is configured to enable a regeneration between the East/South links. However, to accomplish this, note that four ports are utilized on the edge switch for a regeneration. With the configuration shown in Fig. 4.11, only two ports are utilized on the edge switch per regeneration. Additionally, in Fig. 4.13, the edge switch must be capable of switching a WDM-compatible signal; e.g., it could be a MEMS-based fiber cross-connect. An electronic-based switch is not suitable for this application, whereas it would suffice in Fig. 4.11.

Rather than using an edge switch, a degree of configurability can be attained with the non-directionless ROADM-MD architecture through the use of flexible regenerators [SiSa07], similar to the flexible transponders that were discussed in

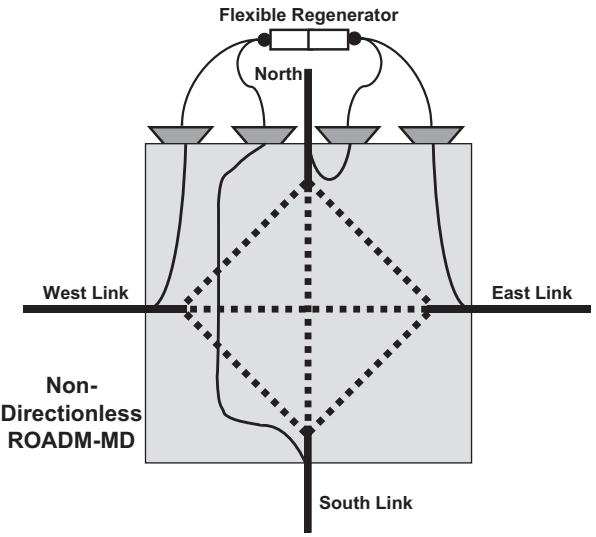


Fig. 4.14 A degree-four non-directionless multi-degree reconfigurable optical add/drop multiplexer (*ROADM-MD*) combined with a flexible regenerator that allows regeneration in either the East/West, East/South, North/West, or North/South directions. An optical backplane can be used to eliminate complex cabling. (Adapted from Saleh and Simmons [SaSi06]. © 2006 IEEE)

Sect. 2.9.4.1. Refer to the flexible regenerator shown in Fig. 4.14 with a degree-four non-directionless ROADM-MD. One side of the regenerator is connected to the East and North links, and the other side is connected to the West and South links. Thus, this one regenerator allows regeneration to occur in either the East/West, East/South, North/West, or North/South directions; i.e., four of the six possible directions through the node are covered. Furthermore, it allows a regenerated signal to be sent out on two simultaneous links; e.g., a signal entering from the East link can be regenerated and sent out on both the West and South links. This is useful for multicast connections, as was discussed in Sect. 3.10.

4.7.3 All-Optical Regeneration

All-optical regeneration has been proposed as an alternative to O-E-O regeneration [LLBB03, LeJC04, Ciar12, Mats12]. If this becomes a commercially viable technology, then due to the scalability of optics, it is expected that the cost of all-optical regenerators will scale well with increasing line rate. For example, there may be only a small price premium for a 100-Gb/s regenerator as compared to a 40-Gb/s regenerator. Furthermore, all-optical regenerators are expected to consume less power as compared to their electronic counterparts, which should improve nodal scalability.

All-optical regenerators may not fully replace electronic regeneration. Some all-optical regenerators provide only 2R regeneration, i.e., reamplification and reshaping, as opposed to 3R, which includes retiming as well. Thus, a combination of all-optical and electronic regeneration may be needed. The bulk of the regeneration

can be performed all-optically, with electronic regenerators used intermittently to clean up the timing jitter.

Early all-optical regeneration techniques were compatible only with simple modulation formats such as OOK. However, this is not sufficient for many of today's networks where more advanced modulation formats are used. Furthermore, modulation formats will continue to grow more complex in order to meet the capacity requirements of future networks. All-optical regeneration that is compatible with advanced modulation formats is an area of current research, e.g., Croussore and Li [CrLi08], Kakande et al. [KBSP10], and Sygletos et al. [SFGE12]. Many of the solutions that have been proposed are fiber based, which will result in a bulky design. It is desirable that solutions that are more integratable be developed, e.g., the PIC-based approach of Andriolli et al. [AFLB13].

All-optical regenerators, at least initially, are likely to operate on a per-wavelength basis, similar to electronic regenerators; i.e., Fig. 4.12 holds for all-optical regenerators as well. It is expected that these all-optical regenerators will provide complete flexibility with respect to wavelength conversion, allowing any input/output wavelength combination.

The economics could potentially improve further if all-optical regeneration can be extended to multiple wavelengths, where a band of wavelengths is processed by a single regenerator [CXBT02, PaVL10, PPPR12, SFGE12]. One of the major challenges is that the nonlinear effects typically used for single-channel all-optical regeneration cause deleterious interchannel crosstalk effects when extended to multiple-channel regeneration. Most of the all-optical multi-wavelength regeneration experiments reported thus far operate on a very small number of channels or utilize large channel spacing. Additionally, it is not clear whether multichannel regeneration can support arbitrary wavelength conversion.

Overall, all-optical regeneration is still an area of active research on many fronts, and requires many technical hurdles to be overcome before this technology could become commercially viable.

4.8 Exercises

4.1 What percentage improvement corresponds to a 1-dB improvement? A 3-dB improvement? A 10-dB improvement? Using these results, to what percentage improvement does a 14-dB improvement correspond?

4.2 Consider two adjacent links, both with a NF of 20 dB and a net gain of 0 dB. (a) What is the NF (in dB) of the two-link path (ignore any network element at the junction of the two links)? (b) In general, if the two links have a NF of L dB (and 0-dB net gain), what is the NF of the two concatenated links? (c) How about if M links each with a NF of L dB (and 0-dB net gain) are concatenated?

4.3 Consider a fiber span with a loss of L dB. Assume that an erbium amplifier with a NF of F dB is placed at the end of the span. What is the formula for the total NF (in dB) of the amplified span? Hint: The NF of a length of fiber equals the loss of the fiber.

4.4 Equation 4.2 is the general NF concatenation formula for two links. Use this to derive the general NF concatenation formula for three links.

4.5 Let $\oplus$ represent the concatenation of two NFs, as given by Equation 4.2, for the general case where the net gain of each link is not equal to 1 (in linear units). (a) Is the concatenation formula associative, i.e., does $A \oplus (B \oplus C)$ equal $(A \oplus B) \oplus C$? (b) Is the concatenation formula "right monotonic"; i.e., is A always less than or equal to $A \oplus B$? (c) Is the concatenation formula commutative; i.e., does $A \oplus B$ equal $B \oplus A$? (d) Is the concatenation formula "right isotonic"; i.e., does $A < B$ imply $A \oplus C < B \oplus C$? (e) Would NF (where the net gain on each link is not necessarily equal to 1) be suitable as a link metric for the Dijkstra algorithm? (f) We know that if there is no net gain on any link, then NF is suitable as a link metric. Is there a more general condition on the gain that allows NF to be suitable as a link metric? Hint: See Yang and Wang [YaWa08].

4.6 Define the OSNR for a link to be $OSNR\,link = \frac{\text{Signal Power Level}}{\text{ASENoise of Link}}$. Assume that a network has a net gain on any span of 0 dB, such that the signal power level in the link OSNR formula can be treated as a constant. Show that using $\frac{1}{\text{OSNR link}}$ as the link metric in a shortest-path routing algorithm is equivalent to using the link NF as the routing metric. Hint: The ASE noise of each link is additive.

4.7 Assume that a series of five links each have a NF of 20 dB. Assume that the system engineering rules allow for one link in this series to have a net gain of 2 dB, as long as the following link has a net gain of −2 dB. (a) Assuming that it is the first link in the series that is overamplified by 2 dB, by how much does the NF improve at the end of the five links? (b) Does it make a difference if it had been the second link in the series that was over-amplified instead of the first? (c) What is the NF at the end of the five links if the first link is *under*-amplified by 2 dB and the second link is over-amplified by 2 dB? Comparing this result with that of part (a), is it better from a NF perspective to under-amplify and then over-amplify, or over-amplify and then under-amplify?

4.8 Assume that a network has span distances of 80 km, and assume that Raman amplification is being used, where the Raman gain is set equal to the span loss. Assume that the overall NF for an amplified span with 20-dB loss is 18.5 dB, and assume that the NF decreases linearly by 0.75 dB for every 1-dB decrease in span loss. Assume that the system can tolerate a cumulative NF of 33 dB before requiring regeneration. (a) Assume that legacy fiber is deployed, with a loss of 0.23 dB/km. What is the optical reach on this fiber (ignore any effects other than fiber loss)? (b) Assume that new fiber is deployed, with a loss of 0.19 dB/km. What is the optical reach on this new fiber?

4.9 Assume that a system has a hybrid two-stage amplifier, where the first stage is Raman based, with a maximum gain of 18 dB, and the second stage is EDFA based, with a maximum gain of 7 dB. Assume that the amplifier is placed at the end of a span that has a total loss of 20 dB, and assume that the net gain, after both stages of amplification, should be 0 dB. The Raman amplification is distributed over the fiber span that precedes it (i.e., treat the fiber span and the Raman amplifier as one stage). At 13-dB Raman gain, the NF for the first

stage is 21 dB; assume that the NF decreases linearly by 0.25 dB for every 1-dB increase in Raman gain. The NF of the EDFA stage is fixed at 6 dB regardless of its gain. What should the gain settings be for the Raman and EDFA portions of the amplifier to minimize the overall NF, and what is the overall NF of the two-stage amplifier?

4.10 (a) Assume that the fiber deployed in a network has a PMD coefficient of $1\ ps/\sqrt{km}$, and that the system DGD tolerance is 50 ps. If no PMD compensation is utilized, how far can a signal travel before it reaches the DGD limit? (b) Assume that the fiber has a chromatic dispersion of 17 ps/(nm·km), and that the system dispersion limit is 30,000 ps/nm. On average, how much dispersion compensation must be added per km such that the PMD-limited distance from part (a) can be attained?

4.11 Assume that the fiber deployed in a network has a chromatic dispersion with a constant positive *slope* across the spectral region of interest (e.g., 1,530–1,565 nm), such that its minimum value is 2 ps/(nm·km) and its maximum value is 7 ps/(nm·km). Assume that the system dispersion tolerance is ±10,000 ps/nm. Assume that dispersion compensation, with *constant* dispersion across the spectral band, is utilized. Assume that it is desired to have a dispersion-limited reach of R, where any connection in the spectral region of interest can attain *at least* this reach. (a) On average, how much dispersion compensation per km is needed to maximize R, and what is the resulting R? (b) Repeat part (a), except assume that the absolute value of the residual average dispersion (i.e., with compensation) must be greater than 0.5 ps/(nm·km) across the spectral region to combat nonlinear impairments.

4.12 Consider an optical-bypass-enabled network where the nodes are arranged in a 3 × 6 grid. Each link is 1,000 km in length and the optical reach is 3,000 km. (a) What is the minimum number of regeneration sites that are required to allow all-to-all traffic, with complete flexibility in selecting the paths for routing (assume that regeneration can occur only in these sites)? (b) What is the minimum number of regeneration sites that are required if only minimum-hop paths can be used?

4.13 Consider a 3 × 6 grid network, where each link is 1,000 km in length and the optical reach is 3,000 km. (a) Partition this network into the fewest number of transparent islands. (More than one solution is possible.) Assume that all *inter-island* traffic is regenerated at a border node. Assuming one wavelength of bidirectional traffic between each pair of nodes, how many regenerations are required in your design? (b) With the same traffic, what is the minimum number of regenerations required in a selective-regeneration architecture?

4.14 How many faces are contained within the network of Fig. 4.7? Could the set of three islands shown in this figure have resulted from the island-forming heuristic described in Sect. 4.6.1? *For further research,* investigate improvements to the island-building algorithm. For example, can the design process be improved by considering the expected traffic between the nodes? Is it better to start with a face in the middle of the network and build outwards, or a face at the periphery of the network and build inwards?

4.15 In Sect. 4.6.1, it is assumed that the islands of transparency have nodal boundaries. There have also been designs that designate links, rather than nodes, as the boundary points; i.e., certain links belong to two different islands. Compare the link-based and node-based boundary architectures.

4.16 Consider a dynamic network, where regeneration requests arrive to a network region according to a Poisson process of 20 Erlangs. (a) Assume that the regenerations are split randomly with equal probability between two nodes. How many regenerator cards are needed at these two nodes to yield a blocking probability (due to no available regenerator cards) of less than 10^{-4}? (b) Second, assume that all regenerations occur on just one of the nodes. How many regenerator cards are needed at this one node to yield a blocking probability (due to no available regenerator cards) of less than 10^{-4}? (c) Based on the results in parts (a) and (b), which is the better strategy for minimizing the number of required regenerator cards?

4.17 Consider the nodal architecture of Fig. 4.10a, which allows transponders to be used for regeneration in any direction through the node. Assume that the node is equipped with a broadcast-and-select directionless ROADM, as shown in Fig. 2.12. Assume that the two transponders used for a particular regeneration are located on the same add/drop port of the ROADM. Are there any wavelength constraints imposed by this architecture for the incoming and outgoing wavelengths of the regenerated connection?

4.18 Consider two systems, where in System 1, back-to-back transponder cards are used for regeneration, whereas in System 2, regenerator cards are used for regeneration. Assume that the regenerator card cost is 70% of the cost of two transponder cards. Assume that connection requests that source/terminate (i.e., "true" add/drop) at a node arrive according to a Poisson process of 30 Erlangs, and that regeneration requests at the node arrive according to a Poisson process of 15 Erlangs. (a) Assume that Fig. 4.10a applies to System 1 and Fig. 4.12a applies to System 2. How many transponder cards are required in System 1 and how many transponder cards and how many regenerator cards are required in System 2 to reduce the blocking probability (i.e., due to no available cards) of a true add/drop and of a regeneration to 10^{-4} or less? How do the total costs of the cards in the two systems compare? (b) Next, assume that Fig. 4.11 applies to System 1, and Fig. 4.13 applies to System 2, and assume that ports on the edge switch cost 10% of a single transponder. The traffic and target blocking probabilities remain the same. How do the total costs of the two systems compare? (For part (b), it may be desirable to run a simulation to assist in determining the number of required transponders in System 1. Assume exponential holding times.)

4.19 If regenerator cards were not wavelength tunable, how many different parts would be needed to accommodate any possible combination of input wavelengths and output wavelengths, assuming W wavelengths per fiber? (Assume that the wavelengths in the two directions of a connection can be different.)

References

[ADZS12] W. T. Anderson, C. R. Davidson, H. Zhang, O. Sinkin, B. Bakhshi, A. Lucero, G. Mohs, A. Pilipetskii, N. S. Bergano, Coherent friendly dispersion map for direct detection transmission formats. In *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'12)*, Los Angeles, CA, 4–8 Mar 2012, Paper JW2 A.54

[AFLB13] N. Andriolli, S. Faralli, X. J. M. Leijtens, J. Bolk, G. Contestabile, Monolithically integrated all-optical regenerator for constant envelope WDM signals. J. Lightwave. Technol. **31**(2), 322–327 (15 Jan 2013)

[AKMC09] S. Azodolmolky, M. Klinkowski, E. Marin, D. Careglio, J. Pareta, I. Tomkos, A survey on physical layer impairments aware routing and wavelength assignment algorithms in optical networks. Comput. Netw. **53**(7), 926–944 (May 2009)

[BaKi02] P. Bayvel, R. Killey, in *Optical Fiber Telecommunications IV B*, ed. by I. Kaminow, T. Li. Nonlinear optical effects in WDM transmission, (Academic Press, San Diego, 2002), pp. 611–641

[BBSB09] A. Bononi, M. Bertolini, P. Serena, G. Bellotti, Cross-phase modulation induced by OOK channels on higher-rate DQPSK and coherent QPSK channels. J. Lightwave. Technol. **27**(18), 3974–3983 (15 Sept 2009)

[BRCM12] O. Bertran-Pardo, J. Renaudier, G. Charlet, H. Mardoyan, P. Tran, M. Salsi, S. Bigo, Overlaying 10 Gb/s legacy optical networks with 40 and 100 Gb/s coherent terminals. J. Lightwave. Technol. **30**(14), 2367–2375 (15 July 2012)

[BSCF13] B. G. Bathula, R. K. Sinha, A. L. Chiu, M.D. Feuer, G. Li, S. L. Woodward, W. Zhang, R. Doverspike, P. Magill, K. Bergman, Cost optimization using regenerator site concentration and routing in ROADM networks. In *Proceedings, 9th International Conference on Design of Reliable Communication Networks (DRCN'13)*, Budapest, Hungary, 4–7 Mar 2013, pp. 139–147

[CaCH04] H. S. Carrer, D. E. Crivelli, M. R. Hueda, Maximum likelihood sequence estimation receivers for DWDM lightwave systems. In *Proceedings, IEEE Global Telecommunications Conference* (*GLOBECOM'04)*, Dallas, TX, 29 Nov–3 Dec 2004, vol 2, pp. 1005–1010

[ChGn06] S. Chandrasekhar, A. H. Gnauck, Performance of MLSE receiver in a dispersion-managed multispan experiment at 10.7 Gb/s under nonlinear transmission. IEEE Photonics Technol. Lett. **18**(23), 2448–2450 (1 Dec 2006)

[ChOM10] F. Chang, K. Onohara, T. Mizuochi, Forward error correction for 100 G transport networks. IEEE Commun. Mag. **48**(3), S48–S55 (March 2010)

[Ciar12] E. Ciaramella, Wavelength conversion and all-optical regeneration: Achievements and open issues. J. Lightwave Technol. **30**(4), 572–582 (15 Feb 2012)

[CMSG04] T. J. Carpenter, R. C. Menendez, D. F. Shallcross, J. W. Gannett, J. Jackel, A. C. Von Lehmen, Cost-conscious impairment-aware routing. In *Proceedings, Optical Fiber Communication (OFC'04)*, Los Angeles, CA, 22–27 Feb 2004, Paper MF88

[Conr02] J. Conradi, in *Optical Fiber Telecommunications IV B*, ed. by I. Kaminow, T. Li, Bandwidth-efficient modulation formats for digital fiber transmission systems, (Academic Press, San Diego, 2002), pp. 862–901

[CRBT09] G. Charlet, J. Renaudier, P. Brindel, P. Tran, H. Mardoyan, O. Bertran Pardo, M. Salsi, S. Bigo, Performance comparison of DPSK, P-DPSK, RZ-DQPSK and coherent PDM-QPSK at 40 Gb/s over a terrestrial link. In *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'09)*, San Diego, CA, 22–26 Mar 2009, Paper JWA40

[CrLi08] K. Croussore, G. Li, Phase and amplitude regeneration of differential phase-shift keyed signals using phase-sensitive amplification. IEEE J. Sel. Top. Quantum Electron. **14**(3), 648–658 (May/June 2008)

[CSGJ03] T. Carpenter, D. Shallcross, J. Gannett, J. Jackel, A. Von Lehmen, Maximizing the transparency advantage in optical networks. In *Proceedings, Optical Fiber Communication (OFC'03)*, Atlanta, GA, 23–28 Mar 2003, Paper FA2

[CXBT02] N. Chi, L. Xu, K. S. Berg, T. Tokle, P. Jeppesen, All-optical wavelength conversion and multichannel 2R regeneration based on highly nonlinear dispersion-imbalanced loop mirror. IEEE Photonics Technol. Lett. **14**(11), 1581–1583 (Nov 2002)

[Desu94] E. Desurvire, *Erbium-Doped Fiber Amplifiers: Principles and Applications* (Wiley, New York, 1994)

[FoTC97] F. Forghieri, R. W. Tkach, A. R. Chraplyvy, in *Optical Fiber Telecommunications III A*, ed. by I. Kaminow, T. Koch Fiber nonlinearities and their impact on transmission systems, (Academic Press, San Diego, 1997), pp. 196–254

[GBSE10] S. Gringeri, B. Basch, V. Shukla, R. Egorov, T. J. Xia, Flexible architectures for optical transport nodes and networks. IEEE Commun. Mag. **48**(7), 40–50 (July 2010)

[GnJo97] A. H. Gnauck, R. M. Jopson, in *Optical Fiber Telecommunications III A*, ed. by I. Kaminow, T. Koch Dispersion compensation for optical fiber systems, (Academic Press, San Diego, 1997), pp. 162–195

[GrBX12] S. Gringeri, E. B. Basch, T. J. Xia, Technical considerations for supporting data rates beyond 100 Gb/s. IEEE Commun. Mag. **50**(2), S21–S30, (Feb 2012)

[GuKh98] S. Guha, S. Khuller, Approximation algorithms for connected dominating sets. Algorithmica. **20**(4), 374–387 (1998)

[Hans12] P. Hansen, *The case for coherent-transponder subsystems*, Lightwave, (Mar/Apr 2012), pp. 22–25

[Haus00] H. A. Haus, Noise figure definition valid from RF to optical frequencies. IEEE J. Sel. Top. Quantum Electron. **6**(2), 240–247 (Mar/Apr 2000)

[HBPS07] J. He, M. Brandt-Pearce, Y. Pointurier, S. Subramaniam, QoT-aware routing in impairment-constrained optical networks. In *Proceedings, IEEE Global Communications Conference* (*GLOBECOM'07)*, Washington, DC, 26–30 Nov 2007, pp. 2269–2274

[ILBK08] E. Ip, A. P. T. Lau, D. J. F. Barros, J. M. Kahn, Coherent detection in optical fiber systems. Opt. Express. **16**(2), 753–791 (21 Jan 2008)

[IpKa10] E. M. Ip, J. M. Kahn, Fiber impairment compensation using coherent detection and digital signal processing. J. Lightwave Technol. **28**(4), 502–519 (15 Feb 2010)

[Isla02] M. Islam, Raman amplifiers for telecommunications. IEEE J. Selec. Top. Quantum Electron. **8**(3), 548–559 (May/Jun 2002)

[KaAr04] E. Karasan, M. Arisoylu, Design of translucent optical networks: Partitioning and restoration. Photonic Netw. Commun. **8**(2), 209–221 (Mar 2004)

[KaSG04] G. S. Kanter, A. K. Samal, A. Gandhi, Electronic dispersion compensation for extended reach. In *Proceedings, Optical Fiber Communication (OFC'04)*, Los Angeles, CA, 22–27 Feb 2004, Paper TuG1

[KBSP10] J. Kakande, A. Bogris, R. Slavík, F. Parmigiani, D. Syvridis, P. Petropoulos, D. J. Richardson, First demonstration of all-optical QPSK signal regeneration in a novel multi-format phase sensitive amplifier. In *Proceedings, European Conference on Optical Communication (ECOC'10)*, Turin, Italy, 19–23 Sept 2010

[KTMT05] P. Kulkarni, A. Tzanakaki, C. M. Machuka, I. Tomkos, Benefits of Q-factor based routing in WDM metro networks. In *Proceedings, European Conference on Optical Communication (ECOC'05)*, vol 4, Glasgow, Scotland, 25–29 Sept 2005 pp. 981–982

[Kurt93] C. Kurtzke, Suppression of fiber nonlinearities by appropriate dispersion management. IEEE Photonics Technol. Lett. **5**(10), 1250–1253 (Oct 1993)

[LeJC04] J. Leuthold, J. Jaques, S. Cabot, All-optical wavelength conversion and regeneration. In *Proceedings, Optical Fiber Communication (OFC'04)*, Los Angeles, CA, 22–27 Feb 2004, Paper WN1

[LLBB03] O. Leclerc, B. Lavigne, E. Balmefrezol, P. Brindel, L. Pierre, D. Rouvillain, F. Seguineau, Optical regeneration at 40 Gb/s and beyond. J. Lightwave Technol. **21**(11), 2779–2790 (Nov 2003)

[Mats12] M. Matsumoto, Fiber-based all-optical signal regeneration. IEEE J. Selec. Top. Quantum Electron. **18**(2), 738–752 (Mar/Apr 2012)

[MBLA08] A. Morea, N. Brogard, F. Leplingard, J.-C. Antona, T. Zami, B. Lavigne, D. Bayart, QoT function and A* routing: An optimized combination for connection search in translucent networks. J. Opt. Netw. **7**(1), 42–61 (Jan 2008)

[MKCV10] K. Manousakis, P. Kokkinos, K. Christodoulopoulos, E. Varvarigos, Joint online routing, wavelength assignment and regenerator allocation in translucent optical networks. J. Lightwave Technol. **28**(8), 1152–1163 (15 Apr 2010)

[MORC05] J. McNicol, M. O'Sullivan, K. Roberts, A. Comeau, D. McGhan, L. Strawczynski, Electrical domain compensation of optical dispersion. In *Proceedings, Optical Fiber Communication (OFC'05)*, Anaheim, CA, 6–11 Mar 2005, Paper OThJ3

[MSMK12] T. Mizuochi, T. Sugihara, Y. Miyata, K. Kubo, K. Onohara, S. Hirano, H. Yoshida, T. Yoshida, T. Ichikawa, Evolution and status of forward error correction. In *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'12)*, Los Angeles, CA, 4–8 Mar 2012, Paper OTu2 A.6

[MSTT07] G. Markidis, S. Sygletos, A. Tzanakaki, I. Tomkos, Impairment aware based routing and wavelength assignment in transparent long haul networks. In *Proceedings, Conference on Optical Network Design and Modeling (ONDM'07)*, Athens, Greece, 29–31 May 2007, pp. 48–57

[OSul08] M. O'Sullivan, Expanding network applications with coherent detection. In *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'08)*, San Diego, CA, 24–28 Feb 2008, Paper NWC3

[PaVL10] P. G. Patki, M. Vasilyev, T. I. Lakoba, Multichannel all-optical regeneration. *IEEE Photonics Society Summer Topicals*. 19–21 July 2010, Playa del Carmen, Mexico, 172–173, Paper WC2.2

[Pers73] S. D. Personick, Receiver design for digital fiber optic communications systems. Bell System Technical Journal, **52**(6), 843–886 (July/Aug 1973)

[PoNa97] C. D. Poole, J. Nagel, in *Optical Fiber Telecommunications III A*, ed. by I. Kaminow, T. Koch. Polarization effects in lightwave systems, (Academic Press, San Diego ,1997), pp. 114–161

[PPPR12] F. Parmigiani, L. Provost, P. Petropoulos, D. J. Richardson, W. Freude, J. Leuthold, A. D. Ellis, I. Tomkos, Progress in multichannel all-optical regeneration based on fiber technology. IEEE J. Selec. Top. Quantum Electron. **18**(2) 689–700, (Mar/Apr 2012)

[Rahb12] A.G. Rahbar, Review of dynamic impairment-aware routing and wavelength assignment techniques in all-optical wavelength-routed networks. IEEE Communications Surveys & Tutorials. **14**(4), 1065–1089, Fourth Quarter, (2012)

[RaSS09] R. Ramaswami, K. N. Sivarajan, G. Sasaki, *Optical Networks: A Practical Perspective*, 3rd edn. (Morgan Kaufmann Publishers, San Francisco, 2009)

[Robe11] K. Roberts, 100G—Key technology enablers of 100Gbit/s in carrier networks. In *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'11)*, Los Angeles, CA, 6–10 Mar 2011, Paper NWA1

[RoSt02] K. Rottwitt, A. Stentz, in *Optical Fiber Telecommunications IV A*, ed. by I. Kaminow and T. Li. Raman amplification in lightwave communication systems, (Academic Press, San Diego, 2002), pp. 213–258

[Sale98a] A. A. M. Saleh, Islands of transparency—an emerging reality in multiwavelength optical networking. In *Proceedings, IEEE/LEOS Summer Topical Meeting on Broadband Optical Networks and Technologies*, Monterey, CA, 20–24 July 1998, p. 36

[Sale00] A. A. M. Saleh, Transparent optical networking in backbone networks. In *Proceedings, Optical Fiber Communication (OFC'00)*, Baltimore, MD, 7–10 Mar 2000, Paper ThD7

[SaSi06] A. A. M. Saleh, J. M. Simmons, Evolution toward the next-generation core optical network. J. Lightwave Technol. **24**(9), 3303–3321, (Sept 2006)

[Savo07] S. J. Savory, Coherent detection—Why is it back?. In *Proceedings, 20th Annual Meeting of the IEEE LEOS*, Lake Buena Vista, FL, 21–25 Oct 2007, Paper TuH1

[SFGE12] S. Sygletos, P. Frascella, F. C. Garcia Gunning, A.D. Ellis, Multi-wavelength regeneration of phase encoded signals based on phase sensitive amplifiers. In *Proceedings, International*

Conference on Transparent Optical Networks (ICTON'12), United Kingdom, 2–5 July 2012, Paper We.B1.4

[ShSS11] G. Shen, Y. Shen, H. P. Sardesai, Impairment-aware lightpath routing and regenerator placement in optical transport networks with physical-layer heterogeneity. J. Lightwave Technol. **29**(18), 2853–2860 (15 Sept 2011)

[ShST09] G. Shen, W. V. Sorin, R. S. Tucker, Cross-layer design of ASE-noise-limited island-based translucent optical networks. J. Lightwave Technol. **27**(11), 1434–1442 (1 June 2009)

[Simm05] J.M. Simmons, On determining the optimal optical reach for a long-haul network. J. Lightwave Technol. **23**(3), 1039–1048 (Mar 2005)

[SiSa07] J. M. Simmons, A. A. M. Saleh, Network agility through flexible transponders. IEEE Photonics Technol. Lett. **19**(5) 309–311 (1 Mar 2007)

[Tayl10] M. G. Taylor, Algorithms for coherent detection. In *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'10)*, San Diego, CA, 21–25 Mar 2010, Paper OThL4

[TCFG95] R.W. Tkach, A. R. Chraplyvy, F. Forghieri, A. H. Gnauck, R. M. Derosier, Four-photon mixing and high-speed WDM systems. J. Lightwave Technol. **13**(5), 841–849 (May 1995)

[TkCh94] R.W. Tkach, A. R. Chraplyvy, Dispersion and nonlinear effects in lightwave systems. In *Proceedings, 7th Annual Meeting of the IEEE LEOS*, vol 1, Boston, MA, 31 Oct–3 Nov 1994, pp. 192–193

[VaAa87] P. J. M. van Laarhoven, E. H. L. Aarts, *Simulated Annealing: Theory and Applications*, (D. Reidel Publishing Co., Boston, 1987)

[VSAJ09] D. van den Borne, V. A. J. M. Sleiffer, M. S. Alfiad, S. L. Jansen, T. Wuth, POLMUX-QPSK modulation and coherent detection: The challenge of long-haul 100G transmission. In *Proceedings, European Conference on Optical Communication (ECOC'09)*, Vienna, Austria, 20–24 Sept 2009, Paper 3.4.1

[Way12] W. I. Way, Optimum architecture for M × N multicast switch-based colorless, directionless, contentionless, and flexible-grid ROADM. In *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'12)*, Los Angeles, CA, 4–8 Mar 2012, Paper NW3F.5

[Winz12] P. J. Winzer, High-spectral-efficiency optical modulation formats. J. Lightwave Technol. **30**(24), 3824–3835 (15 Dec 2012)

[YaRa05a] X. Yang, B. Ramamurthy, Dynamic routing in translucent WDM optical networks: The intradomain case. J. Lightwave Technol. **23**(3), 955–971 (Mar 2005).

[YaWa08] Y. Yang, J. Wang, Design guidelines for routing metrics in multihop wireless networks. In *Proceedings, IEEE INFOCOM 2008*, Phoenix, AZ, 15–17 Apr 2008, pp. 1615–1623

[YeKa03] E. Yetginer, E. Karasan, Regenerator placement and traffic engineering with restoration in GMPLS networks. Photonic Netw. Commun. **6**(2), 139–149 (Sept 2003)

[ZhSB12] J. Zhao, S. Subramaniam, M. Brandt-Pearce, Cross-layer RWA in translucent optical networks. In *Proceedings, IEEE International Conference on Communications (ICC'12)*, Ottawa, Canada, 10–15 June 2012, pp. 3079–3083

Chapter 5
Wavelength Assignment

5.1 Introduction

Wavelength assignment is an integral part of the network planning process in optical-bypass-enabled networks. Its need arises from the wavelength continuity property of optical-bypass elements, where a connection that traverses a node all-optically must enter and exit the node on the same optical frequency. Thus, the wavelengths that are in use on one link may have ramifications for the wavelengths that can be assigned on other links. Effective wavelength assignment strategies must be utilized to ensure that wavelength contention is minimized.

Wavelength assignment is tightly coupled to the routing process, as the selection of the route determines the links on which a free wavelength must be found. The two processes are often referred to together as the routing and wavelength assignment (RWA) problem. However, regeneration, when needed, is just as critical to the process. While the earliest visions of optical-bypass-enabled networks assumed that they would be completely all-optical, with no regeneration, this has not turned out to be the case in practice, especially in regional and backbone networks. The presence of regeneration has a significant impact on wavelength assignment because it allows for a change in wavelength. This is explored in Sect. 5.2.

The previous two chapters looked at the sequential processes of selecting a route followed by selecting regeneration locations. Wavelength assignment can be treated as the third sequential step in the planning process. With this multistep approach, there is no guarantee that the route found will be amenable to a feasible wavelength assignment. Another option is to perform RWA as a single step; this is more complex, but any route that is found is guaranteed to have a feasible wavelength assignment. Multistep and single-step RWA are discussed in Sects. 5.3 and 5.4, respectively.

Wavelength assignment algorithms were one of the first aspects of all-optical networks that were researched heavily. The result is an array of well-studied algorithms that represent different performance/complexity operating points. Some of the strategies that have proved effective in actual network designs are presented in Sect. 5.5. In scenarios where many demands are added at one time to the network,

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4_5, © Springer International Publishing Switzerland 2014

the order in which the connections are assigned wavelengths may affect the performance of the wavelength assignment scheme. Effective ordering strategies are discussed in Sect. 5.6.

One interesting aspect of wavelength assignment relates to the two directions of a bidirectional connection. Scenarios where it may be beneficial to assign different wavelengths to the two directions are covered in Sect. 5.7. Another challenge encountered in some systems is the nonuniformity of the optical reach across the wavelengths in the spectrum. This may be due to, for example, the dispersion properties of the fiber plant. The ramifications for wavelength assignment are discussed in Sect. 5.8.

Another potential issue is dealing with impairments that arise due to adjacent wavelengths propagating on a fiber. In many systems, the impairments are small enough that they can be accounted for by a small penalty that is encompassed in the system margin. In systems with more significant inter-wavelength impairments, it may be desirable to calculate the effect of the impairments more precisely rather than using a worst-case penalty in all scenarios. The problem can be challenging because it depends on the current state of the network, e.g., which of the wavelengths are already carrying live traffic. Furthermore, it is important to not engineer a design so precisely for the current network state that the addition of a new demand causes the performance of existing connections to fall below an acceptable threshold. A particular scenario of interest regarding inter-wavelength impairments arises in mixed line-rate systems where wavelengths of different modulation formats may co-propagate on the same fiber. For some combinations of modulation formats, the performance penalty is significant enough that wavelength assignment schemes that attempt to segregate the various formats are desirable. The topic of inter-wavelength impairments is covered in Sect. 5.9.

Optical-bypass-enabled networks tend to be "closed," where one vendor supplies the amplifiers, reconfigurable optical add/drop multiplexers (ROADMs), and transponders, and all wavelengths are generated from within the system. However, there has been growing interest in "opening up" such systems and providing support for *alien wavelengths*. These may be, for example, wavelengths that originate in the client layer and all-optically enter the wavelength-division multiplexing (WDM) system, or wavelengths that have been all-optically routed between systems from different vendors. The performance of an alien wavelength is unlikely to match that of the wavelengths that are native to the system. This likely has ramifications for the wavelength assignment process. Support for alien wavelengths is addressed in Sect. 5.10.

As this introduction indicates, there are numerous challenges with wavelength assignment that must be addressed. However, many studies have been performed that show the loss of network efficiency due to wavelength contention is generally small, assuming good algorithms are used; live networks exist to further bear this out. Nevertheless, the effect of wavelength contention on the performance of the network continues to be debated in the industry. Section 5.11 presents further results that demonstrate just a small loss of efficiency due to wavelength contention, in both a backbone network and a metro-core network, and offers some insight as to why this is so.

Wavelength assignment is especially important in some optical protection schemes, where the primary and backup paths may be required to be carried on the same wavelength; this is discussed further in Chap. 7. In this chapter, one can assume that any protection is client-based 1 + 1, where there is a primary path and a dedicated backup path, and the network client (e.g., an Internet Protocol (IP) router) determines which of the two paths to use. With this type of protection, wavelengths can be assigned independently to the two paths.

It should be pointed out that the development of cost-effective all-optical wavelength converters is an active area of research. This technology would allow the wavelength of a signal to be changed without requiring optical-electrical-optical (O-E-O) conversion, such that optical bypass would not necessarily imply wavelength continuity. The criticality of wavelength assignment in the overall network planning process would be somewhat abated with the commercial deployment of this technology. However, it is unlikely that the cost of such technology would allow it to be deployed at *all* nodes on *all* wavelengths. Moreover, adding wavelength converters on every wavelength would be antithetical to the optical-bypass paradigm of reducing the amount of equipment in the network. Furthermore, all-optical wavelength converters suffer from many of the same hurdles as all-optical regenerators (see Sect. 4.7.3). Thus, wavelength assignment will remain an important step in the network planning process for optical-bypass-enabled networks for the foreseeable future.

Note that in O-E-O networks, wavelengths can be assigned independently on each link, and thus wavelength contention and wavelength assignment are not major issues.

5.2 Role of Regeneration in Wavelength Assignment

If a demand is carried all-optically from source to destination, then the same wavelength must be used on all links of the path (assuming all-optical wavelength conversion is not available). Consider the connection between Nodes A and Z shown in Fig. 5.1a, which is routed on seven links. If this is an all-optical connection, then one needs to find a wavelength that is free on all seven links. In Fig. 5.1b, this same connection is regenerated at Nodes C and E, thereby creating three subconnections: A-C, C-E, and E-Z. (As a reminder, the term *subconnection* refers to the portions of a connection that fall between two regeneration points or between an end point and a regeneration point.) In this scenario, one needs to find a free wavelength for each of the subconnections, where in most cases there is no requirement that the wavelengths be the same for each subconnection. Finding a free wavelength on a subconnection is clearly an easier problem than finding a free wavelength on the whole end-to-end connection. Thus, the presence of regeneration potentially engenders greater wavelength assignment flexibility.

The importance of fully tunable transponders and regenerators with respect to the wavelength assignment process should be readily apparent. For example, in

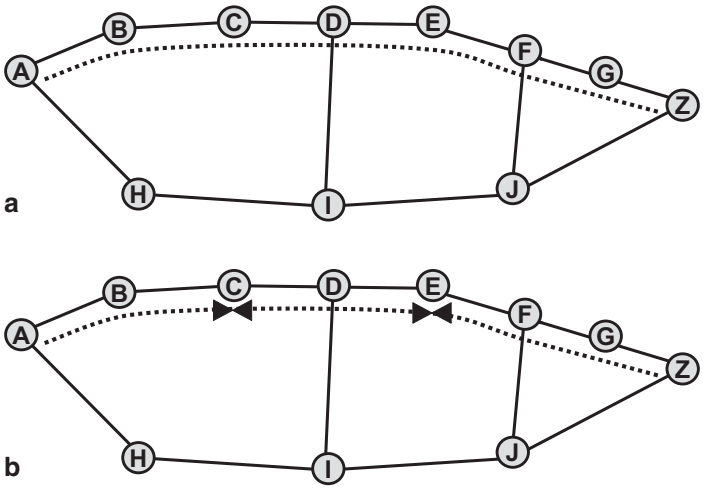


Fig. 5.1 **a** It is necessary to find a wavelength that is free along each of the links of the path between Nodes *A* and *Z*. **b** If regeneration occurs at Nodes *C* and *E,* different wavelengths can be assigned to the three resulting subconnections

an architecture where two transponders are patch-cabled together for regeneration (as is shown in Fig. 4.10), fixed-tuned transponders imply that the wavelengths assigned to consecutive subconnections are determined by the wavelengths that are paired together at the regeneration node. An even more significant constriction arises in systems where the regenerator cards require that the wavelengths on the input and output sides of the regenerator (with respect to a given direction of the connection) be the same. Using such regenerator cards prohibits wavelength conversion at the regeneration node. With this type of equipment, all three subconnections in Fig. 5.1b would need to be assigned the same wavelength, which reduces the flexibility of the wavelength assignment process.

In Chap. 4, O-E-O conversion was discussed in the context of regeneration and optical reach. However, there are functions other than signal regeneration that require O-E-O conversion. For example, in a network with subrate traffic, it is necessary to bundle multiple connections together to utilize more fully the capacity of a wavelength. As is described in Chap. 6, this bundling process is most effective when the traffic can be groomed at various nodes in the network. The grooming process typically is accomplished in the electrical domain, thereby requiring O-E-O conversion. A second driver is shared protection, which is covered in Chap. 7. To effectively share protection bandwidth, O-E-O conversion may be required at the "sharing" points. As these examples indicate, O-E-O conversion is not just a function of optical reach. Thus, even in small metro networks where the path distances do not warrant regeneration based on the optical reach, not every connection may be carried all-optically end-to-end.

Any O-E-O event offers the opportunity to essentially wavelength convert "for free" (again, assuming the equipment permits this flexibility). Wavelength

assignment algorithms should take advantage of this freedom. For simplicity, the remainder of the chapter refers to connections being broken into subconnections due to regeneration; however, keep in mind that it may be due to factors other than optical reach, as discussed above.

5.3 Multistep RWA

When network planning is treated as a multistep process, a route is selected for a connection, the connection is broken into subconnections, if necessary, and each of the subconnections is assigned a wavelength. It is possible that a feasible wavelength assignment will not be found for one or more of the subconnections, requiring some of the steps to be repeated.

Minimizing the occurrence of wavelength contention requires good routing strategies. The discussion here assumes alternative-path routing is used, possibly in combination with dynamic routing (see Sect. 3.5). As presented in Chap. 3, one first generates a set of candidate paths for a particular source/destination combination, where the candidate paths are chosen to minimize cost and to provide good load balancing in the network. As demands are added, the current state of the network is considered when selecting one of the candidate paths to use for a particular demand request. A good strategy is to select the least-loaded candidate path, such that the minimum number of wavelengths that are free on each link of the selected path is maximized. This does not guarantee that the *same* wavelength will be free along the various links; however, it generally improves the chances of finding a feasible wavelength assignment.

If demands are added one at a time to the network, then the algorithm can actually consider which particular wavelengths are free when selecting one of the candidate paths. Thus, the feasibility of the candidate path can be considered when selecting a path for the demand. When multiple demands are added at once to the network, as is often the case in long-term planning, one option is to fully process each demand individually, i.e., route, regenerate, and assign a wavelength for one demand before moving on to the next. This methodology is similar to adding demands one at a time, and allows one to consider the actual free wavelengths on a path when selecting a route. A second option for handling multiple demands, which is typically more advantageous, is to perform the routing and regeneration for all demands *before* the start of the wavelength assignment process. With this strategy, the wavelength assignment algorithm has full knowledge of exactly how many subconnections are routed on each link, and can use this information to better optimize the assignment process. However, this methodology precludes consideration of which wavelengths are free when choosing a path for a demand; i.e., only the current load on each link, due to the demands that have already been routed, is known. (Note that global optimization techniques such as linear programming can process all demands at once while implicitly considering the free wavelengths on a link; see Sect. 5.4.3.)

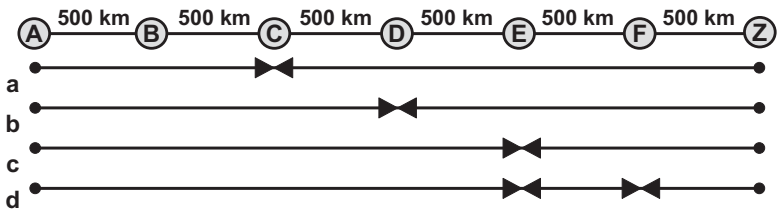


Fig. 5.2 With an optical reach of 2,000 km, one regeneration is required on the path between Nodes *A* and *Z*. The site of the regeneration determines the resulting subconnections, which can affect wavelength assignment. The regeneration is at Node *C* (**a**), Node *D* (**b**), and Node *E* (**c**). A second regeneration is added to break the wavelength contention that is assumed to exist on Links *EF* and *FZ* (**d**)

The multistep methodology of treating routing, regeneration, and wavelength assignment separately usually performs well in practice. However, when the network is very heavily loaded, wavelength contention may occur where subconnections are created for which there are no feasible wavelength assignments. Four strategies for ameliorating this situation are discussed here.

5.3.1 Alleviating Wavelength Contention

First, in a network requiring regeneration, different regeneration sites for a connection potentially can be selected. Consider the connection between Nodes A and Z shown in Fig. 5.2. Assume that the optical reach is 2,000 km, such that one regeneration is required for the connection. The possible locations for the regeneration are Nodes C, D, or E, as shown in Fig. 5.2a, b, and c, respectively. Each regeneration choice creates a different set of subconnections; e.g., regenerating at Node C creates the A-C and C-Z subconnections. If the initial location selected for regeneration leads to wavelength contention, then the planning algorithm can consider a different location. For example, assume that with regeneration at Node C, a free wavelength can be found on the A-C subconnection but not on the C-Z subconnection. One can consider moving the regeneration to either Node D or Node E, to generate different subconnections. Note that selecting a different regeneration location does not incur any additional cost; i.e., the total number of regenerations remains the same.

If this strategy does not work, or if there are no regenerations, then one can consider using a different candidate path. In realistic networks, when wavelength assignment fails, it is typically because of a small number of heavily loaded links. Moving some demands away from these links can be enough to alleviate the wavelength contention problems. The simplest means of achieving this is to pick a different candidate path for some of the demands, where the originally selected path included one or more “bad” links, and where the new path does not include any. Ideally, the newly selected candidate path meets the minimum amount of regeneration possible for the demand so that no additional cost is incurred.

If not enough demands have candidate paths that avoid the bad links, then one can make use of dynamic routing. The links that are causing the problems in the wavelength assignment process can be temporarily eliminated from the topology before calling the shortest-path algorithm. There is no guarantee that this will find a feasible path. Even if a path is found, it may be very circuitous such that it requires several additional regenerations, making it undesirable.

If, after applying the above three strategies, wavelength assignment is still not successful, then one of the candidate paths can be selected and any residual wavelength contention can be alleviated by adding in extra regeneration. For example, in Fig. 5.2, assume that the wavelength contention stems from there being no common free wavelengths on Links EF and FZ. An extra regeneration can be added at Node F, as shown in Fig. 5.2d. This breaks the interdependence between Links EF and FZ for this connection, allowing wavelengths to be assigned independently on these links. Adding in just a few extra regenerations in a network can be quite effective in alleviating wavelength contention. With very little additional cost, the network utilization can be markedly increased. This was demonstrated in Van Parys et al. [VAAD01] and Simmons [Simm02]. It is further explored in Sect. 5.11.

Note that if the network is so full that wavelength contention is causing a great deal of extra regenerations to be added, then it is probably time to add additional capacity to the network.

5.4 One-Step RWA

Rather than relying on techniques to handle infeasible wavelength assignment scenarios when they arise, it is natural to consider methodologies where routing and wavelength assignment are treated as a single problem to ensure feasibility from the start. Various one-step RWA methodologies are discussed below, all of which impose additional processing and/or memory burdens. When a demand is added to a network that is not heavily loaded, the multistep process should have little problem finding a feasible route and wavelength assignment. Thus, under these conditions, the multistep process is favored, as it is usually faster. However, under heavy load, using a one-step methodology can provide a small improvement in performance, as is investigated quantitatively in Sect. 5.11. Furthermore, under heavy load, some of the one-step methodologies may be more tractable, as the scarcity of free wavelengths should lead to lower complexity.

5.4.1 Topology Pruning

One of the earliest advocated one-step algorithms starts with a particular wavelength and reduces the network topology to only those links on which this wavelength is available. The routing algorithm (e.g., a shortest-path algorithm) is run on this

pruned topology. If no path can be found, or the path is too circuitous, another wavelength is chosen and the process run through again on the correspondingly pruned topology. The process is repeated with successive wavelengths until a suitable path is found. With this combined approach, it is guaranteed that there will be a free wavelength on any route that is found. If a suitable path cannot be found after repeating the procedure for all of the wavelengths, the demand is blocked.

In a network with regeneration, using this combined routing and wavelength assignment procedure makes the problem unnecessarily more difficult because it implicitly searches for a wavelength that is free along the whole length of the path. As discussed above, it is necessary to find a free wavelength only along each subconnection, not along the end-to-end connection. One variation is to select ahead of time where the regenerations are likely to occur along a connection, and apply this combined routing and wavelength assignment approach to each expected subconnection individually. However, the route that is ultimately found could be somewhat circuitous and require regeneration at different sites than where was predicted, so that the process may need to be run through again. Overall, this strategy is less than ideal.

5.4.2 Reachability Graph Transformation

A more direct unified RWA approach is to create a transformed graph whenever a new demand request arrives, where the transformation is similar to what was discussed in Sect. 3.6.2. In long-term planning, every network node appears in the transformed graph; in real-time planning, only nodes with available regeneration equipment, plus the source and the destination, are added to the transformed graph. A link is added between a pair of nodes in the transformed graph only if there exists a regeneration-free path between the nodes in the true topology, *and* there exists a wavelength that is free along the path. (Even if there are multiple regeneration-free paths between a node pair, or multiple wavelengths free on a path, at most one link is added between a node pair.) We refer to the transformed graph as the reachability graph. (In Sect. 3.6.2, which dealt with real-time routing, the requirement of a free wavelength was not enforced when creating the reachability graph.)

An example of such a transformation is shown in Fig. 5.3. The true topology is shown in Fig. 5.3a, where the wavelengths that are assumed to still be available on a link are shown. The optical reach is assumed to be 2,000 km. Additionally, it is assumed that the transformation is being performed as part of a long-term planning exercise, so that the available regeneration equipment at a node is not a factor. The demand request is assumed to be between Nodes A and Z. The corresponding reachability graph is shown in Fig. 5.3b. All of the original links appear in this graph except for Link AF, which has no available wavelengths. In addition, Links AC, AD, and BD are added because the respective associated paths, A-B-C, A-B-C-D, and B-C-D, are less than 2,000 km and have a free wavelength (i.e., on each of these paths λ_6 is free). Note that no link is added to represent the path E-F-G even though

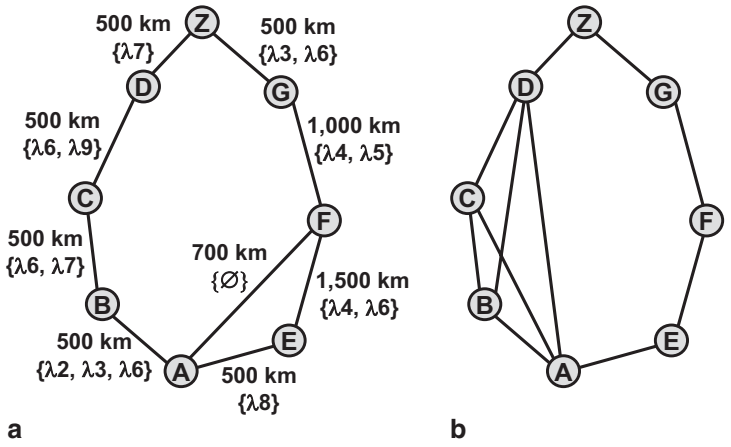


Fig. 5.3 **a** The true network topology where it is assumed that the optical reach is 2,000 km. The wavelengths listed next to each link are the wavelengths that are assumed to be free on the link. **b** The reachability graph, where a link is added between a node pair if there is a regeneration-free path between the nodes with at least one available wavelength along the path

λ_4 is available on this path because the path distance is 2,500 km, which is greater than the optical reach.

In a real network with many nodes and wavelengths, creating this reachability graph can potentially be a time-consuming procedure. For each pair of nodes, say Nodes X and Y, that possibly have a regeneration-free path between them, a search is performed to find a regeneration-free path where some wavelength is available along the whole path. To do this, one could use the topology-pruning approach described in Sect. 5.4.1, where the true topology is pruned down to just those links that have a particular wavelength free. A shortest-path algorithm is run on the pruned topology to search for a regeneration-free path between Nodes X and Y. The process is repeated for each wavelength, until a regeneration-free path is found. (This one-step approach is better suited for generating the links in the reachability graph than it is for finding an end-to-end path; however, it may still be slow.) Alternatively, one can run a *K*-shortest-paths algorithm on the true topology, where *K* is large enough such that any regeneration-free path between Nodes X and Y is found. The paths are then checked for a free wavelength. If a regeneration-free path with a free wavelength is found, a link is added between Nodes X and Y in the reachability graph. The algorithm should keep track of the free wavelength associated with each link in the reachability graph.

To reduce the time to form the reachability graph, one can maintain a list of a few regeneration-free paths for each node pair that has at least one such regeneration-free path between them. (This is no worse than storing a set of candidate paths for alternative-path routing.) If a free wavelength can be found on any of the paths, then a link is added in the reachability graph for the corresponding node pair. This eliminates multiple calls to a shortest-path routine every time a graph transformation is

needed. This method is not guaranteed to find a feasible regeneration-free path if one exists, although in practice it usually does.

Once the reachability graph is formed, a shortest-path algorithm is run from the demand source to the demand destination to find the path in this graph with the fewest hops, where each hop corresponds to a subconnection in the true topology. If a path is found, then it is guaranteed to have the fewest number of feasible regenerations, and each resulting subconnection is guaranteed to have an available wavelength. In the example of Fig. 5.3, path A-D-Z is found, which corresponds to the subconnections A-B-C-D and D-Z in the true topology. These subconnections are assigned λ_6 and λ_7, respectively.

One caveat with this one-step method should be noted. The path produced may include a link that is common to more than one subconnection comprising the path, and where the same free wavelength is associated with the overlapping subconnections. Assuming there is just one fiber pair per link, this would result in the same wavelength being assigned multiple times on a fiber, which is not permitted. A few strategies can be attempted to remedy the situation. Assume that there are two subconnections that overlap. A different free wavelength could be searched for on either of the subconnections. If that is not successful, one of the subconnections can be routed differently, where the subconnection path has the same endpoints, but the overlapping link is avoided. If this is also not successful, then the link associated with one of the overlapping subconnections can be removed from the reachability graph, and another search performed to find a new end-to-end path.

This type of problem is more likely to occur if the regeneration-free paths represented by the links in the reachability graph are "meandering." This can be minimized by using the methodology described above where a small number of fairly direct, regeneration-free paths are maintained for each nodal pair. Only these paths are considered when forming the reachability graph. This methodology was used in the study that is reported on in Sect. 5.11, and no problems with overlapping subconnections were encountered.

If *non-tunable* transponders or regenerator cards are used, then a more complex transformation may be needed with real-time planning to ensure a feasible path is found using one-step RWA. For example, consider a reachability graph that includes only those nodes with regeneration equipment, plus the source and destination. A link is added between a pair of nodes (say A and B) in the reachability graph for *each* wavelength λ_i if there is a regeneration-free path between Nodes A and B on which λ_i is available, *and* there is an available regenerator card (or transponder) at both Nodes A and B with wavelength λ_i. Assume that such a link in the reachability graph is called *ABi*. *Turn constraints* are imposed to ensure that a path can go from link *ABi* to *BCj,* for some nodes A, B, and C, and some wavelengths *i* and *j,* only if there is a regenerator card (or transponder pair) at Node B that interconnects wavelengths *i* and *j*. A shortest-path algorithm that enforces turn constraints is run on the reachability graph (turn constraints were discussed in Sect. 3.6.1). (If the nodes are equipped with non-directionless ROADM-MDs, then additional constraints must be enforced where the regenerators/transponders are tied to certain links.) The number of links in the reachability graph may be quite large if the number of wavelengths

in the network is large. However, such detailed modeling is generally only needed when the network is heavily loaded and feasible paths are difficult to find. At that stage, it is expected that the number of free wavelengths and regenerators are small so that the reachability graph is not excessively large.

5.4.3 Flow-Based Methods

Global optimization techniques can be applied to the one-step RWA problem as well. As discussed in Sect. 3.9, integer linear programming (ILP) approaches are generally not scalable for networks of practical size; however, linear programming (LP) techniques may be feasible. Relaxing the integrality constraints enables more rapid convergence. Various perturbation and rounding techniques are applied to ultimately produce a (possibly nonoptimal) integer solution.

Similar strategies as described in Sect. 3.9 for routing can be used for combined routing and wavelength assignment. RWA for a set of demands can be modeled as a flow problem, where additional variables and constraints are needed to enforce wavelength continuity [OzBe03, ChMV08]. An integer solution to the problem is typically desired, which corresponds to routing each demand over just one path, using a single wavelength on a link. The additional complexity of wavelength assignment adds to the run time. One approach to speed up the process is to input a set of possible paths that can be followed by a demand between any given source and destination. This is analogous to calculating a set of candidate paths for alternative-path routing; the strategies described in Sect. 3.4 for generating the path set can be applied here as well. Restricting the LP to a set of candidate paths, as opposed to allowing the LP to freely select the paths, may result in a less than optimal solution; however, with a good choice of candidate paths, the effect should be small. Another benefit to preselecting the paths is that the regeneration sites can be selected up front. This allows the wavelength continuity constraint to be specified on a per-subconnection basis rather than requiring that wavelength continuity be enforced end-to-end.

A cost function is used to encourage load balancing. Also, the cost function is input as a piecewise linear function with integer breakpoints as another means of pushing the LP towards an integer solution. (This is similar to what was described in Sect. 3.9 for the routing-only problem.)

The results of Christodoulopoulos et al. [ChMV08] indicate that many of the RWA instances tested resulted in integer solutions using just perturbation techniques (see Sect. 3.9). The remaining instances required additional approximations. The reported run times are just a few seconds, but the test network is small, with few demands. The same technique is also used in portions of Christodoulopoulos et al. [ChMV10]. The same network is used, but the number of demands is quadrupled, resulting in about a 20-fold increase in run time. Further studies using larger networks are desirable.

As suggested earlier, using a one-step approach such as an LP methodology may be more expedient when adding a set of demands to a highly loaded network. At that stage, there are few available wavelengths on each link, such that the solution space is much smaller. This should allow the LP to converge more quickly.

It is interesting to compare the results of the one-step LP-based RWA approach to those of a multistep approach, where an LP methodology is used just for the routing portion and a commonly used graph-coloring algorithm is used for wavelength assignment. The results of Christodoulopoulos et al. [ChMV08] showed that the performances of the two approaches are similar (depending on the cost functions used in the LPs), indicating that good results can be obtained using the simpler multistep approach. Furthermore, the run time of the multistep approach was an order of magnitude faster.

Another flow-based one-step RWA methodology, designed for scenarios where one demand is added at a time, is investigated in Gurzi et al. [GNCS09]. Here, the wavelength continuity constraint is taken into account by forming a transformed graph with multiple wavelength layers. Assume that W wavelengths are supported on a fiber. Then each link in the true topology is represented by up to W links in the transformed graph, one for each wavelength that is still available. Each node is represented by W nodes, except at nodes with wavelength conversion ability, where the links converge on just a single node. A maximum-flow algorithm is run on the transformed graph, for the source/destination pair of interest (this is not a multicommodity flow problem and is thus easier to solve). From the possible maximum-flow paths, one is selected for the new demand. (Note that the maximum-flow algorithm will produce integer solutions.) The blocking probability of this strategy was compared to various multistep RWA schemes. In the region of interest to most carriers, say below 1 % blocking, the maximum-flow scheme performed the same as the best multistep scheme. It was only at blocking rates of about 5 % where the maximum-flow scheme outperformed the best multistep method, although the differences were still not large. This reinforces the view that multistep methods can perform well, and that one-step methodologies are best suited for very high load scenarios.

5.4.4 ILP-Based Ring RWA

Although ILP formulations have generally been considered too slow for practical RWA, a scalable ILP methodology has been proposed for ring topologies [YeLR11]. Attempts to find a scalable ILP approach for ring RWA have been ongoing for almost two decades. Yetginer et al. [YeLR11] presents a decomposition approach that is optimal, fast for any reasonably sized ring, with a run time that is essentially independent of the amount of traffic on the ring.

The first step is to enumerate all paths that could possibly be used by the demand set. (Note that for any demand, there are only two possible paths on the ring; one in the clockwise direction and the other in the counterclockwise direction.) Next, a graph is created, where each path in the path set is represented by a node in the

graph. An adjacency is added between any two nodes of this graph if the two corresponding paths have one or more links in common. The ring links are treated as unidirectional; thus, a link from Node B to C is different from the link from Node C to B. The restriction that two overlapping ring paths cannot be assigned the same wavelength corresponds to the restriction that two adjacent nodes in the graph cannot be assigned the same color. This formulation corresponds to the classic graph-coloring problem [CLRS09].

Next, consider enumerating all *maximal independent sets* (MISs) in this graph (heuristics exist for this process, e.g., Bron et al. [BrKe73]). An *independent set* is a collection of nodes in the graph that are pairwise nonadjacent. An MIS is an independent set such that any node not in the MIS is adjacent to at least one of the nodes in the MIS. Note that the set of nodes in an MIS can be assigned the same color; i.e., the corresponding paths in the ring can be assigned the same wavelength without any conflicts.

One can then employ an ILP formulation where there is a variable corresponding to each possible MIS in the graph [RaSi95]. By selecting a collection of MISs, the ILP implicitly selects a path and assigns a wavelength for each traffic demand on the ring. However, the number of possible MISs in a graph grows exponentially with graph size, such that this version of the MIS methodology is not scalable.

If one looks at the MISs in the graph formed from the ring paths, any MIS will consist of a set of nodes that represent paths in the clockwise direction of the ring and a set of nodes that represent paths in the counterclockwise direction (due to clockwise links being distinct from counterclockwise links). Furthermore, these two sets of nodes must themselves be an MIS with respect to the clockwise and counterclockwise directions. Wavelengths can be assigned independently on the two directions of the ring. Thus, it is sufficient to consider all of the clockwise MISs and all of the counterclockwise MISs, rather than all of the MISs in the overall graph. By performing a clockwise/counterclockwise decomposition, the number of variables in the ILP is on the order of $2M$ rather than M^2, where M is the number of MISs in either direction of the ring. This two-way decomposition represents a significant run-time improvement. In some sense, using variables to represent all MISs in the overall graph adds a lot of redundancy; decomposing the problem removes some of this redundancy.

Further decompositions are presented in Yetginer et al. [YeLR11]. In addition to the clockwise/counterclockwise decomposition, consider splitting the ring in half, say a right half and a left half. Any paths that lie completely within the left half are independent from the paths that lie completely within the right half. There is another set of paths that span both halves. Any MIS in a given direction of the graph can be represented by a set of nodes from the left half, a set of nodes from the right half, and some group of "core nodes" from the spanning set (see Yetginer et al. [YeLR11] for details on forming the sets). This recognition further reduces the number of required ILP variables, at the expense of some added constraints. Using this four-way decomposition removes more redundancy, which translates to an even faster run time.

One can continue this decomposition process, although the marginal benefits are small after the four-way decomposition. The results in Yetginer et al. [YeLR11] indicate that this methodology can optimally solve the RWA problem in a 16-node ring in a few seconds (16 nodes is often the largest-sized ring used in carriers due to the original Synchronous Optical Network and Synchronous Digital Hierarchy (SONET/SDH) specification that includes only four bits for the node number). Furthermore, the number of ILP variables is independent of the traffic, such that solution times are essentially constant for any amount of traffic.

This ILP methodology is thus a scalable one-step RWA approach for realistic ring problem instances. Further research is needed to determine if the decomposition procedure can be extended to arbitrary mesh topologies.

5.5 Wavelength Assignment Strategies

The specific wavelength assignment strategy that is used can affect the performance of both multistep and one-step RWA.

With multistep RWA, a route is selected and then broken into subconnections based on where regenerations are needed. If no regeneration is needed, then the subconnection equals the whole connection; this is still referred to as a subconnection here. The wavelength assignment strategy determines the order in which wavelengths are considered when assigning a wavelength to each of the subconnections.

For one-step RWA, the wavelength assignment strategy determines the order in which wavelengths are considered when pruning the topology (Sect. 5.4.1) or the order in which wavelengths are considered when adding links to the reachability graph (Sect. 5.4.2). A particular wavelength assignment strategy is not typically explicitly enforced in the flow routing methodology, although the cost function may influence the choice of wavelength.

With multistep RWA, if there is no wavelength that is free along a subconnection, then one of the methods described in Sect. 5.3.1 is used to ease the wavelength constraints, or the corresponding demand is declared blocked. With one-step RWA, if the associated methodology fails to find a solution, the demand is declared blocked.

Wavelengths must be assigned such that the same wavelength is not used more than once on any fiber. To clarify this restriction, refer to Fig. 5.4. Link AB in Fig. 5.4a is populated with one fiber pair, where one fiber carries traffic from A to B, and the other fiber carries traffic from B to A. A given wavelength can be assigned once on the A-to-B fiber and assigned once on the B-to-A fiber. Link CD in Fig. 5.4b is populated with three fiber pairs. Three of the fibers carry traffic from C to D, and three from D to C. A particular wavelength can be assigned three times in each direction of the link, where each assignment is carried by a different fiber.

As these examples illustrate, it is possible that a given wavelength is used multiple times on a *link,* but it can be assigned only once per *fiber*. One exception to this rule is a system that supports bidirectional transmission of the same wavelengths on

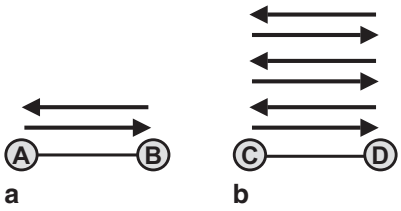


Fig. 5.4 A wavelength can be assigned to at most one subconnection per fiber. **a** There is one fiber pair on the link; the same wavelength can be used in both directions (once per fiber). **b** There are three fiber pairs per link; the same wavelength can be used three times in each direction (once per fiber)

a single fiber [Obar07], which is rarely implemented. For wavelength assignment purposes, the one fiber can be treated like a fiber pair.

Numerous wavelength assignment schemes have been devised over the years, where the differences in performance among the schemes are fairly small. Reference Zang et al. [ZaJM00] provides a good overview of the various schemes along with some performance curves. Here, the focus is on three particular schemes that have proved to be effective when preparing designs for actual carrier networks: *First-Fit, Most-Used,* and *Relative Capacity Loss* (RCL). First-Fit and Most-Used were first proposed for wavelength assignment in Chlamtac et al. [ChGK89], although the algorithms were not given these names until later. RCL was proposed in Zhang et al. [ZhQi98]. The three schemes are described below, followed by a qualitative comparison. All three of these schemes are suitable for any topology. Furthermore, the schemes can be applied whether there is a single fiber pair or multiple fiber pairs on a link.

Note that there are schemes specifically designed for the multiple fiber-pair scenario, most notably the *Least-Loaded* scheme [KaAy98]. This is shown in Zang et al. [ZaJM00] to perform better than the three aforementioned schemes when there are several fiber pairs per link. As fiber capacities have rapidly increased, however, systems with several fiber pairs on a link have become a less common occurrence (although this may change in the future as discussed in Chap. 9).

5.5.1 First-Fit

First-Fit is the simplest of these three wavelength assignment schemes. Each wavelength is assigned an index from 1 to *W,* where *W* is the maximum number of wavelengths supported on a fiber. No correlation is required between the order in which a wavelength appears in the spectrum and the assigned index number. The indices remain fixed as the network evolves. Whenever wavelength assignment is needed, the search for an available wavelength proceeds in an order from the lowest index to the highest index. The first available wavelength found is selected.

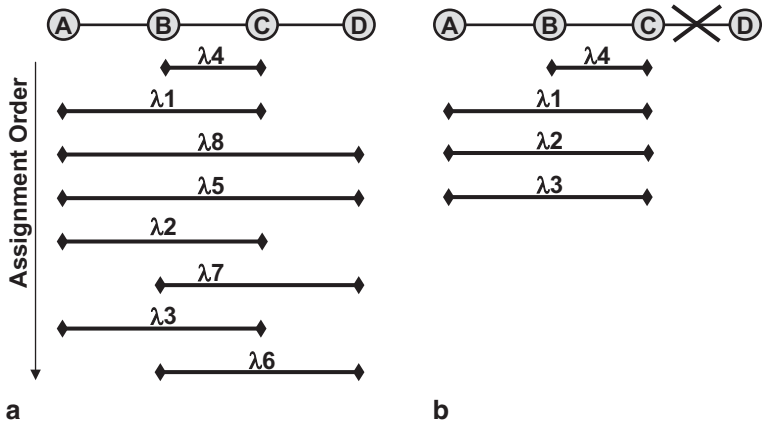


Fig. 5.5 **a** Eight connections are added one-at-a-time, in the order shown. **b** After Link *CD* fails, only four connections remain

It is reasonable to consider using the First-Fit indexing order to guide the network growth in a manner that may potentially improve network performance. However, due to network churn (i.e., the process of connections being established and then later torn down), the interdependence of wavelength assignment across links, and the presence of failure events, the *indexing ordering* does not guarantee the actual *assignment order* on a link. Consider the example shown in Fig. 5.5. Assume that it is desirable for wavelengths to be assigned on links such that there is a good spread across the spectrum, i.e., assign some wavelengths in the middle, some at the low end, and some at the high end of the transmission band. (The motivation for this is that some Raman amplifiers perform better when the power levels are fairly evenly distributed across the transmission band.) For simplicity, assume that there are only eight wavelengths in the system, and assume that the index order for First-Fit is: λ_4, λ_1, λ_8, λ_5, λ_2, λ_7, λ_3, and λ_6. With a focus on the three links shown in Fig. 5.5a, assume that eight connections are added to the network in the order shown in the figure. The figure indicates the wavelengths that are assigned to each connection, based on the indexing scheme specified above. Even though the index order is consistent with a good spread across the transmission band, the wavelengths assigned on Link CD are λ_8, λ_5, λ_7, and λ_6. These wavelengths are all in the upper half of the spectrum as opposed to being spread across the band.

Furthermore, assume that Link CD fails; the remaining connections are shown in Fig. 5.5b. Thus, while Link AB originally was populated with wavelengths spread across the band, the wavelengths remaining after the failure are λ_1, λ_2, and λ_3. Again, the wavelengths are clustered towards one end of the band.

As another example of trying to use the First-Fit index to enhance performance, He et al. [HeBr06] proposes using the indexing scheme to minimize crosstalk among the wavelengths; i.e., the indices are ordered in an attempt to delay the point at which adjacent wavelengths are assigned on a link. However, because of network churn and the interdependence of wavelength assignment across links, adjacent

wavelengths may need to be assigned prior to a link being half full. Furthermore, the transmission system should not be designed such that it works only if nonadjacent wavelengths are assigned on a link. An established subconnection should not fail due to any other subconnection using any other wavelength being added to the network. Requiring an established demand to be rerouted due to the addition of a new demand is undesirable.

Thus, the indexing strategy of First-Fit may be viewed as a short-term means of potentially providing additional system margin under certain conditions, but it should not be relied on to enforce a critical system constraint, because network growth or network churn may cause the benefits to be lost.

5.5.2 Most-Used

The second wavelength assignment scheme considered here, Most-Used, is more adaptive than First-Fit, although it requires more information. Whenever a wavelength needs to be assigned, a wavelength order is established based on the number of link-fibers on which each wavelength has already been assigned. The wavelength that has been assigned to the most link-fibers already is given the lowest index, the wavelength that has been assigned to the second-most link-fibers is given the second lowest index, etc. After the wavelengths have been indexed, the assignment procedure proceeds as in First-Fit. The motivation behind this scheme is that a wavelength that has already been assigned on a lot of fibers will be more difficult to use again. Thus, if a scenario arises where a heavily used wavelength can be used, it should be assigned.

5.5.3 Relative Capacity Loss

RCL is more complex than either of the previous schemes. The idea is that wavelength assignment should take into account how much "harm" it is doing to demands that may be added in the future; i.e., how likely will it cause wavelength contention for future demands. This scheme is more amenable to the multistep RWA process; thus, this is the focus of the discussion.

The first step in RCL is to generate the set of possible paths in the network, e.g., based on a traffic forecast. When regeneration is present in a network, it is actually the set of possible subconnections that needs to be enumerated. If some subconnections are expected to arise more than others, as is likely to be the case, then the subconnections should be added to the list multiple times to reflect their expected relative frequency. This enumeration step is done prior to any traffic being added to the network, although the list can be updated if necessary as the network evolves.

As demand requests enter the network, a connection path is selected, and the new connection is partitioned into subconnections. Each new subconnection must then be assigned a wavelength. Consider one such new subconnection, which is

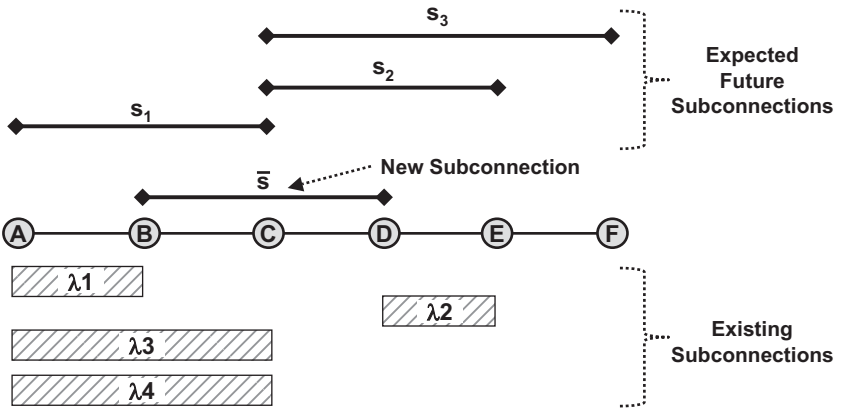


Fig. 5.6 Setup to illustrate RCL wavelength assignment where the shaded boxes indicate the links on which a wavelength has already been assigned. λ_1 is selected for $\overline{s}$ because this assignment is less "harmful" to the expected future subconnections

denoted here by $\overline{s}$. A set S is generated that contains all forecasted subconnections that have at least one link in common with $\overline{s}$. For each subconnection s_j in S, it is determined how many wavelengths are currently available to be assigned to it. Let this quantity be represented by N_j. Note that if there are multiple fiber pairs per link and a wavelength is available on f fiber pairs on each link of s_j, then the quantity N_j takes this into account.

Next, for each wavelength λ_i that could possibly be assigned to $\overline{s}$, where $1 \le i \le W$, it is determined whether or not s_j is affected if $\overline{s}$ were assigned wavelength λ_i (i.e., if $\overline{s}$ were assigned wavelength λ_i, does that reduce N_j by one). Let I_{ij} be 1 if s_j is affected, and let I_{ij} be 0 if s_j is not affected. Then for each wavelength λ_i that is available to be assigned to $\overline{s}$, the following sum is calculated (where the sum is calculated over all s_j in S):

$$C_i = \sum_j \frac{I_{ij}}{N_j} \tag{5.1}$$

The wavelength λ_i with the minimum C_i is selected as the wavelength to assign to $\overline{s}$. If there is a tie among the λ_i's for the lowest C_i, then the one with the lowest index is selected. (Note that C_i equals 0 if the selection of λ_i does not affect any other subconnection.) This procedure is run through whenever a wavelength needs to be assigned to a new subconnection.

This algorithm is illustrated using the example of Fig. 5.6. Assume that there is a maximum of four wavelengths per fiber, and one fiber pair per link. The shaded boxes in the figure indicate the links on which wavelengths have already been assigned. The subconnection of interest, $\overline{s}$, extends between Nodes B and D. The set S is composed of subconnections s_1, s_2, and s_3. Subconnection s_1, which extends between Nodes A and C, has only one available wavelength, λ_2. Thus, N_1 is 1.

Subconnections s_2 and s_3 each have three available wavelengths, so that N_2 and N_3 are both 3. The possible wavelengths that could be assigned to $\overline{s}$ are λ_1 and λ_2. If λ_1 were assigned to $\overline{s}$, then only s_2 and s_3 are affected, because s_1 already cannot use λ_1. Thus, C_1 is $1/3+1/3=2/3$. If λ_2 were assigned to $\overline{s}$, then only s_1 is affected; C_2 thus equals 1. C_1 is lower than C_2, resulting in λ_1 being assigned to $\overline{s}$. Even though the λ_1 assignment affects two subconnections and the λ_2 assignment would affect only one, the s_1 subconnection has fewer options, and is "hurt" more if λ_2 were assigned to $\overline{s}$.

The RCL scheme is especially suitable when a whole set of demands is added at once. In this scenario, with multistep RWA, all subconnections are known before the start of the wavelength assignment phase (thus, the step where the set of expected subconnections is enumerated based on a forecast is unnecessary). When assigning a wavelength to a subconnection and determining the relevant set *S,* only those subconnections that have not been assigned a wavelength yet need to be considered for inclusion in *S*.

5.5.4 Qualitative Comparison

All three wavelength assignment schemes provide relatively good performance in realistic networks. For example, wavelength contention does not generally become an issue until there are at least a few links in the network with roughly 85 % of the wavelengths used. RCL often performs somewhat better than the other two schemes in minimizing wavelength contention; however, it is also the most complex to implement. It is more difficult to rank the relative performance of First-Fit and Most-Used, as which performs better depends on the network topology and the traffic. In any case, the differences in performance are small. One advantage to First-Fit is that, in contrast to the other two schemes, it does not require any global knowledge to operate.

In long-term planning with a set of demands being added, and multistep RWA being used, it is possible to try more than one scheme on the set of subconnections. For example, one could first attempt to use either First-Fit or Most-Used to assign wavelengths to the set of subconnections. If this does not result in feasible wavelength assignments for all of the subconnections, then one could restart the wavelength assignment process using RCL.

Another factor that has an impact on the success of the assignment scheme is the order in which the subconnection set is processed. This is discussed next.

5.6 Subconnection Ordering

Consider adding multiple demands to the network at once and assume that multistep RWA is used. Assume that all routing and regeneration occurs prior to the wavelength assignment step, such that all subconnections have already been created.

In the previous section, different strategies were presented for the order in which wavelengths should be considered for assignment to a particular subconnection. This section discusses the order in which the subconnections are processed.

Finding an available wavelength for a subconnection is more difficult when the links on which a subconnection is routed are congested and when the subconnection is routed over many links. Thus, link load and number of subconnection hops are two important factors that may be considered when determining the order in which subconnections should be assigned wavelengths.

A natural first step is to determine the most heavily loaded link in the network. (It is assumed here that the heaviest load is no larger than the number of wavelengths supported on a fiber. If there are more subconnections routed on a link than there are wavelengths, then clearly a feasible assignment cannot be found. In this case, the routing process needs to be redone.) If there are multiple links tied for the heaviest load, then the link with the longest subconnections on it, in terms of hops, should be selected. Designate this as the "worst link." The next step is to assign wavelengths to all subconnections that are routed on the worst link, given that these are likely to be the most difficult on which to find an assignment; no wavelength conflicts can occur in this stage. When the assignment process first starts, all wavelengths can be considered equivalent (although see Sects. 5.8 and 5.9). Thus, the wavelengths can be assigned arbitrarily to the subconnections on the worst link; any assignment can be mapped into any other.

The next step is to order the remaining subconnections based on factors such as the load of the links traversed by the subconnections and the number of hops in the subconnections. Various heuristic ordering schemes can be devised; three examples are presented here.

In one scheme, if the load along a whole subconnection is low, then the assignment order is based solely on the number of hops in the subconnection. If, however, a subconnection traverses some heavily loaded links, then the hop metric is artificially inflated to reflect the expected difficulty of finding an available wavelength for that subconnection. The subconnections are then processed for wavelength assignment in an order from the largest hop metric to the smallest hop metric.

In a second scheme, a metric is devised for each link that reflects the load, where the metric is less than unity, and the heavier the load the lower the metric. The metric for the subconnection is the product of its corresponding link metrics, thus taking into account both load and number of hops. This metric is then used to determine the subconnection wavelength assignment order, where a lower metric results in earlier assignment.

A third scheme selects a subconnection for wavelength assignment based on the number of available wavelengths that could possibly be assigned to it. The number of available wavelengths for a subconnection monotonically decreases as wavelengths are assigned to other subconnections. Thus, as the wavelength assignment process progresses, the number of available wavelengths needs to be updated for each of the remaining unassigned subconnections. At each step, the subconnection with the fewest wavelength possibilities is processed next. If there is a tie, then the subconnection that overlaps (i.e., shares at least one hop) with the most other

unassigned subconnections is selected. If there is still a tie, then the subconnection with the most number of hops is selected. This is a natural ordering scheme to use with the RCL wavelength assignment algorithm because RCL already tracks the number of wavelengths that can possibly be assigned to each remaining subconnection. We refer to this ordering scheme as Min-WL-Remain.

If a particular subconnection ordering does not produce a feasible assignment, then the wavelength assignment process can be restarted using a different ordering strategy. Furthermore, various combinations of subconnection ordering and wavelength assignment schemes can be considered.

Note that one of the advantages of multistep RWA when working with a set of demands is the order in which demands are routed and the order in which subconnections are assigned wavelengths can be completely decoupled, allowing more design flexibility. Also note that when using global optimization techniques such as an LP approach, ordering of subconnections is not necessary.

5.6.1 Graph Coloring

As has been alluded to already, the wavelength assignment problem is analogous to the graph-coloring problem, where each vertex in a graph must be assigned a color, subject to the restriction that vertices connected by an edge cannot be assigned the same color [CLRS09]. In the corresponding wavelength assignment problem, the vertices represent subconnections and the colors represent wavelengths; two vertices are connected if the subconnections have at least one link in common. (This is sometimes referred to as the *conflict graph*.) Any solution to the graph-coloring problem corresponds to a valid solution to the wavelength assignment problem.

Graph coloring (on large graphs) is known to be a difficult problem for which to find an optimal solution, i.e., where the number of colors needed is minimized. However, there are known bounds on the minimum number of required colors. (The minimum number of required colors is referred to as the *chromatic number* of the graph.) If the largest vertex degree in the graph to be colored is D, then it is possible to color the graph using *at most* $D+1$ different colors; i.e., at most $D+1$ different wavelengths are needed in the corresponding wavelength assignment problem. For simple connected graphs, other than fully connected graphs and cycles with an odd number of vertices, the upper bound on the number of required colors is D. Furthermore, if the order of the largest clique[1] in the graph to be colored is C, then *at least* C different colors (wavelengths) are required to color the graph.

Numerous graph-coloring heuristics have been developed, where the heuristics generally differ in the order in which the vertices are assigned a color. This is analogous to the order in which subconnections are assigned a wavelength. Many of the ordering schemes are based on the degree of the vertex, which corresponds to the number of other subconnections with which a particular subconnection has at least one common link. The ordering is then typically combined with a coloring scheme such as First-Fit.

[1] A clique is a set of fully connected vertices.

Any of the graph-coloring heuristics can be used to order subconnections for wavelength assignment. The heuristics can be enhanced by using factors such as link load and number of subconnection hops as tiebreakers in the ordering process. One graph-coloring heuristic of note is the *Dsatur* strategy [Brel79]. The coloring order in this heuristic is based primarily on which vertex has the fewest choices of feasible colors remaining. This directly corresponds to the Min-WL-Remain ordering scheme described above, which in each iteration selects the subconnection with the fewest number of feasible wavelengths remaining.

Various graph-coloring schemes are investigated in Exercises 5.9 through 5.12, including the *Dsatur* scheme.

5.7 Bidirectional Wavelength Assignment

Most demands in carrier networks are bidirectional, where a connection from Node A to Node Z implies a connection from Node Z to Node A. Usually, both directions of the connection are routed over the same path and regenerated at the same sites, thereby yielding identical subconnections in the two directions. An interesting question is whether the same wavelength should be assigned to each pair of subconnections; i.e., if a particular wavelength is assigned to a subconnection extending from Node X to Node Y, should the same wavelength be assigned to the reverse subconnection extending from Node Y to Node X.

From a network management point of view, it may be most expedient to simply assign the same wavelength to the subconnection pair. However, there are scenarios where assigning different wavelengths can improve the efficiency of the network. Consider the very simple four-node topology of Fig. 5.7, and assume that a fiber supports just two wavelengths. Furthermore, assume that the optical reach is longer than any of the possible connection paths, such that each connection can be potentially established without any regeneration.

Assume that one bidirectional connection is established between A and C, and another between A and D. In Fig. 5.7a, λ_1 is assigned to both directions of the AC connection, and λ_2 is assigned to both directions of the AD connection. With this wavelength assignment, it is not possible to add a bidirectional connection between C and D without converting the wavelength at node B. There is no wavelength that is free along both links of the CD connection. The wavelength conversion required at Node B incurs the cost of an extra regeneration.

In Fig. 5.7b, different wavelengths are assigned to the two directions of connections AC and AD. This allows the CD connection to be added without any need for wavelength conversion, as shown in the figure. Note that different wavelengths are assigned to the two directions of CD. This wavelength assignment scheme thus provides a lower-cost solution as compared to that of Fig. 5.7a.

While this is just a small example, this type of situation does arise in the design of real networks. Even when the number of wavelengths is large, there are scenarios where using different wavelengths in the two directions of a subconnection can

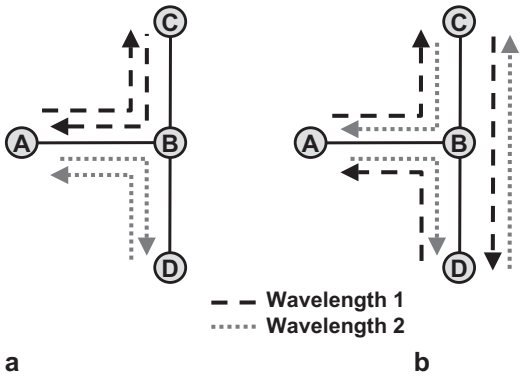


Fig. 5.7 Assume that there are only two wavelengths supported on a fiber in this small network. **a** The wavelength assignments for connections *AC* and *AD* are both bidirectionally symmetric. If a connection between *C* and *D* is added, that connection must undergo wavelength conversion at Node *B* in order to avoid wavelength conflicts with the existing connections. **b** Different wavelengths are assigned to the two directions of connections *AC* and *AD*. The connection between *C* and *D* can be added without any wavelength conversion, as shown. (Adapted from Simmons [Simm06]. © 2006 IEEE)

result in a lower-cost network, due to less wavelength contention. It occurs most commonly in real networks when there are degree-three nodes, with a lot of bypass traffic in all three directions through the node. It also may occur in higher-order odd-degree nodes, although this situation rarely arises in practice.

Any of the wavelength assignment schemes described above can be readily modified to produce different wavelength assignments in the two directions of a subconnection. For example, consider a scheme where odd and even wavelengths are used for the two directions of a subconnection. Assume that First-Fit runs through the odd-numbered indexed wavelengths, yielding an assignment of λ_5 for a particular subconnection. Then the subconnection in the reverse direction can be assigned λ_6. If the traffic set consists of unidirectional demands as well, where traffic goes in just one direction, then more care needs to be taken because wavelengths are not always assigned in pairs.

While assigning different wavelengths to the two directions of a connection was considered above for cost reasons, there are some scenarios that *require* the wavelengths to be different. These scenarios typically arise due to certain protection architectures, as noted in Sect. 7.4.1.

5.8 Wavelengths of Different Optical Reach

When assigning wavelengths to subconnections, another factor that needs to be considered in some systems is that not all wavelengths have the same optical reach. This phenomenon is often dependent on the type of fiber that is installed in the

network. As discussed in Chap. 4, some amount of chromatic dispersion is desirable in a fiber as it helps to minimize the effects of nonlinear optical impairments. However, some fiber types have a region of very low dispersion that partially overlaps with the portion of the spectrum used for transmission. The wavelengths that fall in this low-dispersion region suffer from greater impairments, resulting in reduced reach. These "reduced-reach" wavelengths usually represent a small percentage of the overall wavelengths.

This effect needs to be considered when performing wavelength assignment. For example, consider an 80-wavelength system where the nominal optical reach is 2,000 km, but where 5 of the 80 wavelengths have an optical reach of only 1,000 km. It is desirable to assign these five wavelengths early in the process. Otherwise, these wavelengths may be left to the end of the wavelength assignment process, in which case extra regenerations may be needed to chop the remaining subconnections into even shorter subconnections. In a First-Fit wavelength assignment scheme, the five reduced-reach wavelengths can be assigned low indices. When looking for a free wavelength, the assignment process must check that the reach of the wavelength is suitable for the subconnection being considered. If the reach of the wavelength is too short, that wavelength is passed over. With such a strategy, there typically is no need to proactively chop up a connection into shorter subconnections in order to utilize the reduced-reach wavelengths. There are usually enough short subconnections that are naturally formed as part of the design process.

Some carriers have a mix of fiber types in their network where just a subset of the links have fiber with low-dispersion regions. In this scenario, the wavelength assignment process must first consider the fiber type of the links on which a subconnection is routed. If a subconnection is partially on "good" fiber and partially on "bad" fiber, then it is up to the system engineers to develop rules for the optical reach of the wavelengths. This also could be a factor in determining where to regenerate a connection in the scenario where regeneration is needed but it can be located at one of several nodes in the path. It may be desirable to pick the regeneration location such that the resulting subconnections are routed on homogeneous fiber types, if possible.

Note that with new builds, carriers are careful to install fiber with an appropriate level of dispersion across the transmission spectrum to avoid this problem. Thus, this issue will become less important with time.

Even if there are no dispersion issues, there still may be small differences in the optical reach of the wavelengths due to other factors. For example, a Raman amplifier may have a lower noise figure at the longer wavelengths, such that these wavelengths have somewhat longer optical reach. If one wants to squeeze every advantage out of an optical-bypass-enabled system, then one can use the fullest reach of each wavelength. However, there are some complications with this approach. First, it may ultimately lead to more wavelength contention. The length of the subconnection becomes the prime determinant of which wavelength is assigned to it; the orderly wavelength assignment schemes, such as the ones described in Sect. 5.5, are less relevant. Second, a particular connection may be regenerated differently based on the wavelength(s) that will carry it. This results in less alignment of the

subconnection endpoints, which can result in more wavelength contention. Third, it makes multistep RWA more challenging because selecting the regeneration sites for a connection may depend on which wavelength is assigned to it, so that these two steps must be coupled. Furthermore, as is discussed in Chap. 10, after some point, the marginal benefits of increased optical reach begin to rapidly diminish. Maximizing the reach of each individual wavelength may only result in a small cost savings, which may not justify the additional complexity. It may be more desirable to set an optical reach that almost all wavelengths can attain, with perhaps a small number of wavelengths relegated to shorter reach due to low-dispersion (or other) issues.

5.9 Nonlinear Impairments Due to Adjacent Wavelengths

In many long-reach systems, the transmission system is designed such that the power levels are low enough, or the dispersion levels are high enough, so that impairments due to adjacently propagating wavelengths are relatively small. However, there may be transmission systems where relatively high power levels are required (e.g., to obtain extended optical reach at high line rates), leading to scenarios where populating adjacent, or nearly adjacent, wavelengths in the spectrum produces non-negligible nonlinear impairments, most notably cross-phase modulation (XPM). In such systems, the quality of transmission (QoT) for a given connection may depend on what other wavelengths are in use on the same fibers.

There are two methods for dealing with this scenario. The first strategy is to ensure that connections are established with enough system margin to tolerate the *worst-case* impairments that could possibly arise from populating adjacent wavelengths with other connections. (One may have to consider more than just the immediately adjacent wavelengths; e.g., wavelength i could suffer impairments due to wavelengths $i-2$, $i-1$, $i+1$, $i+2$. The effects of wavelengths outside of this range are likely to be negligible.) For example, the system rules may require an extra N dB of optical signal-to-noise ratio (OSNR) to take the worst-case impairment scenario into account. This allows connections to be assigned to wavelengths without concern over inter-wavelength impairments. If a particular connection is deemed feasible at the time of its establishment, it should remain feasible regardless of what other connections may later be added.

In the second strategy, the effects of inter-wavelength impairments are calculated more precisely. The optical reach of a particular available wavelength along a given path is determined at the time a demand request is received, based on the state of the adjacent wavelengths. Consider assigning wavelength i to a new connection on a given path. If wavelengths $i-1$ and $i+1$ are not being used on the fibers that comprise this path, then wavelength i may have additional optical reach as compared to the case where a worst-case reach assumption is used. This could lead to fewer required regenerations for the new connection. The drawback, of course, is that if future connections populate wavelength $i-1$ and/or $i+1$, the performance of wavelength i may degrade below an acceptable QoT, forcing the associated connec-

tion to be assigned to a different wavelength or be rerouted. Rerouting/reassigning an active connection is possible using "make-before-break" techniques, but it is undesirable. If modifying an existing connection is not permitted, then the strategy of maximizing the reach of wavelength i could result in future connections being blocked from using wavelengths $i-1$ or $i+1$.

The decision as to which of the two strategies to use may depend on how regenerations are handled. If regeneration is permitted in the network, then the impact of using a worst-case impairment assumption will likely be extra regenerations, because the system optical reach will effectively be reduced. However, as shown in a case study in Chap. 10, as long as the optical reach is reasonably long, small changes in the reach do not have a large impact on the amount of required regeneration and the overall network cost. (For example, the benefit of a 2,800-km reach as opposed to a 2,500-km reach in a continental-scale network is likely to be less than a 3 % reduction in network transmission costs.)

If, however, the system requires that connections be *truly* all-optical, with no regeneration, then the policy for handling impairments may have an impact on blocking. For example, the end-to-end path distance of a new demand may be very close to the nominal optical reach. Establishing the new connection on a wavelength that is distant from any populated wavelengths may allow the connection to be successfully deployed, whereas the worst-case impairment assumption would dictate that it be blocked. This effect was examined more fully in Christodolopoulos et al. [CKMV09], for a backbone network of relatively small geographic extent where no regeneration was permitted. The two strategies outlined above were compared; i.e., either assume worst-case inter-wavelength impairments or calculate the inter-wavelength impairments more accurately based on the actual network state. In either strategy, moving an existing connection to a different path or wavelength was not permitted. Thus, a new connection could not be added if it would result in an unacceptable QoT for an existing connection (this is not an issue in the worst-case impairments strategy). The results indicated that when inter-wavelength impairments were more precisely calculated, the blocking rates were reduced by about an order of magnitude, due to there being a larger set of feasible paths from which to choose. However, this type of pure all-optical scenario would not arise in a network of large geographic extent, because some regeneration is needed regardless of how inter-wavelength impairments are treated. Inter-wavelength impairments would also unlikely be an issue in a metro network, where the optical reach, even with worst-case assumptions, is typically longer than any path. Thus, the benefit of more precisely calculating inter-wavelength effects may not be very significant in most practical networks.

However, *if* it is desirable to take inter-wavelength impairments into account when performing RWA for a new demand request, then one can utilize a cost-vector approach to routing [MKCV10], as described in Sect. 4.4. Various per-wavelength components can be included in the cost-vector that is used for "shortest-path" routing. For example, the cost-vector could include the noise variance of a "1" and of a "0" for each one of the available wavelengths, where the noise variance captures inter-wavelength impairments such as crosstalk, XPM, and four-wave mixing. For

each available wavelength on a link, the cost component for that wavelength-link combination is calculated based on the wavelengths that are already populated on that link. As detailed in Sect. 4.4, a modified Dijkstra routing algorithm is run with the cost vector, using the principle of dominated paths. Additionally, as the routing algorithm progresses, the Q-factor (a performance measure correlated to bit error rate) corresponding to each available wavelength is calculated. If the Q-factors for all of the available wavelengths on a path fall below the acceptable threshold, then that path can be eliminated from further consideration. When the routing algorithm terminates, a scalar-generating function is applied to the final cost vector for each remaining feasible path/wavelength combination to determine which one to use.

5.9.1 Mixed Line-Rate Systems

An important scenario that may warrant accounting for inter-wavelength impairments more precisely is when multiple line rates and modulation formats co-propagate on a single fiber. This scenario was discussed in Sect. 4.2.6. For example, a mixed line-rate system may include 10-Gb/s on–off keying (OOK) wavelengths along with 40-Gb/s and 100-Gb/s dual-polarization quadrature phase-shift keying (DP-QPSK) wavelengths. As described earlier, experiments have shown that DP-QPSK wavelengths suffer performance penalties due to XPM from nearby co-propagating 10-Gb/s OOK wavelengths. The penalties are worse for 40-Gb/s wavelengths than for 100-Gb/s wavelengths. Furthermore, the penalties are severe enough that leaving enough system margin to account for the worst-case XPM would be too detrimental to the system reach.

An alternative strategy is to ensure that the OOK and DP-QPSK wavelengths are sufficiently separated from each other. To reduce the performance penalty to an acceptable level requires a guardband of roughly 300 GHz between the 40- and 10-Gb/s wavelengths; a guardband of roughly 150 GHz suffices between the 100- and 10-Gb/s wavelengths [BRCM12]. Wavelengths are typically spaced at every 50 GHz in a backbone network; thus, such large guardbands represent a significant loss of available fiber bandwidth. The wavelength assignment process should take this into account to minimize the use of such guardbands.

A “soft” partitioning can be enforced, where the 40- and 100-Gb/s wavelengths are assigned from one end of the spectrum, whereas the 10-Gb/s wavelengths are assigned starting at the other end. Note that adjacent co-propagating 40- and 100-Gb/s DP-QPSK wavelengths suffer little performance penalty. Additionally, relatively *short* 40- and 100-Gb/s subconnections, which can tolerate the performance penalty of adjacent 10-Gb/s wavelengths, can be proactively assigned wavelengths from the buffer area between the two portions of the spectrum. (This is similar to the principle discussed in Sect. 5.8 for proactively assigning wavelengths that have reduced reach to short subconnections.) Note that we are not advocating a *fixed* partitioning of resources among the line rates, as fixed partitioning generally leads to higher blocking.

This wavelength assignment strategy will have a tendency to segregate the conflicting line rates, to minimize the need for guardbands. As the network fill-rate increases, and the high and low spectral ranges approach each other, the cost-vector approach discussed above could be used to capture the penalties associated with adding a particular wavelength of a particular modulation format to a given link.

Once a path/wavelength is selected for a new demand, it is important to verify that existing connections will remain feasible. This is especially important when a new 10-Gb/s wavelength is added near existing 40-Gb/s wavelengths.

5.10 Alien Wavelengths

As was illustrated in Fig. 2.1, the client layer (e.g., an IP router) typically interfaces to the optical layer via a 1,310-nm signal, which is received by the short-reach interface of the WDM transponder. The WDM transponder converts the 1,310-nm signal to one that is compatible with the particular WDM system. In order to eliminate the cost of the short-reach interface, there have been attempts to directly integrate WDM transceivers in the client-layer equipment, as described in Sect. 2.13. For the most part, this has been possible when the vendors of the client-layer equipment and the optical system are either the same or have collaborated, to ensure that the WDM signal generated by the client is compatible with the optical transport system.

WDM wavelengths that are generated outside of the optical transport system are referred to as *alien wavelengths*. (In contrast, *native wavelengths* are generated by WDM transponders that are part of the optical transport system.) There has been a push, especially by IP router vendors, for optical systems to more broadly support alien-wavelength transport (i.e., without requiring collaboration among vendors), to make integration between layers easier to achieve [GCPW09]. It is also envisioned that alien-wavelength support may allow all-optical routing between the equipment of two different system vendors, thereby eliminating the need to create O-E-O demarcated vendor-specific islands [FaSk13].

As a first step in this direction, the International Telecommunication Union (ITU) has provided optical interface specifications for systems with 50- or 100-GHz spacing and line rates of 2.5 or 10 Gb/s (the specifications are primarily intended for metro network applications) [ITU09]. The portion of the optical system that lies between the entry and exit points of an alien wavelength is treated as a "black link" (analogous to a "black box"). The ITU specifications include acceptable values for parameters such as mean channel output power of the alien transmitter and OSNR tolerance of the alien receiver.

Note that implementing alien wavelengths removes the O-E-O monitoring point between the client and optical layers, thereby posing numerous operational challenges [MBLV09]. For example, it may be more difficult for the optical layer to monitor or control these wavelengths. If faults arise in the connections carried by the alien wavelengths, it may not be readily determined in which layer the problem originates.

Another challenge is that the performance of alien wavelengths is unlikely to match that of native wavelengths, due to, for example, different tolerances or different forward error correction (FEC) implementations on the alien transponders. This may necessitate somewhat different rules for regeneration and wavelength assignment. As indicated in Sect. 5.8, some wavelengths in the spectrum may have somewhat longer reach than others. It may be desirable to preferentially assign these wavelengths to alien wavelengths, if they are present.

Other problems may arise due to potentially deleterious interactions between the alien and native wavelengths. This may require that the different wavelength types be quasi-segregated, similar to what was described for mixed line-rate systems.

Any special wavelength assignment rules would need to be communicated to the system in which the alien wavelengths originate.

5.10.1 Analog Services

As was illustrated in Fig. 1.3, it is possible for wavelengths from the high-level application layer to bypass the lower electronic layers and directly access the optical layer. This is another scenario where alien wavelength support is required from the optical layer. In principle, the optical layer can carry such wavelengths, independent of the particular format, including analog signals. This potential capability with respect to analog services is referred to as *analog transparency*.

While transparency to various digital signal formats is possible, transparency to analog signals represents a significant challenge [Phil04]. The analog signals need to be carried all-optically end-to-end and be delivered with acceptable fidelity. This capability is likely to be desired by only a very small number of users, e.g., those involved with special types of sensor networks. Judicious wavelength assignment may be advantageous in this scenario, where a small number of wavelengths are designated for analog services. By selecting portions of the spectrum with minimal impairments, increasing the wavelength spacing in these spectral regions, and giving preferential signal-to-noise-ratio treatment to these wavelengths within the optical amplifiers and ROADMs [SaSi06], end-to-end transparency for analog signals *may* be attainable, but it is likely to be limited to metro-scale networks.

5.11 Wavelength Contention and Network Efficiency

Even with good wavelength assignment algorithms, wavelength contention is likely to occur when a network becomes heavily loaded. Any contention can be alleviated by adding more regeneration to reduce the wavelength dependencies among the links. However, in real-time planning, there may not be equipment available for regeneration at the desired network nodes. In addition, it is undesirable to add a significant amount of extra regeneration because it will increase the cost of the

network. If wavelength contention cannot be resolved, then a demand may be blocked even though there is available capacity to carry the demand.

Wavelength contention is not an issue in pure O-E-O-based networks, where wavelengths can be assigned independently on each link. This raises the question of what impact wavelength contention has on the performance of optical-bypass-enabled networks. Many studies have been performed to investigate this question. The conclusion of most of the studies is that, assuming good algorithms are used, just a small amount of wavelength conversion is needed to approximate the performance of an O-E-O system. Some of these studies can be found in Subramaniam et al. [SuAS96], Karasan et al. [KaAy98], Van Parys et al. [VAAD01], and Simmons [Simm02].

Nevertheless, there is not a unanimity of opinion in the industry regarding this question. It is possible to produce studies that indicate wavelength contention has a significant negative effect on network performance. However, this is often due to the choice of algorithm. For example, some studies do not take advantage of regeneration as a chance to perform wavelength conversion; i.e., these studies unnecessarily require that the same wavelength be used end-to-end even when regeneration occurs along the path. Even if wavelength conversion is permitted, some studies do not allow extra regenerations to be added to alleviate wavelength contention. Conversely, another example of a less-than-ideal strategy is adding enough regenerations to an optical-bypass-enabled network in order to attain an efficiency that is *identical* to that of an O-E-O network. This approach may be too extreme and result in an excessive number of added regenerations. A slight reduction in network efficiency is acceptable with an optical-bypass-enabled network because the potential cost savings due to reduced electronics is still very significant.

As the topic of wavelength contention remains somewhat of a controversial issue, a small study is presented here for both a backbone network and a metro-core network. This study shows, again, that assuming a relatively small amount of extra regeneration can be added to the network to alleviate wavelength contention, then wavelength constraints have a very small impact on the network performance. Furthermore, the discussion provides a rationale for why this is so.

5.11.1 Backbone Network Study

Reference Network 2, from Sect. 1.10, is used for the backbone network portion of the study. This topology has 60 nodes, 77 links, and an average nodal degree of 2.6. (Reference Simmons [Simm02] includes a similar study performed on several other backbone networks of various sizes; the results are consistent across the networks.) In the study, there was one fiber pair per link with a maximum of 80 wavelengths per fiber. A realistic traffic set was used where roughly 20% of the nodes could be considered major nodes that generated a significant amount of the traffic. All demands were at the line rate such that no grooming was required; all demands were unprotected. Furthermore, the traffic was modeled as being totally dynamic

(which is an extreme assumption), where the demands arrived one-by-one according to a Poisson process with holding times that were exponentially distributed.

The two architectures compared were an O-E-O network where it was assumed that the optical reach was long enough (i.e., 1,200 km) such that no regenerations were required in the middle of a link, and an optical-bypass-enabled network with ROADMs/ROADM-MDs at all nodes and an optical reach of 2,500 km. In both scenarios, the load on the network was increased until the desired steady-state blocking probability was reached. The study focused on the 0.1 and 1.0% blocking scenarios. (For each offered load level, several simulations were run where the system was allowed to reach steady state; the averages were obtained using the *replication/deletion* approach [LaKe91].)

The key parameter used for evaluation is the average network utilization at these blocking levels. Utilization is measured as the bandwidth-distance product of the successfully routed demands. The distance used in this calculation is the shortest possible distance for a demand source/destination pair, which is not necessarily the route taken by the demand. Thus, the utilization measure cannot be artificially increased by circuitous path routing.

Another important statistic is the average number of transponders needed per demand. The minimum is two, i.e., the transponders at the end points. Additionally, any regeneration along a connection counted as another two transponders.

It was assumed that equipment was available when needed. Furthermore, at a regeneration point, it was assumed that the wavelength of the incoming subconnection could be changed to any wavelength for the outgoing subconnection. In many early studies of wavelength contention, limited wavelength conversion was assumed, where a given wavelength could be converted to only a small set of other wavelengths. However, current tunable transponders and regenerators are generally tunable across the whole transmission band (all-optical wavelength converters, if they ever become commercially available, are expected to also have full conversion capabilities); thus, this restriction is generally no longer applicable.

The O-E-O network regenerated every path at every node, and thus there were no wavelength contention issues. In the optical-bypass-enabled network, wavelength contention occurred whenever a feasible wavelength assignment could not be found for a subconnection. Regenerations were judiciously added to alleviate wavelength contention. If more than one regeneration needed to be added to a subconnection in order to find a feasible wavelength assignment, the associated demand was blocked. This resulted in a lower utilization than was actually possible, but moderated the number of added regenerations.

The results from the study are shown in Table 5.1. The average network utilization is normalized to 1.0 for the O-E-O architecture for both of the blocking probabilities of interest. In absolute terms, the average network utilization with 1.0% blocking was roughly 10% higher than with 0.1% blocking.

Two transponder statistics are provided in the table. The first statistic listed, the average number of transponders per demand, reflects the actual transponder count in the design. The second transponder statistic indicates the number of transponders that would have been needed if no regenerations were added to alleviate wavelength

Table 5.1 Results from the backbone network study with 100% dynamic traffic

		O-E-O network (1,200 km reach)	Optical-bypass-enabled network (2,500 km reach)
Results at 0.1% blocking	Normalized average utilization	1.0	0.98
	Average number of transponders per demand	8.42	2.60
	Average number of transponders per demand (not counting the excess regeneration)	8.42	2.57
Results at 1.0% Blocking	Normalized average utilization	1.0	0.98
	Average number of transponders per demand	8.47	2.64
	Average number of transponders per demand (not counting the excess regeneration)	8.47	2.56

contention; i.e., this is the number of transponders required simply based on optical reach. The difference in the two numbers is a measure of how much wavelength contention was encountered.

The optical-bypass-enabled network achieved 98% of the utilization of the O-E-O network, indicating that wavelength contention caused little excess blocking. (The 90% confidence intervals are on the order of ±1%.) To attain this high utilization, a relatively small amount of regeneration was added to alleviate wavelength contention. With 0.1% blocking, the average number of transponders per demand increased from 2.57 to 2.60 due to the added regeneration, which is about a 1% increase. For 1.0% blocking, the average number of transponders increased by 3% due to the added regeneration. (With a higher allowable blocking rate, the network is, on average, more heavily loaded, such that wavelength contention arises more often. Thus, a greater percentage of extra regenerations were needed at 1.0% blocking as compared to 0.1% blocking.)

Even with the extra regeneration, the optical-bypass-enabled network required less than one third of the number of transponders per demand needed in the O-E-O network. The average path distance of the successfully routed demands was about 1% longer in the optical-bypass-enabled network as compared to the O-E-O network, indicating that there were not major differences with respect to fairness of the demands that were accepted. (The path distance used in this calculation is the shortest possible path between the source and destination, which may be different from the path actually followed by a demand.)

5.11.2 Metro Network Study

A similar study was performed on a metro-core network with an interconnected-ring topology, as shown in Fig. 5.8. (For simplicity, we use the terminology “metro

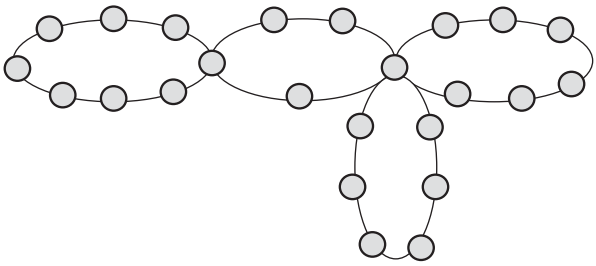


Fig. 5.8 Metro-core network used in the study

network" in the remainder of the section.) In this study, there was one fiber pair per link with a maximum of 40 wavelengths per fiber. As with the backbone study, the traffic was modeled as unprotected and at the line rate. Approximately 35 % of the traffic was inter-ring, with the remainder intra-ring. The traffic was again assumed to be completely dynamic.

An O-E-O-based design and an optical-bypass-enabled design were performed for the metro network, where the latter design assumed that the optical reach was long enough to eliminate all required regeneration. The criteria for comparison are again the network utilization, as defined for the backbone study, and the average number of transponders per demand. The study focused on the 0.1 and 1.0 % blocking scenarios. In the optical-bypass-enabled design, up to one regeneration could be added per connection for purposes of alleviating wavelength contention.

The results of the study are shown in Table 5.2, where the average network utilization is normalized to 1.0 for the O-E-O designs. (In absolute terms, the average network utilization with 1.0 % blocking was roughly 12 % higher than with 0.1 % blocking.) The optical-bypass-enabled network achieved 94–98 % of the utilization, again indicating that wavelength contention caused little excess blocking. (The 90 % confidence intervals are on the order of ± 1 %.) Regeneration due to wavelength contention resulted in a 3 % increase in the average number of transponders per demand at 0.1 % blocking, and a 6 % increase at 1.0 % blocking. Even with this increase, the optical-bypass-enabled network required less than 40 % of the number of transponders per demand needed in the O-E-O network.

(Note that, in reality, metro-network traffic is likely to require some amount of grooming. Grooming normally occurs in the electrical domain, thereby requiring O-E-O conversion. Thus, some of the regenerations that were added to alleviate wavelength contention potentially would be needed anyway for grooming.)

To demonstrate that optical-bypass-enabled networks can achieve the same level of utilization as an O-E-O network by adding in more regeneration, another optical-bypass-enabled design was performed for the same metro network, with 1 % blocking. Enough regenerations were added in the optical-bypass-enabled design to produce a normalized utilization of 1.0 (instead of 0.94 as in the original design). The average number of transponders per demand increased to 2.23 (from 2.12 in the original design). In this particular scenario, the increase in required transponders was fairly moderate. However, in general, adding enough regeneration to achieve

Table 5.2 Results from the metro network study with 100% dynamic traffic

		O-E-O network	Optical-bypass-enabled network
Results at 0.1% blocking	Normalized average utilization	1.0	0.98
	Average number of transponders per demand	5.50	2.05
	Average number of transponders per demand (not counting the excess regeneration)	5.50	2.00
Results at 1.0% blocking	Normalized average utilization	1.0	0.94
	Average number of transponders per demand	5.72	2.12
	Average number of transponders per demand (not counting the excess regeneration)	5.72	2.00

parity with the O-E-O design, in terms of utilization, may not be a desirable strategy because it requires more equipment and reduces some of the operational advantages afforded by optical bypass.

5.11.3 Study Conclusions

For both the backbone and metro studies, the loss in network utilization due to wavelength contention was relatively small, i.e., on the order of 5% or less. The amount of regeneration added to produce these high utilizations was also relatively small, although it was somewhat higher for the metro case where it was assumed that there were no regenerations needed based on path distances.

To provide insight into why wavelength contention is not a large problem, it is helpful to examine the number of hops in a subconnection. In the backbone network study, the average and maximum number of hops per subconnection were 3.5 and 9, respectively; in the metro study, the corresponding numbers were 3.3 and 9. (These statistics are the number of hops prior to dividing up subconnections due to wavelength contention.) With less than four hops in the average subconnection, it is not very difficult to find a wavelength that is free on all of the hops.

To explore this further, the backbone network study was repeated, but with the restriction that wavelength conversion could not occur when a connection was regenerated. Thus, the same wavelength needed to be used on each link of the end-to-end path. The average and maximum number of hops in an end-to-end path increased to 4.8 and 16, respectively. With this large number of hops on which to find a free wavelength, combined with no longer having the option to add regenerations to alleviate wavelength contention, the normalized average utilization with 1% blocking dropped from 0.98 to 0.72, which is a significant drop-off. This demonstrates that optical-bypass-enabled networks can perform poorly if proper design strategies are not employed.

Another factor in achieving high utilization, albeit a much more minor one, is making use of one-step RWA to find a feasible path when the optical-bypass-enabled network was heavily loaded. In the studies, a transformation similar to that described in Sect. 5.4.2 was employed when the network was heavily loaded; i.e., the reachability graph was formed. To speed up the transformation process, rather than searching for a set of regeneration-free paths each time a demand was added, the algorithm maintained a list of up to four possible regeneration-free paths for each node pair (in the backbone network, some node pairs had no such paths). Under heavy load, whenever a demand was added, these predetermined paths were checked for an available wavelength. If a wavelength was found, a corresponding link was added to the reachability graph. The reachability graph was then used to determine the minimum-regeneration path that had a feasible wavelength assignment. This process yielded an approximately 2% increase in the utilization of both the backbone network and the metro network, as compared to a design that always uses a multistep RWA approach; the number of transponders needed was almost identical. (This improvement in utilization is included in the results shown in Tables 5.1 and 5.2.) Thus, a small benefit can be achieved with the one-step approach. (The one-step approach might produce more significant benefits in real-time planning, where the pre-deployed equipment must be taken into account, especially if non-tunable transponders are used.)

It is also worth mentioning that the First-Fit wavelength assignment algorithm was used in the studies. Thus, a very simple assignment strategy is able to produce high utilizations.

The overall conclusion is that algorithms are important in maximizing the performance of a network design. While optical-bypass-enabled networks may require more algorithms than O-E-O networks, these algorithms have already been developed and incorporated in commercial design tools; furthermore, they are not overly complex. These algorithms enable similar utilization as an O-E-O architecture, but with significantly fewer transponders. Because most system vendors provide design tools as part of their system, the burden of handling wavelength assignment is largely removed from the network operator.

5.12 Exercises

5.1. Consider an optical-bypass-enabled network that consists of a linear chain of nodes, where no regeneration is required between any pair of nodes. Assume that a fiber supports W wavelengths. Prove that for any traffic pattern where the routing results in no more than W wavelengths routed on any fiber, it is possible to find a valid wavelength assignment (without using wavelength conversion).

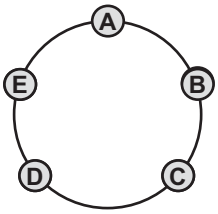
5.2. Consider a network similar to that of Exercise 5.1, except that the topology is a ring rather than a linear chain. Prove that for any traffic pattern where the routing results in no more than W wavelengths routed on any fiber, it is

possible to find a valid wavelength assignment where, at most, one of the nodes is capable of wavelength conversion.

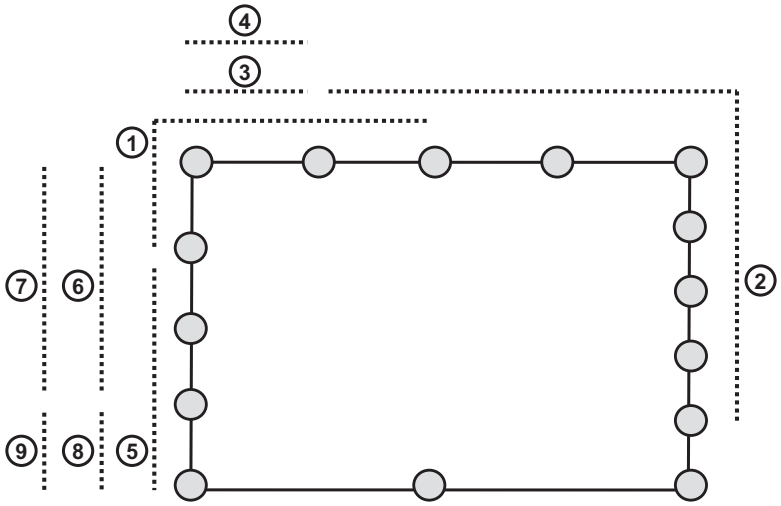
5.3. Assume that a fiber supports W wavelengths and that no more than W wavelengths are routed on any fiber. (a) Propose an algorithm for optimal wavelength assignment on a linear chain of nodes. (The solution to Exercise 5.1 should be helpful here.) (b) Adapt this algorithm for wavelength assignment on a ring (not necessarily an optimal algorithm), where wavelength conversion is *not* permitted. (c) In the ring topology, if no more than W wavelengths are routed on any fiber, is it always possible to find a valid wavelength assignment (again, assuming no wavelength conversion)? How about if each connection is routed over the minimum-hop path?

In Exercises 5.4 through 5.12, assume that the networks are optical-bypass-enabled. Assume that no regeneration is required and that wavelength conversion is not permitted, *unless* otherwise specified.

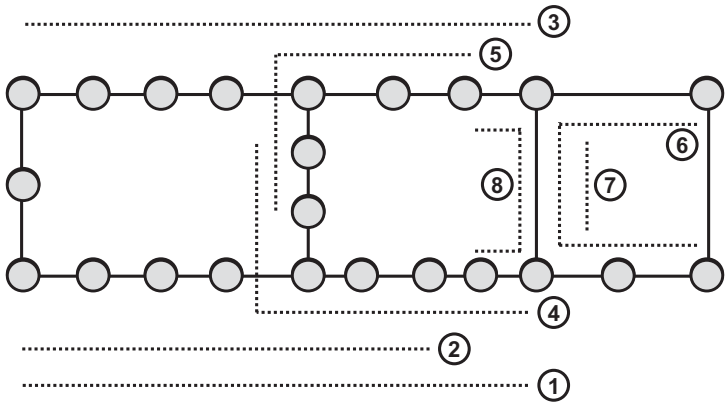
5.4. Consider the five-node ring shown below. Assume that there is one wavelength of bidirectional traffic between each pair of nodes, where the demands arrive in the following order: (B-D), (A-E), (A-D), (C-D), (D-E), (A-B), (B-E), (C-E), (B-C), and (A-C). Assume that the same wavelength is assigned in both directions of a connection. Assume that there is a maximum of three wavelengths per fiber. (a) Assume that wavelengths are assigned as each demand arrives, using First-Fit wavelength assignment. What is the result? (b) If Most-Used wavelength assignment is used instead, what is the result? (c) Next, assume that all of these demands arrive in one batch, such that they can be sorted prior to assigning wavelengths. Assume that they are sorted based on hops, where the demands with the most hops are assigned wavelengths first (break ties alphabetically, i.e., A-C, A-D, B-D,…). Use First-Fit wavelength assignment. What is the result?



5.5. Assume that the network shown below supports a maximum of three wavelengths per fiber. Eight connections are shown, numbered by their wavelength assignment order; i.e., Connection 1 is assigned a wavelength first. Apply the First-Fit and Most-Used wavelength assignment schemes to this example, and report the wavelength assignment results. Based on the results, can you suggest any improvements to the Most-Used algorithm?

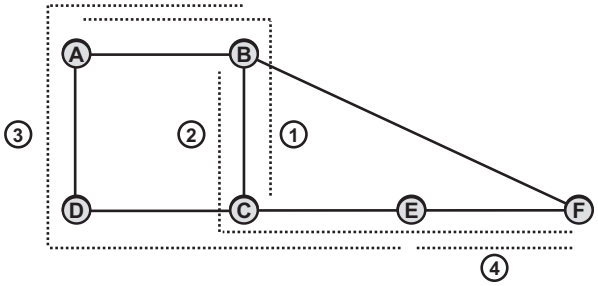


5.6. Assume that the network shown below supports a maximum of three wavelengths per fiber. The eight demands shown are received in one batch. The connections are numbered by their wavelength assignment order; i.e., Connection 1 is assigned a wavelength first. Apply the First-Fit, Most-Used, and RCL wavelength assignment schemes to this example, and report the wavelength assignment results. (In any of the schemes, if there is a tie regarding which wavelength to select, choose the lowest-indexed one.)

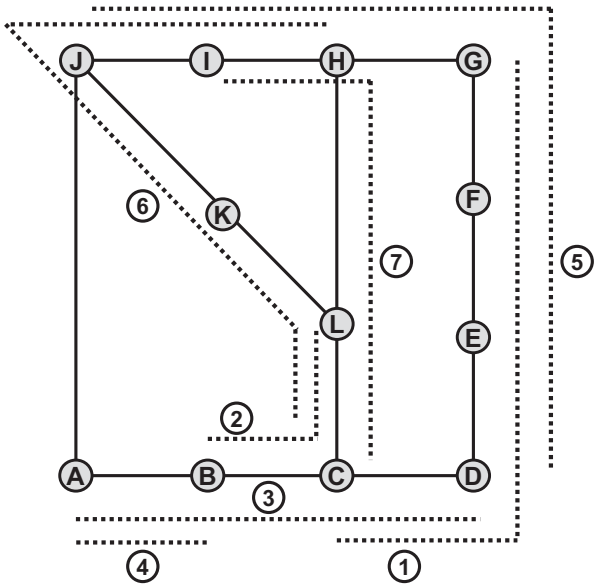


5.7. (a) Transform the wavelength assignment problem for the connection pattern shown below into a graph-coloring problem, where each of the four connections is represented by a node in the graph. (b) Based on the topology (e.g., maximal clique, nodal degrees) of the graph produced in part (a), what

is the minimum number of wavelengths needed for wavelength assignment? (c) Assume that wavelength conversion occurs at Node B. Draw the corresponding graph-coloring problem. (d) From the topology of the graph produced in part (c), can we determine the minimum number of wavelengths needed for wavelength assignment?



5.8. Transform the wavelength assignment problem for the connection pattern shown below into a graph-coloring problem, where each of the seven connections is represented by a node in the graph. This graph will be used in several of the exercises below.

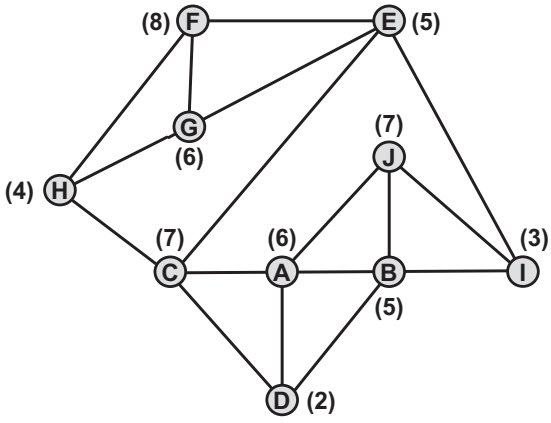


5.9. Apply the *Largest First* graph-coloring scheme to the graph produced in Exercise 5.8. In this scheme, the node with the largest degree is colored first.

That node and its links are then removed from the topology. The node with the largest degree in the remaining topology is colored next. This process continues until all nodes are colored. If multiple nodes are tied for the largest degree, then the node corresponding to the connection with the most hops is selected. Assume that there is a maximum of three wavelengths per fiber (i.e., three colors). In what order are the nodes assigned colors? Combine this ordering with the First-Fit wavelength assignment scheme—what is the result?

5.10. Repeat Exercise 5.9 (i.e., color the graph produced in Exercise 5.8), except use the Dsatur scheme for ordering/coloring the nodes. Start off with the node with the largest degree. If there is a tie, select the node corresponding to the connection with the most hops. Color that node, where coloring is based on First-Fit. Then select the node that has the fewest choices of colors remaining. If there is a tie, pick the node with the largest degree in the "uncolored" topology (i.e., the topology that remains if the colored nodes are removed). If there is still a tie, pick the node corresponding to the connection with the most hops. Color that node, using First-Fit wavelength assignment. Continue the process of selecting the node with the fewest choices of colors remaining. In what order are the connections assigned wavelengths? What wavelength is assigned to each connection? Compare the results to those of Exercise 5.9.

5.11. The figure below represents a node coloring graph corresponding to ten connections (labeled *A* through *J*). The numbers in parentheses indicate the number of hops that are in the corresponding connections. Assume that there is a maximum of three wavelengths per fiber (i.e., three colors). Order the nodes according to the *Smallest Last* graph-coloring scheme. In this scheme, the assignment order of the nodes is built in reverse, from the last node to the first. The node with the smallest degree in the graph is colored *last*. If there are ties, the node corresponding to the connection with the fewest hops is selected. That node is removed from the topology. The node with the smallest degree in the remaining topology is colored second to last. This process continues until all nodes are ordered. What ordering is produced when applied to the figure below? Combine this ordering with the First-Fit wavelength assignment scheme—what is the result?



5.12. Repeat Exercise 5.11, except use the Dsatur scheme for ordering/coloring the nodes (see Exercise 5.10). What is the node ordering, and what wavelength is assigned to each connection? Compare the results to those of Exercise 5.11.
5.13. From the perspective of wavelength assignment, is it better for an optical-bypass-enabled network to be populated with one fiber pair per link with 40 wavelengths per fiber or two fiber pairs per link with 20 wavelengths per fiber?
5.14. In Sect. 5.7, it was advantageous to assign different wavelengths in the two directions of the bidirectional demands that optically bypassed the degree-three node. Why is this phenomenon more of an issue for nodes of odd degree as opposed to nodes of even degree?
5.15. The ILP-based methodology of Sect. 5.4.4 is based on maximal independent sets. If one has a heuristic to enumerate the maximal cliques of a graph, how can it be used to find the maximal independent sets? (A clique is a set of nodes that is fully connected.)
5.16. *Research Suggestion:* In Sect. 4.6.2.1, several strategies were presented for selecting the nodes to use as regeneration sites. All of these strategies only considered regeneration due to optical reach. However, regeneration can also be used for wavelength conversion. Develop a strategy for selecting regeneration sites where both optical reach and the need for wavelength conversion are taken into account.

References

[BRCM12] O. Bertran-Pardo, J. Renaudier, G. Charlet, H. Mardoyan, P. Tran, M. Salsi, S. Bigo, Overlaying 10 Gb/s legacy optical networks with 40 and 100 Gb/s coherent terminals. J. Lightwave Technol. **30**(14), 2367–2375 (15 July 2012)

[Brel79] D. Brelaz, New methods to color the vertices of a graph. Commun. ACM. **22**(4), 251–256 (April 1979)

[BrKe73] C. Bron, J. Kerbosch, Algorithm 457: Finding all cliques of an undirected graph. Commun. ACM. **16**(9), 575–577 (Sept 1973)

[ChGK89] I. Chlamtac, A. Ganz, G. Karmi, Purely optical networks for terabit communication. *Proceedings, IEEE INFOCOM 1989*, vol. 3, Ottawa, 23–27 April 1989, pp. 887–896

[ChMV08] K. Christodoulopoulos, K. Manousakis, E. Varvarigos, Comparison of routing and wavelength assignment algorithms in WDM networks. *Proceedings, IEEE Global Communications Conference (GLOBECOM'08)*, New Orleans, 30 Nov–4 Dec 2008

[ChMV10] K. Christodoulopoulos, K. Manousakis, E. Varvarigos, Offline routing and wavelength assignment in transparent WDM Networks. IEEE/ACM Trans. Netw. **18**(5), 1557–1570, (Oct 2010)

[CKMV09] K. Christodoulopoulos, P. Kokkinos, K. Manousakis, E.A. Varvarigos, Cross layer RWA in WDM networks: Is the added complexity useful or a burden? *Proceedings, International Conference on Transparent Optical Networks (ICTON'09)*, Ponta Delgada, 28 June–2 July 2009, Paper Tu.A3.3

[CLRS09] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, *Introduction to Algorithms*, 3rd edn. (MIT Press, Cambridge, 2009)

[FaSk13] A.M. Fagertun, B. Skjoldstrup, Flexible transport network expansion via open WDM interfaces. *Proceedings, International Conference on Computing, Networking and Communications (ICNC'13)*, San Diego, 28–31 Jan 2013

[GCPW09] O. Gerstel, R. Cassata, L. Paraschis, W. Wakim, Operational solutions for an open DWDM layer. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'09)*, San Diego, 22–26 March 2009, Paper NThF1

[GNCS09] P. Gurzi, A. Nowe, W. Colitti, K. Steenhaut, Maximum flow based routing and wavelength assignment in all-optical networks. *Proceedings, International Conference on Ultra Modern Telecommunications (ICUMT'09)*, St. Petersburg, 12–14 Oct 2009

[HeBr06] J. He, M. Brandt-Pearce, RWA using wavelength ordering for crosstalk limited networks. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'06)*, Anaheim, 5–10 March 2006, Paper OFG4

[ITU09] International Telecommunication Union, Amplified Multichannel Dense Wavelength Division Multiplexing Applications with Single Channel Optical Interfaces, ITU-T Rec. G.698.2, Nov 2009

[KaAy98] E. Karasan, E. Ayanoglu, Effects of wavelength routing and selection algorithms on wavelength conversion gain in WDM optical networks. IEEE/ACM Trans. Netw. **6**(2), 186 – 196 (April 1998)

[LaKe91] A.M. Law, W.D. Kelton, *Simulation Modeling and Analysis*, 2nd edn. (McGraw-Hill, Inc., New York, 1991)

[MBLV09] S. Melle, G. Bennett, C. Liou, C. Villamizar, V. Vusirikala, Alien wavelength transport: An operational and economic analysis. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'09)*, San Diego, 22–26 March 2009, Paper NThF2

[MKCV10] K. Manousakis, P. Kokkinos, K. Christodoulopoulos, E. Varvarigos, Joint online routing, wavelength assignment and regenerator allocation in translucent optical networks. J. Lightwave Technol. **28**(8), 1152–1163 (15 April 2010)

[Obar07] H. Obara, Bidirectional WDM transmission technique utilizing two identical sets of wavelengths for both directions over a single fiber. J. Lightwave Technol. **25**(1), 297–304 (Jan 2007)

[OzBe03] A. E. Ozdaglar, D. P. Bertsekas, Routing and wavelength assignment in optical networks. IEEE/ACM Trans Netw. **11**(2), 259–272 (April 2003)

[Phil04] M. R. Phillips, Analog optical fiber transmission systems: A comparison with digital systems. *Proceedings, 17th Annual Meeting of the IEEE LEOS*, Puerto Rico, 7–11 Nov 2004, Paper TuB1

[RaSi95] R. Ramaswami, K. Sivarajan, Routing and wavelength assignment in all-optical networks. IEEE/ACM Trans. Netw. **3**(5), 489–500 (Oct 1995)

[SaSi06] A.A.M. Saleh, J.M. Simmons, Evolution toward the next-generation core optical network. J. Lightwave Technol. **24**(9), 3303–3321 (Sept 2006)

[Simm02] J.M. Simmons, Analysis of wavelength conversion in all-optical express backbone networks. *Proceedings, Optical Fiber Communication (OFC'02),* Anaheim, 17–22 March 2002, Paper TuG2

[Simm06] J.M. Simmons, Network design in realistic 'all-optical' backbone networks. IEEE Commun. Mag. **44**(11), 88–94 (Nov 2006)

[SuAS96] S. Subramaniam, M. Azizoglu, A.K. Somani, All-optical networks with sparse wavelength conversion. IEEE/ACM Trans. Netw. **4**(4), 544–557 (Aug 1996)

[VAAD01] W. Van Parys, P. Arijs, O. Antonis, P. Demeester, Quantifying the benefits of selective wavelength regeneration in ultra long-haul WDM networks. *Proceedings, Optical Fiber Communication (OFC'01)*, Anaheim, 19–22 March 2001, Paper TuT4.

[YeLR11] E. Yetginer, Z. Liu, G.N. Rouskas, Fast exact ILP decompositions for ring RWA. J. Opt. Commun. Netw. **3**(7), 577–586 (July 2011)

[ZaJM00] H. Zang, J.P. Jue, B. Mukherjee, A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks. Opt. Net. Mag. **1**(1), 47–60 (Jan 2000)

[ZhQi98] X. Zhang, C. Qiao, Wavelength assignment for dynamic traffic in multi-fiber WDM networks. *Proceedings, International Conference on Computer Communications and Networks (ICCCN'98)*, Lafayette, 12–15 Oct 1998, pp. 479 – 485

Chapter 6
Grooming

6.1 Introduction

As networking technology and services have evolved, one characteristic that has persisted is that much of the traffic requires a service rate that is less than that of a full wavelength. For example, while many backbone networks support 40 or 100 Gb/s wavelengths, most client demands require rates of 10 Gb/s or lower. Furthermore, the wavelength line rate is expected to increase to 400 Gb/s and higher, whereas it is forecast that, for the foreseeable future, more than 90 % of client demands will require rates of 10 Gb/s or below, with almost half of them requiring rates of 2.5 Gb/s or below [Infi12]. Demands at the bit rate of a wavelength are referred to as *line rate traffic* or *wavelength services*; demands at a lower bit rate are referred to as *subrate traffic*.

With Synchronous Optical Network (SONET)/Synchronous Digital Hierarchy (SDH) framing at the physical layer, the wavelength line rate has evolved in accordance with the SONET/SDH rate hierarchy (see Sect. 1.4.1). When used to carry subrate SONET/SDH services, the wavelength partitioning is straightforward. For example, with a SONET-based system, a line rate of OC-*N* carries a maximum of *N* OC-1 (51.8 Mb/s) units. Thus, one OC-192 wavelength can carry, for example, a combination of three OC-48s and four OC-12s. Similarly, an SDH line rate of STM-*N* carries a maximum of *N* STM-1 (155.5 Mb/s) units.

Optical Transport Network (OTN) transport frame rates run from OTU1 to OTU4, which correspond to approximately 2.5, 10, 40, and 100 Gb/s, respectively (see Sect. 1.4.2). These line rates are used to carry OTN services, which range in rate from ODU0 to ODU4. For example, four ODU2s (i.e., 4×10 Gb/s) are mapped into an OTU3 (i.e., 40 Gb/s) frame. The finest-granularity OTN service, ODU0, corresponds to 1.25 Gb/s. To provide a more efficient mechanism for carrying a range of services, OTN also supports ODU-Flex. The more common flavor of ODU-Flex is ODU-Flex-GFP, which supports service rates of $N \cdot 1.25$ Gb/s, for integer N. ODU-Flex-CBR supports an arbitrary client bit rate.

With Internet Protocol (IP) services, there is not a set of fixed rates. The traffic rate can be arbitrary, with fine granularity. Additionally, IP services typically

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4_6,

include bursty best-effort traffic, where there is no pre-negotiated bandwidth dedicated to carrying the traffic.

There are several options for carrying subrate traffic in a network. The simplest approach utilizes a full wavelength to carry a subrate demand, thereby wasting the remaining capacity of the wavelength. The percentage of waste can be quite large; for example, carrying one 10 Gb/s demand in a 100 Gb/s wavelength wastes 90 % of the wavelength capacity. This level of inefficiency is untenable in a network.

The preferred solution is to carry multiple subrate traffic demands in a single wavelength. In one such strategy, known as *end-to-end multiplexing*, subrate demands that have the same source and destination are bundled together to better fill a wavelength. The demands are then routed as a single unit from source to destination. While multiplexing improves the network efficiency, it may still be inefficient if the level of traffic between node pairs is small. A more effective technique is *grooming*, where traffic bundling occurs not only at the endpoints of the demands, but also at intermediate points. Demands may ride together on the same wavelength even though the ultimate endpoints are not the same, providing opportunities for more efficient wavelength packing. The relative merits of multiplexing and grooming are discussed in Sects. 6.2 and 6.3, respectively.

While grooming is an effective means of transporting subrate traffic, it can also be costly. Switches that perform grooming may be expensive and may present challenges in power consumption, heat dissipation, and physical space, which will only be exacerbated as the network traffic increases. This affects how such switches are architecturally deployed in a node, as is covered in Sect. 6.4. Furthermore, for cost and architectural reasons, grooming switches may be deployed in only a subset of the network nodes. Methodologies for selecting the grooming nodes, and strategies for delivering traffic to these sites from the non-grooming nodes, are discussed in Sects. 6.5 and 6.6, respectively.

Given a set of subrate demands, there typically is no single optimal design to groom the demands into wavelengths. For example, one design could favor minimizing cost at the expense of routing demands over very circuitous paths, whereas another design could place a greater emphasis on minimizing path length and reserving capacity for future subrate demands. Such trade-offs are explored in Sect. 6.7.

Similarly, there is no single heuristic grooming algorithm that always produces the “best” results. Rather than covering the spectrum of grooming algorithms that have been developed over time, Sect. 6.8 focuses on one grooming methodology that has produced cost-effective and capacity-efficient designs when applied to realistic optical networks, while maintaining a rapid run time even with a very large number of subrate demands.

Much of this chapter is applicable to both optical-electrical-optical (O-E-O) networks and optical-bypass-enabled networks. As grooming is typically accomplished in the electrical domain, it is advantageous to take regeneration into account when selecting grooming sites in a network with optical bypass. For example, the grooming algorithm should favor grooming a connection at sites where regeneration is required anyway. This philosophy is incorporated in the methodology of Sect. 6.8. In Sect. 6.9, a network study is performed to illustrate various grooming

properties. One of the main results is that even as the amount of grooming increases, a significant amount of optical bypass is attainable, indicating that processes such as O-E-O grooming are compatible with an optical-bypass-enabled network.

There are a few notable trends in the field of grooming research. As indicated above, the scalability of grooming switches has become a challenge, especially for IP routers. While cost and physical size are concerns, the largest impediment is power consumption. One approach to dealing with these scalability issues is by implementing architectural paradigms that decrease the amount of required grooming. A second approach is to move at least some of the grooming from the electrical domain to the optical domain to take advantage of the lower energy-per-bit that typically is provided by optics. Section 6.10 provides an overview of some of the proposed techniques, along with a discussion of the geographic tiers in which such techniques may be best suited. This section is intended to provide a glimpse of how grooming may evolve in the future, as opposed to being a definitive discourse on what carriers will actually implement.

One technique that has been proposed as a means of reducing the amount of required grooming is to divide up the spectrum into arbitrarily sized bandwidth "slices" that better match the client service rates [Jinn08]. This approach falls more broadly under the category of flexible networks; spectral slicing is discussed in Chap. 9.

While this chapter addresses subrate traffic, it is also possible that a network service could require a bit rate that is higher than what it is supported by a wavelength. For example, an IP router may generate 40 Gb/s outputs while the wavelengths carry only 10 Gb/s. This type of bit-rate mismatch is handled via *inverse multiplexing,* where the traffic is carried over more than one wavelength. In many inverse multiplexing implementations, the wavelengths supporting the traffic demand must be contiguous in the spectrum and must be routed over the same path. However, *Virtual Concatenation* (VCAT), the inverse multiplexing scheme approved by the ITU, is more flexible. It allows an aggregate signal to be broken up and routed on different paths on any set of wavelengths [Choy02, BCRV06]. (Multipath routing was covered in Sect. 3.11.)

Note that to avoid using terms specific to a particular service (e.g., OC-192), this chapter, for the most part, will use the explicit service data rate, e.g., a 10 Gb/s demand, or simply a "10G."

6.2 End-to-End Multiplexing

End-to-end multiplexing, where traffic demands with the same source and destination are packed into wavelengths, is a simple means of grouping subrate traffic to better utilize network capacity. Once the subrate demands have been grouped into a wavelength, they can be treated as if they are a single demand; i.e., routing, regeneration, and wavelength assignment can be performed on the bundle, as opposed to considering the individual demands comprising the bundle.

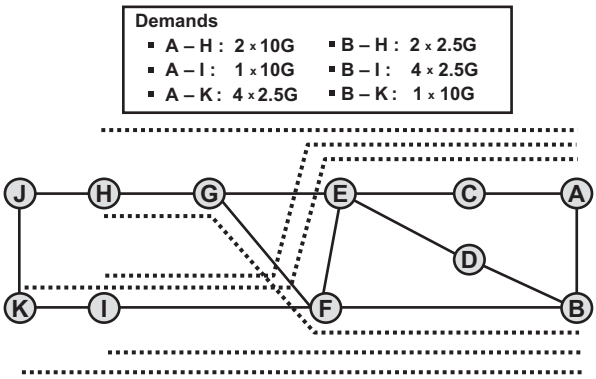


Fig. 6.1 The line rate is 40 Gb/s and the subrate demands are as shown. With end-to-end multiplexing, demands with the same source and destination are bundled together. Six wavelengths are required to carry the traffic in this example, as shown by the *dotted lines*

The network of Fig. 6.1 is used to illustrate end-to-end multiplexing. The line rate is assumed to be 40 Gb/s, and the subrate demands are as shown in the box at the top of the figure. Each source/destination pair is grouped separately, yielding the six connections shown by the dotted lines in the figure. For example, one wavelength carries the two 10Gs between Nodes A and H and is thus only 50% full. On average, the six wavelengths are 27% full. Note that if no multiplexing were used, such that each subrate demand is carried on a separate wavelength, the average wavelength fill rate would be roughly 12%.

The multiplexing function is most commonly accomplished via a wavelength-division multiplexing (WDM) transponder equipped with multiple client-side feeds. This is referred to as a multiplexing transponder, or simply, a *muxponder*. For example, many system vendors offer a "quad" muxponder, e.g., a 40 Gb/s transponder with four 10 Gb/s client-side feeds. The price premium for a muxponder versus a regular transponder is typically small. Thus, multiplexing is a relatively cost-effective operation.

When multiplexing traffic together, it is important to ensure that the individual subrate demands are compatible. For example, for purposes of meeting a certain quality of service (QoS), some demands may have a requirement that they not be routed on certain links in the network. If demands are multiplexed together where each has a different set of "forbidden" links, then it may be difficult to find a path that satisfies all of the constituent demands.

As a second example, typically some demands require protection whereas others do not. If such demands are multiplexed together and the multiplexed unit is protected, then all of the demands in the bundle will be protected, whether or not it is required. It may be ultimately more efficient to reserve space in a partially filled protected wavelength for future protected demands rather than mix in unprotected demands.

Consider the scenario where multiple subrate demands are added at one time to the network. If the amount of subrate traffic between a given source and destination fills more than one wavelength, then a bin-packing algorithm can be applied to judiciously pack the traffic onto wavelengths. The first step is to sort the subrate

demands by source/destination pair and then by their protection level (or other QoS parameter). Within each source/destination/protection class, the demands are then sorted in order from highest bit rate to lowest bit rate. Assume that the demands in a particular class will be bundled into K groups, numbered 1 through K, where each group can contain no more than a line-rate's worth of traffic. Each demand, starting with the highest bit rate demand, is added to the lowest numbered group that still has room for it. Each resulting group is then multiplexed onto a wavelength.

This multiplexing strategy is equivalent to the *First Fit Decreasing* bin-packing methodology [GaJo79]. With SONET/SDH rates, where each successively higher data rate is an integral multiple of the previous one, this strategy packs the traffic onto the minimum number of wavelengths (also see Exercise 6.2). With more arbitrarily sized subrate demands, as is characteristic of IP services, this strategy typically produces no more than about 25 % more than the minimum number of wavelengths (although in the worst case, it can produce close to twice the minimum) [Dosa07].

After the groups are formed, the network planner may choose to combine groups with the same source and destination but with different protection levels if the fill levels of the groups are low. Again, while this may be more efficient for the current set of demands, it may be ultimately less efficient when demands are added in the future.

The efficacy of multiplexing clearly depends on how much traffic there is between each pair of nodes relative to the line rate. If the level of traffic is low, then the wavelengths will be poorly filled, resulting in inefficient network utilization. Even with high levels of traffic between node pairs, there may be inefficiently filled wavelengths. For example, assume that a node pair generates nine 10 Gs in a system with a 40 Gb/s line rate. With two wavelengths 100 % full, one wavelength will be only 25 % full. Furthermore, as networks evolve, more node pairs may begin to generate traffic; the initial traffic level between these node pairs may be small, leading to poorly filled wavelengths.

Another potential source of multiplexing inefficiency is traffic churn, where demands are periodically established and torn down. Consider two 40 Gb/s quad-muxponder cards at a node, where four 10 G clients feed each card. Assume that all eight 10 Gs have the same destination. If two 10 Gs on each card are torn down, the 40 Gb/s wavelengths are only half full. One could combine the four remaining 10 Gs onto a single wavelength; however, this requires either manual intervention or an edge switch in order to move two of the 10 G clients to the other muxponder. (Note that moving a 10 G service to a different card would disrupt live traffic unless a "make-before-break" operation can be implemented.)

In summary, while end-to-end multiplexing is a simple and relatively cost-effective option, the resulting network efficiency may be diminished by its relative inflexibility. Furthermore, carriers tend to dislike the use of muxponders due to their rigid partitioning of traffic, their need for manual intervention, and their lack of an inherent protection mechanism. All of these concerns are addressed by the grooming process and the associated grooming switches, as covered next.

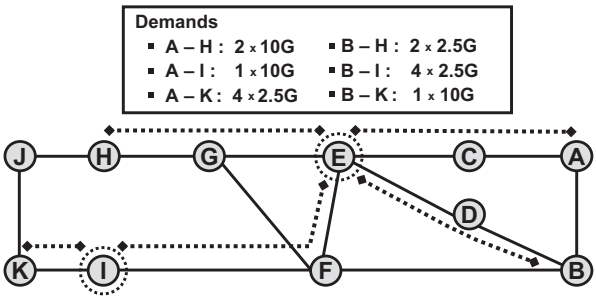


Fig. 6.2 The network and demand set are identical to that of Fig. 6.1. Grooming is used, with intermediate grooming at Nodes *E* and *I*, to pack the wavelengths more efficiently. Five grooming connections are formed as shown by the *dotted lines*

6.3 Grooming

While the multiplexing process bundles demands into wavelengths end-to-end, grooming allows re-bundling of wavelengths to occur at intermediate nodes of a connection. Grooming attempts to form well-packed wavelengths between two particular grooming sites as opposed to between the ultimate source and destination of the subrate demands. Thus, subrate demands with different endpoints may be bundled onto the same wavelength. Furthermore, the other subrate demands with which a given subrate demand is bundled may change at various points along its path.

Grooming is illustrated in Fig. 6.2, where the network and the demands are identical to what was shown in Fig. 6.1. One possible grooming strategy is illustrated in the figure. A single wavelength is used to carry all of Node A's demands to Node E, regardless of the ultimate destination. Similarly, a single wavelength carries all of Node B's demands to Node E. It is assumed that grooming equipment is deployed at Node E, such that the wavelengths can be "broken apart" and then reconstituted using different groupings. One wavelength produced by Node E carries all of the demands destined for Node H, regardless of the original source. A second wavelength from Node E carries all of the demands destined for either Node I or Node K. At Node I, which is also equipped with grooming equipment, further processing occurs. The demands with a destination of Node I are dropped at this node, whereas the remaining demands are packed into a wavelength and transmitted to Node K.

Several measures can be used to compare this grooming design with the multiplexed design of Fig. 6.1. First, the grooming design requires five connections in contrast to the six connections required for multiplexing. Second, the groomed wavelengths are on average 75% filled, in comparison to 27% for the multiplexed wavelengths. Finally, in terms of wavelength-link units, grooming requires 9 units whereas multiplexing requires 21 units. (One wavelength utilized on one link constitutes one wavelength-link unit.) By any of these measures, grooming produces a more efficient design. Further comparisons of multiplexing and grooming efficiency are included as part of the network study in Sect. 6.9.

Grooming is accomplished through the use of specialized grooming switches, which provide more flexibility and numerous operational advantages as compared to muxponders. A grooming switch allows the subrate traffic that originates at a node to be packed into wavelengths along with subrate traffic that is transiting the

node. The switch can automatically repack service demands into wavelengths as traffic patterns change. Furthermore, grooming switches typically have built-in protection mechanisms.

The required grooming equipment depends on the type of traffic. SONET and SDH traffic demands are groomed in SONET and SDH grooming switches, respectively, where the switch granularity depends on the vendor and the application. For example, two common granularities for SONET grooming switches have historically been OC-1 and OC-48, where the line rate is typically OC-192 or OC-768. (Switches with an OC-1 granularity are often referred to as operating at the DS-3 level. A DS-3 is a 45 Mb/s signal that is used to carry telephony traffic; it is mapped into a 51.8 Mb/s OC-1 signal.) The switch granularity is the smallest data rate at which the subrate demands can be "mixed-and-matched." Consider a SONET switch with an OC-48 granularity and assume that four OC-12s are delivered to the switch as a single OC-48 bundle. Because the switch granularity is coarser than an OC-12, the four OC-12s must stay bundled together; it is not possible to swap out some of the OC-12s and combine them into different OC-48 bundles. A grooming switch with OC-1 granularity, however, would allow such an operation. In general, finer switch granularity improves the network utilization, but also results in switches that are larger and more costly.

OTN switches are analogous to SONET/SDH switches, with the switch granularity typically being an ODU0.

A very different flavor of grooming device is used for IP traffic, i.e., an IP router. IP routers generally operate on the granularity of a packet or a flow, where, for simplicity, a flow can be considered a consecutive sequence of packets between the same two router endpoints. IP routers are more complex than SONET/SDH or OTN grooming switches because, in addition to grooming, they also perform per-packet or per-flow routing. In comparison, SONET/SDH and OTN switches are circuit based, where the switch ports are configured for the duration of the connection.

While grooming can significantly increase network efficiency and automate subrate-traffic management, the equipment needed to implement grooming is significantly more complex and costly than a muxponder. The economic and scalability issues affect how grooming equipment is deployed both within a node and across a network. In addition to cost and operational issues, grooming also requires algorithms in order to be effective. These aspects of grooming are explored in the next several sections.

6.4 Grooming-Node Architecture

This section examines how a grooming node should be architected, taking into account cost and scalability issues. For simplicity, the term "grooming switch" is used in the text to represent any grooming device, including an IP router; however, the figures are labeled with "grooming switch or router" to emphasize that the architecture applies to multiple types of grooming devices.

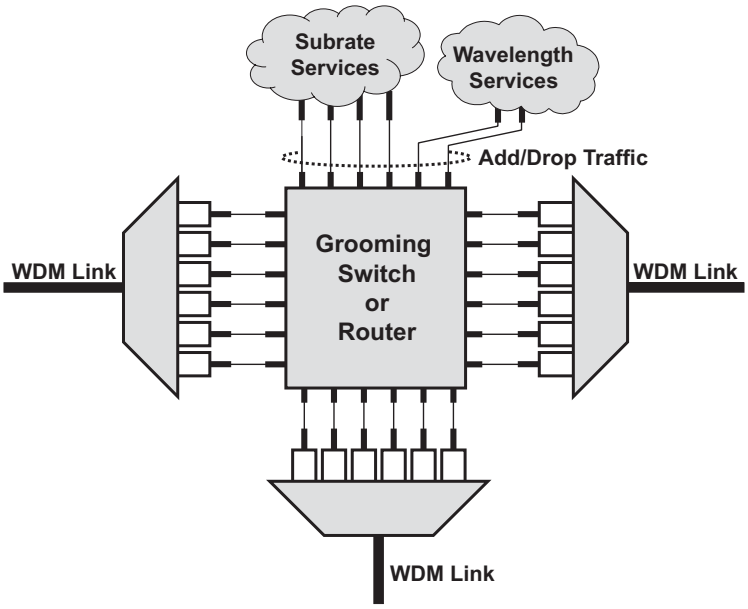


Fig. 6.3 *Grooming switch or router* at the nodal core. All traffic entering the node, whether transiting traffic or add/drop traffic, is processed by the grooming switch or router

6.4.1 Grooming Switch at the Nodal Core

In the first nodal architecture considered here, the grooming switch serves as the "core" network switch. In this architecture, all network traffic entering the node is processed by the grooming switch, as illustrated in Fig. 6.3. The grooming switch operates in the electrical domain such that all of the switch ports are equipped with short-reach interfaces. (Grooming in the optical domain is discussed in Sect. 6.10.) WDM transponders are required for all network traffic entering the node, regardless of whether the traffic is just transiting the node. This architecture is similar to the O-E-O architecture discussed in Chap. 2 (see Fig. 2.6), and has all the attendant scalability issues discussed there, e.g., cost, physical size, power consumption, and heat dissipation.

Furthermore, the fact that the grooming switch operates on a granularity that is typically much finer than a wavelength exacerbates the situation. Not only does the transiting traffic undergo O-E-O conversion, but it also unnecessarily "burns" the grooming resources. Consider the case where the grooming device in a node is an IP router, and consider a wavelength that is carrying IP traffic, none of which is destined for the IP router at the node. With the architecture of Fig. 6.3, not only are transponders and router ports needed for this traffic, but also the contents of the wavelength are unnecessarily processed by the IP router. The need to process all traffic entering the node results in an excessively large IP router (as well as excess

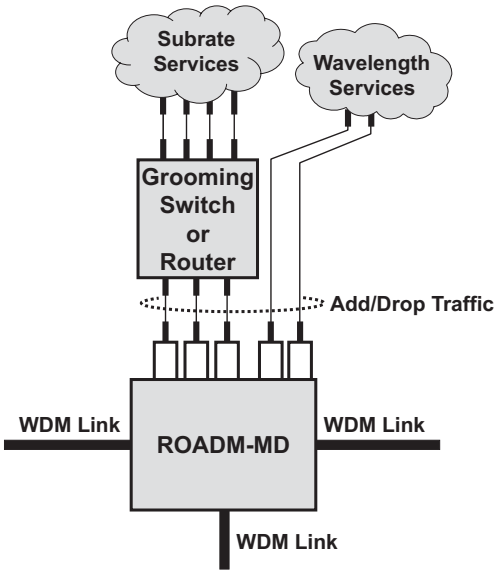


Fig. 6.4 *Grooming switch or router* deployed at the nodal edge with a wavelength-level switch at the core. The wavelength-level switch can provide optical bypass, as shown here. Only the *subrate services* that need to be groomed at the node or that need to be added/dropped at the node are processed by the *grooming switch or router,* yielding a more scalable architecture

delay). Routers, and grooming devices in general, tend to be costly, such that this architecture is likely untenable as the network traffic level increases.

A further inefficiency arising from this architecture is that all nodal add/drop traffic enters the grooming switch, even the wavelength services. Such services do not require grooming because they already fill a wavelength; thus, some amount of grooming resources are wasted on such traffic.

6.4.2 Grooming Switch at the Nodal Edge

A more scalable grooming-node architecture is shown in Fig. 6.4. Here, the core switch at the node is a wavelength-level switch, with the grooming switch serving as an edge switch. In the configuration shown in the figure, the wavelength-level switch enables optical bypass; i.e., it is a reconfigurable optical add/drop multiplexer (ROADM) or multi-degree ROADM (ROADM-MD). This allows any traffic transiting the node to remain in the optical domain so that no transponders or electronic switch ports are required for this traffic.

Wavelengths carrying subrate demands that are either destined for the node or being further groomed at the node are directed by the core switch to the grooming switch (both types of traffic can be considered "drop" traffic, as shown in the figure). Thus, the grooming switch is used only for the traffic that actually needs to be groomed at the node. As compared with Fig. 6.3, the grooming switch is appreciably smaller in this architecture, yielding a more cost effective and scalable node. Depending on the traffic pattern, there may be greater than a 50% reduction in the required switch capacity and number of grooming ports.

Unless otherwise stated, this is the grooming-node architecture assumed in the remainder of this chapter.

Note that the wavelength services originating at the node feed directly into the core switch, thereby avoiding the grooming switch. If the core switch is a *non-directionless* ROADM-MD, then, as discussed in Sect. 2.9.4, edge configurability is not supported. The direction in which a wavelength service can be routed is solely determined by the transponder into which it feeds. If greater routing flexibility is required, then the wavelength services could be passed through the grooming switch so that these services can be directed to different WDM transponders as required. It is not ideal to "burn" grooming switch ports for wavelength services, but if the amount of traffic needing this flexibility is moderate, this may be acceptable. Another option is to deploy a small wavelength-level edge switch, e.g., a fiber cross-connect, to provide edge configurability for these wavelength services. It may also be desirable to pass the output of the grooming switch through the fiber cross-connect, as that may allow fewer grooming ports to be deployed. The fiber cross-connect would essentially break the one-to-one relationship between grooming ports and transponders such that a grooming port could access any transponder (also see Fig. 2.19a and the associated discussion).

If the core switch is a directionless ROADM-MD, then edge configurability is built-in and this particular issue does not arise.

Note that, in principle, the core switch could be an O-E-O switch with wavelength granularity. This still allows the transiting traffic and the add/drop wavelength services to bypass the grooming switch. Although the combination of an O-E-O wavelength-level switch and a grooming switch is more scalable than the architecture of Fig. 6.3, the amount of electronics is still likely to be an impediment to continued network growth, as was discussed in Chap. 2.

One possible enhancement to the architecture of Fig. 6.4 is to integrate the WDM transceivers on the grooming switch, as shown in Fig. 6.5. Integrating transceivers with a general switching element was discussed in Sect. 2.13. Clearly, the outputs of the transceivers must meet the technical specifications of the transmission system. With the integrated-transceiver solutions that are commercially available, either a single vendor provides both the grooming switch and the transmission system, or separate vendors collaborate to ensure compatibility.

The integration could be carried one step further such that the wavelength-level switch and the grooming switch are integrated in one box, i.e., a switch with dual switch fabrics. This is exemplified by the Packet-Optical Transport Platform discussed in Sect. 2.14. One fabric operates at the wavelength level and the other at the subrate level. Ideally, the switch ports are not tied to a particular fabric, so that a port can flexibly direct a wavelength to either fabric depending on its contents.

6.4.3 Intermediate Grooming Layer

The nodal architectures discussed in Sects. 6.4.1 and 6.4.2 are shown again in Fig. 6.6a, b, respectively, as part of an evolutionary sequence of grooming architec-

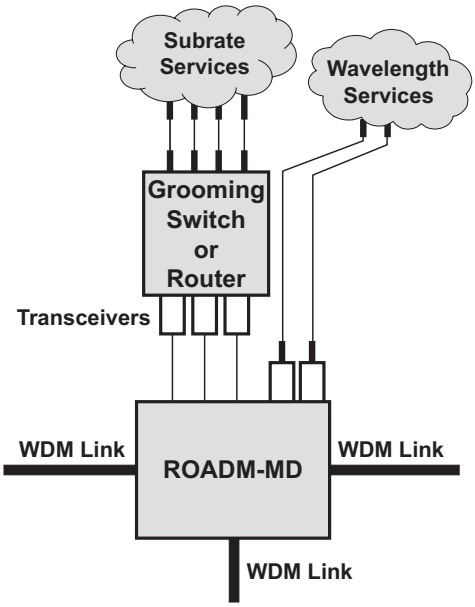


Fig. 6.5 This architecture is similar to that of Fig. 6.4 except that the wavelength-division multiplexing (*WDM*) *transceivers* are integrated on the *grooming switch or router*

tures for IP traffic. With respect to IP grooming, another architecture is emerging among carriers, where an intermediate grooming layer is placed in between the IP and the optical layers [Elby09a, MaDo09]. The resulting nodal architecture is illustrated in its most basic form in Fig. 6.6c, where the intermediate grooming layer is assumed to be OTN. The philosophy is that, in most circumstances, IP traffic is processed by an IP router when it enters and exits the network, but any further grooming of the traffic occurs in the OTN switch. Thus, an end-to-end connection

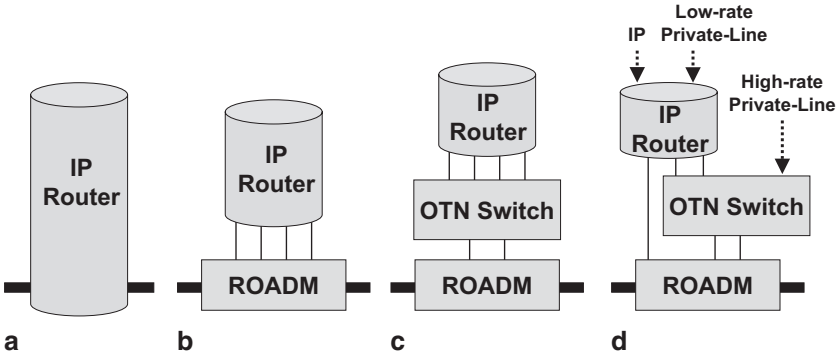


Fig. 6.6 Nodal evolution: **a** All traffic is processed by the *IP router*. **b** A *ROADM* allows optical bypass of the *IP router*. **c** Much of the grooming is moved to an intermediate *OTN switch*. **d** Improved layering to reduce the amount of traffic passing through both an *IP router* and an *OTN switch*. (Wavelength services are not shown)

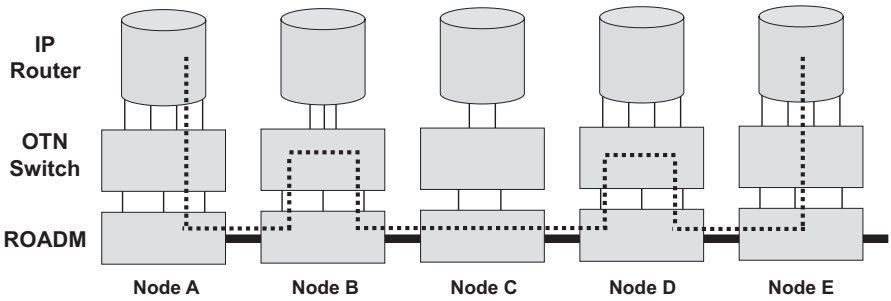


Fig. 6.7 A connection passes through the *IP Router,* the *OTN switch,* and the *ROADM* at the ingress and egress points. At *Nodes B* and *D,* the *OTN switch* is used to bypass the *IP router.* At *Node C,* the *ROADM* is used to bypass both the *OTN switch* and the *IP router*

may look like that shown in Fig. 6.7. The connection originates at Node A, passing through both the IP router and OTN switch. The connection bypasses the IP layer at Nodes B, C, and D; additionally, it bypasses the OTN layer at Node C. Finally, it terminates in the OTN switch and the IP router at Node E.

The main motivations for this architecture are reduced cost and power consumption. OTN switches tend to be much less costly than IP routers. Furthermore, on a per bit/sec basis, an OTN switch consumes about 20 % of the power that an IP router does [TaHR10]. Thus, the goal is to push as much of the grooming burden into the OTN layer as possible, to reduce the load of the IP layer. A ROADM provides further cost and power benefits. For example, the power consumption of a ROADM, on a per bit/sec basis, is roughly two orders of magnitude lower than that of an OTN switch [Tuck11b]. Thus, bypassing both the IP and the OTN layers is favored when possible, subject to the increase in capacity requirements that may result as discussed in Sects. 6.7.2 and 6.7.3.

There is another, more subtle, benefit to adding a sub-wavelength layer below the IP layer, which involves the granularity of IP adjacencies. (An *adjacency*, or an *IP link*, exists between two IP routers if they are neighbors in the IP *virtual* topology.) Because of the distributed nature of IP Protocols, it is important to maintain a relatively fixed IP virtual topology. Adding and removing IP adjacencies may destabilize the network performance. If the IP layer sits directly on top of the optical layer, then creating an adjacency between two routers requires that a wavelength be routed between them. Thus, establishing an "express IP link" between two routers that may not be physically adjacent (in order to bypass any intermediate IP routers) is cost effective only if there is enough traffic to justify the use of this wavelength and the associated router ports.

An intermediate sub-wavelength layer, however, provides a means of creating finer-granularity adjacencies between routers. For example, if OTN is the intermediate layer, the granularity of a router adjacency can be as fine as an ODU0 (i.e., 1.25 Gb/s). Consider the scenario shown in Fig. 6.8, and assume that the wavelength line rate is 40 Gb/s. Assume that it is desired that an IP adjacency be established between each of Nodes A, B, and C and Node E. Additionally, assume that the amount

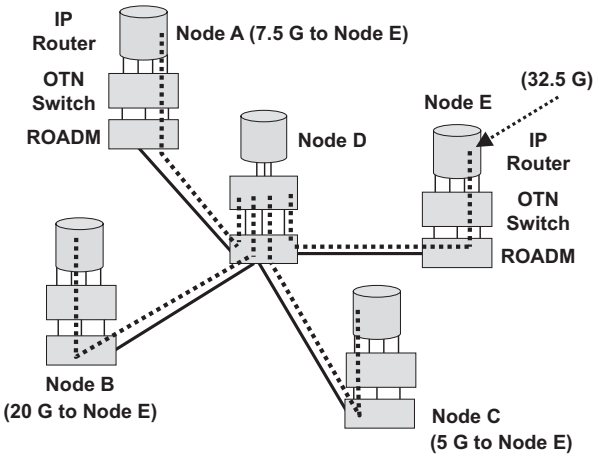


Fig. 6.8 Assume that the line rate is 40 Gb/s. *Nodes A, B,* and *C* exchange less than a wavelength's worth of traffic with *Node E.* However, by using the *OTN switch* at *Node D* to groom this traffic together onto a single wavelength, *Nodes A, B,* and *C* can establish efficient IP adjacencies with *Node E*

of IP traffic that is currently exchanged between Nodes A and E, Nodes B and E, and Nodes C and E does not justify running a full wavelength between any of these node pairs (the assumed traffic amounts are indicated in the figure). Instead, Nodes A, B, and C send their traffic to Node D. The OTN switch at Node D is used to bypass the IP router at this node and grooms all of the traffic destined for Node E onto a single wavelength. This traffic is ultimately delivered to the IP router at Node E (utilizing just one router port at this node), thereby creating express IP links between each of Nodes A, B, and C and Node E. Thus, for example, even though Node C currently exchanges just 5 Gb/s of IP traffic with Node E, an efficient IP adjacency (of size 4×1.25 Gb/s) has been established. As the traffic between Nodes C and E changes, the capacity of the adjacency can be adjusted accordingly, subject to the 1.25 Gb/s granularity. As this example illustrates, the presence of the OTN layer allows the establishment of fine-granularity long-lived IP adjacencies, which engenders greater stability of the IP virtual topology.

Another advantage of adding the intermediate sub-wavelength layer is that utilizing a time-division-multiplexed (TDM) switch (e.g., an OTN switch) for grooming, as opposed to a packet router, should result in less end-to-end latency and jitter. (Latency is the end-to-end transmission delay of a connection; jitter is the variation in this delay over time. The latency and jitter produced by an IP router are typically much more significant than that produced by a TDM switch.) The trade-off historically has been inefficiencies in the TDM layer due to the burstiness of the IP traffic. However, current line rates are so large relative to the service rates that a large number of services are multiplexed together in a circuit. This has an overall smoothing effect on the traffic, which improves the efficiency of TDM with respect to bursty traffic.

One disadvantage of the intermediate sub-wavelength architecture is the added complexity of having to manage another layer. Ideally, there is unified control across the three layers [FHAT12] to coordinate and optimize functions such as routing, bypass, and protection. This is one of the rationales of the Packet-Optical Transport

Platform discussed in Sect. 2.14, where multiple switching fabrics are included in one box. It is also one of the drivers for *Software-Defined Networking*, one of the major topics of Chap. 8.

Another disadvantage of adding another layer is the inefficiency of passing traffic through both an IP router and an OTN switch at the ingress and egress points of the connection (and possibly at some of the intermediate nodes). Figure 6.6d illustrates a possible alternative architecture that improves upon the layering strategy. In this architecture, services may be handled differently depending on their traffic type and rate. For example, high-rate private-line traffic feeds directly into the OTN layer, whereas lower-rate private-line services enter the IP router. (Private-line services typically require a guaranteed-bandwidth, high-availability connection.) IP traffic enters the network via the IP router; traffic that fills an entire wavelength is sent from the IP router directly to the ROADM. The remaining IP traffic is sent from the router to the OTN layer for further grooming with the transiting traffic. This architecture significantly reduces the amount of traffic processed by both the IP and the OTN layers.

Note that in the architectural evolution depicted in Fig. 6.6, i.e., from (a) to (d), the required size of the IP router decreases relative to the amount of network traffic. However, with network traffic continuing to exhibit explosive growth, IP router scalability is still a challenge, as covered in Sect. 6.10.

6.5 Selection of Grooming Sites

For economic reasons, most carriers do not deploy a grooming switch in every network node. The nodes without a grooming switch typically must *backhaul* their subrate traffic to nearby grooming nodes. If too few grooming sites are deployed in a network, there may be excessive backhauling, leading to circuitous end-to-end paths. Furthermore, the few grooming switches may be quite large and the links feeding into the grooming sites may become congested with traffic. If too many nodes have grooming switches, there are likely to be underutilized switches, resulting in unnecessary cost. From experience with actual metro-core and backbone networks, selecting about 20–40% of the nodes to be grooming sites produces designs that are efficient from both a cost and a network-utilization perspective. However, carriers may deploy switches at more nodes to provide greater flexibility, forecast tolerance, and reliability.

Selecting the nodes in which to deploy grooming switches is usually performed as part of the initial network design phase, before any traffic is provisioned in the network. The network topology and the traffic forecast are used to assist in selecting the grooming sites. Several factors should be considered in this process. First, a node that generates a lot of subrate traffic is a natural location at which to put a grooming switch. Otherwise, there will be a large amount of traffic to backhaul to other sites, which may be inefficient. Another important factor is the geographic location of the node. Nodes near the center of the network or nodes that lie along heavily trafficked routes are favored for grooming, as it is likely to be efficient to direct subrate traffic to these sites. Furthermore, higher-degree nodes (i.e., those

with several incident links) are also good candidates for grooming. Such junction sites provide a good opportunity to "mix-and-match" the subrate traffic coming from many links so that efficiently packed wavelengths are produced. One strategy for indirectly capturing these various criteria is to first perform a test routing of the forecast subrate traffic using, for example, shortest-path routing. The total subrate traffic tentatively routed on the links incident on a node can then be used as one of the metrics to assist in determining the set of grooming nodes. (This is related to the *betweenness centrality* metric, which measures the fraction of shortest path routes that pass through a particular node [GOJK02].)

With optical-bypass-enabled networks, another factor that should be considered is the amount of regeneration that is likely to occur at a node. Grooming is normally performed in the electrical domain, so that traffic that is groomed is automatically regenerated as well. By deploying grooming switches at sites where a large amount of regeneration may be required anyway, the overall amount of electronics in the network can be reduced further.

One should also consider the proximity of the non-grooming nodes to those that do support grooming. For example, a possible goal for an optical-bypass-enabled network design may be to select the grooming nodes such that the non-grooming nodes are able to backhaul their traffic without requiring any regeneration along the backhaul path. Thus, in deciding whether a node should be a grooming site, one can consider the number of other nodes that can be reached from it via a regeneration-free path (although this is typically not the most critical factor in selecting grooming sites).

Another factor that may be considered is redundancy, especially with regard to IP routers. Some carriers choose to designate backbone nodes as IP grooming sites in pairs; i.e., two geographically close nodes are both equipped with core routers. (This is in addition to having two core routers *in* each grooming node for purposes of redundancy and facilitation of maintenance activities.)

Each node in the network can be ranked with respect to the above criteria. Nodes that are ranked highly in two or more categories or that are ranked very highly in one category are generally good nodes to choose as grooming sites.

These criteria are used to generate an initial list of grooming sites. As part of the network design process, a few iterations can be run, where a small number of nodes are added or removed as grooming sites to check their effect on network cost and capacity (using the traffic forecast). The results can be used to fine-tune the final selection of grooming nodes.

Two other strategies for selecting the grooming nodes, which are adapted from hierarchical grooming proposals, can also be considered. These are described next.

6.5.1 Hierarchical Grooming

Selecting just a subset of the nodes to be equipped with grooming switches implicitly establishes a grooming hierarchy in the network. Hierarchical grooming has been more formally proposed as an effective grooming architecture in Chen et al. [ChRD08] and Chen et al. [ChRD10]. In these proposals, a grooming hierarchy is

explicitly created, where the network nodes are partitioned into clusters, with one node in each cluster selected as the hub. The bulk of the *inter-cluster* subrate traffic is first directed to the hub corresponding to the source node's cluster; this traffic is then routed to the hub corresponding to the destination node's cluster and from there to the ultimate destination. Most of the *intra-cluster* subrate traffic is groomed in the hub as well. Some of the inter-cluster and intra-cluster traffic can be routed more directly if the amount of traffic between nodes is high enough.

In the hierarchical schemes considered in Chen et al. [ChRD08] and Chen et al. [ChRD10], all nodes are equipped with grooming switches. Nevertheless, there is a clear parallel between the hierarchical grooming methodology and the architecture where only a subset of the nodes have grooming switches. Thus, the two strategies investigated in Chen et al. [ChRD08] and Chen et al. [ChRD10] for selecting the set of hub nodes can be adapted for selecting the set of grooming sites.

6.5.1.1 K-Center Approach

The hub-selection scheme proposed in Chen et al. [ChRD08] is based on the *K*-center problem, where the objective is to find a set of *K* nodes in a graph (call them centers) that minimizes the maximum distance between any non-center site and the nearest center site. Greedy algorithms are often used to find approximate solutions to this problem. At each stage of the algorithm, let D_i equal the distance of non-center node *i* to the closest center node that has been selected thus far. The non-center node with the maximum D_i is added to the list of center nodes. The algorithm can be run multiple times, where in each run, a different node is selected as the very first center; the best solution over all of the runs is then chosen.

For the purposes of selecting nodes to equip with grooming switches, we are interested in factors other than just distance to the grooming nodes. Thus, the algorithm can be modified such that at each step, the non-grooming nodes that rank in the top, say, 10% of D_i are candidate nodes to be added as grooming sites. Of these nodes, a metric is used that captures: the amount of subrate traffic at the node, the degree of the node, the geographic suitability of the node, and the number of other nodes that have a regeneration-free path to this node. One can devise various metrics based on these factors. This process continues until a preset number of grooming nodes have been selected.

If the goal is to ensure that every non-grooming site has a regeneration-free path to a grooming site, then a dominating-set-based approach to the *K*-center problem (e.g., [MiRo05]) can be adapted for this purpose. (This is related to the connected-dominated-set methodology for determining regeneration sites; see Sect. 4.6.2.)

6.5.1.2 Link Capacity Approach

In Chen et al. [ChRD10], there is a greater emphasis on link capacity to determine the grooming hubs, to ensure that there is not excessive congestion around the hubs.

Consider choosing a node to serve as a hub on which to build a new cluster. Of the nodes that have not already been assigned to a cluster, the algorithm selects the one with the most remaining capacity on its links (remaining after taking into account any directly routed traffic). With this node now designated as a hub, the next step is to build up its associated cluster, using the following strategy. Consider all unassigned nodes that are directly attached to the cluster being created. Add the node that will maximize the ratio of intra-cluster traffic to inter-cluster traffic. If there is a tie, select the node that results in the smallest cluster diameter. (The diameter is the maximum number of hops in the shortest-hop path between any two nodes in the cluster.) The cluster can keep growing until some maximum number of nodes have been added, or until the ratio of intra-cluster to inter-cluster traffic falls below a threshold, or until the inter-cluster traffic exceeds some threshold relative to the link capacity connecting the cluster to the remainder of the network.

This methodology can be similarly applied to the problem of selecting the set of nodes to equip with grooming switches. It also can be used to determine to which grooming site a non-grooming node should backhaul its traffic. Because of the emphasis on link capacity, the scheme tends to favor selecting nodes with high degree as grooming sites. More effective designs are produced if both the link capacity and the expected amount of subrate traffic on each link (based on a test routing of the forecast traffic) are considered.

6.5.1.3 Hierarchical Grooming and Optical Bypass

Implicitly employing hierarchical grooming through the designation of a limited number of nodes as grooming sites, or explicitly implementing hierarchical grooming, is advantageous for achieving efficiently packed wavelengths while still maintaining a high degree of optical bypass. This can be appreciated by examining the idealized network shown in Fig. 6.9. The network is composed of 36 nodes, arranged in a 6×6 grid. It is assumed that the line rate is 40 Gb/s and that there is one 2.5 Gb/s demand between every pair of nodes. The network is partitioned into "supernodes" composed of four nodes each, as indicated by the dashed-line circles in the figure. The notion of a supernode was introduced in Simmons et al. [SiGS98] and Simmons and Saleh [SiSa99] and is analogous to the clusters of Chen et al. [ChRD08]. Note that exactly 40 Gb/s worth of traffic is exchanged between each supernode. Thus, the inter-supernode traffic can be perfectly packed on a wavelength, such that no further grooming is needed. These wavelengths can optically bypass any intermediate node, subject to the optical reach. One node (or possibly more) within each supernode is designated as the grooming site (or the hub). While there is little to no optical bypass for the intra-supernode traffic, this represents only a small proportion of the total wavelength-links of traffic in a large network.

Although a simple example, it demonstrates that even in a network with relatively low-rate traffic, electronic grooming can be compatible with optical bypass. This is further exemplified in the study using a realistic network in Sect. 6.9.

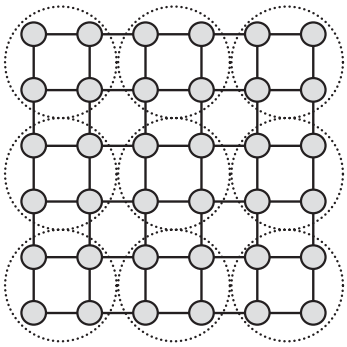


Fig. 6.9 The 36-node grid is partitioned into 9 supernodes, each with 4 nodes. Assuming a line rate of 40 Gb/s and one 2.5 Gb/s demand between every pair of nodes, then exactly one wavelength's worth of traffic is exchanged between each pair of supernodes. Thus, the inter-supernode traffic is well packed with no need for intermediate grooming, thereby providing opportunities for optical bypass

6.6 Backhaul Strategies

If only a subset of the network nodes are equipped with a grooming switch, then the remaining nodes with subrate traffic either use end-to-end multiplexing to carry their subrate traffic or they backhaul their subrate traffic to a grooming node. If the latter option is used, the non-grooming node is said to "*home*" on a grooming node; the grooming node is referred to here as the "*parent*" node. It is important to consider how the subrate traffic is being delivered from the higher networking layers (i.e., the client layers) to the optical network. If the traffic is packed into wavelengths on the client side without any regard to the ultimate destination, then the non-grooming node will generally send all of its subrate traffic to one particular grooming node (or two such grooming nodes for improved reliability, as discussed below). If the higher networking layer performs some grooming of its own such that the subrate traffic enters the optical network already having been grouped according to its intended parent node, then the non-grooming site may distribute the traffic to multiple grooming nodes. Additionally, the non-grooming site may use end-to-end multiplexing if there is a large amount of subrate traffic that is destined for a particular destination node.

There are several criteria that may be used to determine on which node, or nodes, a non-grooming node should home. (Some of these were enumerated in Sect. 6.5.1.2, in relation to forming clusters in hierarchical grooming.) Distance is certainly one key criterion, where the shorter the backhaul distance, the more favored a grooming node is as a parent node. The expected destination of the subrate traffic may play a role as well. If the bulk of the subrate traffic at a non-grooming node is destined for sites to the West, then selecting a parent grooming node to the West may be advantageous to produce more efficient routing. The maximum size of a grooming switch at

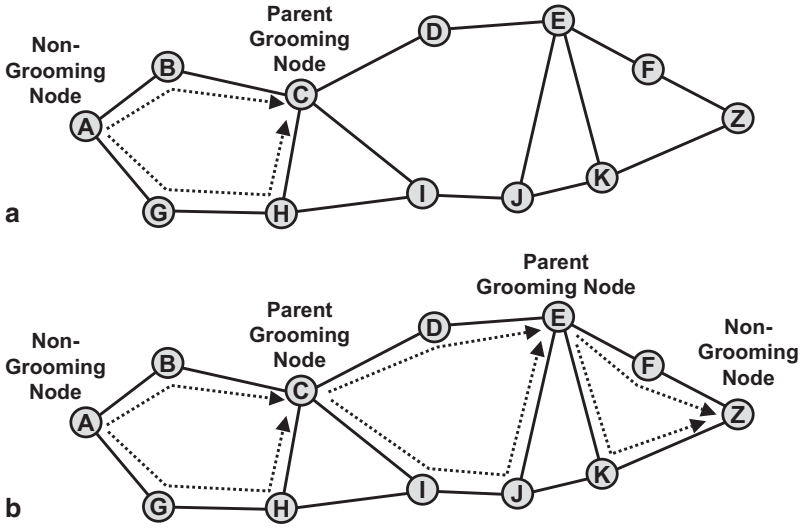


Fig. 6.10 **a** Node *A,* which is not equipped with grooming equipment, homes on just a single grooming node, Node *C*. **b** An end-to-end path from Node *A* to Node *Z,* both of which home on just a single *parent grooming node,* is protected against failures except at Nodes *C* and *E*

a node may also need to be considered. If too many non-grooming nodes home on the same grooming node, the required grooming switch size may be too expensive.

Reliability is another key consideration in backhauling the subrate traffic. There are two schemes that are generally used to provide protection for the backhauled traffic. In one scheme, a non-grooming node homes on a single parent node, but the subrate traffic is routed on diverse paths to the parent node. This is illustrated in Fig. 6.10a where Node A is a non-grooming site that homes on Node C. After the traffic is delivered to Node C, it can be treated as if the traffic originated at that node. This is a relatively simple scheme to implement, and it does provide protection for the path between Nodes A and C. However, the traffic is vulnerable if Node C, or the grooming switch at Node C, fails.

Using this backhauling scheme, a protected end-to-end path may look as shown in Fig. 6.10b. The path extends from Node A to Node Z, where Node Z is a non-grooming node that homes on Node E. Both Nodes C and E are points of vulnerability in this scheme.

A more robust scheme is to backhaul the traffic to diverse parent nodes, as shown in Fig. 6.11a. Here, Node A sends traffic to both of its parent nodes, C and H, over diverse paths. Note that this is an example of where diverse routing from one source to two destinations is desired, as covered in Sect. 3.7.3.

The advantage of the scheme in Fig. 6.11 is that protection is provided against a grooming-node failure. The disadvantage is that the grooming costs are greater due to the redundancy. It also results in two independent end-to-end paths for the subrate traffic. For example, a protected end-to-end path between non-grooming nodes A and Z is shown in Fig. 6.11b. Node Z homes on Nodes E and K. One path

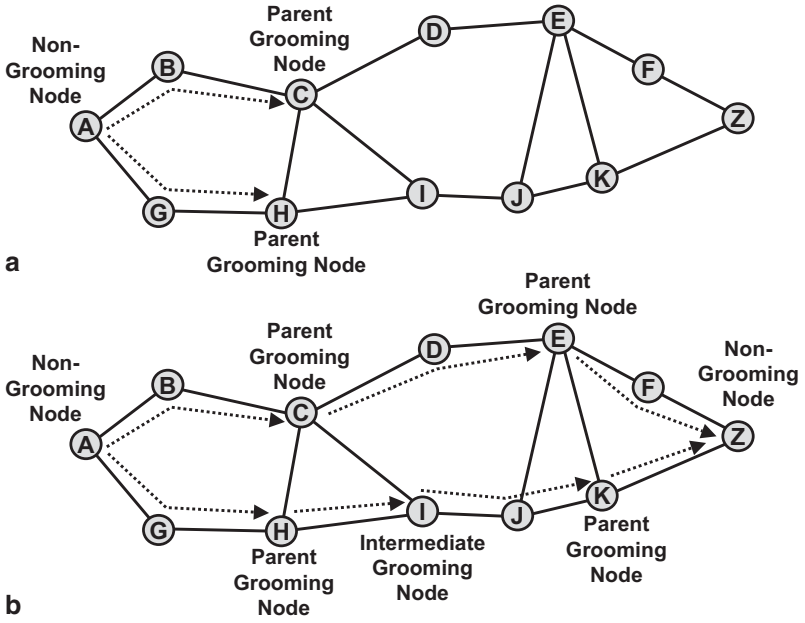


Fig. 6.11 **a** Node *A* homes on both Nodes *C* and *H* and delivers its subrate traffic to both parents over diverse paths. **b** An end-to-end protected path from Node *A* to Node *Z.* There are no single points of failure along the path. It is assumed that Node *I* is used for intermediate grooming

makes use of grooming nodes C and E, and the other path utilizes grooming nodes H, I, and K (it is assumed that Node I is used for intermediate grooming). This may be more difficult to manage as compared to having only one set of grooming nodes for the connection.

Protection of subrate demands is covered more fully in Chap. 7.

6.7 Grooming Trade-offs

The process of grooming subrate traffic often presents design trade-offs in factors such as capacity and cost. The yardstick with which a grooming design is evaluated may depend on the preferences of the carrier or may depend on the circumstances under which the design is being performed. For example, in a network that is very heavily loaded, link capacity may be the most important factor. Adding a small amount of extra grooming equipment may be justified if it results in not needing to add a second fiber pair along a link. As another example, a carrier may issue a network-planning exercise in order to evaluate the equipment costs of various system vendors. In this scenario, from the viewpoint of the vendors, producing the lowest cost design may be the most important factor.

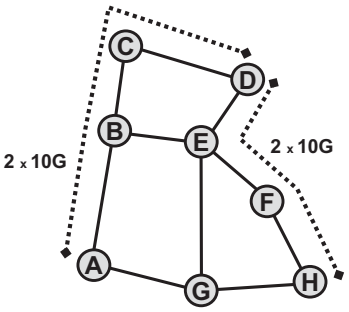


Fig. 6.12 The line rate is 40 Gb/s. One wavelength between Nodes *A* and *D* carries two *10G*s. One wavelength between Nodes *D* and *H* also carries two *10G*s. If a new 10G demand is added between Nodes A and H, it can potentially be carried in the existing wavelengths, rather than establishing a new connection directly along *A–G–H*. This is the lower cost option (at least in the short term) but utilizes a longer path that burns future capacity

6.7.1 Cost Versus Path Distance

The first grooming trade-off illustrated here is between network cost and the distance over which a subrate demand is routed. Consider the network shown in Fig. 6.12 and assume that the network line rate is 40 Gb/s and assume that all nodes are equipped with grooming switches. Assume that there are four existing 10Gs provisioned in this network. Two 10Gs are between Nodes A and D, and are routed in a single wavelength along the path A-B-C-D. The other two 10Gs are between Nodes D and H, and are routed in a single wavelength along the path D-E-F-H. Both of these wavelengths are 50% full.

Assume that a new 10G demand request arrives, between Nodes A and H. One option is to route this new demand along the most direct route A-G-H. This option utilizes a grooming switch port at both Nodes A and H, and utilizes a wavelength along the A-G-H path, which would be 25% full. (When counting grooming switch ports in this section, only the network-side ports are included, not the client-side ports.) If the network is O-E-O based, then there is also a regeneration required at Node G.

The second option is to carry the new 10G demand using the two wavelengths that have already been deployed. One wavelength carries the demand from Node A to Node D. At Node D, the traffic enters a grooming switch that directs this 10G to the wavelength running between Nodes D and H. This solution does not utilize any additional grooming switch ports or transponders, and is thus of lower cost than the first option. Additionally, it does not require provisioning any new wavelengths, so from that viewpoint, it requires less capacity; i.e., there are a total of six wavelength-links occupied with this option as compared to eight wavelength-links with the first option.

However, if capacity is evaluated based on a finer granularity than a wavelength, e.g., a 10G, then the solution that directly routes the new demand over A-G-H utilizes a total of 14 10G-links whereas the second option utilizes 18 10G-links. If it is expected that there will be future subrate demands that will require the bandwidth between A and D or between D and H, then it may be desirable to directly route the 10G over A-G-H, with the expectation that ultimately this will result in a lower-cost network.

Another factor to consider is that the second option routes the new demand over a longer path and requires one intermediate grooming. Thus, this option is somewhat more vulnerable to failure with respect to the new demand and will result in greater latency. Furthermore, if this traffic is IP, such that this option results in the traffic being processed by an intermediate IP router at Node D, then this may result in additional jitter.

A carrier would need to weigh these various factors to determine how the new demand should be carried.

6.7.2 Cost Versus Capacity

The second grooming trade-off considered here is between network cost and capacity. Figure 6.13 illustrates a small linear network, where it is assumed that the line rate is 40 Gb/s. Assume that two new 10Gs are being added to the network, one between Nodes A and C and one between Nodes B and D.

One grooming option is shown in Fig. 6.13a, where each 10G is simply assigned to a new wavelength. This utilizes one grooming switch port at each of Nodes A, B, C, and D. The number of utilized wavelength-links is four.

A second option is shown in Fig. 6.13b. Here, Nodes B and C are used as intermediate grooming sites. The 10G from Node A is delivered to the grooming switch at Node B, which bundles it with the 10G originating at Node B. Both 10Gs are carried on a single wavelength to Node C, where they are delivered to the grooming switch at that node. The 10G destined for Node C drops at the node, whereas the remaining 10G is carried in a wavelength to Node D. This utilizes a total of six grooming switch ports: one at Node A, two each at Nodes B and C, and one at Node D. Thus, two more ports are utilized as compared to the first option. However, the number of utilized wavelength-links is three as opposed to four, because just one wavelength needs to be provisioned on Link BC as opposed to two. If Link BC is heavily loaded, such that reducing the number of utilized wavelengths is important, then the second option, though more costly, may be preferred. Protection resources may also need to be considered, as discussed in Sect. 6.7.3.

As mentioned in the previous section, the addition of intermediate grooming along the paths of the two 10Gs is another factor to consider. The extra grooming of Fig. 6.13b potentially reduces the reliability of the circuits. Additionally, if the intermediate grooming occurs in an IP router, then there may be additional latency or jitter.

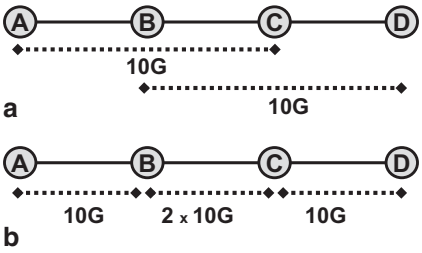


Fig. 6.13 The line rate is 40 Gb/s. Two *10G* demands are added, one between *A* and *C,* and one between *B* and *D.* **a** The two *10G*s are carried in separate connections between their respective endpoints. This option requires four grooming ports and utilizes two wavelengths on Link BC. **b** Nodes *B* and *C* are used to groom the traffic such that there is a single connection between *B* and *C* carrying two *10G*s. This option occupies just one wavelength on Link BC, but it requires a total of six grooming ports

6.7.3 Cost Versus Protection Capacity

The example of Sect. 6.7.2 illustrates only one component of the potential cost versus capacity trade-off inherent in deciding whether to bypass the grooming switch at a node. As Fig. 6.13 shows, utilizing more grooming may lead to better-packed wavelengths, but at the expense of more grooming ports. Another factor to consider when bypassing a grooming switch is the effect on the required protection resources. We specifically focus on IP traffic for this discussion.

In Fig. 6.14a, an IP link is established only between neighboring routers. (As a reminder, an IP link is an adjacency between IP routers; i.e., it is a link in the IP virtual topology, corresponding to one or more physical links in the optical-layer topology.) Consider the MPLS-based Fast Reroute mechanism for IP protection, which makes use of Next-Hop (NHOP) tunnels for link protection, and Next-Next-Hop (NNHOP) tunnels for link and node protection [PaSA05]. Assuming NNHOP is implemented, then if the series of *IP nodes* through which IP traffic is routed is N_1, N_2, etc., and the IP link between N_i and N_{i+1} fails, then the traffic is restored locally on a path from N_i to N_{i+2} (unless N_{i+1} is the destination node, in which case the traffic is restored locally on a path from N_i to N_{i+1}). This is illustrated in Fig. 6.14b, where it is assumed that (physical) Link CD fails, and the traffic that had been routed on the IP link from Node C to Node D is restored from Node C to Node E. The protection path is not explicitly shown in the figure, but is represented by the dashed line.

Next, consider Fig. 6.14c, where in addition to the IP links between physically adjacent routers, there are also two "express IP links" that bypass one or more IP routers, including the router at Node C. If Link CD fails, as shown in Fig. 6.14d, multiple IP links fail, requiring multiple restoration paths. Thus, creating express IP links likely results in greater required protection resources.

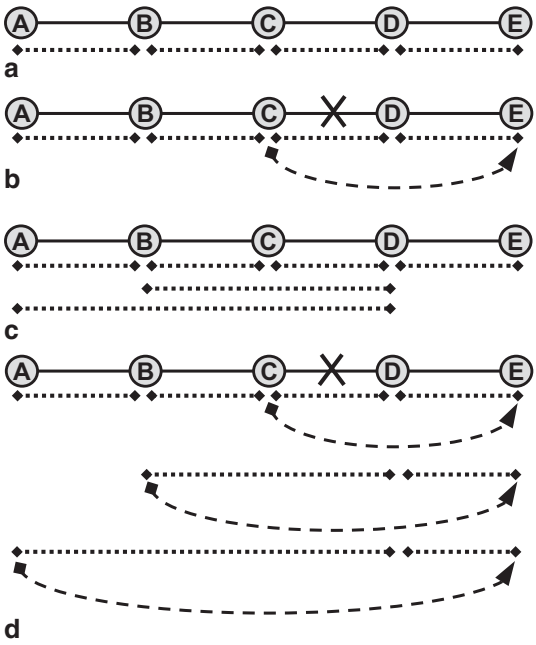


Fig. 6.14 **a** IP links, shown by the *dotted lines,* are established only between neighboring routers. **b** When physical Link CD fails, only one IP link fails. The restoration path (in the *A* to *E* direction), using NNHOP, is represented by the *dashed line.* **c** Two express IP links are created, both of which bypass the router at Node *C.* **d** When physical Link CD fails, three IP links fail, requiring three different NNHOP restoration paths

To evaluate the overall cost versus capacity trade-off further, a network study was performed using Reference Network 1 of Sect. 1.10 [CCCD12]. An IP-over-optical design was performed, with 60% of the IP traffic requiring protection; the remainder was best-effort traffic. An IP link was established between each pair of physically adjacent routers. In addition, express IP links were gradually added to the design, based on the traffic levels between nodes. As expected, the addition of the express IP links resulted in fewer required IP ports, but more required capacity (both working capacity and protection capacity). Using a cost equivalence of one router port to 770 wavelength-km of transport capacity, the minimum cost was produced with 73 express IP links. With this design, the number of router ports decreased by 32% and the wavelength-km of capacity increased by 15%, as compared to a design with no express links. Adding additional express IP links resulted in a more costly design; for example, with 107 express links, the cost was 5% higher.

This study included only capital costs (i.e., the cost of the equipment), not operational costs (i.e., the cost to run the equipment). If operational costs are considered, or if the size of the IP routers is a bottleneck, due to, for example, power consumption, then more express IP links may be warranted.

6.7.4 Grooming Design Guidelines

To better control the grooming process, a carrier may specify certain design guidelines. For example, a limit may be imposed on the number of intermediate grooming

switches through which any given demand can be processed. This may be done for reliability or latency/jitter reasons. Second, a limit may be imposed on the allowable path "circuitousness" for any demand. For example, there could be a guideline that the end-to-end path over which a demand is routed should be no greater than P% longer than its most direct path, for some positive value P. This can be imposed for latency or reliability reasons. Alternatively, it can be specified to serve as a guideline as to the desired balance between reducing cost in the current design at the expense of occupying bandwidth that may be needed for future demands. The higher the value of P, the greater the emphasis placed on reducing current cost. Other restrictions may be placed on the types of traffic that can be carried together in a single wavelength. There could be segregation based on service types, protection types, certain customers, etc.

It is important that the grooming algorithms be flexible enough to enforce any such rules.

6.8 Grooming Strategies

Grooming algorithms have evolved in concert with the network topology and the network traffic. Some of the earliest work on grooming specifically addressed minimizing cost in ring topologies, e.g., Simmons et al. [SiGS98], Gerstel et al. [GeRS98], and Simmons et al. [SiGS99]. However, networks have evolved to mesh topologies, requiring grooming algorithms that work on arbitrary topologies. The algorithms must be flexible with respect to the demand granularities as well. Furthermore, the number of individual subrate demands that may need to be carried by the network can be in the tens of thousands, thereby requiring efficient grooming techniques.

One particular general grooming strategy that has produced cost-effective and wavelength-efficient designs for realistic networks is presented next. More general coverage of grooming can be found in texts such as Zhu et al. [ZhZM05] and Dutta et al. [DuKR08] or tutorial papers such as Dutta and Rouskas [DuRo02], Zhu and Mukherjee [ZhMu03], and Huang and Dutta [HuDu07].

6.8.1 Initial Bundling and Routing

Assume that there are multiple new subrate demands being added to the network at once. The first step is to group the new subrate demands into bundles that contain at most one wavelength's worth of traffic, where all of the demands in a given bundle have the same source and destination nodes. The First Fit Decreasing bin packing scheme described in Sect. 6.2 in relation to end-to-end multiplexing can be used for this purpose. The algorithm must ensure that the demands that are bundled together are compatible; e.g., they cannot have conflicting QoS requirements. If there is just a single new subrate demand, then it is placed in a "bundle" by itself.

After the bundles are formed, they are routed end-to-end using the standard techniques for routing a wavelength service, e.g., alternative-path routing, as described in Sect. 3.5.2. (The source and destination nodes of the demands within the bundle are the endpoints of the bundle as well.) Network load can be used as one criterion for selecting which alternative path to select for a given bundle. The load on a link can be approximated by the wavelengths that have already been provisioned on the link for existing traffic and by the new bundles that have already been routed. (It is only the approximate load because the amount of grooming that will ultimately occur is not known at this point.)

Another factor that needs to be considered when selecting a path is whether the endpoints of the bundle have grooming switches. If one or both of the bundle endpoints do not have grooming capability, then the path chosen must pass through the appropriate "parent nodes" that do have grooming switches. If none of the predetermined candidate paths pass through the desired parent nodes, then the routing can be done in steps. For example, assume that the bundle endpoints are Nodes A and Z, and assume that Node A does not have a grooming switch. The routing must go from Node A to the parent node of A, and then from the parent node of A to Node Z.

At the end of this phase of the grooming algorithm, all bundles have been routed over a tentative path. The tentative path for the bundle can be considered the baseline path for each demand in the bundle. If a different path is considered for a demand in order to improve the grooming, as described below, the new path can be compared to the baseline path to determine whether it is excessively long.

Some of the bundles that are formed may contain a full wavelength, or very close to a full wavelength, of traffic. No further grooming operations need to be done with these bundles. As the amount of traffic grows in the network, the number of full bundles increases as well, so that the grooming process remains scalable.

6.8.2 Grooming Operations

At this point, all bundles are routed end-to-end, similar to what would be done if the traffic were simply being multiplexed. The term *grooming connection* (GC) is used here to refer to a path that is terminated at both ends on a grooming switch. This is illustrated in Fig. 6.15, where a GC extends between Nodes A and E, both of which are assumed to have grooming switches. The intermediate nodes may have grooming switches as well; however, they are not used to process this particular GC. Additionally, note that there may be regeneration along the path of a GC, even in an optical-bypass-enabled network. In the figure, regeneration is occurring at Node C. (If a portion of a bundle's path extends from a non-grooming node to a parent grooming node, then that portion is not a GC and does not need to be considered for the GC combination operations described below.)

Two types of GCs are distinguished here. First, the *existing GCs* encompass those GCs that have already been established in the network. While new subrate demands can be added to an existing GC, subject to its maximum capacity, it is

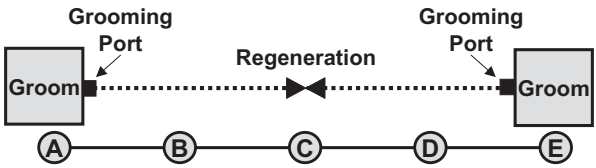


Fig. 6.15 The grooming connection (GC) represented here by the *dotted line* terminates on the grooming switches at Nodes *A* and *E*. A GC may need to be regenerated, as shown at Node *C*. The intermediate nodes, *B, C,* and *D,* may contain grooming switches as well; however, this GC is not processed by them

assumed that an existing GC cannot be rerouted, as that would disrupt existing traffic. Any existing GC that is already filled to capacity can be ignored for purposes of further grooming. Second, there are *new GCs*, formed from routing the bundles containing the new subrate demands. There is more flexibility with the new GCs: New demands can be moved into or out of a new GC; a new GC can be routed over a different path; and a new GC can be split into multiple shorter GCs. (If the grooming switch supports a "make-before-break" feature, then moving or rearranging existing GCs may be possible without bringing down the corresponding demands. For example, if it is desired that an existing GC be re-routed, then a duplicate GC is created and sent over the new route. After a short period of time, the GC on the original route is removed. With this feature, the above distinction between new and existing GCs may not be necessary.)

Each GC occupies one wavelength along each hop of the GC, and utilizes a grooming switch port at either endpoint. Reducing the number of GCs can be beneficial, as it frees up capacity and switch ports, and possibly removes some regeneration equipment. In order to reduce the number of GCs, the next step is to perform various "combination operations." Typically, the operations proceed starting with the new GCs that have relatively low fill and that extend over several hops. In all of the operations described below, for simplicity, it is assumed that the line rate is 40 Gb/s and the demands are 10Gs; clearly, the operations hold for more general scenarios. In all of the examples, *it is assumed that GC 1 is a new GC* (i.e., GC 1 contains subrate demands that have not been provisioned yet), whereas the other GCs can be either new or existing. (Again, as mentioned above, if the grooming switch supports the make-before-break feature, then GC 1 could be an existing GC. With this feature, it may be desirable to combine existing GCs due to network churn that results in partially full GCs.)

The first operation considered is where all of the demands in a new GC are moved into another GC, where both GCs have the same path. This simple operation, illustrated in Fig. 6.16a, allows one GC to be removed. This type of operation often occurs after another operation "chops" a longer GC into smaller GCs, where a resulting GC now aligns with another GC.

A similar operation is where all of the demands from one GC are moved into another GC that has the same endpoints, but a different path. This is shown in

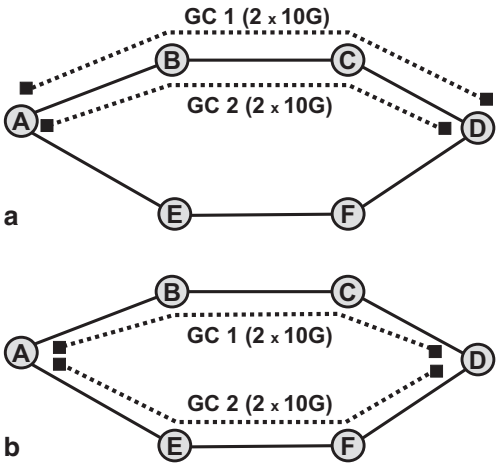


Fig. 6.16 Combining *GC*s with the same endpoints to allow one of the *GC*s to be removed. **a** The two *GC*s have the same path. The two *10G*s from *GC 1* can be moved to *GC 2*. **b** The two *GC*s have different paths. The two *10G*s from *GC 1* can be moved to *GC 2,* assuming this new path is satisfactory for the demands in *GC 1*

Fig. 6.16b. GC 1 is routed along the path A-B-C-D with two 10Gs, and GC 2 is routed along path A-E-F-D with two 10Gs. The demands from GC 1 can be merged in with GC 2, such that GC 1 is removed. In performing this operation, it is necessary to check that the new path is satisfactory for the demands in GC 1.

It may be necessary to perform multiple operations in order to remove a GC. In Fig. 6.17, GC 1 is routed along the path A-B-C-D with two 10Gs, GC 2 is routed along the same path with three 10Gs, and GC 3 is routed along path A-E-F-D with three 10Gs. One 10G from GC 1 is moved into GC 2 and the other 10G is moved into GC 3, allowing GC 1 to be removed. (This assumes that the demands in GC 1 do not need to be carried in the same wavelength and do not need to be routed over the same path.)

In the next operation, a new GC is "split" at an intermediate point, and the demands moved into two shorter GCs. This is illustrated in Fig. 6.18. The demands from GC 1 are placed into both GC 2 and GC 3. While the path is the same, this operation adds an intermediate grooming point for the demands from GC 1; thus, it must be verified that the maximum number of grooming points for a demand, if specified, is not violated. In a variation of this operation, GC 2 and/or GC 3 do not lie along the same path as GC 1. This both adds an extra grooming point and modifies the path. These operations can be taken a step further such that a new GC is split at two points and the demands are moved into three shorter GCs, resulting in two additional intermediate grooming points for the demands and possibly a different path.

The operation that is shown in Fig. 6.19 reduces the capacity requirements, although not necessarily the switch port requirements. Figure 6.19a shows the original setup with GC 1 and GC 2. This requires one wavelength along Links AB and BC and two wavelengths along Links CD and DE, and one switch port at Node A, one at Node C, and two at Node E. Figure 6.19b shows the result of the operation. GC 1 is shortened such that it extends only from Node A to Node C, and the demands in

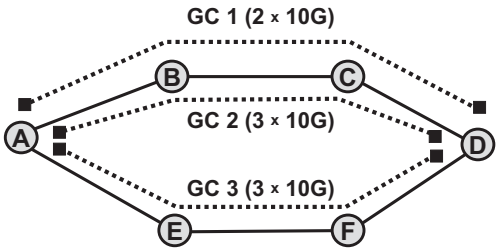


Fig. 6.17 One *10G* from grooming connection (*GC*) *1* can be moved to *GC 2* and the other *10G* can be moved to *GC 3,* allowing *GC 1* to be removed

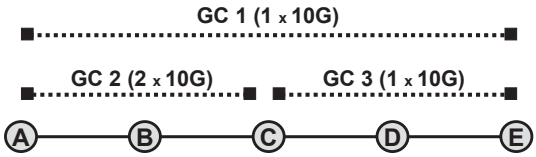


Fig. 6.18 The demands from grooming connection (*GC*) *1* can be moved into both *GC 2* and *GC 3.* This adds another grooming point for the demands in *GC 1*

GC 1 are now carried in both GC 1 and GC 2. This requires one wavelength along each link, although it still requires a total of four switch ports. It also adds another grooming point for the demands in GC 1.

In an optical-bypass-enabled network, it is preferable to perform this operation such that regenerations are eliminated, when possible. For example, in Fig. 6.19, assume that the optical reach is 1,000 km, such that GC1 originally requires one regeneration at Node C. By terminating GC 1 at Node C, this regeneration is removed (or, more precisely, the regeneration is occurring in concert with grooming), thus reducing the network cost. Consider performing an alternative grooming operation on the network of Fig. 6.19, where it is now assumed that a half-filled GC, GC 3, extends from Node D to Node E. Assume that GC 1 is terminated at Node D (instead of Node C), and its demands carried in both GC 1 and GC 3 (instead of GC 1 and GC 2). This is shown in Fig. 6.19c. With this operation, one regeneration is still needed along GC 1, resulting in a more costly arrangement than in Fig. 6.19b. Thus, as this example illustrates, regeneration should be a factor in selecting which grooming operations to perform.

Another operation that saves capacity, albeit at added expense, was illustrated in Fig. 6.13 in Sect. 6.7.2, where the overlapping portions of two GCs are combined. This operation is more favorable to perform when the capacity is tight in the network.

The grooming algorithm can make several passes through the GCs to perform these various operations. It is generally preferable to perform the operations that do not change the path prior to those that do change the path. This allows the demands that ideally should be routed along a certain path to use the GCs that lie along this path, as opposed to using GCs that result in a circuitously routed demand.

In scenarios where demands are added one at a time, the algorithm can be less aggressive in shifting demands to longer paths, in anticipation that future demands will be better suited to be routed along some of these links. Furthermore, it may be

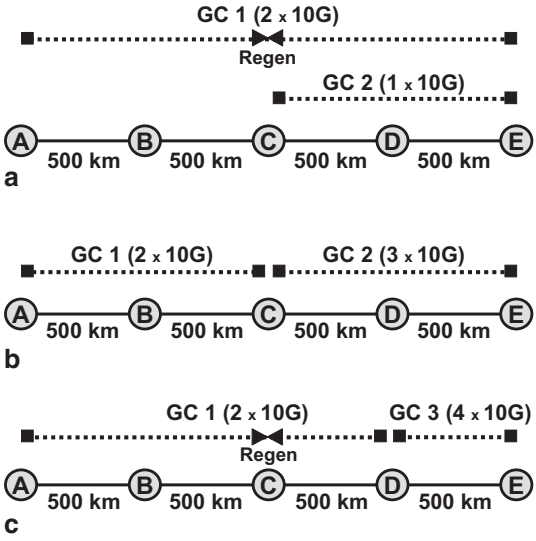


Fig. 6.19 **a** In the original setup, two wavelengths are utilized along *C-D-E.* **b** After grooming connection (*GC*) *1* is shortened and its demands also placed in *GC 2,* only one wavelength is required along *C-D-E.* Assuming an optical reach of 1,000 km, the regeneration at Node *C* is removed. **c** If *GC 1* had been terminated at Node *D* instead of Node *C* and the demands carried in both *GC 1* and *GC 3,* one regeneration would still be required

desirable to proactively divide a new GC into two smaller GCs, even though it provides no current benefit, so that future demands may be more efficiently groomed. These decisions can be initially guided by the traffic forecast and later by the network's traffic history.

The run time of this grooming scheme is very manageable. For example, in a series of tests where these grooming operations were performed on the 60-node Reference Network 2 with 10,000 subrate demands, the run time was a few seconds on a 1.6 GHz PC. As an indicator of the effectiveness, the initial average GC fill rate was 23 % (this corresponds to end-to-end multiplexing); after completing the grooming operations, the average GC fill rate was 85 %. With the number of subrate demands doubled, the grooming run time increased by about 80 %.

After the grooming operations are complete, the new GCs that remain can be treated like wavelength-level end-to-end demands. Regeneration sites on each GC are selected, if necessary, to break the GCs up into subconnections, as described in Chap. 4. Wavelengths are then assigned to the subconnections using the techniques described in Chap. 5.

The grooming methodology described above holds for unprotected and protected demands. When performing the various grooming operations, which may entail shifting the path of a protected demand, it is necessary to ensure that the working and protect paths of the demand remain routed over diverse paths. Furthermore, depending on the protection scheme, the grooming combination operations may allow the working paths of some subrate demands to be bundled with the protection paths of other subrate demands. Protection of subrate demands is specifically addressed in Sect. 7.12.

In real-time grooming scenarios where equipment may be limited, it may be necessary to use a graph model that captures the available equipment and available

wavelengths [ZZZM03]. This is similar to the graph transformation discussed in Sect. 3.6.2, where a graph is constructed to represent the available equipment and capacity in the network. The existing GCs with spare capacity can also be included in such a graph. A grooming design can then be performed by running a shortest-path algorithm on the graph. As discussed previously, this type of detailed resource modeling is more helpful when there are few available resources in the network and it becomes very difficult to find satisfactory paths with the requisite grooming switches, regeneration equipment, and wavelengths. Moreover, when the amount of available resources is relatively small, the size of the transformed graph is accordingly smaller, making these modeling methods more tractable.

6.9 Grooming Network Study

To investigate various quantitative trends related to grooming, a network study was performed using Reference Network 2 of Sect. 1.10 (a 60-node backbone network). The results of any grooming exercise heavily depend on the traffic distribution among nodes, the network topology, and the network line rate. The numbers presented here are indicative of relative trends as opposed to absolute statistics that hold across all networks.

SONET-based traffic was assumed, with a line rate of OC-192, and all demands at either OC-48, OC-12, OC-3, or DS-3 rates. (Note that SONET-specific terminology is used in this section.) Typically, network traffic would include at least some wavelength services as well; however, to focus on the grooming aspects, such services were not a part of this study. An optical-bypass-enabled network was assumed, with an optical reach of 2,500 km. In the grooming procedure, demand paths were allowed to be up to 30 % longer than their baseline (i.e., directly-routed-path) distance or were allowed to be up to 1,000 km in length, whichever was longer.

Three different aggregate network demand scenarios were considered: a total of 1.2 Tb/s of traffic, a total of 2.5 Tb/s of traffic, and a total of 5 Tb/s of traffic. In all scenarios, 50 % of the traffic required protection, which was implemented with 1 + 1 dedicated protection (the protection was implemented at the subrate level, as described in Sect. 7.12.2; two diverse end-to-end paths were allocated for each protected demand). The aggregate network demand is calculated by summing the total bidirectional traffic sourced in the network. Protected demands were counted twice; e.g., a protected OC-48 demand contributed 5 Gb/s to the aggregate demand.

Two different strategies were considered for grooming-switch deployment. In Sect. 6.9.1, all nodes are equipped with a grooming switch, corresponding to a “flat” grooming architecture. In Sect. 6.9.2, only 25 % of the nodes have grooming switches, corresponding to a more hierarchical grooming approach. In either scenario, the architecture of the grooming nodes was assumed to be that shown in Fig. 6.4. As the results indicate, the hierarchical approach, which is more in line with current carrier practice, is much more efficient in packing the traffic into wavelengths.

6.9.1 Grooming Switch at All Nodes

In the first design, a grooming switch was deployed at each node in the network. Figure 6.20 shows a plot of the average number of wavelengths utilized per link as a function of the allowable number of intermediate grooming points per demand, for each of the traffic scenarios. First, consider the results for the 5 Tb/s aggregate demand scenario. With no intermediate grooming, an average of 190 wavelengths were utilized per link. As expected, the number decreased as the allowable number of intermediate grooming points increased. For example, with up to five intermediate grooming points per demand, an average of 42 wavelengths were utilized per link. The average fill rate of the resulting GCs, shown next to each data point on the graph, ranged from 24% with no intermediate grooming to 88% with up to five intermediate grooming points. Most of the grooming benefit was achieved by allowing just two intermediate grooming points per demand, which produced an average fill rate of 76%.

With just 2.5 Tb/s of aggregate demand, more intermediate grooming was required to achieve a similar fill rate. For example, up to three intermediate grooming points were required to achieve an average fill rate of 75%. With 1.2 Tb/s of aggregate demand, up to five intermediate grooming points were required to achieve this fill rate. As could be expected, lower traffic levels required more grooming to produce well-packed wavelengths.

Note that zero intermediate grooming is equivalent to end-to-end multiplexing, indicating the potential inefficiency of this scheme. With no intermediate grooming, the average fill rates were 10, 15, and 24% for the 1.2 Tb/s, 2.5 Tb/s, and 5 Tb/s aggregate demand scenarios, respectively.

Given that grooming occurs in the electrical domain, it is interesting to examine the average optical-bypass percentage in the network as the amount of grooming increased. This statistic represents the percentage of wavelengths entering a node that traverse the node in the optical domain. The average optical-bypass percentage for the 5 Tb/s scenario is shown by the top curve in Fig. 6.20. The percentage of optical bypass decreased by a small amount, from 75 to 70%, as the amount of grooming increased. (Though not shown in the figure, the percentage dropped from 75 to 68% for the 2.5 Tb/s scenario, and from 75 to 64% for the 1.2 Tb/s scenario.) This indicates that much of the O-E-O conversion required by the additional grooming was offset by the reduced need for regeneration due to shorter GCs. Thus, efficient grooming is quite compatible with an optical-bypass-enabled network.

One can also consider an architecture where all subrate traffic is passed through a grooming switch at *every* intermediate node, i.e., the architecture of Fig. 6.3. Thus, on each link, the wavelengths are filled as much as possible. In the 5 Tb/s scenario, this reduced the average utilization by 10% as compared to the scenario where at most five intermediate grooming points are allowed. This indicates that limited intermediate grooming achieves close to the optimal packing for high traffic levels. In the 1.2 Tb/s scenario, grooming at every node resulted in a 30% savings in average link utilization. However, if a network only needs to support this relatively low level of demand, then the network fill rate is low enough that efficient wavelength

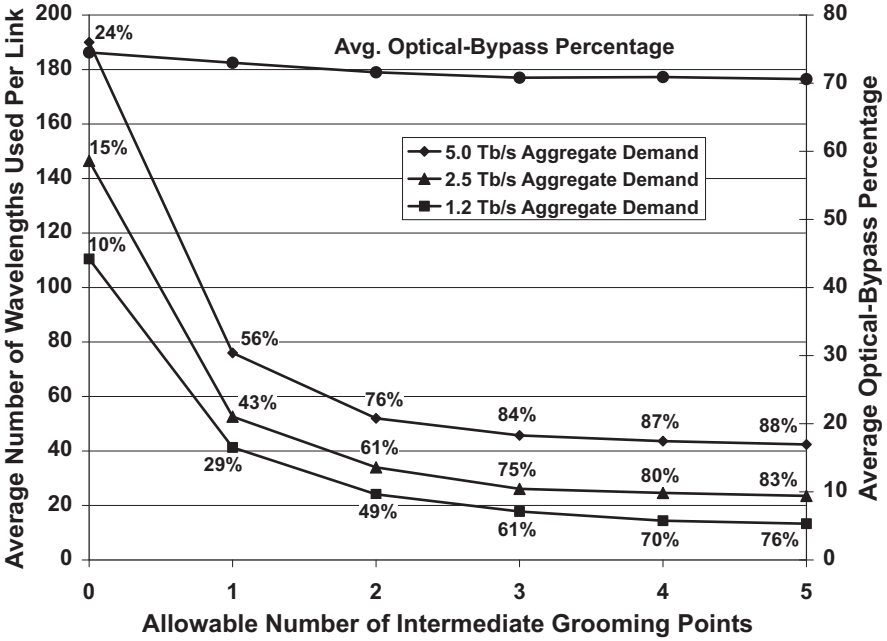


Fig. 6.20 Average link utilization (with 1.2, 2.5, and 5 Tb/s of aggregate demand) and average optical-bypass percentage as a function of the allowable number of intermediate grooming points per demand. The percentages specified next to the data points are the average fill rates of the resulting GCs. The optical-bypass curve is for the 5 Tb/s scenario; the percentages were slightly lower for the other scenarios. All 60 network nodes were equipped with grooming switches in this scenario

packing may not be critical. Thus, the extra cost of grooming at every intermediate node is likely not justified.

6.9.2 Grooming Switch at a Subset of the Nodes

The grooming study was repeated for the same network topology and traffic, but where only 15 of the 60 nodes were equipped with grooming switches. The factors enumerated in Sect. 6.5 were used to select the grooming nodes. The 45 non-grooming nodes backhauled their unprotected traffic to a single grooming node and their protected traffic to two diverse grooming nodes, as in Fig. 6.11. (The backhauling was not counted as intermediate grooming.) The results are shown in Fig. 6.21. After backhauling, the traffic is concentrated at a relatively small number of nodes, resulting in well-packed wavelengths even without any further grooming. For example, in the 5 Tb/s aggregate demand scenario, the *GCs* were 84 % filled, on average, with just end-to-end multiplexing between the grooming sites. This increased to 92 % with up to one intermediate grooming point allowed per demand.

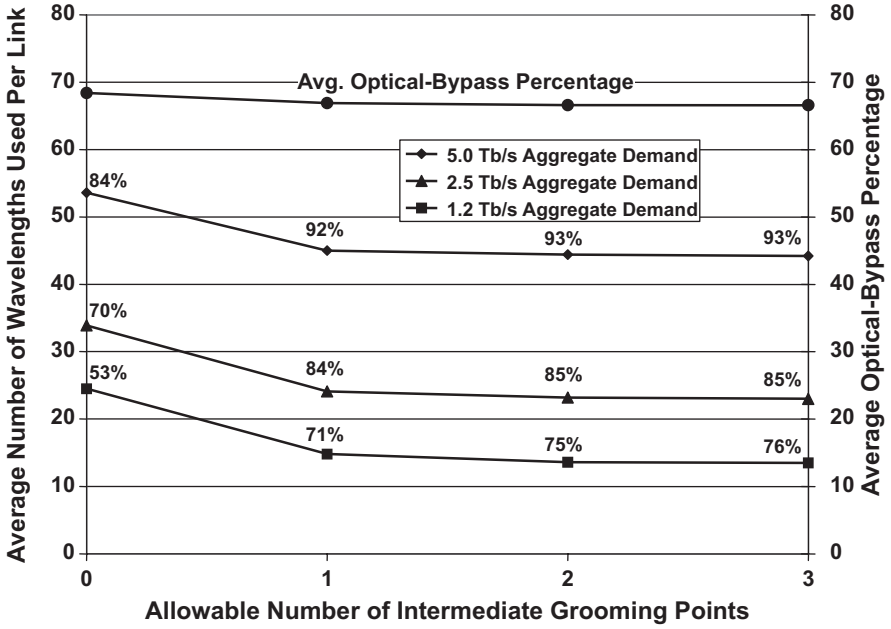


Fig. 6.21 Same as Fig. 6.20, except that only 15 of the 60 network nodes were equipped with grooming switches. Less intermediate grooming was required to achieve high GC fill rates. (Note that the scale of the *left*-hand axis is different from that in Fig. 6.20)

The average optical-bypass percentage was roughly five percentage points lower as compared to the design where all nodes had grooming switches, due to backhauling traffic from a non-grooming node to a nearby grooming node; however, the amount of optical bypass was still high. (The optical-bypass curve shown in Fig. 6.21 holds approximately for all three aggregate demand scenarios.)

The chief motivation for concentrating the grooming at a relatively small number of nodes is cost. It is generally more cost effective to have a small number of switches where the ports are used for well-packed wavelengths rather than have many switches where the ports are used inefficiently. Moreover, the first-deployed cost of some of the switches may not be justified by the level of grooming required at the corresponding nodes.

Note, however, that backhauling results in longer end-to-end path distances. In the study, the paths were roughly 10% longer when grooming switches were deployed in just 15 nodes as compared to when they were deployed in all 60 nodes. Additionally, while it did not occur in this study, concentrating the grooming in a relatively small number of nodes could lead to poor load balancing, where the links near the grooming nodes would be more heavily utilized. Favoring placing grooming switches at nodes with relatively high degree partly mitigates this effect, as it allows the traffic directed to the grooming switch to be spread out over more links.

Deploying grooming switches in just a subset of the nodes leaves the network somewhat vulnerable to major changes in the traffic pattern. For example, if a non-grooming node ends up with a significant amount of subrate traffic, a lot of back-hauling would be needed. It may be necessary to deploy grooming switches at more nodes as the traffic levels increase and traffic patterns change.

6.10 Evolving Techniques for Addressing Power Consumption in the Grooming Layer

As indicated earlier in this chapter, grooming switches are becoming an impediment to scaling up the network traffic, due to issues with cost, power consumption, heat dissipation, and physical size. This is especially true with regard to the power consumption of IP routers. For example, a fully equipped core router, with over 100 Tb/s of capacity, requires hundreds of kilowatts of power [TaHR10]. It is estimated that IP routers are responsible for roughly 80% of the power consumption in core networks (transponders account for roughly 15%) [Gree13]. The problems go beyond simply the cost of the electricity. Continued improvement in system density, which is critical to keeping equipment affordable as bandwidth needs grow, has also been adversely affected by the large power demands. Thermal density is now a limiting design criterion, which has become a significant barrier to improving equipment costs [UCSB13]. Furthermore, the challenges are growing worse: Energy efficiency in IP routers is estimated to be improving at a rate of 10–15% per year [TaHR10], while network traffic continues to grow at more than 30% per year [Cisc13]. For detailed analyses of the energy challenges in today's networks, see Tamm et al. [TaHR10], Zhang et al. [ZCTM10], Tucker [Tuck11a], Tucker [Tuck11b], and Kilper et al. [KGHA12].

The concern over the overall energy consumption of communications equipment has led to an interest in alternative networking paradigms, with IP routing being the target of much of the research. The required size of the IP router can be reduced through the use of optical bypass and/or the addition of an intermediate sub-wavelength grooming layer, as previously discussed. These are approaches that have already been implemented or that are being actively discussed for near-term deployment. This section is focused on long-term grooming research that is more of a departure from how networks are implemented today, to enable continued scalable growth of data traffic.

One avenue of pursuit addresses the scalability challenge architecturally, e.g., with schemes that reduce the amount of required grooming. Sections 6.10.1 through 6.10.3 fall under this category. A second approach is to replace electronics with optics when possible. Optics is more agnostic to data rates than electronics. Thus, as data rates increase, an optical approach may require lower energy-per-bit as compared to the electronic analog. The difficulty lies in that some functions are best performed in the electrical domain, such that the overall benefit may be somewhat

curbed. Sections 6.10.4 through 6.10.7 consider methodologies for grooming in the optical domain.

Note that historically one of the inefficiencies of grooming IP traffic has stemmed from the burstiness of the traffic. The average fill rate of a wavelength may be purposely kept low to leave "headroom" to accommodate the peak burst rates. However, with line rates increasing to 40 Gb/s and higher, a very large number of IP flows can be statistically multiplexed on one wavelength. This has the effect of smoothing the aggregate traffic, thereby allowing much higher average fill rates (also see Sect. 9.2.5 and Exercises 6.6 and 6.7). Some of the schemes discussed below were originally targeted at addressing the issues of bursty data flows. These same schemes tend to reduce the amount of required grooming. With the focus now on energy conservation, these schemes have taken on a new purpose.

It is emphasized that this section is not intended as a predictor of what may actually be implemented in carrier networks in the future. Some of the methodologies below have been studied for more than a decade and are likely still a long way from practical deployment. Additionally, as indicated in the introduction of this chapter, the spectral slicing proposal for reducing the amount of required grooming in the network is covered in Chap. 9 as opposed to here.

The various schemes are, for the most part, discussed in the context of IP traffic; however, some of the schemes hold for more general traffic types that require grooming.

6.10.1 Routing and Grooming with Energy Considerations

A relatively new avenue of research takes energy concerns into account when routing traffic to grooming nodes. For example, assume that solar energy is being used as a partial power source for some nodes in the network. During the daylight hours relative to these nodes, it may be desirable to favor these nodes for grooming traffic. More transport resources may be utilized to accomplish this routing, but with transmission equipment one to two orders of magnitude more efficient than IP routers, on an energy per bit basis, a net benefit should be realized [KGHA12].

Another proposal is to attempt to route traffic away from certain nodes so that some of the deployed grooming equipment can be powered down or put in a "sleep mode" [ChMN12, Idzi13]. This assumes that the network has been dimensioned to meet peak demands, such that there are unused resources during non-peak times. One concern is that excessive power cycling may be detrimental to the lifetime of the equipment. An alternative is to proactively route the traffic away from certain grooming nodes to take advantage of rate adaptation. With reduced load, the grooming switch can *potentially* run at a more energy-efficient rate. This may be preferable to powering down the switch entirely.

Another strategy is to decrease the amount of grooming without modifying how traffic is routed in the optical layer. At times of relatively low load, the amount of optical bypass at a node can be increased such that some subset of the traffic no longer is processed by the grooming switch at that node. While this will increase the

number of utilized wavelengths on certain links, it will utilize fewer grooming ports such that a net power benefit should be realized. (The smallest unit that can be powered down is likely a card, which typically contains multiple grooming ports. Thus, to optimize this strategy, traffic should be assigned to ports to maximize the number of cards that can be powered down under low-load conditions [LuSS13].) The resulting increase in link load should not create bottlenecks, as this technique would be performed only at times of reduced network traffic. (However, if the grooming layer is IP, then it may be desirable to limit such reconfigurations if they change the IP virtual topology.)

All of these strategies are consonant with dynamic networking, which is the topic of Chap. 8.

6.10.2 Selective Randomized Load Balancing

The next approach is a departure from typical network routing that results in less required grooming. The *Selective Randomized Load Balancing (SRLB)* scheme [ShWi06] is based on a traffic model where the endpoints of the traffic streams may rapidly change, but the aggregate amount of traffic sourced/sunk at each node remains fairly constant. This is known as the *hose* traffic model [DGGM02]. Consider a network with N nodes, where M of these nodes are selected as hub (grooming) sites. In SRLB, the traffic sourced at any node is randomly delivered to one of the M hub sites, independent of the ultimate destination (although, all traffic in a given IP flow is routed to the same hub). The traffic is groomed at a hub node and then sent to the destination.

One advantage of this two-phase routing approach is its smoothing effect on variable traffic, thereby making the network amenable to slow circuit switching. With the hose-traffic-model assumption, the connections between any of the N nodes to any of the M hubs are relatively constant in size even though the endpoints of the traffic may be highly variable. Additionally, because traffic passes through just one hub, the aggregate size of the IP routers in the network is smaller as compared to a scheme that utilizes several stages of intermediate grooming. Furthermore, network jitter is likely reduced because a flow is processed by just one router. The disadvantage of the scheme is that the end-to-end path may be significantly longer than the shortest possible path. Additionally, the performance of the scheme is tied to the validity of the hose model.

6.10.3 Optical Flow Switching

Optical flow switching (OFS), as described in Chan et al. [ChWM06] and Chan [Chan12], is another architecture that reduces the amount of required electronic grooming. End-to-end wavelength connections are requested from the network to carry a data flow. Scheduling mechanisms are used to coordinate the assignment

of resources to meet these connection requests. Once the connection is established, there is no need for further processing of the data flow within the network. More specifically, this traffic does not consume IP routing resources within the network. It is assumed that the flow duration is long enough such that the process of scheduling the connection and configuring the resources prior to the flow transmission does not represent a significant inefficiency.

To better ensure that the bandwidth of the wavelength connection is well utilized, aggregation in the optical domain can occur in the access or metro-core portions of the network. The aggregation network could be based on a switched architecture, requiring very fast switches. Alternatively, a passive broadcast solution could be implemented, where a group of users have access to the same wavelength; a media access control (MAC) protocol is used to avoid collisions among the users sharing the wavelength. Overall, this scheme advocates fast access reconfiguration and slow core reconfiguration.

A somewhat related scheme, proposed in Saleh and Simmons [SaSi06], also utilizes optical grooming at the network edge, for a subset of the traffic, in order to reduce the need for IP grooming in the network core. This scheme is analyzed in more detail in Sect. 10.7.

6.10.4 Optical Burst Switching

One proposal for grooming in the optical domain is *optical burst switching (OBS)* [QiYo99, ChQY04]. OBS operates on the granularity of a data burst rather than a circuit or a packet. The data bursts are assembled at the edge of the network using electronic buffers. A separate control packet is sent to the destination a short time ahead of the data burst, reserving the necessary resources for the burst at each intermediate switch along the path. This allows the data burst to be immediately switched in the optical domain upon its arrival at an intermediate node, without requiring any buffering, assuming the control packet was successful in scheduling the resources. The data burst is sent without waiting for an end-to-end path to be established; thus connections with a very short duration can be handled efficiently.

In one OBS implementation, *Just In Time* (JIT), the required resources are reserved for a data burst as soon as the associated control packet arrives at an intermediate node [TeRo03]. In another OBS variant, *Just Enough Time* (JET), the resources are not actually configured until right before the data burst arrives; i.e., in between the control packet arrival and the data burst arrival at a node, the resources can be used for other bursts in order to improve the system efficiency [ChQY04].

One drawback to OBS, however, is the potential for contention when reserving resources. A data burst may be sent partially along its path only to encounter a node where the required resources were not able to be reserved. The burst must then be sent on an alternate path or be dropped (in this context, “dropped” indicates that the burst is lost). In order to avoid a high drop rate, it may be necessary to operate the network at a relatively low level of utilization [Chan12, WLWZ13].

Contention is likely worse as the geographic size of the network increases. Thus, OBS may be better suited to smaller metro-core or regional applications as opposed to backbone networks. By playing a grooming role at the edge of the backbone network, optical grooming techniques such as OBS potentially can reduce the required size of the electronic grooming boxes in the core of the network.

To better deal with contention, *Labeled OBS with Home Circuits* (LOBS-H) has been proposed [QGSL10]. In this variant, bandwidth is reserved between each source/destination pair based on the expected traffic. All bursts emanating from the same source node over the same "partial path" can share the reserved bandwidth, regardless of the ultimate destination. For example, if traffic from Node A to Node Y follows path A-B-C-D-E-Y and traffic from Node A to Node Z follows path A-B-C-F-G-Z, then the bursts from these two source/destination pairs can share the reserved bandwidth on A-B-C. Additionally, if any of the reserved bandwidth for a given source is unused, it can be used for traffic from a different source. The motivation for LOBS-H is reduced contention through bandwidth reservation, while still achieving some degree of statistical multiplexing.

6.10.5 TWIN

The *Time-Domain Wavelength Interleaved Networking* (TWIN) approach [WSGM03] can be considered a form of OBS. TWIN creates optical multipoint-to-point trees to each destination node, where any node that directly communicates with the destination node is a member of the tree. Each "destination-tree" is associated with a particular wavelength. The branching points of the tree are equipped with switches that are capable of being configured as merging devices; i.e., the signals from multiple network input ports are directed to the same output port.

The traffic source transmits a data burst to a particular destination by tuning its transmitter to the appropriate wavelength. By making use of rapidly tunable transmitters, bursts from multiple sources are multiplexed together. A MAC protocol is used to schedule the sources so that the bursts do not collide on the tree. As with OBS, TWIN may be more suitable for a regional or metro-core network as opposed to a large backbone network.

6.10.6 Lighttrail

The *lighttrail* scheme [GuCh03] accomplishes optical-domain grooming by utilizing the drop-and-continue functionality of broadcast-and-select ROADMs, where a wavelength can both drop at a node and traverse the node. A wavelength connection is first created between two nodes, say Nodes A and Z. The intermediate nodes along the path are configured to allow optical bypass of the wavelength. However, the ROADMs allow any of these intermediate nodes to access the wavelength as well. Essentially, a bus network is established on the wavelength between Nodes A and Z, with any two nodes along the path able to grab the bandwidth at a given time

in order to communicate. This allows the wavelength to be shared in time among the nodes on the bus. A MAC protocol is used to mediate access to the wavelength.

6.10.7 Optical Packet Switching

In *optical packet switching (OPS)*, IP routers distributed across the network continue to groom the IP traffic similar to the current routing paradigm; however, the data packets are switched all-optically rather than electronically (although the packet header is likely to be processed with electronics) [Blum04, YaYo05]. The motivation is to maintain packet-level granularity for purposes of efficiency, but perform the switching in optics for purposes of scalability. As analyzed in Tucker et al. [TPBH09], the two major functional blocks of IP routers that potentially can be moved to the optical domain are the switch fabric and the buffers. However, it is estimated that these components account for only 15 % of the total power consumption in today's electronic routers. Thus, implementing these functions in optics will not significantly reduce the overall power requirements. (The largest contributor to power consumption is the forwarding engine, which implements functions such as pattern matching and complex searches to manage data flow to the switch fabric and the buffers.) A further impediment to OPS is that optical buffers remain very challenging [TuMH07]; see Exercise 6.15.

6.11 Exercises

6.1. Assume that the wavelength line rate is 100 Gb/s, and assume that the following six demands between a pair of nodes need to be multiplexed onto wavelengths: 2×40 Gb/s and 4×30 Gb/s. (a) Using *First Fit Decreasing* bin packing, how many wavelengths are required? (b) Is this optimal, in terms of the number of required wavelengths?

6.2. When multiplexing SONET/SDH traffic onto 40 Gb/s wavelengths, where the line rate and service-rate hierarchy are integer multiples of each other, *First Fit Decreasing* bin packing yields the minimum number of wavelengths. Does it produce the minimum number of wavelengths when packing SONET/SDH service demands onto 100 Gb/s wavelengths (the issue is that 100 Gb/s is not an integer multiple of 40 Gb/s)? Can you come up with a general rule for when First Fit Decreasing bin packing is optimal?

6.3. Assume that due to scalability issues, a grooming switch is composed of three interconnected smaller switches rather than one large switch. Assume that 60 % of the traffic that enters any of the smaller switches can be processed solely within that switch; the remaining 40 % is sent to the other two switches (20 % to each) to be groomed with traffic in those switches. Assume that each

of the three switches has a capacity of 50 Tb/s. What is the effective overall capacity of the interconnected switches?

6.4. Assume that 50 % of the traffic that enters a node is bypass traffic (i.e., this traffic is not sourced/terminated at the node and does not require grooming at the node). Assume that 20 % of the *non-bypass* traffic represents wavelength services that are sourced/terminated at the node. Assume that 50 % of the *non-bypass subrate* traffic is sourced/terminated at the node, with the remainder of the *non-bypass subrate* traffic being groomed in the node on its way to its final destination. (a) With this traffic profile, what is the ratio of the grooming-switch network-side traffic in the architecture shown in Fig. 6.3 as compared to the architecture shown in Fig. 6.4? What is the ratio of the grooming-switch client-side traffic in these same two architectures? (b) If the size of the grooming switch is determined by the sum of the network-side and client-side traffic, what is the ratio of the grooming switch size in the two architectures?

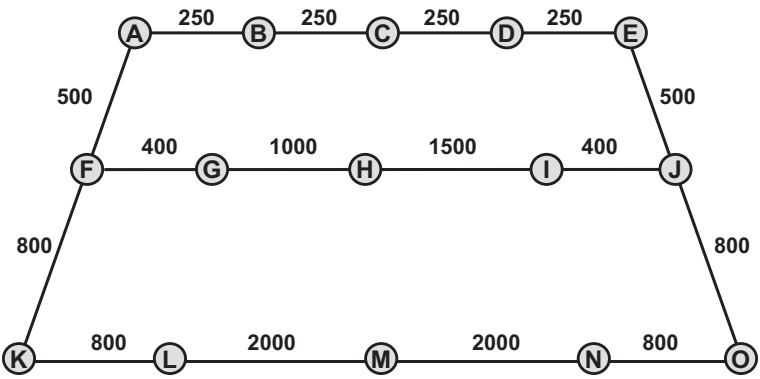
6.5. Consider the IP-over-OTN-over-Optical nodal architecture of Fig. 6.6c. Assume that 50 % of the traffic that enters the node remains in the optical layer; i.e., it optically bypasses both the OTN and IP layers. Of the traffic that is dropped from the optical layer, 50 % of it can bypass the IP layer; i.e., it is only groomed by the OTN switch. The remainder is delivered to the IP router. What cost ratio between the IP ports and the OTN ports justifies this architecture from a cost basis, as compared to the IP-over-Optical nodal architecture of Fig. 6.6b, where all of the non-optical-bypass traffic is delivered to the IP router? (Assume that all of the costs of the IP routers and OTN switches are in the ports. Assume that the OTN network-side ports and OTN client-side ports have the same cost.)

6.6. Assume that 135 (identical and independent) services are multiplexed onto a 100 Gb/s wavelength. Each service can be represented by an ON/OFF model, where a service is ON with probability 0.6. When the service is ON, the requested service rate is 1 Gb/s. (a) What is the probability that the intended offered load exceeds the wavelength bit rate? Let P equal this probability. (b) On average, how full is the 100 Gb/s wavelength? (c) Next, consider the scenario where these same services are multiplexed onto 10 Gb/s wavelengths. How many services can be multiplexed onto one 10 Gb/s wavelength such that the probability that the intended offered load exceeds the wavelength bit rate is no higher than P? (d) With this number of services on a 10 Gb/s wavelength, on average, how full is the wavelength? (e) What is the statistical multiplexing gain in the 100 Gb/s scenario versus the 10 Gb/s scenario (for the level of P calculated above)?

6.7. Repeat Exercise 6.6 except instead of 135 services that request a service rate of 1 Gb/s when in the ON state, assume that there are 756 services that request a service rate of 200 Mb/s when in the ON state. All other assumptions remain the same. Answer the same questions as in Exercise 6.6. Compare the statistical multiplexing gains in the two exercises.

6.8. Consider using the *First Fit Decreasing* bin packing scheme when multiplexing bursty IP services onto a wavelength. Even if it produces the minimum number of wavelengths, what is a possible disadvantage of this strategy?

6.9. Consider the optical-bypass-enabled network shown below, where the link labels indicate the link distances. Assume that the wavelength line rate is 40 Gb/s, the optical reach is 2,000 km, regeneration is performed with back-to-back transponders, all nodes are equipped with a grooming switch, and muxponders are not used. For cost purposes, assume that the following are equivalent: one network-side grooming port; 200 wavelength-km of capacity; and 4 WDM transponders (ignore all other costs). Assume that there are three 10 Gb/s service demand requests: AE, GI, and LN. (Note that to carry one or more 10 Gb/s demands end-to-end on a 40 Gb/s wavelength, the demands must enter a grooming switch at the two endpoints. Assume that the grooming-port cost includes the cost of a transceiver. Thus, transponders are needed only for regeneration.) Produce a routing/grooming design for each of the following criteria: (a) requires the fewest number of network-side grooming ports (if multiple designs are tied, select the one with the lowest cost); (b) yields the lowest cost; (c) requires the least amount of capacity, as measured by wavelength–km; and (d) requires the least amount of capacity, as measured by 10 Gb/s–km.



6.10. Consider a 14-node ring (with nodes numbered sequentially around the ring), with a line rate of 40 Gb/s and with one bidirectional 10 Gb/s demand between every pair of nodes. Assume that all link distances are equal, all nodes have ROADMs, and no regenerations are required. Apply the supernode approach to this ring (see Sect. 6.5.1.3). Assume that only one node per supernode (i.e., the even-numbered node) has grooming capabilities, and thus serves as the hub for the supernode. Also assume that shortest-path routing between the supernodes is used. (a) How many wavelengths need to be supported on the most heavily loaded links in the ring? (b) How

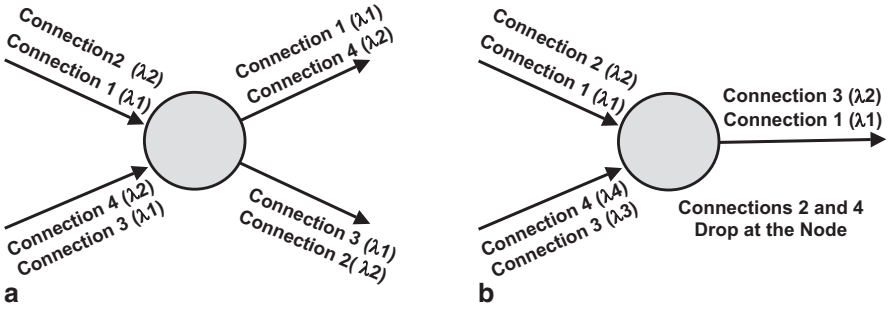
many electrical terminations are required? (c) What is the average nodal drop ratio in this network? This ratio is defined as

$$\frac{\sum_i \text{Number of Wavelengths that Drop at Node } i}{\sum_i \text{Number of Wavelengths that Enter Node } i}$$

6.11. Consider the same network as in Exercise 6.10, but assume that a hubbed architecture is utilized, where Node 1 is designated as the hub and all traffic must be processed by the hub; i.e., two non-hub nodes cannot communicate directly. Use shortest-path routing. (Split the traffic to/from Node 8 evenly in the two directions around the ring.) Answer the same questions as in Exercise 6.10, and compare the results. Repeat this for the scenario where there are two hubs, Nodes 1 and 8. Assume that the non-hub nodes send their traffic to the closest hub. The two hubs communicate directly, with the inter-hub traffic evenly split as much as possible (assume that the "extra" inter-hub wavelength is routed on the link between Nodes 8 and 9).

6.12. For an IP network composed of N routers, where all traffic transiting a node enters a router, some carriers historically have used the rule-of-thumb that each router should be physically connected to *approximately* $\sqrt{N}$ neighbors. (a) What might be the rationale for this design guideline? (b) How might this guideline change if optical bypass is implemented (i.e., where traffic can transit a router-equipped node without necessarily entering the router)?

6.13. In a system based on wavebands, packing the wavelengths into wavebands is a form of grooming. Assume that a system supports a total of four wavelengths per fiber, partitioned into two wavebands of two contiguous wavelengths each. Assume that the nodes are equipped with a two-level (waveband and wavelength) hierarchical switch. Draw diagrams of how the following two waveband grooming operations are implemented in the hierarchical switch: (a) Two wavebands enter a node on two different fibers, and two wavebands exit that node on two different fibers but with a different grouping of connections (see left side of figure below). (b) Two wavebands enter a node on two different fibers; one waveband, with a different grouping of connections, exits that node; two of the connections are dropped at that node (see right side of figure). Note: Wavebands were discussed in Sect. 2.9.7; a hierarchical switch was discussed in Sect. 2.11. Wavelength converters may be used, as shown in Fig. 2.30.



6.14. In Sect. 6.10.1, an energy-saving scheme is described where the level of optical bypass is increased during times of low traffic to reduce the amount of traffic that is groomed in the higher layers. Assume that the network has an IP-over-Optical architecture. Provide an example where such an increase in optical bypass results in a change to the IP virtual topology and another example where it does *not* change the IP virtual topology.

6.15. Assume that a single-pass fiber delay line is being used as an optical buffer in an OPS router. Assume that the port speed is 100 Gb/s, and that each port is equipped with a buffer large enough to hold 100 packets of average size 250 bytes. (a) What length of fiber is needed to provide a port buffer of this size? (b) If the capacity of the router is 100 Tb/s (i.e., 1,000 ports), how much total fiber is needed? (Assume that the speed of light in fiber is 2×10^8 m/sec.)

6.16. *Research Suggestion*: Various factors and methodologies for selecting the nodes to equip with grooming switches were outlined in Sect. 6.5. Investigate this further by developing alternative effective metrics and strategies for choosing the grooming nodes.

References

[BCRV06] G. Bernstein, D. Caviglia, R. Rabbat, H. Van Helvoort, VCAT-LCAS in a clamshell. IEEE Comm. Mag. **44**(5), 34–36 (May 2006)

[Blum04] D.J. Blumenthal, Optical packet switching. *Proceedings, 17th Annual Meeting of the IEEE LEOS*, Puerto Rico, 7–11 Nov 2004, Paper ThU1

[CCCD12] A.L. Chiu, G. Choudhury, G. Clapp, R. Doverspike, M. Feuer, J.W. Gannett, J. Jackel, G.T. Kim, J.G. Klincewicz, T.J. Kwon, G. Li, P. Magill, J.M. Simmons, R.A. Skoog, J. Strand, A. Von Lehmen, B.J. Wilson, S.L. Woodward, D. Xu, Architectures and protocols for capacity efficient, highly dynamic and highly resilient core networks. J. Opt. Commun. Netw. **4**(1), 1–14 (Jan 2012)

[Chan12] V.W.S. Chan, Optical flow switching networks. Proc. IEEE. **100**(5), 1079–1091 (May 2012)

[ChMN12] L. Chiaraviglio, M. Mellia, F. Neri, Minimizing ISP network energy cost: Formulation and solutions. IEEE/ACM Trans. Netw. **20**(2), 463–476 (Apr 2012)

[Choy02] L. Choy, Virtual concatenation tutorial: Enhancing SONET/SDH networks for data transport. J. Opt. Netw. **1**(1), 18–29 (Jan 2002)

[ChQY04] Y. Chen, C. Qiao, X. Yu, Optical burst switching: A new area in optical networking research. IEEE Netw. **18**(3), 16–23 (May/June 2004)

[ChRD08] B. Chen, G.N. Rouskas, R. Dutta, On hierarchical traffic grooming in WDM networks. IEEE/ACM Trans. Netw. **16**(5), 1226–1238 (Oct 2008)

[ChRD10] B. Chen, G.N. Rouskas, R. Dutta, Clustering methods for hierarchical traffic grooming in large scale mesh WDM networks. J. Opt. Commun. Netw. **2**(8), 502–514 (Aug 2010)

[ChWM06] V.W.S. Chan, G. Weichenberg, M. Médard, Optical flow switching. *Workshop on Optical Burst Switching (WOBS)*, San Jose, Oct 2006

[Cisc13] Cisco Visual Networking Index: Forecast and Methodology, 2012–2017, White Paper (29 May 2013)

[DGGM02] N.G. Duffield, P. Goyal, A. Greenberg, P. Mishra, K.K. Ramakrishnan, J.E. van der Merwe, Resource management with hoses: Point-to-cloud services for virtual private networks. IEEE/ACM Trans Netw. **10**(5), 679–692 (Oct 2002)

[Dosa07] G. Dósa, The tight bound of first fit decreasing bin-packing algorithm is $FFD(I) \leq 11/9OPT(I)+6/9$. *International Symposium on Combinatorics, Algorithms, Probabilistic and Experimental Methodologies*, Hangzhou, 7–9 Apr 2007, pp. 1–11

[DuKR08] R. Dutta, A.E. Kamal, G.N. Rouskas (eds.), *Traffic Grooming for Optical Networks: Foundations, Techniques and Frontiers* (Springer, New York, 2008)

[DuRo02] R. Dutta, G.N. Rouskas, Traffic grooming in WDM networks: Past and future. *IEEE Netw.* **16**(6), 46–56 (Nov/Dec 2002)

[Elby09a] S. Elby, Bandwidth flexibility and high availability—and return on invested capital. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'09)*, San Diego, 22–26 March 2009, Service Provider Summit

[FHAT12] M.Z. Feng, K. Hinton, R. Ayre, R.S. Tucker, Network energy efficiency gains through coordinated cross-layer aggregation and bypass. J. Opt. Commun. Netw. **4**(11), 895–905 (Nov 2012)

[GaJo79] M.R. Garey, D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (W.H. Freeman and Co., New York, 1979)

[GeRS98] O. Gerstel, R. Ramaswami, G. Sasaki, Cost effective traffic grooming in WDM rings. *Proceedings, IEEE INFOCOM 1998,* San Francisco, 29 March–2 April 1998, pp. 69–77

[GOJK02] K.I. Goh, E. Oh, H. Jeong, B. Kahng, D. Kim, Classification of scale-free networks. *Proc. Natl. Acad. Sci.* U S A. **99**(20), 12583–12588 (1 Oct 2002)

[Gree13] GreenTouch, GreenTouch Green Meter Research Study: Reducing the Net Energy Consumption in Communications Networks by up to 90% by 2020, GreenTouch White Paper, Version 1.0, 26 June 2013

[GuCh03] A. Gumaste, I. Chlamtac, Mesh implementation of light-trails: A solution to IP centric communication. *Proceedings, International Conference on Computer Communication and Networks (ICCCN'03)*, Dallas, 20–22 Oct 2003, pp. 178–183

[HuDu07] S. Huang, R. Dutta, Dynamic traffic grooming: The changing role of traffic grooming. IEEE Commun. Surv. Tutori. **9**(1), 32–49 (First Quarter 2007)

[Idzi13] F. Idzikowski et al., TREND in energy-aware adaptive routing solutions. IEEE Commun. Mag. **51**(11), 94–104 (Nov 2013)

[Infi12] Infinera, Network efficiency quotient, White Paper WP-EQ-06-2012, 2012

[Jinn08] M. Jinno et al., Demonstration of novel spectrum-efficient elastic optical path network with per-channel variable capacity of 40 Gb/s to over 400 Gb/s. *Proceedings, European Conference on Optical Communication (ECOC'08)*, Brussels, 21–25 Sep 2008, Paper Th.3.F.6

[KGHA12] D. Kilper, K. Guan, K. Hinton, R. Ayre, Energy challenges in current and future optical transmission networks. Proc. IEEE. **100**(5), 1168–1187 (May 2012)

[LuSS13] Y. Lui, G. Shen, W. Shao, Design for energy-efficient IP over WDM networks with joint lightpath bypass and router-card sleeping strategies. J. Opt. Commun. Netw. **5**(11), 1122–1138 (Nov 2013)

[MaDo09] P. Magill, R. Doverspike, The core photonic networks—where are things heading? *Proceedings, European Conference on Optical Communication (ECOC'09)*, Vienna, 20–24 Sep 2009, Paper 4.6.1

[MiRo05] J. Mihelic, B. Robic, Solving the K-center problem efficiently with a dominating set algorithm. J. Comput. Inf. Technol. **13**(3), 225–233 (Third Quarter 2005)

[PaSA05] P. Pan, G. Swallow, A. Atlas, Editors, Fast Reroute Extensions to RSVP-TE for LSP Tunnels. Internet Engineering Task Force, Request for Comments (RFC) 4090, May 2005

[QGSL10] C. Qiao, M. A. González-Ortega, A. Suárez-González, X. Liu, J. C. Lopéz-Ardao, On the benefit of fast switching in optical networks. *Proceedings, Optical Fiber Communication/ National Fiber Optic Engineers Conference (OFC/NFOEC'10),* San Diego, 21–25 March 2010, Paper OWR2

[QiYo99] C. Qiao, M. Yoo, Optical Burst Switching (OBS)—A new paradigm for an optical internet. J. High Speed Netw. **8**(1), 69–84 (Jan 1999)

[SaSi06] A.A.M. Saleh, J.M. Simmons, Evolution toward the next-generation core optical network. J. Lightwave Technol. **24**(9), 3303–3321 (Sept 2006)

[ShWi06] F.B. Shepherd, P.J. Winzer, Selective randomized load balancing and mesh networks with changing demands. J. Opt. Netw. **5**(5), 320–339 (May 2006)

[SiGS98] J.M. Simmons, E.L. Goldstein, A.A.M. Saleh, On the value of wavelength-add/drop in WDM rings with uniform traffic. *Proceedings, Optical Fiber Communication (OFC'98)*, San Jose, 22–27 Feb 1998, Paper ThU3

[SiGS99] J.M. Simmons, E.L. Goldstein, A.A.M. Saleh, Quantifying the benefit of wavelength add–drop in WDM rings with distance-independent and dependent traffic. J. Lightwave Technol. **17**(1), 48–57 (Jan 1999)

[SiSa99] J.M. Simmons, A.A.M. Saleh, The value of optical bypass in reducing router size in gigabit networks. *Proceedings, IEEE International Conference on Communications (ICC'99)*, vol. 1, Vancouver, 6–10 June 1999, pp. 591–596

[TaHR10] O. Tamm, C. Hermsmeyer, A.M. Rush, Eco-sustainable system and network architectures for future transport networks. Bell Labs Tech. J. **14**(4), 311–327 (Feb 2010)

[TeRo03] J. Teng, G.N. Rouskas, A comparison of the JIT, JET, and Horizon wavelength reservation schemes on a single OBS node. *Proceedings, The First International Workshop on Optical Burst Switching (WOBS)*, Dallas, 16 Oct 2003

[TPBH09] R.S. Tucker, R. Parthiban, J. Baliga, K. Hinton, R.W.A. Ayre, W.V. Sorin, Evolution of WDM optical IP networks: A cost and energy perspective. J. Lightwave Technol. **27**(3), 243–252 (1 Feb 2009)

[Tuck11a] R.S. Tucker, Green optical communications—Part I: Energy limitations in transport. IEEE J. Sel. Top. Quantum Electron. **17**(2), 245–260 (March/April 2011)

[Tuck11b] R.S. Tucker, Green optical communications—Part II: Energy limitations in networks. IEEE J. Sel. Top. Quantum Electron.**17**(2), 261–274 (March/April 2011)

[TuMH07] R.S. Tucker, S.S. Mughal, K. Hinton, In search of the elusive all-optical packet buffer. *Proceedings, International Conference on Photonics in Switching*, San Francisco, 19–22 Aug 2007, pp. 3–4

[UCSB13] UC Santa Barbara, The Institute for Energy Efficiency, ICT Core Networks: Towards a Scalable, Energy-Efficient Future, Roundtable Report, June 2013

[WLWZ13] M. Wang, S. Li, E.W.M. Wong, M. Zukerman, Evaluating OBS by effective utilization. IEEE Commun. Lett. **17**(3), 576–579 (March 2013)

[WSGM03] I. Widjaja, I. Saniee, R. Giles, D. Mitra, Light core and intelligent edge for a flexible, thin-layered, and cost-effective optical transport network. IEEE Commun. Mag. **41**(5), S30-S36 (May 2003)

[YaYo05] H. Yang, S. J. B. Yoo, All-optical variable buffering strategies and switch fabric architectures for future all-optical data routers. J. Lightwave Technol. **23**(10), 3321–3330 (Oct 2005)

[ZCTM10] Y. Zhang, P. Chowdhury, M. Tornatore, B. Mukherjee, Energy efficiency in telecom optical networks. IEEE Commun. Surv. Tutori. **12**(4), 441–458 (Fourth Quarter 2010)

[ZhMu03] K. Zhu, B. Mukherjee, A review of traffic grooming in WDM optical networks: Architectures and challenges. Opt. Netw. Mag. **4**(3), 55–64 (Mar/Apr 2003)

[ZhZM05] K. Zhu, H. Zhu, B. Mukherjee, *Traffic Grooming in Optical WDM Mesh Networks* (Springer, New York, 2005)

[ZZZM03] H. Zhu, H. Zang, K. Zhu, B. Mukherjee, A novel generic graph model for traffic grooming in heterogeneous WDM mesh networks. IEEE/ACM Trans. Netw. **11**(2), 285–299 (Apr 2003)

Chapter 7
Optical Protection

7.1 Introduction

Any network is subject to failures, whether it is due to fiber cuts, equipment failures, software errors, technician errors, environmental causes, or malicious attacks. Protection against failures, by providing alternative paths or backup equipment, is a necessary component of network design. One of the key design decisions is selecting the networking layer, or layers, in which to implement protection. For example, higher-layer protocols, such as Internet Protocol (IP), typically have standardized protection mechanisms. However, these mechanisms usually operate on a relatively fine traffic granularity; as traffic levels increase, implementing failure recovery solely in these layers may be too slow. Optical protection, which operates on the granularity of a wavelength (or even a waveband or a fiber), has received growing attention, largely due to its ability to scale more gracefully with increasing traffic levels. As the wavelength bit rate increases, e.g., from 40 to 100 Gb/s and beyond, the amount of network traffic that can be restored by rerouting a wavelength grows accordingly.

There are numerous optical protection schemes, where the mechanisms differ in the amount of spare capacity and equipment required, the speed of recovery, the number of concurrent failures from which recovery is possible, and the operational complexity. It is possible to support a combination of protection mechanisms in a network, where the protection scheme used for a particular demand largely depends on the required *availability* for that demand. Availability is defined as the probability that a connection is in a working state at a given instant of time. Such requirements are usually specified as part of the service level agreements (SLAs) between a carrier and its customers. For example, some demands may have very stringent requirements such that recovery from a failure must be almost immediate (e.g., in less than 50 ms). At the other extreme, some demands may be contracted as best effort, where no resources are specifically allocated for their protection.

Section 7.2 through Sect. 7.5 describe some of the major classes of protection and their inherent trade-offs. Specifically, these sections probe dedicated versus shared protection, client-side versus network-side protection, ring versus mesh protection, and failure-dependent versus failure-independent protection. In addition to

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4_7,

describing the basic properties of these protection classes, the discussion will address how the presence of optical bypass affects the efficacy of the various schemes.

Note that protection schemes are used for *failure recovery,* where a connection is restored after it has failed. This is in contrast to *failure repair,* which refers to actually fixing what has failed; e.g., repairing a fiber cut or replacing a failed piece of equipment. Failure recovery generally occurs on the order of seconds or less, whereas failure repair may take several hours. Because of the length of time required to repair a failure, additional failures may occur such that a network is affected by multiple concurrent failures. This is especially true in networks of large geographic extent, or in networks deployed in hostile environments. Thus, for some demands, it may be necessary to allocate enough spare resources such that recovery from any two (or possibly more) concurrent failures is supported. Recovery from multiple failures, including an analysis of catastrophic failures, is covered in Sect. 7.6.

In Sect. 7.7, the relationship of optical bypass and optical protection is probed in greater depth. There are specific properties of optical-bypass-enabled networks that must be considered when developing a protection scheme, most notably the sensitivity to optical amplifier transients that arise from sudden changes of the optical power level on a fiber. This favors employing protection schemes that do not require rapid turning on or off of lasers or rapid switching of wavelengths.

Section 7.8 through Sect. 7.10 discuss three specific protection methodologies in more detail, all of which are applicable to shared protection for general mesh topologies. Section 7.8 describes protection based on pre-deployed subconnections. The scheme is notable because it avoids issues with optical amplifier transients, and is thus well suited to optical-bypass-enabled networks. Section 7.9 discusses protection schemes based on "pre-cross-connected" bandwidth. These schemes, though potentially complex to design, provide relatively fast recovery speed while remaining efficient with respect to the required spare capacity. Section 7.10 addresses protection through *network coding,* which represents a major departure from traditional protection schemes. By utilizing processing in the nodes, the speed of recovery is on par with the fastest conventional protection schemes, but with potentially less required spare capacity. Network coding and, to a lesser degree, pre-cross-connected protection challenge the conventional wisdom that protection schemes must trade off capacity for speed.

This is followed by a discussion of protection planning methodologies in Sect. 7.11. The various protection types discussed in this chapter require different design techniques and different optimizations. Rather than cover the full gamut of protection planning, this section mainly focuses on design methodologies for shared mesh protection.

Chapter 6 addressed multiplexing and grooming of subrate demands. Section 7.12 specifically addresses some of the options for protecting such demands, where protection can be provided at the wavelength level (i.e., optical layer), at the subrate level (i.e., grooming layer), or both. While wavelength-level protection can be rapid, it is not necessarily sufficient for recovering from all network failures. Subrate-level protection may be more efficient and may provide more fault coverage, but it also tends to be slower. Multilayer protection combines the approaches to

achieve the benefits of both layers; however, it does present challenges in coordination of the layers.

The last section of this chapter deals with performance monitoring and determining the location of a failure, where again the focus is on optical-bypass-enabled networks. Fault localization is somewhat more challenging in the presence of optical bypass because per-connection performance monitoring in the electrical domain does not occur at every node.

It is necessary to clarify some of the terminology that is used in this chapter. The connection path from source to destination that is used under the condition of no failures is referred to as the *working path* or the *primary path*. The alternative path that is used after a failure occurs is referred to as the *protect path,* the *backup path,* or the *secondary path*. The network capacity that is allocated for protection is referred to here as the *protection capacity* or the *spare capacity*. In the past, the terms *protection* and *restoration* have been used to distinguish the relative speed of the recovery method, where protection mechanisms are generally preplanned and fast, whereas restoration schemes require calculations at the time of failure and are thus relatively slow. However, because the differences between such schemes have become blurred and because there are no universally accepted definitions, the terms will be used interchangeably here.

Most of this chapter discusses recovery from a link, node, or transponder failure; more general equipment protection is not addressed. In some networks, it is assumed that node failures are so infrequent that nodal protection is not required; however, unless otherwise noted, it is assumed here that nodal protection is necessary. The major difference is that the protection mechanism looks for link-and-node-disjoint paths as opposed to just link-disjoint paths. Note that the term "node-disjoint paths" refers to paths with no *intermediate* nodes in common; the endpoints of the paths can be the same (it is assumed that if a demand endpoint fails, the connection cannot be recovered).

Finally, it should be emphasized that optical protection is a very rich topic. This chapter covers the major points, with an emphasis on optical-bypass-enabled networks. There are several books that are dedicated to the topic of protection, e.g., Grover [Grov03], Vasseur et al. [VaPD04], Ou and Mukherjee [OuMu05], and Bouillet et al. [BELR07]. Additionally, Gerstel and Ramaswami [GeRa00] and Ellinas et al. [EBRL03] are good tutorial papers.

7.2 Dedicated Versus Shared Protection

One of the basic dichotomies among protection schemes is whether the protection is *dedicated* or *shared*. This dichotomy exists with protection at any network layer, not just optical protection. For ease of discourse, this section compares dedicated and shared protection in relation to link/node failures (as opposed to, e.g., equipment failures). Furthermore, it assumes that the protection mechanism is path based, where recovery is provided by moving the connection to an alternative

end-to-end path; however, the dedicated versus shared paradigm applies to more general protection schemes. (Path-based protection is discussed further in Sect. 7.5.2.)

7.2.1 Dedicated Protection

In dedicated protection, spare resources are specifically allocated for a particular demand. If a demand is brought down by a failure, it is guaranteed that there will be available resources to recover from the failure, assuming the backup resources have not failed also.

Dedicated protection generally falls into one of two categories. In 1 + 1 dedicated protection, the backup path is "active"; i.e., there are two live connections between the source and destination, and the destination is equipped with decision circuitry to select the better of the two paths. In contrast, in 1:1 dedicated protection, the backup path does not become active until after a failure has occurred on the primary path. After the failure is repaired, the connection may remain on the backup path (i.e., non-revertive mode) or may return to the primary path (i.e., revertive mode).

There are several advantages to operating dedicated protection in a 1 + 1 mode. First, recovery from a failure can be almost immediate. As soon as the receiver detects that the primary path has become unsatisfactory, it can switch over to the secondary path. (There is usually a small synchronization delay due to the transmission latency of the two paths being different.) The 1:1 mode is slower, as the failure must first be detected (usually by the destination), and then the source must be notified of the failure so that it can begin to transmit over the backup path. Another advantage to 1 + 1 is that failures on the backup path can be detected when they occur. With 1:1, a "silent failure" can occur on the backup path, such that the failure is not detected until the backup path is actually needed. One disadvantage to 1 + 1 protection is that it may require more equipment at the connection endpoints, to support two active paths (this depends on the multicast capabilities of the endpoints).

The downside of dedicated protection, whether 1 + 1 or 1:1, is the large amount of spare capacity that it generally requires. In typical networks, the ratio of the dedicated backup capacity to the working capacity is often on the order of 2 to 1. However, with 1:1 protection, the spare capacity can potentially be used to carry low-priority traffic that is preempted when the capacity is needed for failure recovery. This has the added advantage of reducing the likelihood of a silent failure on the backup path.

7.2.2 Shared Protection

In contrast to dedicated protection, shared protection potentially requires significantly less spare capacity by allowing the protection resources to be used for multiple working paths. The working paths that share protection capacity should have no links or intermediate nodes in common so that a single network failure does not

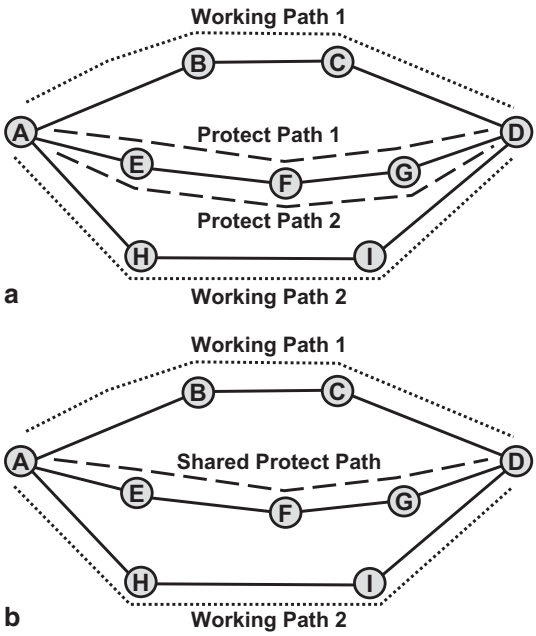


Fig. 7.1 **a** Dedicated protection where a protect path is reserved for each working path. **b** Shared protection where the two working paths share the protection resources

bring down more than one of the paths. (If the endpoint of one working path is an intermediate node of another working path, then whether to allow the two paths to share protection resources depends on the scheme; see Ellinas et al. [EBRL03] and Exercise 7.4.)

While sharing protection resources improves the capacity efficiency, one drawback is that contention for the resources may arise if there are multiple concurrent failures. For a given network failure rate, the required availability of a connection determines whether shared protection is suitable, and if so, what level of sharing is acceptable. Another potential drawback is that shared protection usually requires greater coordination in the network to respond to a failure. Additionally, recovery using shared protection may be relatively slow as it may require that switches be reconfigured in order to form the desired protect path. This is discussed further in Sect. 7.2.3.

Note that shared protection generally operates in a revertive mode, such that the protection resources are released by the connection after the failure is repaired.

7.2.3 Comparison of Dedicated and Shared Protection

Dedicated and shared protection are illustrated in Fig. 7.1 for two wavelength-level demands. In Fig. 7.1a, the working paths are routed over paths A-B-C-D and A-H-I-D. Both of the paths are protected with spare capacity allocated along the path A-E-F-G-D. Note that two wavelengths of spare capacity are allocated along this path, one wavelength dedicated to each of the working paths.

Figure 7.1b illustrates shared protection for the same two working paths. Only one wavelength of spare capacity is allocated along A-E-F-G-D. The two working paths are link-and-node disjoint and thus can share this spare capacity. As illustrated by this small example, shared protection usually requires significantly less resources than dedicated protection. In typical networks, shared protection requires roughly 50–70 % less spare capacity than dedicated protection.

In addition to saving capacity, shared protection may reduce the network cost, although the cost savings is usually much more significant in an optical-electrical-optical (O-E-O)-based network than in an optical-bypass-enabled network. In Fig. 7.1, shared protection allows one protection wavelength along A-E-F-G-D to be removed. In an O-E-O network, this also removes the three regenerations that would occur along this path, providing a cost savings as well. However, with optical bypass, if the distance along path A-E-F-G-D is less than the optical reach, then no regenerations are saved by implementing shared protection. In fact, shared protection may be more expensive than dedicated protection in an optical-bypass-enabled network, as is illustrated next.

In Fig. 7.1b, the protection capacity was shared by working paths with the same endpoints. More generally, however, working paths with different endpoints can share portions of the protection capacity. Consider the network shown in Fig. 7.2, which is assumed to be optical-bypass enabled with an optical reach of 2,000 km. Two working paths are shown, A-B-C-D and A-I-J, neither of which requires regeneration. With dedicated protection, assume that the two corresponding protect paths are A-E-F-G-D and A-E-F-H-J; neither protect path requires regeneration.

With shared protection, the spare capacity along A-E-F can be used to protect either working path. This saves one wavelength of capacity on these links. However, because the dedicated protect paths did not require any regeneration, no regenerations are saved by using shared protection. Furthermore, with shared protection, switching is required at Node F to establish the desired protect path in response to whichever working path has failed. For example, if working path A-I-J fails, then the switch at Node F is configured such that it concatenates the protection capacity along A-E-F and F-H-J to form the A-E-F-H-J protect path. In an optical-bypass-enabled network, there are two options for implementing this switch operation.

First, consider switching in the optical domain at Node F, where the multi-degree reconfigurable optical add/drop multiplexer (ROADM-MD) at this node is used to establish the desired protect path. With this option, no electronics are needed along the protection capacity, so that this solution would be similar in cost to dedicated protection (the actual cost difference would depend on how transponders are deployed at the connection endpoints for dedicated and shared protection, as discussed in Sect. 7.3). However, the wavelength continuity constraint can make this challenging to design because the same wavelength would need to be assigned along each of the six protection links shown in the figure. Furthermore, there are operational issues with utilizing switching in the optical domain for protection (e.g., optical amplifier transients), as covered in Sect. 7.7.

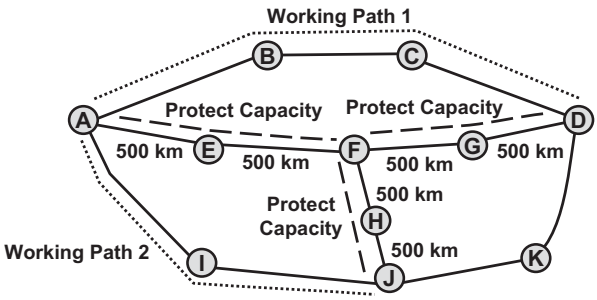


Fig. 7.2 The two working paths share the protection capacity along *A-E-F*. Switching is required at Node *F* to configure the proper protect path

To avoid these issues, it may be desirable to switch the protection capacity at Node F in the electrical domain. The wavelengths corresponding to the A-E-F, F-G-D, and F-H-J protection segments would be dropped from the optical layer at Node F. At least two transponders are needed at Node F (possibly three transponders, if the segments must be kept lit to avoid optical amplifier transients). A corresponding number of electronic switch ports are utilized as well. Thus, because of the required electronics at Node F, this shared protection configuration would likely be *more* costly than dedicated protection (although, again, there could be fewer transponders required at the source and destination with shared protection, which would partially offset the switching cost). As this example illustrates, in an optical-bypass-enabled network, the switching required by shared protection may increase the cost of the network, especially if the optical reach is long enough that few regenerations are needed for the dedicated protect paths.

Using the same example of Fig. 7.2, consider the comparison between shared and dedication protection in an O-E-O-based system, where regeneration is required at every intermediate node. Sharing the protection capacity eliminates one wavelength along A-E-F and, thus, one regeneration at Node E. Furthermore, with shared protection, only three transponders are required at Node F (one to terminate each of the A-E-F, F-G-D, and F-H-J protection segments), as opposed to four transponders with dedicated protection (two for regeneration of A-E-F-G-D and two for regeneration of A-E-F-H-J). There may be additional transponder savings at the endpoints depending on the shared protection scheme. While switching is needed at Node F to allow sharing of the A-E-F segment, the O-E-O network may already make use of electronic switches at each node. Thus, overall, using shared protection instead of dedicated protection should reduce the cost of an O-E-O network.

Another characteristic of shared protection, which holds for either O-E-O or optical-bypass-enabled networks, is that it is generally slower to recover due to the amount of required switching. In Fig. 7.2, restoration from a failure on either of the working paths requires that Node F be notified of the failure and the switch at that node be reconfigured (assuming it is not already in the desired configuration). This delays the restoration process and adds to the complexity of the restoration operation. Many shared protection schemes take hundreds of milliseconds (or even multiple seconds) to restore all failed demands. However, recent research on shared

protection has focused on minimizing the amount of operations required at the time of failure, thereby conceivably reducing the restoration time to within 100 ms for a continental-scale backbone network; this is discussed further in Sect. 7.6.4 and Sect. 7.8 through Sect. 7.10.

To summarize, shared protection is more capacity efficient than dedicated protection. In O-E-O networks, shared protection is typically less costly as well; with optical bypass, the cost savings, if any, is typically small (or shared protection may actually cost more). Shared protection is generally slower to recover from a failure than dedicated protection and is more complex to implement. Finally, shared protection leaves the network more vulnerable to a second failure. Clearly, the major impetus behind shared protection is its capacity efficiency.

7.3 Client-Side Versus Network-Side Protection

Client-side and network-side protection refers to where the protection mechanism is triggered, i.e., at the clients at the connection endpoints (client side) or in the optical layer (network side). From a cost perspective, the main impact of the two schemes is the amount of equipment required at the connection endpoints. Either type of protection can be operated in a dedicated or shared mode; however, for illustrative purposes, the discussion in this section assumes dedicated protection.

First, consider client-side 1 + 1 dedicated protection, where the client is responsible for generating two copies of the signal, both of which are transmitted over the optical layer. A connection endpoint with this type of protection is illustrated in Fig. 7.3a, where an IP router is serving as the client. The router sends two copies of the signal to the optical layer, which routes them over disjoint paths. In the reverse direction, the router receives two copies of the signal and is responsible for selecting the better of the two copies.

This configuration utilizes two transponders at either endpoint of the connection. The working and protect paths do not have to be assigned the same wavelengths. In addition to protecting against link/node failures (assuming the two copies of the signal are routed over diverse paths), this architecture also provides protection against a transponder failure, a client-port failure, or a failure of the interconnect between the client and optical layers.

Another client-side protection configuration is shown in Fig. 7.3b, where the client generates just one copy of the signal, which passes through a passive splitter that feeds two transponders. The optical layer routes the resulting two signals over disjoint paths. In the receive direction, there is decision circuitry to select the better of the two paths. This configuration is of lower cost because it utilizes only one client port, but this also leaves it more vulnerable to failure.

A connection endpoint with network-side dedicated protection is illustrated in Fig. 7.3c, where it is assumed that the optical-layer switch is directionless (i.e., a transponder can access any of the network ports). First, consider 1:1 protection. Under the no-failure condition, the signal is transmitted over the working path; upon

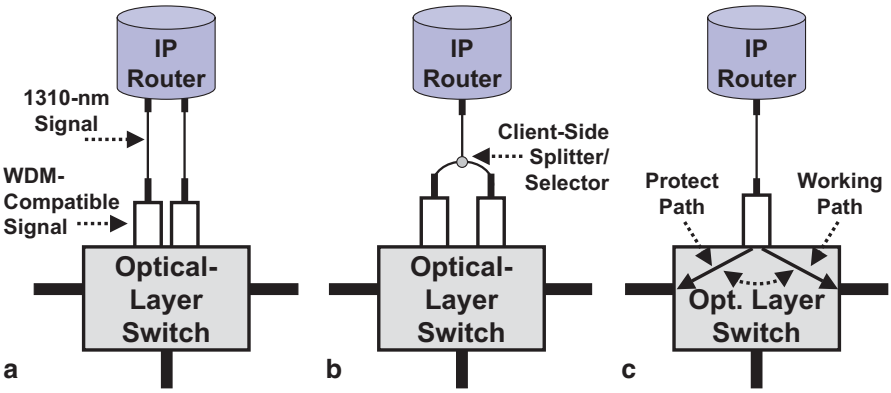


Fig. 7.3 **a** Client-side protection where the client (here an IP router) delivers the working and protect signals to the optical layer in the transmit direction and selects the better of the two signals in the receive direction. **b** Client-side protection where a splitter feeds the client signal into two transponders, and a selector chooses the better of the two paths in the receive direction. **c** Network-side protection where the optical-layer switch either multicasts the signal to both the working and protect paths (1 + 1) or switches from the working path to the protect path at the time of failure (1:1). If 1 + 1 protection is employed, the transponder selects the better of the two paths in the receive direction

failure, the optical-layer switch is reconfigured to direct the signal to the protect path. If the optical-layer switch is multicast-capable, then 1 + 1 protection could also be implemented. In this scenario, the signal is multicast over both the working and protect paths. In the receive direction, the transponder would need to be equipped with decision circuitry to select the better of the two paths and send it to the client. If the switch is not multicast-capable, then a second option for 1 + 1 implementation is to use a two-way flexible transponder, which is capable of simultaneously transmitting a signal to two network ports. Again, in the receive direction, the flexible transponder would need to be capable of selecting the better of the two received signals. (Flexible transponders were discussed in Sect. 2.9.4.1.)

7.3.1 Transponder Protection

The advantage of the configuration of Fig. 7.3c, as compared to (a) and (b), is that only a single transponder is required at either endpoint instead of two. However, this leaves a connection vulnerable to a failure of the transponder itself. Thus, this type of protection is typically used in conjunction with 1:*N* shared protection of the transponders, where every group of *N* transponders is protected by one spare (or, more generally, *M:N* shared protection is used, where there are *M* spares to protect *N* transponders). As *N* increases, the protection efficiency improves; however, the vulnerability of the scheme to multiple failures increases. Three architectures for 1:*N* transponder protection are shown in Fig. 7.4. Other 1:*N* transponder protection architectures are possible as described in Gerstel and Ramaswami [GeRa00].

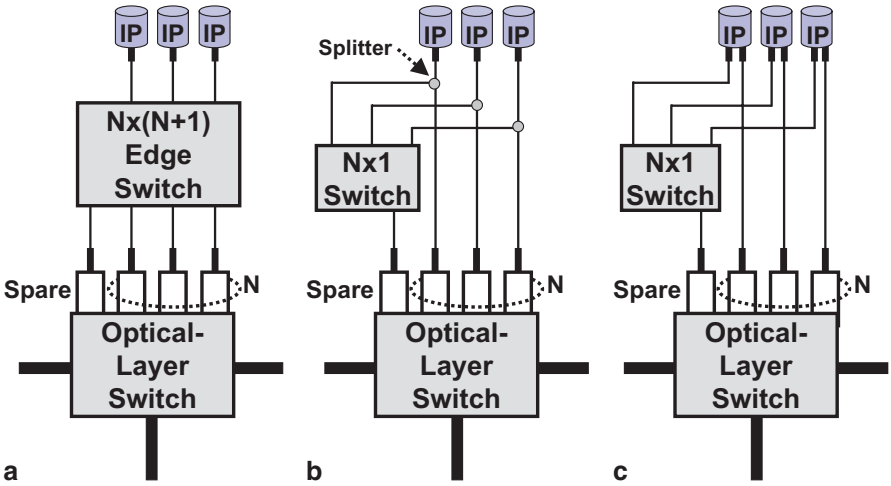


Fig. 7.4 1:*N* transponder protection. **a** An edge switch directs the client signal associated with the failed transponder to the spare transponder. **b** A passive splitter directs the client signal to its primary transponder and to an $N \times 1$ switch that feeds the spare transponder. **c** Dual client signals are sent to the primary transponders and the $N \times 1$ switch that feeds the spare transponder. In all three illustrations, the edge switch depicted is a photonic switch; an O-E-O switch could also be used

In Fig. 7.4a, the client signal enters an edge switch (e.g., a fiber cross-connect) that directs the signal to its primary transponder, or if that has failed, to the spare transponder. This architecture is well suited for a node where an edge switch is deployed anyway for purposes of node flexibility. In Fig. 7.4b, a passive splitter is used to direct the client signal to its primary transponder and to an $N \times 1$ switch. If one of the primary transponders fails, the $N \times 1$ switch is configured to select the client signal associated with the failed transponder and direct it to the spare transponder. A $1 \times N$ switch is used in the receive direction. The architectures of Fig. 7.4a, b are still vulnerable to a client-port failure. Figure 7.4c is one method of addressing this vulnerability, where the client uses two ports to send two copies of the signal, one of which enters the $N \times 1$ switch feeding the spare transponder. In any of these architectures, the spare transponder is ideally tunable so that it can tune to the wavelength of whichever transponder has failed, thereby allowing the affected connection to continue using the same wavelength.

7.3.2 Wavelength Assignment with Network-Side Protection

When operated in the $1+1$ mode, the network-side protection configuration of Fig. 7.3c (i.e., with just one transponder at the endpoints) poses interesting wavelength-assignment challenges in an optical-bypass-enabled network. A transponder typically has a single laser, such that the working and protect paths in this architecture are launched on the same wavelength. Similarly, assume that the receiver

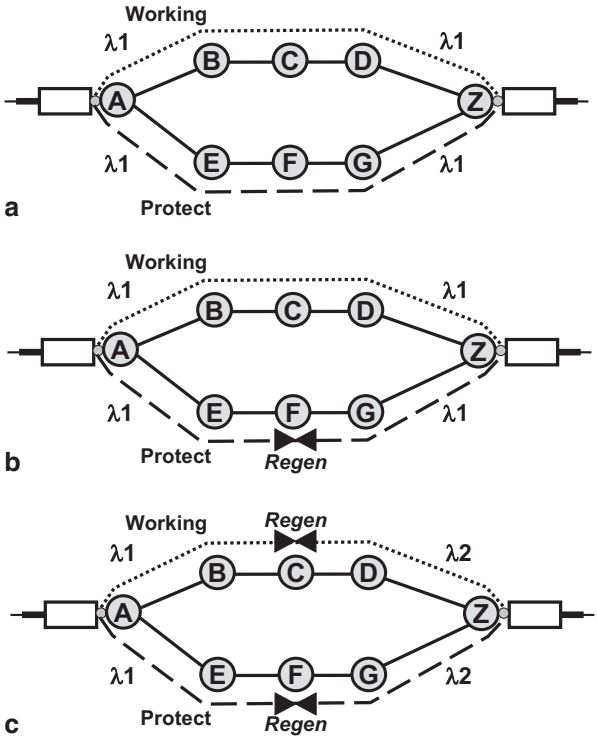


Fig. 7.5 Possible wavelength constraints with network-side 1 + 1 protection. **a** With no regenerations, the same wavelength must be assigned on all links of both the working and protect paths. **b** Even though the protect path has a regeneration, it cannot be used to change the wavelength, due to the constraints posed by the working path. **c** With a regeneration in both the working and protect paths, the wavelength used in one half of the working and protect paths can be different from the wavelength used in the other half

is such that the two received paths must be at the same wavelength. The resulting wavelength constrictions depend on the amount of regeneration along the two paths, as illustrated in Fig. 7.5.

In Fig. 7.5a, there is no regeneration in either the working or the protect paths, thereby requiring that the same wavelength be assigned along both of the paths. In Fig. 7.5b, there is one regeneration on the protect path and no regenerations on the working path. Normally, a regeneration affords the opportunity to change the wavelength; however, in this scenario, wavelength conversion on the protect path is not possible because it is not possible on the working path. Thus, again, the same wavelength must be assigned along both of the paths. In Fig. 7.5c, there is one regeneration on both the working and protect paths. With this configuration, the same wavelength must be used on both A-B-C and A-E-F, but a different wavelength could be used on C-D-Z and F-G-Z. If there were additional regenerations along the paths, there would be further freedom in assigning the wavelengths (as long as the portions of the paths emerging from Node A are carried on the same wavelength, and the portions of the paths converging at Node Z are carried on the same wavelength).

Note that if the configuration of Fig. 7.3c operates in the 1:1 mode, and the transponder is not tunable, then the wavelength constraints depicted in Fig. 7.5 apply. If the transponder is tunable, then the working and protect paths may be carried on different wavelengths, requiring that the transponder be retuned at the time of failure.

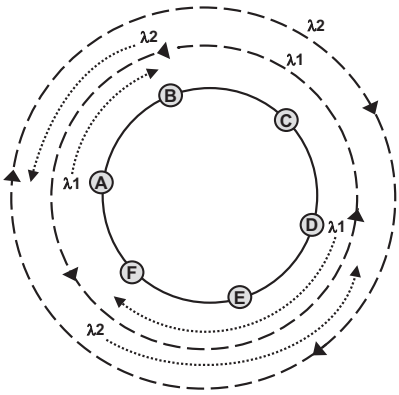


Fig. 7.6 Shared optical ring protection for two demands, *AB* and *DF,* under the no-failure condition. As shown, the wavelengths assigned to the two directions of a bidirectional connection are not the same

7.4 Ring Protection Versus Mesh Protection

Protection can be implemented using ring topologies or arbitrary mesh topologies. Ring protection tends to be simpler, although its additional constraints generally result in more required spare capacity as compared to mesh protection.

7.4.1 Ring Protection

In ring protection, the protection capacity is organized into ring structures. A ring is a simple survivable topology, such that if a single link or node failure occurs, all traffic that was routed over the failed link/node is routed in the reverse direction around the ring to avoid the failure. Note that a network with a mesh topology can use ring-based protection; a working path that traverses multiple rings is protected by spare capacity in each of the individual rings. There are many variations of ring-based optical protection, as described in Li et al. [LSTN05]. For example, the implementations may differ on whether they use two-fiber rings or four-fiber rings, or whether they restore each path individually or several multiplexed wavelengths at once.

Dedicated ring protection is fairly straightforward: A bidirectional working path combined with its protection capacity occupies one wavelength around both directions of the ring. Shared ring protection allows greater capacity efficiency, although it may be more challenging to implement, as described next.

Network-side shared ring protection is described here in more detail to illustrate some of the interesting features, especially with regard to optical bypass. An example of this type of protection is shown in Fig. 7.6, where the ring is assumed to be a two-fiber ring. One fiber carries traffic in the clockwise direction, the other in the counterclockwise direction. In the example, there are two bidirectional working paths on the ring, between Nodes A and B and between Nodes D and F, as shown by the dotted lines. Assume that shared protection is used to protect the two working

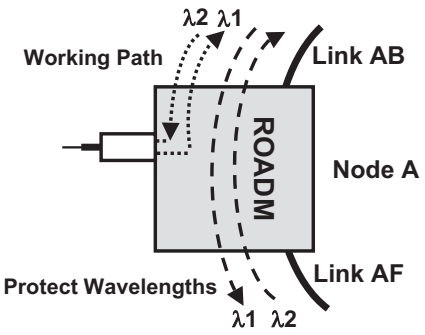


Fig. 7.7 Details of *Node A* from Fig. 7.6, under the no-failure condition. It is assumed that the reconfigurable optical add/drop multiplexer (*ROADM*) is directionless

paths. The protection capacity extends all the way around the ring, as indicated by the dashed lines. (Note that the working paths between A and B and between D and F could be portions of end-to-end paths in a mesh network where multiple rings are used for protection.)

It is assumed that each node is equipped with a directionless ROADM, as shown in more detail for Node A in Fig. 7.7. Furthermore, it is assumed that the transponders are not tunable, or that the tuning time is too slow for failure recovery. With this assumption, the wavelength used to carry the working path must also be used for the associated protect path. Assume that the working path from A to B has been assigned λ_1. Then, this same wavelength must be assigned to protect the connection along the path A-F-E-D-C-B. Because this protection wavelength is also shared by the demand from D to F, λ_1 must be used for the working path along D-E-F. Thus, λ_1 is used in the clockwise direction on one fiber to carry both working paths, and in the counterclockwise direction on the other fiber to carry the associated protection capacity. (If there are enough regenerations, this requirement can be relaxed to a degree; however, the most restrictive case is considered here.)

Now, consider the working paths in the reverse direction, e.g., B to A. Because λ_1 is used in the counterclockwise direction for protection, this same wavelength cannot be used to carry the working paths in this direction. Thus, the two directions of a connection (e.g., A to B vs. B to A) are forced to use different wavelengths. In Fig. 7.7 (and Fig. 7.6), it is assumed that λ_2 carries the counterclockwise working paths and the corresponding clockwise protection.

When a failure occurs, the two endpoints of the affected demand reconfigure their ROADMs to transmit/receive on the protect path. For example, Fig. 7.8 illustrates the configuration after Link EF has failed, and the connection between D and F has been restored. Nodes D and F now transmit and receive on the protect path, whereas Nodes A and B continue to use their working path.

The configuration of the protection capacity under the no-failure condition is also of interest in an optical-bypass-enabled network. If all ROADMs along the ring are configured to allow the protection wavelength to optically bypass, and assuming no regenerations are deployed on the protection wavelength, then *lasing* may occur where noise present in the corresponding portion of the spectrum continues to loop

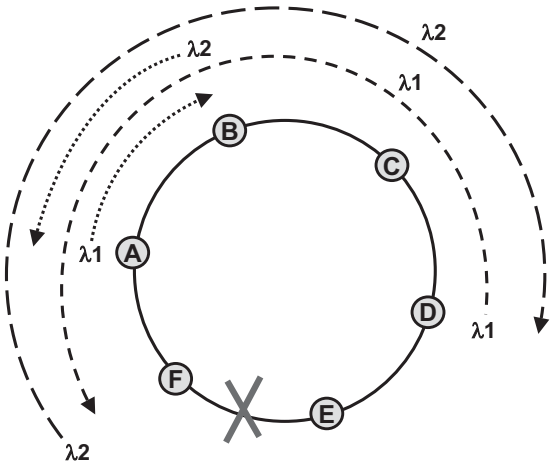


Fig. 7.8 Shared optical ring protection after Link *EF* fails. The reconfigurable optical add/drop multiplexers (*ROADMs*) at Nodes *D* and *F* are reconfigured to restore the connection over the protect wavelengths

around the ring, leading to system instabilities. One solution is to have at least one of the ROADMs along the ring configured to block the protection wavelength. In Fig. 7.6, assume that the ROADM at Node A is used for this purpose. If a failure occurs such that the protection wavelength is needed, but where Node A is not an endpoint of the failed connection, then the ROADM at Node A must be reconfigured to allow the protection wavelength to pass. This slows down the restoration process.

A second approach is to install one regeneration at some node in the protection ring (even if it is not required based on optical reach), such that the protection wavelength enters the electrical domain at this node, breaking up the purely all-optical loop. Assume that the regeneration is placed at Node A in Fig. 7.6. If a failure occurs where Node A is not an endpoint of the failed connection, then no action is needed at this node. If, however, the demand between A and B fails, then Node A must simultaneously turn off the back-to-back transponders used for regeneration and reconfigure its ROADM to transmit on the protect path. This method incurs the cost of a regeneration; however, it has the advantage that the nodes not involved in the failure do not need to be reconfigured. Note that if wavelength contention can occur on the ROADM add/drop ports (see Sect. 2.9.5), then the transponders used for the connection and the transponders used for the regeneration will need to be located on different add/drop ports.

The type of optical ring protection described above is also called two-fiber Optical-Channel Shared Protection Ring (OCh-SPRing) protection [LSTN05]. The reference to an “optical channel” emphasizes that each wavelength is restored individually. This is in contrast to two-fiber Optical Multiplex Section Shared Protection Ring (OMS-SPRing) protection [LSTN05], which restores all failed wavelengths on a fiber at once, as illustrated in Fig. 7.9. Assume that the odd-numbered wavelengths on the clockwise fiber carry working paths and the even-numbered wavelengths on that fiber carry the protect paths, with the opposite convention on the counterclockwise fiber. As shown in the figure, when a link fails, the two nodes adjacent to the failure, i.e., Nodes A and F, reconfigure a fiber switch to

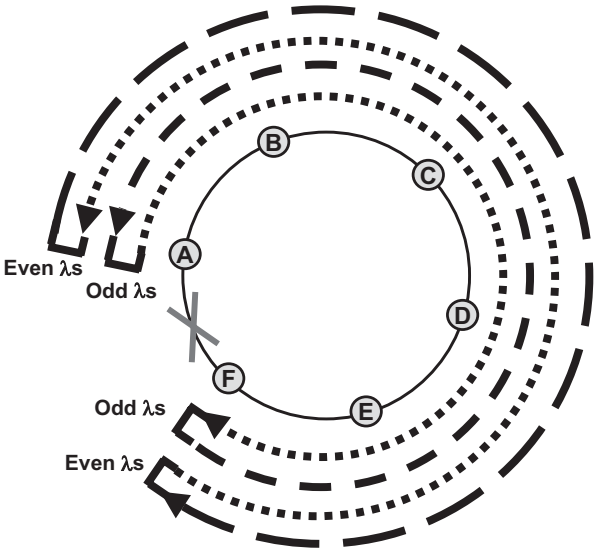


Fig. 7.9 OMS-SPRing protection where a fiber carrying traffic in one direction is looped back to the fiber carrying traffic in the opposite direction to avoid the failed link. The *dotted lines* indicate the working wavelengths and the *dashed lines* indicate the protect wavelengths

interconnect the two fibers. Thus, a working path that had been routed clockwise along E-F-A-B is now routed E-F-E-D-C-B-A-B. The same wavelength can be used along the whole path, due to the assignment of odd and even wavelengths in the two fibers. The benefit of this type of protection is that many demands are restored at once; the downside is that the post-failure path may be excessively long. (Note that OMS-SPRing protection is analogous to Synchronous Optical Network (SONET) Bi-directional Line-Switched Ring (BLSR) protection.)

7.4.2 Mesh Protection

Mesh protection allows the protect path to be routed in an arbitrary fashion rather than having to follow a precise topology such as a ring. The greater freedom generally translates to greater capacity efficiency; i.e., mesh protection requires on the order of 20–60% less spare capacity as compared to rings [GeRa00]. However, mesh protection requires more sophisticated tables to track the backup paths, and may require more communication among the nodes, especially with shared mesh protection.

An example of shared mesh protection is illustrated in Fig. 7.2. Node A from Fig. 7.2 is shown in more detail in Fig. 7.10. In Fig. 7.10a, assume that the node is equipped with a *non-directionless* ROADM-MD and an edge switch. The two working paths, which both have Node A as an endpoint, share the protection capacity as well as the protect transponder. The edge switch at the node is used to direct the appropriate client signal to the shared transponder when a failure occurs. (Other configurations are possible.) In Fig. 7.10b, assume that the node is equipped

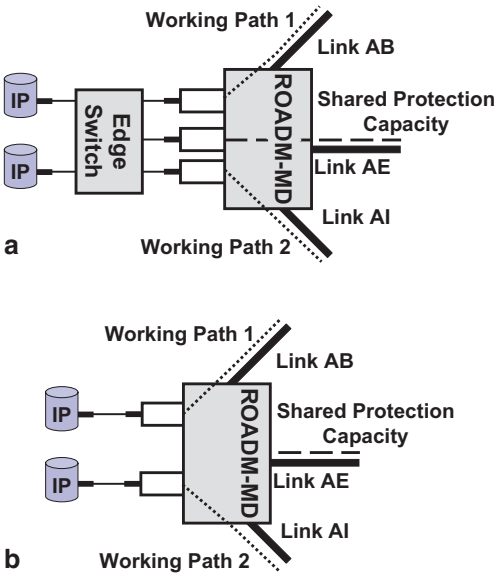


Fig. 7.10 Two shared-mesh protection configurations of Node A from Fig. 7.2 (other configurations are possible). **a** With a non-directionless multi-degree reconfigurable optical add/drop multiplexer (*ROADM-MD*), the edge switch allows the two working paths to share the protection capacity and its associated transponder. **b** If the ROADM-MD is directionless, it can be used to direct the transponder associated with a failed connection to the shared protection capacity

with a *directionless* ROADM-MD. There is no protect transponder associated with the protection wavelength. When a failure occurs, the transponder associated with the failed path is directed by the ROADM-MD to the protection capacity. In this scenario, if the transponders are not tunable (or if the tuning time is too slow for recovery), then the two working paths, as well as the shared protection capacity, must be assigned the same wavelength. If the transponders are tunable, then wavelengths can be assigned to the paths independently.

Shared mesh protection is covered in more detail, in the context of two specific schemes, in Sects. 7.8 and 7.9.

7.5 Fault-Dependent Versus Fault-Independent Protection

Another dichotomy among protection schemes is whether the protection mechanism for a connection depends on where the failure has occurred along its path. In *fault-dependent* schemes, the protection used depends on where the failure has occurred; in *fault-independent* schemes, the same protection mechanism is used regardless of the fault location. In O-E-O-based networks, where a signal can be electronically monitored at each node along a path (e.g., the SONET/Synchronous Digital Hierarchy (SDH) overhead bytes can be monitored for errors), determining the location of a failure is relatively straightforward. In an optical-bypass-enabled network, where the signal is not converted to the electrical domain at each node, fault localization is more challenging and may take longer. (Fault localization is

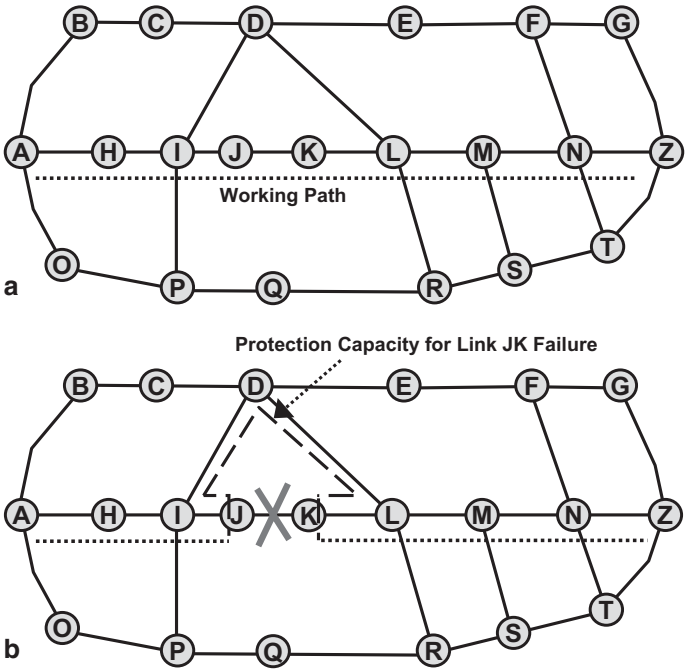


Fig. 7.11 Link-based protection. **a** The path between Nodes *A* and *Z* under the no-failure condition is shown by the *dotted line*. **b** After Link *JK* fails, the protection capacity shown by the *dashed lines* is used to avoid the failed link

covered in Sect. 7.13.) Thus, fault-independent protection schemes are generally favored in the presence of optical bypass.

7.5.1 Link Protection

At one extreme is fault-dependent link-based protection, where the recovery mechanism depends on which specific link has failed. This is illustrated in Fig. 7.11 for a demand from Node A to Node Z. Under the no-failure condition, the demand is routed over the path shown by the dotted line in Fig. 7.11a. If a link fails, the recovery occurs between the two endpoints of the failed link. This is illustrated in Fig. 7.11b, where it is assumed that Link JK has failed, and the protection resources along J-I-D-L-K are used to reroute the connection around the failure, as shown by the dashed line. Link-based protection is potentially very rapid, due to the proximity of the nodes involved with the recovery process.

In an optical-bypass-enabled network, wavelength assignment with link-based protection may be challenging. The simplest solution is to allow wavelength conversion at the endpoints of the link that has failed (e.g., Nodes J and K in Fig. 7.11) to break the interdependence between the wavelength assigned to the working path

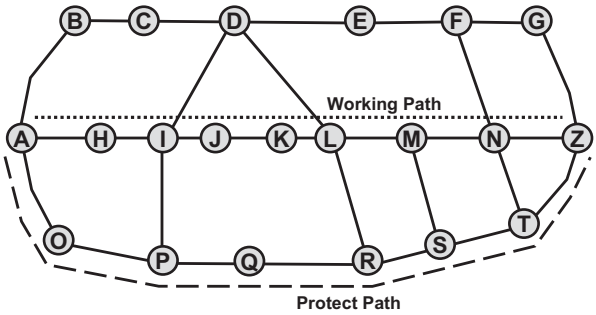


Fig. 7.12 Fault-independent path-based protection, where the same protect path is used regardless of the location of the failure on the working path

and the wavelength assigned to the detour path. However, this would require deploying extra transponders at every node along the working path for purposes of wavelength conversion. Another possibility is to use the same wavelength on both the working path and the detour path. This may lead to interesting wavelength-assignment constraints. For example, in Fig. 7.11, the same wavelength would have to be assigned on *Link IJ* of the working path and *Link JI* of the detour path (also on Link LK of the detour path and Link KL of the working path). If the connection between Nodes A and Z were bidirectional, this would imply that different wavelengths must be assigned to the working paths in the two directions. (Alternatively, all failures can be treated as bidirectional, such that Link JI of the detour path makes use of the wavelength that had been carrying the working path in the Z to A direction.) In addition to wavelength assignment challenges, adding in the detour path extends the length of the path, such that additional regeneration may be required along the new path. Overall, as this discussion elucidates, link-based protection is not well suited for optical-bypass-enabled networks.

7.5.2 Path Protection

At the other extreme is fault-independent path protection, where a diverse backup path running from the source to the destination is used whenever the working path fails, regardless of where the failure has occurred. Furthermore, the wavelength (or wavelengths) utilized by the backup path is independent of the fault location. Path-based protection is illustrated in Fig. 7.12. The demand from Node A to Node Z is routed over the same working path as in Fig. 7.11; the protect path is allocated along A-O-P-Q-R-S-T-Z. When a failure occurs anywhere along the working path, Node Z eventually detects errors in the received signal. It then communicates the failure condition to Node A to trigger the switchover to the protect path. (This assumes that 1 + 1 protection is not being used.) The distance between the demand endpoints can be large, especially in a backbone network; thus, this communication between nodes may take tens of milliseconds, delaying the onset of recovery. However, in a system where immediate fault localization may not be possible, this communication

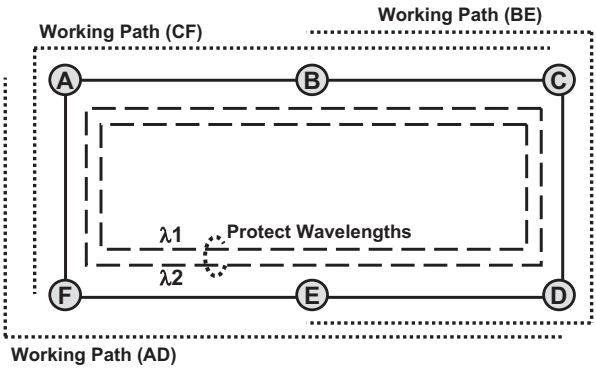


Fig. 7.13 Assume that the *AD* working path is protected with λ_1 and that the *BE* working path is protected with λ_2, and that wavelength conversion is not available on the protect paths. Then, the wavelength used to protect the *CF* working path depends upon which link fails. Thus, the protection mechanism is not failure independent

delay is preferable to suffering an even longer delay to determine the exact location of the failure. Furthermore, if the connection is bidirectional and the failure brings down both directions, then both ends of the connection can initiate a switchover to the protect path, which improves the recovery time. Additionally, path protection is typically more capacity efficient than link protection [IrMG98, AlAy99], further encouraging its use.

Note that path protection does not necessarily imply fault-independent protection. For example, there can be scenarios where the same end-to-end protection path is used for recovery regardless of the failure location, but where the wavelength assigned to the backup path depends upon the location of the failure [EBRL03]. This is illustrated in the ring network of Fig. 7.13, where there are three working paths (AD, BE, and CF) that are pair-wise intersecting. Assume that there are two wavelengths for protection allocated around the ring using λ_1 and λ_2, and assume that wavelength conversion is not available along the protect path. This protection is sufficient to protect against any single failure, depending on the wavelength assignment. Assume that the AD working path is protected using λ_1 and that the BE working path is protected using λ_2. Then, the CF working path is protected using λ_1 if Link BC fails but is protected using λ_2 if Link AF fails (either wavelength can be used if Link AB fails). Because the wavelength assignment depends on the failure location, this is not considered failure-independent protection. A third wavelength would be needed to protect the CF demand to eliminate this dependence. While this example demonstrates that fewer protection resources may be required by delaying wavelength assignment until after the failure occurs, studies have shown that, in practice, the benefit is small [DDHH99, EBRL03].

7.5.3 Segment Protection

A protection scheme that can be considered intermediary with respect to link and path protection is segment protection [HoMo02, GuPM03, XuXQ03, WLYK04]. After a working path is selected for a demand, it is divided up into multiple

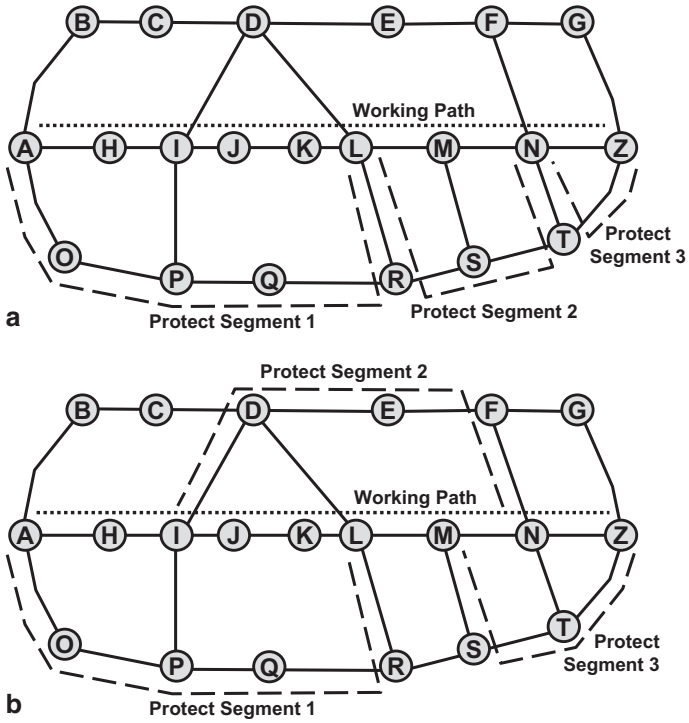


Fig. 7.14 Segment-based protection. **a** Nonoverlapping segments (*A* to *L, L* to *N, N* to *Z*), where the nodes at the intersection of the segments are vulnerable to failure. **b** Overlapping segments (*A* to *L, I* to *N, M* to *Z*) where all of the intermediate nodes are protected

segments, where a backup path is independently provided for each segment. As discussed below, by dividing the path up into multiple shorter segments, the failure recovery time is reduced as compared to path protection.

Nonoverlapping segment protection is illustrated in Fig. 7.14a, where the working path is divided into three segments, A-H-I-J-K-L, L-M-N, and N-Z. The corresponding backup segments are A-O-P-Q-R-L, L-R-S-T-N, and N-T-Z. It is assumed that the endpoint of a segment is capable of detecting a failure that occurs in its associated segment. For example, if Link JK fails, Node L detects that a failure has occurred and signals Node A to switch to the backup segment; no switchover is required in the remaining segments. This should be faster than path protection, where Node Z would have to detect the failure and signal Node A to switch to the backup path. Furthermore, the backup segments typically have fewer hops than an end-to-end backup path, such that fewer switch reconfigurations may be required with shared segment protection.

Depending on the number of segments created, segment protection may be more capacity efficient than path protection, especially when operated in a shared protection mode. With shared *path* protection, in order to share backup resources, the associated working *paths* must be disjoint. In shared *segment* protection, the

associated working *segments* must be disjoint. Segments encompass fewer links than paths, leading to a greater opportunity for sharing. Furthermore, two working segments from the same demand path can share backup resources, assuming the working segments are completely disjoint. However, as the number of segments increases, the excess routing required to provide protection for each small segment begins to nullify the benefits of better sharing. (Note that in the extreme case where each link is a segment, the scheme is equivalent to link-based protection.)

Segment protection also provides protection against more failures than path protection. For example, it can recover from a multiple-failure scenario if no more than one failure occurs in any segment. Path protection fails if there is one failure anywhere on the working path and one failure anywhere on the protect path.

One drawback to the scheme of Fig. 7.14a is that it does not provide protection if a node at a segment boundary fails. This is rectified in Fig. 7.14b, where the three segments (A-H-I-J-K-L, I-J-K-L-M-N, and M-N-Z) are overlapping such that all intermediate node failures are recoverable as well. (Failure on a link that lies in two different segments is discussed below.)

Consider using segment protection in an optical-bypass-enabled network. Because the endpoints of the segments are responsible for detecting failures and triggering the protection mechanism, it is natural to regenerate the working path at these locations [ShGr04, KaAr04]; i.e., with O-E-O regeneration, the signal will be converted to the electrical domain to better enable error detection. Thus, in Fig. 7.14a, the working path would be regenerated at Nodes L and N. (It may be possible to select the segment endpoints based on where the working path needs to be regenerated anyway due to the optical reach, so that no extra regenerations have to be added to the working path.) As noted above, however, the configuration of Fig. 7.14a does not provide protection against failures to Nodes L and N.

If the configuration of Fig. 7.14b is used instead, and the endpoints of the (bidirectional) segments are required to be O-E-O, then the working path would be regenerated at Nodes I, L, M, and N; with the number of regenerations doubled, much of the benefit of optical bypass is negated. Alternatively, regeneration could be implemented at different nodes in the two directions of the connection, where only the downstream segment endpoint is required to be O-E-O to detect failures. In the A to Z direction, regeneration is required at Nodes L and N, whereas in the Z to A direction, regeneration is required at Nodes M and I. The subconnections created by this regeneration pattern would be different in the two directions, which may ultimately lead to more wavelength assignment conflicts in the network.

Furthermore, with Fig. 7.14b, consider a failure on Link JK, which belongs to both segments 1 and 2. Some convention needs to be adopted as to which segment recovers from such a failure. This is relatively simple to implement in an O-E-O network where it is assumed that the failed link can be readily determined. However, in an optical-bypass-enabled network, handling this scenario may be more challenging. If Link JK fails, segments 1 and 2 both detect a segment failure, but the exact location may not be readily determined. If both segments were to switch to their backups, the resulting backup path would be discontinuous. Thus, fault localization, at least to some degree, is needed. In essence, with overlapping segments,

the protection mechanism may become more fault dependent. Given this extra complexity with segment protection, path protection likely remains the more desirable mechanism for optical-bypass-enabled networks, especially if protection against node failures is desired.

Another scheme, sub-path protection, is closely related to segment protection [OZSZ04]. Here, the network is partitioned into multiple areas or domains. For a path that crosses multiple domains, self-contained working and protect paths are found within each traversed domain. It differs from segment protection in that the working and protect paths within a domain can be searched for together, rather than first finding the working path, dividing it up into segments, and then looking for backup paths for each segment. Searching for the working and protect paths together within each domain is often more capacity efficient. The recovery mechanism, however, is similar to segment protection.

7.6 Multiple Concurrent Failures

As discussed in the chapter introduction, while the network may recover from a failure very rapidly, repairing a failure may take several hours. During the repair time, additional failures may occur. Furthermore, networks are vulnerable to catastrophic events such as earthquakes, tornadoes, hurricanes, or hostile acts, which may result in multiple failures.

Protecting a connection from multiple concurrent failures may be costly. Deciding whether this level of protection is worth providing largely depends on the likelihood of multiple concurrent failures and the required availability of the connection. To quantify the first criterion, the next two sections present results on the expected failure statistics in the three reference backbone networks of Sect. 1.10, both with and without catastrophic failures included in the model [Simm12]. (As a reminder, Reference Network 1 has 75 nodes and 99 links, Reference Network 2 has 60 nodes and 77 links, and Reference Network 3 has 30 nodes and 36 links. We will refer to these networks as simply Networks 1, 2, and 3.)

7.6.1 Multiple Concurrent Failures: Without Catastrophes

Most SLAs cover connection downtime due to network infrastructure and equipment failures and network maintenance activities; however, they typically exclude catastrophic events. Thus, carriers are generally more interested in analyzing whether providing protection against multiple concurrent failures is justified if catastrophes are *not* considered. To provide insight into this analysis, link failures were modeled in the three reference backbone networks, using the assumption that catastrophes do not occur.

The most common cause of a link failure is a fiber cut or an amplifier failure. A link may also be brought down due to a maintenance event; however, a carrier

Table 7.1 Average time per year with *N* failed network links: *without* catastrophes

Concurrent link failures	Network 1 (hours)	Network 2 (hours)	Network 3 (hours)
1	404	361	263
2	10	8	4
3	0.1	0.1	0.04

can exert control over when maintenance occurs, and thus such events were not included in the model. (For example, the maintenance activity on one link may be able to be postponed or cut short if another link fails.) In addition to link failures, individual connections are vulnerable to component failures, most notably transponder failures. However, as noted in Sect. 7.3.1, transponders are usually protected *locally* with an *M:N* mechanism, such that other protection resources are generally not required (e.g., the affected connection does not need to be rerouted). Thus, transponder failures were not included in the model.

The fiber-cut rate was assumed to be two cuts per 1,000 miles per year [MaLe03, Feue05]. (This is a realistic fiber-cut rate for US backbone networks; the fiber-cut rate for metro networks is five to ten times higher.) The time to repair a fiber cut was assumed to be uniformly distributed between 6 and 10 h. It was assumed that optical amplifiers have a FIT (failures in 10^9 h) rate of 2,000, with the repair rate uniformly distributed between 3 and 5 h. With these assumptions, extensive simulations were run to analyze the likelihood of multiple concurrent link failures occurring in the network, subject to the assumption of no catastrophic events. The results are shown in Table 7.1, where the table entries indicate the amount of time per year that there are expected to be *N* concurrent link failures in each of the three reference networks. (There was little probability of more than three concurrent link failures occurring in any of the networks.)

As the results indicate, a single link failure is a very common occurrence. Two concurrent link failures occur with enough frequency that demands with high-availability requirements may need protection from this scenario. Three concurrent link failures is a rare event, such that providing protection for this scenario is unlikely to be necessary to meet an SLA. For example, it is expected that Networks 1 and 2 would experience concurrent link failures in any combination of three links for roughly 6 min per year. The expected time per year that *any given demand* is afflicted by three failures is much less than this (because only certain combinations of three link failures will affect a given demand). Thus, 99.999 % availability (i.e., "platinum level" or "five 9" protection), which corresponds to no more than 5 min of downtime per year, does not require that protection against three concurrent link failures be provided.

7.6.2 Multiple Concurrent Failures: With Catastrophes

As shown in the previous section, protection against three concurrent link failures is unlikely warranted for most demands. However, a small subset of the demands

Table 7.2 Average time per year with *N* failed network links: *with* catastrophes

Concurrent link failures	Network 1 (hours)	Network 2 (hours)	Network 3 (hours)
1	425	378	272
2	19	15	7
3	3	2	0.6

in a network may be considered *mission critical,* where any downtime, regardless of the cause, is potentially harmful, e.g., connections vital for national defense. For this type of traffic, catastrophic events should be considered, in addition to fiber cuts and amplifier failures, even though their occurrence may be rare. A description of some real-life catastrophic events that have severely impacted the networking infrastructure can be found in Sterbenz et al. [SHCJ10].

Catastrophes can be modeled as correlated link failures, where the links in the affected region fail with a particular probability [LeML10]. For the purposes of the study presented here, it was assumed that a catastrophe hits, on average, one node of Network 1 each year; the rate was correspondingly lower for Networks 2 and 3, which have fewer nodes. Each node was assumed to have an equal probability of being afflicted. With probability 5 %, it was assumed that the catastrophe results in the whole node failing, which is modeled as all of its incident links failing. For the remaining catastrophes, it was assumed that each link incident on the afflicted node fails independently with probability 35 %. Furthermore, any non-incident link that passes within 35 km of the afflicted node was assumed to fail with probability 10 %. These assumptions, which are reasonable, though arbitrary, result in an average of approximately one failed link per catastrophe. The time to repair a link that has failed due to a catastrophe was uniformly distributed between 1 and 3 days. If multiple links fail due to a catastrophe, they were assumed to be repaired independently.

When catastrophes were added to the failure model of Sect. 7.6.1, the expected amount of time per year with *N* concurrent link failures in each of the three reference networks is as shown in Table 7.2. As compared to the results shown in Table 7.1, the most notable difference is that the fraction of time with three concurrent link failures increases by an order of magnitude. Thus, for mission-critical demands, providing protection against this scenario should be considered.

A few points should be made regarding catastrophic failures. First, providing protection against catastrophic failures is different from providing protection against failures to shared risk link groups (SRLGs). SRLGs typically consist of a small number of links where the associated fiber partially resides in the same conduit, such that the links are vulnerable to a single cut. SRLGs can be taken into account when determining diverse paths for protection, as discussed in Sect. 3.7.4. In contrast, during a catastrophe, *any* of the links within a geographic area may fail, which is more challenging to address.

Second, in Sect. 7.5.3, segment protection was discussed as being less vulnerable to multiple failures as compared to path protection. However, with geographically correlated failures, the links that fail are likely to fall within the same segment, such that segment protection would be vulnerable as well.

Third, there is a range of ongoing research with respect to catastrophic failures; an overview is provided in Habib et al. [HTDM13]. For example, Agarwal et al. [AEGH10], Neumayer et al. [NZCM11], and Rahnamay-Naeini et al. [RPAG11] propose methodologies for determining the points in a network that are most vulnerable to a catastrophe, where vulnerability may be measured by the amount of link capacity that is lost, the amount of carried traffic that is brought down, or the number of source/destination pairs that have been disconnected. The failure regions are typically represented by line segments or disks. In Lee et al. [LeML10] and Diaz et al. [DXMK12], the goal is to minimize the probability that a particular connection is brought down by a catastrophe, by considering link vulnerabilities when selecting the work and protect paths. The general idea is to route around regions of high vulnerability while not selecting paths that are excessively circuitous. Related work in Skorin-Kapov et al. [SkCW10] routes demands to minimize the effect of an attack. *Recovery* from a catastrophe also lends itself to optimization techniques. For example, Wang et al. [WaQY11] examines the order in which resources should be restored to maximize the aggregate prioritized flow in the network over the recovery period, subject to a limited number of repair resources. (Such a technique could possibly extend to recovery from massive electrical grid failures.)

7.6.3 Protection Schemes for Multiple Concurrent Failures

As indicated by the link-failure analyses of the previous two sections, protection against two, and possibly three, concurrent failures may be worthwhile for a subset of the traffic. To make the discussion more concrete, this section specifically discusses protection mechanisms for dual failures. However, the same approaches are generally applicable to three failures as well. In many of the schemes presented below, it is necessary to recalculate the protect path in between failures. It is common to make the assumption that there is sufficient time between restoration from a first failure and the onset of a second failure to allow such a computation [ScAF01, KiLu03, ZhZM06]. One disadvantage of schemes that rely on this assumption is that they may not be able to protect against failures with *simultaneous onset times*; however, this should be a rare occurrence. Even with a catastrophe, multiple links are unlikely to fail at precisely the same moment.

Note that while the protect path for a demand may change after a failure occurs, it is generally not acceptable to modify the working path of a demand that is otherwise unaffected by the failure, unless that demand is preemptible.

7.6.3.1 1+2 Dedicated Protection

Some of the protection methods described in Sect. 7.2 through Sect. 7.5, at least in theory, can be extended to protect against two failures. For example, 1+1 dedicated protection can be enhanced to 1+2 dedicated protection, where one working path is established along with two active backup paths. If all three paths are mutually

Table 7.3 Network connectivity statistics for the three reference backbone networks

	Network 1	Network 2	Network 3
Only 2 link-diverse paths	2,145	1,493	427
Only 3 link-diverse paths	575	271	8
4 or more link-diverse paths	55	6	0

link-and-node-disjoint, then 1+2 protection can recover from any combination of two failures, where failure recovery is almost immediate. With client-side 1+2 protection, three transponders are required at both the source and the destination of the connection. With network-side 1+2 protection, one transponder is required at both the source and the destination (with decision circuitry on the receive side); additionally, a multicast-capable directionless ROADM-MD is required at the endpoints.

One hindrance to 1+2 protection is that in most realistic networks, there are not three totally diverse paths between all source/destination pairs. This is exemplified in Table 7.3, which presents the connectivity statistics for the three reference backbone networks. For example, in the first reference network, 2,145 of the possible source/destination pairs have only two link-diverse paths between them. Thus, 1+2 dedicated protection, over three diverse paths, is precluded for connections between these pairs.

Moreover, even if three diverse paths exist, implementing 1+2 protection would require an excessive amount of protection capacity. Thus, 1+2 dedicated protection is likely only acceptable if the number of demands requiring this level of protection is very small. (Note that the routing algorithm code provided in Chap. 11 is capable of finding *N* link-disjoint or *N* link-and-node-disjoint paths between two nodes, for arbitrary *N*. If such paths do not exist, it can be used to find *N* maximally disjoint paths between the nodes.)

7.6.3.2 1+1(+1) Protection

Another option is to initially establish 1+1 dedicated protection to provide rapid protection against a first failure. Based on where the first failure occurs, a new protection path is set up, such that a new 1+1 dedicated protection scheme is established to allow rapid recovery from a second failure as well. Because the new protection path is established based on the location of the first failure, some sharing of the protection resources is possible, leading to less spare capacity requirements as compared to 1+2 protection. Note that the protection mechanism still operates in a failure-independent end-to-end mode for either the first or second failure. However, between failures, a failure-dependent calculation is performed to determine the new backup path.

7.6.3.3 1+1+ Shared Protection

A related option establishes 1+1 dedicated protection to protect against a first failure. However, after recovering from a first failure, it recalculates a new protect path,

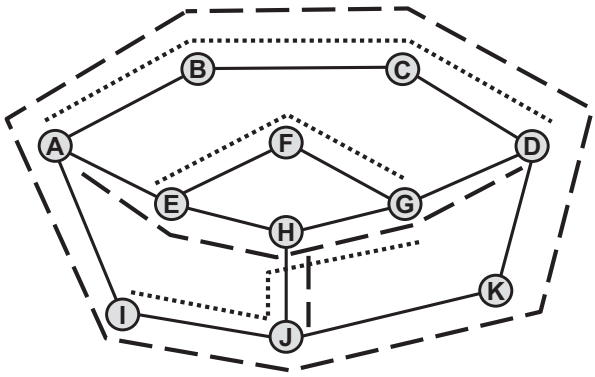


Fig. 7.15 Shared protection example where the working paths are shown by the *dotted lines* and the protection capacity by the *dashed lines.* The *AD* demand requires protection against two failures whereas the other demands require protection against just one failure. Depending on where the first failure occurs, the protect path for the *AD* demand may change

and then relies on shared protection to recover from a second failure. Thus, recovery from the second failure is slower, but requires less spare capacity.

7.6.3.4 Shared Protection

A fourth option is to solely use shared protection to protect against either the first or second failure. This is the most capacity efficient of the options because it enables the most sharing of protection resources, although it is the slowest to restore from the first failure. This option is illustrated in the example of Fig. 7.15. There are three demands, AD, EG, and IG; their respective working paths are routed over A-B-C-D, E-F-G, and I-J-H-G, as indicated by the dotted lines. It is assumed that the AD demand requires protection against any two network failures, whereas the other demands require protection against just one. There is one wavelength of shared protection capacity allocated on all links of the network except Links EF and FG, as indicated by the dashed lines.

With no failures, assume that the protect path for the AD demand is planned as A-E-H-G-D. Assume that a first failure occurs that affects one of the three working paths. If the first failure occurs along A-B-C-D and the AD demand is recovered using A-E-H-G-D, then to prepare for a second failure, the AD protect path is recalculated to be A-I-J-K-D. Next, assume that the first failure occurs along E-F-G, and the EG demand is recovered along E-H-G. Again, the protect path of AD is recalculated to be A-I-J-K-D. Finally, consider the scenario where the first failure occurs along I-J-H-G, and the IG demand is recovered using I-A-B-C-D-G. The AD protect path is now calculated to be A-E-H-J-K-D. Thus, the location of the first failure determines how AD recovers if it is affected by a second network failure. The recovery mechanism still operates in a failure-independent end-to-end mode. However, between failures, a failure-dependent calculation is performed to determine the proper backup path for AD, which allows AD to share its protection capacity with the other two demands.

Even with shared protection, providing protection from two failures for *all* demands requires an excessive amount of spare capacity [ClGr02]. However, using

shared protection to protect against multiple failures can be efficient if only a small subset of the traffic requires this level of protection. A study was performed on Reference Network 1, where 10% of the demands required protection against two concurrent failures, and 2.5% of the demands required protection against three concurrent failures (at most, one of the failures could be a node failure); the remainder of the demands were protected against just one failure. Using shared protection, the amount of required spare capacity increased by about 7% as compared to the scenario where all of the demands required protection against just one failure [CCCD12]. Similarly, modest increases in protection capacity were reported in Clouqueur and Grover [ClGr02] when just a small subset of the traffic required protection against dual failures.

In all of the schemes described above (except for 1+2 protection), there was a degree of dynamism, as a new protect path was calculated based on the location of the first failure. In the example of Fig. 7.15, it is possible to select a new protect path for the AD demand that provides protection against any second failure because there are three diverse paths between Nodes A and D. In comparison, consider the example shown in Fig. 7.16, where the source/destination pair, Nodes A and Z, has only two diverse paths between them. Assume that the working path is established along A-B-Z, with a protect path of A-C-D-E-Z, as shown in Fig. 7.16a. Assume that Link AB fails; the demand is moved to the protect path, and a calculation is performed to determine a new protect path to protect against a second failure. However, it is not possible to select a new protect path that *guarantees* recovery from *any* second failure; i.e., there is no remaining path that is fully disjoint from A-C-D-E-Z. (Clearly, if the second failure occurs on Link AC, recovery is impossible, regardless of the scheme, and thus this scenario is not of interest.) If the new protect path is calculated to be A-C-D-B-Z, as shown in Fig. 7.16b, and the second failure occurs on Link CD, then the demand is not recovered even though a viable path (A-C-F-G-E-Z) does exist. If the new protect path is calculated to be A-C-F-G-E-Z, then the demand is vulnerable if the second failure occurs on Link EZ. Again, a viable path (A-C-D-B-Z) does exist, but is not utilized. (If the original protect path had been chosen to be A-C-F-G-E-Z, this problem does not occur; however, this is assumed to be a much longer path than A-C-D-E-Z, and thus less desirable.) As this example illustrates, it may be desirable to search for a protect path *after* the next failure occurs, when the set of failed links is known, rather than relying on a pre-calculated protect path. An example of such a dynamic protection scheme is discussed next.

7.6.4 Protection through Dynamic Networking

Section 7.6.3.2 through Sect. 7.6.3.4 described schemes where a new protect path is calculated based on the failures that have already occurred, to be prepared for the next failure. As illustrated by the example of Fig. 7.16, however, source/destination pairs with few disjoint paths between them may remain vulnerable to multiple failures. The availability of a connection potentially can be improved if a restoration

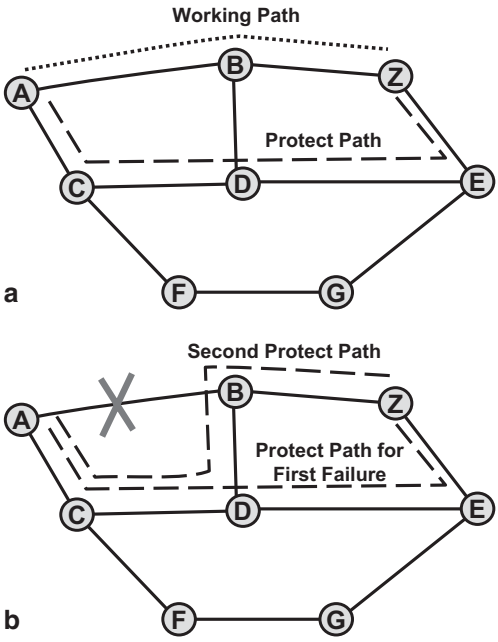


Fig. 7.16 **a** Under no failures, the *AZ* demand is routed along *A-B-Z*; the protect path is pre-calculated to be *A-C-D-E-Z.* **b** After Link *AB* fails, the connection is moved to the protect path. If the new protect path is calculated to be *A-C-D-B-Z,* the connection is vulnerable to a failure of Link *CD* (or *AC*)

path is dynamically searched for *at the time* of the next failure, rather than *before* the next failure. The drawback of this type of approach has always been the slow speed of recovery. However, one of the key developments in the field of dynamic-networking research is the potential to establish a new connection in less than 100 ms in a continental-scale network. This enables dynamic networking to be incorporated as one prong of an efficient protection methodology that can meet all but the most stringent restoration time requirements. Dynamic networking is covered more thoroughly in Chap. 8. Here, we simply assume that a mechanism exists for satisfying a new connection request in less than 100 ms.

Implementing protection via a request for a new connection needs to be done judiciously. If all connections brought down by a link failure rely on issuing a new connection request as their means of failure recovery, the network control plane would be flooded. Furthermore, a significant amount of resource contention would occur if a distributed protocol were responsible for connection setup. A better strategy is to initially pre-calculate two diverse paths for each demand, i.e., one working path and one protect path (the protection can be dedicated or shared). This enables immediate recovery from a first failure. If the demand requires protection against multiple failures, then any further disruptions are handled by issuing a new connection request. (More restrictively, the rule could be that the demand issues a new connection request in response to the Nth failure that affects it if there are fewer than $N+1$ diverse paths between the source and destination. Otherwise, the protect path is pre-calculated.) Thus, when a link fails, it is only those demands for which this is a second (or third) failure, and only those demands that require a high level of

availability, that issue a new connection request. With these stipulations, the number of simultaneous connection requests should be manageable.

Returning to the example of Fig. 7.16, the work and initial protect paths are established as assumed previously. However, after the first failure, the recovery method operates by issuing a new connection request. Thus, if the second failure is on Link CD, then path A-C-F-G-E-Z is established for recovery; if the second failure is on Link DE or EZ, then path A-C-D-B-Z is established.

As a point of interest, using simulation to estimate the availability of a connection that is protected from multiple failures may require a very long run time. A simulation methodology for more rapidly approximating the availability, using importance sampling, is described in Conway [Conw11] for dynamic path restoration.

7.7 Effect of Optical Amplifier Transients on Protection

The previous sections addressed various classes of protection schemes that are relevant in both O-E-O and optical-bypass-enabled networks. This section discusses an operational issue related to protection that chiefly pertains to optical-bypass-enabled networks due to their susceptibility to optical amplifier transients. Such transients occur, for example, when there is a sudden change in the power level on a fiber, as conceptually illustrated in Fig. 7.17. In Fig. 7.17a, three wavelengths are routed all-optically from Link AB to Link BC, via the ROADM at Node B. A fourth wavelength, λ_1, is added at Node B onto Link BC. The power level of λ_1 is plotted as a function of time in Fig. 7.17b. Assume that a failure occurs on Link AB, such that the three wavelengths on that link are suddenly brought down. Because these wavelengths had optically bypassed Node B, the failure brings down these wavelengths on Link BC as well, leaving λ_1 as the remaining wavelength on the fiber. This causes the power level of λ_1 to spike as the optical amplifiers attempt to maintain a constant total power level on the fiber, as shown in Fig. 7.17b.

Transients arise with either erbium-doped fiber amplifier (EDFA) [MoLS02] or Raman [KNSZ04] amplification. Transmission systems typically have dynamic controls to dampen such power variations and return the power level of the surviving wavelength(s) to the desired level. As illustrated in Fig. 7.17b, after a small amount of oscillation, the power level of λ_1 eventually returns to its pre-failure power level.

Excursions in the signal power level, even if brief, are undesirable as they lead to error bursts. Furthermore, in the presence of optical bypass, transients on one link may have a ripple effect, producing transients on other links, thereby causing errors to propagate. While Fig. 7.17 illustrates the effect of wavelengths suddenly being brought down, a similar, though inverse, effect occurs if wavelengths are suddenly added to a fiber.

Optical amplifier transients due to wavelengths being brought down by a failure are unavoidable (although the amplifier control mechanism should reduce their duration as much as possible). However, transients caused by a system operation,

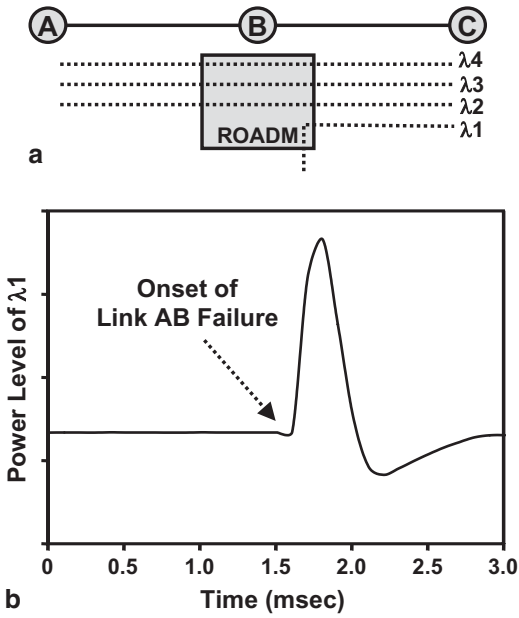


Fig. 7.17 **a** Initially, there are three wavelengths routed all-optically from Link *AB* to Link *BC*; λ_1 is added at Node *B*. **b** The power level of λ_1 is plotted as a function of time. After *Link AB* fails, bringing down the other three wavelengths, the power level of λ_1 spikes

whether it is bringing up a new connection, tearing down an existing connection, or restoring service after a failure, are generally considered unacceptable by carriers. Thus, for optical-bypass-enabled systems, it is important that protection schemes be implemented such that transients are avoided, while still taking advantage of the economic benefits of optical bypass.

For 1 + 1 dedicated protection, transients should not be an issue in the recovery process; the backup path is active even under the no-failure condition, obviating the need for path turnup at the time of failure. Transients are typically more problematic for shared protection (and 1:1 protection) schemes in which the backup path is "lit" after a failure occurs. In order to avoid transients in these schemes, the power level of the protect path needs to be increased slowly, which slows down the restoration process.

To avoid contending with transients, the protection method should not require turning on or off the WDM-compatible lasers, tuning the lasers to different wavelengths, or switching signals on the network side. Switching on the client side, however, is typically acceptable. A capacity-efficient shared mesh protection scheme that satisfies these constraints is described in the next section.

Note that O-E-O networks are more immune to such power-level transients as each link is essentially isolated by the O-E-O regeneration that occurs at each node. Thus, a failure on one link does not cause power-level variations on the adjacent links. For example, if the network of Fig. 7.17a were O-E-O based, λ_2, λ_3, and λ_4 would all be regenerated at Node B. If Link AB fails, the corresponding transponders (or regenerators) would still produce light on Link BC, leaving λ_1 unaffected.

7.8 Shared Protection Based on Pre-deployed Subconnections

The paradigm of pre-deployed subconnections can be used to avoid issues with power-level transients. The notion of a subconnection was introduced in Chap. 4, where regenerations along a path effectively break the end-to-end connection into smaller subconnections. Both ends of a subconnection are terminated in the electrical domain, with optical bypass at the intermediate nodes. In Chap. 4, the design process started with a connection and broke it into subconnections for regeneration and wavelength assignment purposes. With subconnection-based protection, the process is reversed; subconnections are pre-deployed in a network and concatenated as needed to form end-to-end backup paths.

A pre-deployed subconnection refers to a lit wavelength that is routed between two transponders, where the capacity is not currently being used to carry traffic. Thus, the transponders have been pre-deployed in the network and turned on for purposes of future traffic. Using pre-deployed subconnections as a building block for rapidly accommodating dynamic traffic or rapidly recovering from a failure was proposed in Simmons et al. [SiSB01]. Shared mesh protection based on pre-deployed subconnections is described below; further details can be found in Simmons [Simm07].

Consider the network shown in Fig. 7.18a, where it is assumed that the nodes are equipped with ROADM/ROADM-MDs and edge switches. Two working paths are established as indicated by the dotted lines, i.e., along A-B-C-D and A-J-K-I. There are three pre-deployed protection subconnections as indicated by the dashed lines: A-F-G-H, H-D, and H-I. The transponders at the endpoints of the working paths as well as the transponders at the endpoints of the protection subconnections are fed into the edge switch at the respective nodes. The details of Nodes A and H are shown in Fig. 7.18b, c respectively.

The protection subconnection along A-F-G-H is shared by both working paths. If there is a failure along A-B-C-D, the edge switch at Node H concatenates the A-F-G-H subconnection with the H-D subconnection to form a backup path along A-F-G-H-D. In addition, the edge switches at Nodes A and D are reconfigured such that the client (in this case, an IP router) is connected to the backup path. Alternatively, if the failure occurs along A-J-K-I, then Node H concatenates A-F-G-H to H-I to form a protect path along A-F-G-H-I, and Nodes A and I reconfigure their edge switches to direct the client to the protect path. The scheme provides fault-independent path-based protection, such that fault localization is not needed to initiate recovery.

The most salient features of this scheme are that the transponders at either end of the protection subconnections are always on and at the desired wavelength, and that any switching occurs in the edge switch as opposed to in the ROADM-MD (i.e., client-side signals are switched, not network-side signals). Thus, the power levels on the fibers do not change as the protect path is formed, thereby avoiding issues with optical amplifier transients. After a destination detects that a connection has failed, the speed of the protection mechanism depends on the time it takes to notify

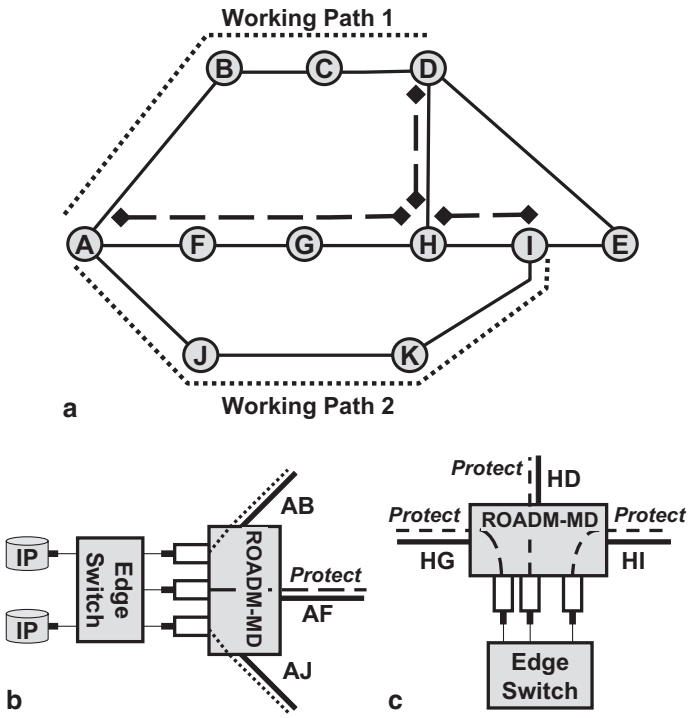


Fig. 7.18 **a** Shared protection based on pre-deployed subconnections. The three protection subconnections are indicated by the *dashed lines*. **b** Details of Node *A*. **c** Details of Node *H*. In both nodes, the edge switch as shown is photonic

the source and the intermediate switching locations (e.g., Node H) of the failure, and the time required to reconfigure the edge switches. No turning on, retuning, or switching of WDM-compatible signals is required. In a continental-scale network, recovery on the order of 100 ms should be possible.

In addition, the scheme is compatible with restoration signaling architectures that already have been developed by carriers. For example, the Robust Optical-Layer End-to-End X-Connection (ROLEX) signaling mechanism [DSST99, CDLS09] can be used to progress from one end of the restoration path to the other, selecting the subconnection to use and requesting the desired cross-connection in the edge switch. If both directions of a bidirectional connection fail, then two-ended ROLEX can be utilized, where recovery is initiated at both ends. Eventually, the processes meet at some intermediate node, such that the end-to-end bidirectional restoration path is established.

Furthermore, this protection scheme is well suited for the hierarchical protection paradigm [Simm99], such that it takes good advantage of optical bypass even for the protect wavelengths. In hierarchical protection, a subset of the nodes are chosen as “high-level” nodes, where the bulk of the protection capacity extends between these nodes, optically bypassing the “low-level” nodes. Nodes that generate a lot of

protected traffic, nodes with a high degree, and nodes that are located strategically in the network (e.g., for regeneration) are generally favored as high-level nodes. Applying this paradigm to the subconnection scheme described above, the majority of the protection subconnections are pre-deployed with high-level nodes as endpoints (a small number of subconnections need to terminate on low-level nodes in order to provide protection for the demands that originate at these nodes). This allows a significant amount of optical bypass to be realized as most of the protection capacity transits the low-level nodes.

The possible disadvantage of this scheme is the requirement for an edge switch at some, or all, of the nodes. However, as has been emphasized previously, an edge switch can improve the flexibility of a node, such that it may be desirable to deploy such switches anyway. Another pre-deployed-subconnection-based shared mesh protection scheme, which does not require an edge switch, was proposed in Li et al. [LiCS05]. The scheme uses the combination of tunable regenerator cards and directionless ROADM-MDs to concatenate the protection subconnections; the scheme is not compatible with a non-directionless ROADM-MD. Failure recovery requires retuning some transponders and regenerators, turning off some regenerators, and reconfiguring the ROADM-MDs at the connection endpoints. Thus, while the requirement for the edge switch is eliminated, the restoration process is somewhat slower and does not avoid the transient issue. As systems evolve to better manage optical amplifier transients, e.g., Zhou et al. [ZhFB07], such schemes will be more viable. Furthermore, if optical amplifier transients are managed to the point that they are a nonfactor, then the subconnection paradigm can remain in place but the requirement that they be pre-lit can be removed. This allows transponders to be shared among the subconnections at a node (i.e., a transponder is not assigned to a subconnection until it is actually needed), thereby reducing the number of transponders that must be pre-deployed. It also provides the opportunity to *all-optically* connect two subconnections, assuming that they both support the same wavelength and the concatenated subconnection does not violate the optical reach [CCCD12].

7.8.1 Cost Versus Spare Capacity Trade-off

The pre-deployed-subconnection protection scheme inherently poses a trade-off of cost versus capacity. To achieve better sharing of the protection capacity, shorter subconnections are pre-deployed (i.e., subconnections with fewer hops). This translates into a greater number of required protection subconnections, where each subconnection incurs the cost of two transponders and two edge-switch ports. Figure 7.19 illustrates this trade-off. The dotted lines represent the working paths, and the dashed lines represent the protection subconnections. The same three working paths are shown in Fig. 7.19a, 7.19b: A-E, A-G-D, and C-G. In Fig. 7.19a, there are three protection subconnections, whereas in Fig. 7.19b there are four. Either configuration is sufficient to provide protection from a single link or node failure.

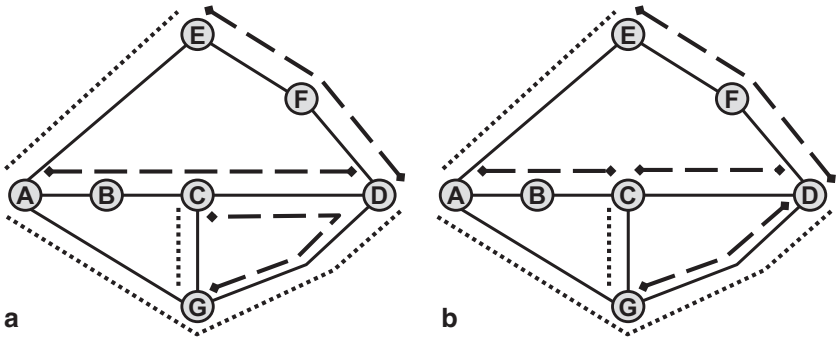


Fig. 7.19 The working paths are indicated by the *dotted lines,* and the pre-deployed subconnections are indicated by the *dashed lines*. **a** Three pre-deployed subconnections, requiring seven wavelength-links of capacity and six transponders. **b** Four pre-deployed subconnections, requiring six wavelength-links of capacity and eight transponders

Figure 7.19a requires six protect transponders and seven wavelength-links of protection capacity, whereas Fig. 7.19b requires eight protect transponders but requires only six wavelength-links of protection capacity. Thus, Fig. 7.19b is more capacity efficient, but more costly. By dividing the A-B-C-D protection subconnection into A-B-C and C-D, as in Fig. 7.19b, the C–D protection subconnection can be shared by all three working paths.

A study was performed to investigate the cost versus capacity trade-off further, using Reference Network 2. The network was assumed to be optical-bypass enabled, with an optical reach of 2,500 km. Several shared mesh protection designs were performed, where in each design an increasing number of network nodes were selected as *protection hubs*. The protection hubs are akin to the "high-level" nodes in the hierarchical protection scheme described earlier, where protection capacity is "chopped" into subconnections at the hubs. Thus, the greater the number of hubs, the shorter the resulting protection subconnections, yielding more opportunities for sharing, but resulting in higher cost. (The working paths can optically bypass the hubs, however.)

Demands requiring shared protection were added one by one to the network, with no knowledge of future demands. Enough demands were added such that the resulting capacity requirement on the most heavily loaded link was on the order of 100 wavelengths. All demands were at the line rate (i.e., no traffic grooming was needed). The paths of the demands were selected with an emphasis on sharing the existing protection capacity.

Varying the number of protection hubs produces the "cost" versus capacity curve plotted in Fig. 7.20. (Each point in the curve represents an average of several runs; the variance among the runs was very small.) The primary *y*-axis is the normalized total number of transponders required for the working and protect paths; this is used as a rough measure of network cost. (Any regeneration was tallied as two transponders.) The *x*-axis is the normalized total required capacity for the working and protect paths, measured in wavelength-km. The percentages

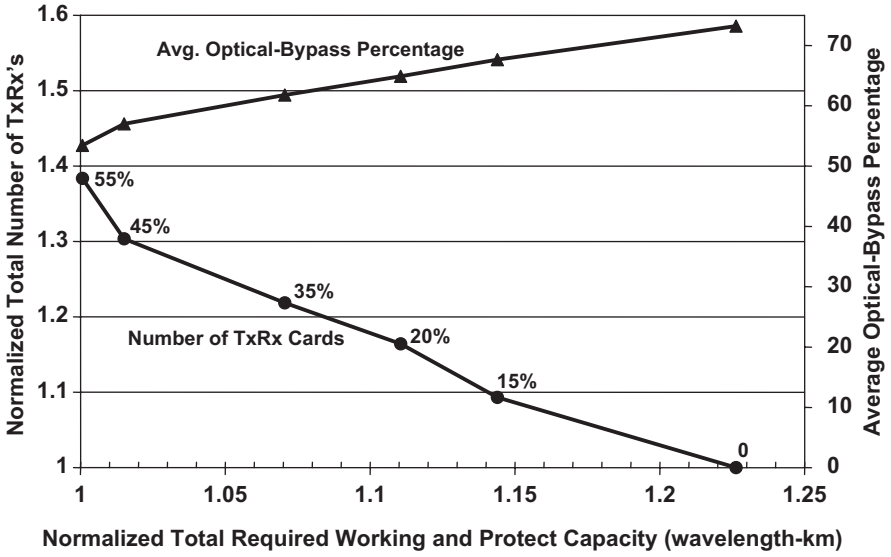


Fig. 7.20 The *lower curve* represents "cost" versus capacity for shared mesh protection based on pre-deployed subconnections, in Reference Network 2. The total number of transponders (*TxRxs*) required for the working and protect paths is used as a rough measure of cost. The percentages next to the data points indicate the percentage of nodes selected as protection hubs. The *upper curve* is the average optical-bypass percentage achieved in the network, assuming 2,500-km optical reach

next to the data points indicate the percentage of nodes that were selected as hub nodes. As expected, as the number of hubs decreases, the required total capacity increases but the cost decreases. From this graph, selecting roughly 15–20 % of the nodes as protection hubs represents a good trade-off point, where the number of transponders and the required capacity are both within ~ 15 % of their minimums. A study was performed for several other networks in Simmons [Simm07], producing similar results. Note that, while not shown on the graph, selecting 100 % of the nodes to be protection hubs reduces the total required capacity by less than 1 % and increases the total number of transponders by almost 10 %, as compared to the scenario where 55 % of the nodes are protection hubs; thus, this is not an attractive option.

Given that the protection subconnections require transponders at the endpoints, and hence O-E-O conversion, it is interesting to investigate the amount of optical bypass attainable in the network. The top curve in Fig. 7.20 plots the average optical bypass in the network (this is the percentage of working and protect wavelengths that enter a node that optically bypass the node). As the number of hubs decreases, the average optical bypass increases because the protection capacity is being electronically terminated less frequently. With 15–20 % of the nodes as hubs, the average optical bypass is about 65 %, indicating that this shared protection scheme is able to take good advantage of optical bypass.

7.9 Shared Protection Based on Pre-Cross-Connected Bandwidth

The pre-deployed-subconnection protection scheme avoids problems with optical amplifier transients; however, it does require a small amount of switching to concatenate subconnections together to form the appropriate backup path. It is worthwhile to discuss a class of shared protection schemes based on pre-cross-connected bandwidth, where the need for *any* switching at the time of failure is eliminated except at the endpoints of the failed link or path. Because of the minimal amount of switching required, these schemes are likely to be somewhat faster than the subconnection method, although the issue of optical amplifier transients does need to be addressed in these schemes.

A variety of pre-cross-connected protection structures have been proposed. We focus on two specific structures below: cycles and trails. (More general structures are investigated in Sebbah and Jaumard [SeJa12].) In either case, designing the protection for a given set of traffic demands can be challenging, due to the large number of potential cycles and trails that exist in a network. However, *column generation decomposition* has been shown to be an effective technique in the design process [KiAJ09, KSCA11].

7.9.1 P-Cycles

The origin of this protection class is pre-connected protection cycles, or *p-cycles,* where the spare capacity is pre-connected to form cycles [GrSt98]. Each cycle protects against failures on the cycle itself, as well as failures on links that straddle the cycle. The initial p-cycle proposal considered only link-based protection; however, the approach was later extended to path-based protection in Kodian and Grover [KoGr05]. The key feature is that restoration requires switching only at the endpoints of the failed link (in link-based protection) or at the endpoints of the failed connection (in path-based protection).

P-cycle link-based shared protection is illustrated in Fig. 7.21, where there are two working paths as shown by the dotted lines: B-F-E-D and B-F-G. Only one cycle of protection capacity is shown, A-B-C-D-E-F-A, as indicated by the dashed line. This one cycle is not sufficient to protect against all possible working-path failures; however, for simplicity, the other cycles are not shown. (Note that because this is a closed protection ring, *lasing* could be an issue in an optical-bypass-enabled network. Mitigating techniques such as adding a regeneration somewhere along the ring could be used, as described in Sect. 7.4.1.)

If Link FE fails, affecting the B-F-E-D connection, then the switches are configured at Nodes F and E to direct the connection around the protection cycle to avoid the failed link; i.e., the new path is B-F-A-B-C-D-E-D. This is similar to typical link-based ring recovery. However, the p-cycle can also be used to protect against failures on chords of the protection cycle. For example, if Link BF fails, affecting

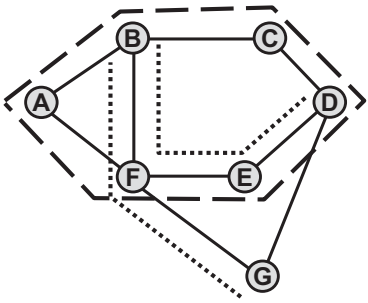


Fig. 7.21 P-cycle link-based shared protection where just one of the required protection cycles is shown as indicated by the *dashed lines*. It can protect against failures to links on the cycle (e.g., Link *FE*) and failures to links that are chords of the cycle (e.g., Link *BF*)

both working paths, the switches at Nodes B and F are reconfigured to redirect the two connections; one path is rerouted over B-A-F-G and the other over B-C-D-E-F-E-D. Taking advantage of chordal protection allows the spare capacity requirements of the scheme to be similar to that for mesh protection, while requiring switching only at the endpoints of the failure. As noted in Grover and Stamatelakis [GrSt98], the scheme combines the speed of ring protection with the capacity efficiency of mesh protection. Extensive research has been performed on p-cycles, as summarized in Kiaei et al. [KiAJ09]. For example, the p-cycle approach can be extended to protect against dual failures [ScGC04].

Nevertheless, as discussed for general link-based protection in Sect. 7.5.1, p-cycle link-based protection poses challenges for optical-bypass-enabled networks (see Exercise 7.12). These complications can be mitigated by using path-based p-cycles; however, this variant was shown to be significantly less capacity efficient [GGCS07].

7.9.2 Pre-cross-connected Trails

Restricting the pre-cross-connected structure to be a cycle is somewhat restrictive. Thus, the idea was extended to more general protection topologies in a path-based shared protection scheme called *pre-cross-connected trails* (PXT) [ChCF04]. The protection capacity is laid out in arbitrary formations similar to more general shared mesh protection schemes, with the stipulation that all "branch points" be eliminated. In the pre-deployed-subconnection scheme of Fig. 7.18, Node H represents a branch point; i.e., it is an intersection point of multiple protection subconnections, such that Node H is required to switch depending on which working path has failed. Thus, the arrangement of the protection capacity in Fig. 7.18 is not suitable for PXT protection. Figure 7.22 illustrates this same network, with the same two working paths, using PXT protection. Here, the protection capacity is routed over A-F-G-H-I-E-D; this path is considered a PXT. Under any failure scenario, only the endpoints of the failed connections need to participate in the recovery. For example, if the working path along A-B-C-D fails, the switches at Nodes A and D are reconfigured to direct the client traffic to the protection trail, with no other switching needed.

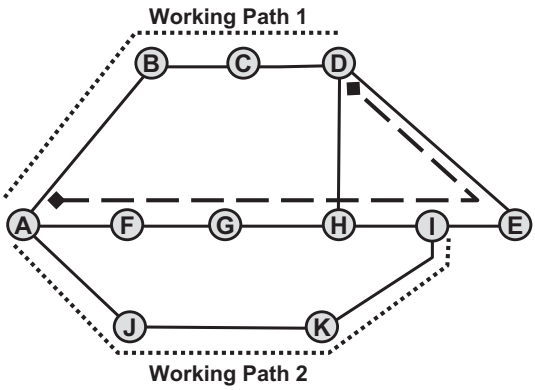


Fig. 7.22 PXT path-based shared protection. The PXT, shown by the *dashed line,* can protect either of the two working paths with switching required only at the connection endpoints

Similarly, if the working path along A-J-K-I fails, switch reconfigurations occur only at Nodes A and I.

Various experiments in Chow et al. [ChCF04] showed PXT protection to be as efficient as more general shared mesh protection schemes. Direct comparisons between PXTs and link-based p-cycles have yielded mixed results, depending on the traffic distribution and the algorithms used to enumerate the cycles and the trails [GGCS07, KSCA11]. For example, PXTs are more efficient than p-cycles when traffic on the links is unbalanced. P-cycles are not well suited to this scenario because protection is provided around the whole cycle even though some of the protection capacity may be unnecessary. (This shortcoming of p-cycles is addressed in a more flexible, path-based, *virtual cycle* protection scheme [EiLS11] .)

Both PXTs and p-cycles potentially have issues with optical amplifier transients, even if the protection structure is pre-lit prior to the failure. Depending on the failure scenario and the nodal architecture employed at the point of recovery, the restoration process may require that ROADM-MDs be reconfigured and/or that transmitters be turned on at the time of failure. (This was covered in more detail in the first edition of this text; also see Exercise 7.13.) These operations need to be performed gradually to avoid optical amplifier transients. While this does not preclude such schemes from being employed in optical-bypass-enabled networks, it must be factored in when estimating the time for recovery.

7.10 Network Coding

In current carrier networks, the only means of providing near instantaneous recovery from a network failure is through 1 + 1 dedicated protection, as discussed in Sect. 7.2.1. The disadvantage of this approach is the large amount of capacity required to provide a dedicated backup connection. While the pre-cross-connected protection class is more capacity efficient than 1 + 1 protection, and faster than general shared-mesh protection, these schemes still do not provide near-immediate

recovery. They require some amount of switching at the time of the failure; furthermore, the switching may have to be performed gradually to avoid optical amplifier transients.

Recently, the theory of *network coding* [ACLY00] has been extended to the realm of failure recovery to provide both near-immediate recovery *and* capacity efficiency [KoMe03]. In this approach, the network nodes are used to process a set of signals such that a destination receives independent, typically linear, combinations of signals over diverse paths. The combinations are such that if one signal is lost due to a failure, it can be recovered (almost) immediately from the other signals that are received. By *combining* signals at specific nodes, rather than *duplicating* the signals end to end, network resources may be used more efficiently.

Network coding is compared with more conventional protection schemes in Fig. 7.23. Assume that there are two connections that require protection, AZ and BZ. In all panels of the figure, the working paths are shown by the dotted lines, and the protection capacity is indicated by the dashed lines. Figure 7.23a illustrates 1 + 1 protection for the two connections, requiring six wavelength-links of protection capacity. Shared mesh protection is shown in Fig. 7.23b; it requires only four wavelength-links of protection capacity, but provides slower recovery times. The network coding solution is shown in Fig. 7.23c. The AZ and BZ protect signals are combined in Node C, such that one signal, represented by AZ⊗BZ, is transmitted from Node C to Node Z. This solution requires four wavelength-links of protection capacity *and* provides near-immediate recovery. (The small recovery delay is due to the different path latencies and the processing required to recover the lost signal.) The operations that are employed by Node C to combine the AZ and BZ signals must allow recovery of the individual signals. For example, if there is a failure on Link DZ, Node Z still receives the BZ signal and the AZ⊗BZ signal, which allows it to reconstruct the AZ signal. Similarly, if there is a failure on Link EF, Node Z recovers the BZ signal from the AZ and AZ⊗BZ signals. (Note that the bits of the two received signals need to be aligned as they were at Node C in order to reconstruct the lost signal.)

In optical-bypass-enabled networks, an interesting question is whether the processing of the signals at an intermediate node (e.g., Node C in Fig. 7.23c) can be performed all-optically [MDXA10, LLLC12]. It is likely that only simple coding can be performed in the optical domain. For example, Menendez and Gannett [MeGa08] propose using photonic bitwise *exclusive-or* (XOR) to combine two signals. However, this type of simple operation is too restrictive to obtain the full benefits of network coding in more general scenarios [KiMO09]. If coding must be performed in the electrical domain, then the number of required transponders may increase, unless it is performed in conjunction with regeneration that is required anyway. Comparing the number of transponders needed for protection in each of the solutions in Fig. 7.23, we see that 1 + 1 protection requires four protect transponders; shared mesh protection requires six protect transponders; and network coding, with all-optical coding at Node C, requires three protect transponders. If the coding is performed in the electrical domain at Node C, then three additional transponders are required at this node. However, if electronic coding is required, a better option

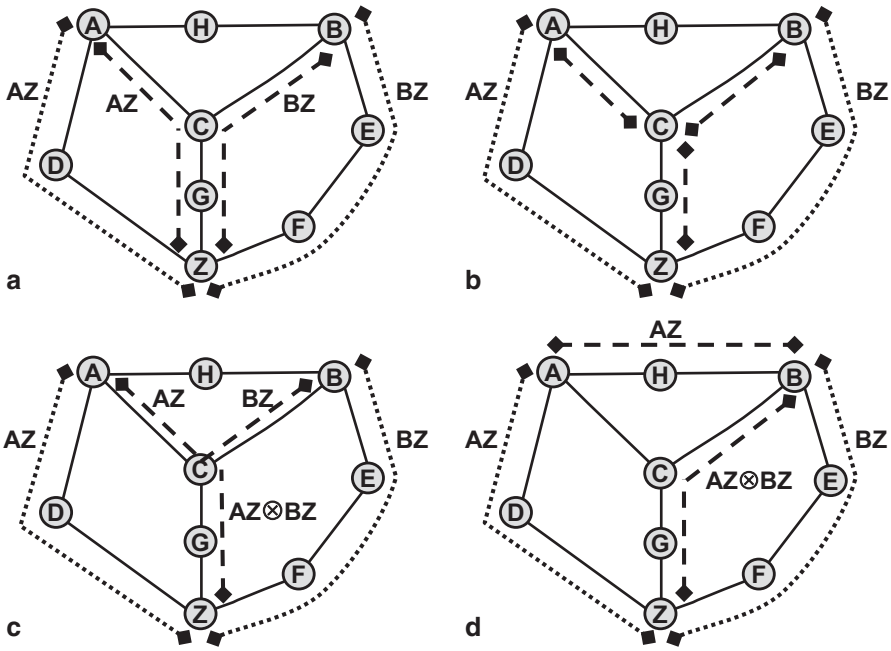


Fig. 7.23 The working paths are shown by the *dotted lines,* the protection capacity by the *dashed lines.* **a** 1 + 1 protection. **b** Shared-mesh protection. **c** Network coding, with all-optical coding at Node *C.* **d** Network coding, with electronic coding at Node *B*

may be for Node A to transmit the AZ protect signal to Node B, and have Node B perform the coding operation, as shown in Fig. 7.23d. This requires four protect transponders instead of six, but requires five wavelength-links of protection capacity instead of four. (In all of these examples, it is assumed that separate transponders are needed for the working paths and the protection capacity.)

To mine the bandwidth benefits of network coding, there must be multiple signals that can be advantageously combined [Kama08]. This often arises with multicast, which is a prime target of network coding research [MeGa08, KiMO09, MDXA10]. Figure 7.24 demonstrates the use of network coding to protect two multicast connections, M_A and M_B, which originate at Nodes A and B, respectively, and terminate on common destination nodes X, Y, and Z. The working paths are shown by the dotted lines, and the protection capacity by the dashed lines. Coding is performed at Node C, with the resulting combination $M_A \otimes M_B$ sent to all three destinations, thereby protecting both multicast connections from any single network failure.

Network coding does not always provide a bandwidth advantage as compared to other protection strategies that only require switching at the destination node. Clearly, it depends upon the traffic pattern and the network topology. For example, Kim [KiMO09] simulated many multicast scenarios on three small networks, randomly selecting one source and three destinations from among the nodes. Network coding provided a bandwidth advantage in a small percentage of the tested scenarios.

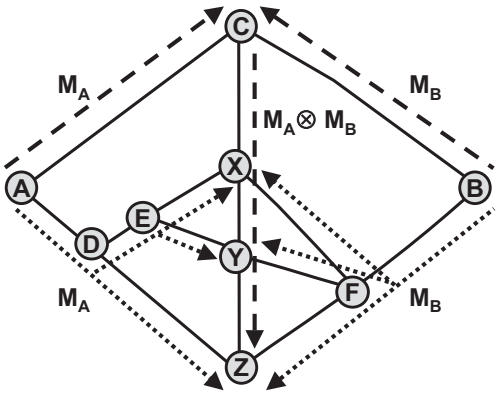


Fig. 7.24 Node *A* multicasts connection M_A to destination nodes *X, Y,* and *Z.* Node *B* multicasts M_B to the same destinations. Protection is provided by the combined $M_A \otimes M_B$ signal, which is sent to all three destinations

(Interestingly, the strategies used to find good network coding designs could be applied to more conventional protection schemes as well.)

Another important consideration is the reliability of the coding operation. If the failure probability of the equipment used to combine signals is similar to the failure probability of a link, then the overall availability should approach that of 1 + 1 protection [MeGa08].

More research is needed to determine the practicality and applicability of network coding in carrier networks.

7.11 Protection Planning Algorithms

This chapter has presented numerous optical protection schemes that differ with respect to capacity efficiency, recovery speed, fault coverage, equipment requirements, wavelength assignment restrictions, and design complexity. Different network planning algorithms are needed to optimize the various schemes. This section addresses protection planning algorithms at a high level, to illustrate general techniques that may be used in the design process.

Section 3.7 covered routing algorithms that find link-and-node-disjoint paths between a source and destination, i.e., shortest pair of disjoint paths (SPDP) algorithms. If completely disjoint paths are not possible, such algorithms can find the maximally disjoint set of paths. SPDP algorithms are used in many protection design strategies.

Section 3.7 also covered the notion of shared risk link groups (SRLGs), where, for example, a single failure may cause multiple links to fail due to shared conduit. More generally, shared risk groups (SRGs) refer to any set of network resources that are part of the same failure group. Diversity with respect to SRGs is desirable as part of protection planning. As described in Sect. 3.7, there are various graph transformations that can be used to handle the most common shared-risk configurations.

7.11.1 Algorithms for Dedicated Protection

First, consider network planning with dedicated protection. Routing can be performed using the techniques of Chap. 3, where an SPDP algorithm is used to find a set of candidate working and protect paths that are link/node disjoint. As demand requests enter the network, one of the candidate path pairs is selected, typically based on the current network state. Regeneration sites are chosen for the working and protect paths, if needed, and the resulting subconnections are assigned wavelengths.

Depending on the protection scheme, the wavelength assignment process may need to be modified to account for additional constraints. For example, with the network-side 1 + 1 protection scheme described in Sect. 7.3, wavelengths cannot be assigned independently to the working and protect paths. One method of enforcing these constraints is to form "subconnection groups" that are composed of the subconnections that must be assigned the same wavelength. A subconnection group can then be treated as if it were one large subconnection composed of all of the hops that are included in each of the individual subconnections in the group.

It is also possible to use one-step RWA methods with dedicated protection. However, as pointed out in Sect. 3.7.4, the graph transformations that accompany some one-step methods may produce SRLG-like situations that need to be handled when running an SPDP algorithm on the transformed graph. The transformation procedures also need to be modified to enforce situations where portions of the working and protect paths are required to be assigned the same wavelength.

7.11.2 Algorithms for Shared Protection

In contrast to dedicated protection, shared protection requires more advanced algorithms because sharing creates an interdependence among the protected demands. In general, two demands may share protect capacity only if their working paths are disjoint. Thus, the selection of the working path for a demand may affect the capacity required for protection. For example, a longer working path may be selected in order to make better use of the protection capacity already deployed. Consider the example of Fig. 7.25, where there is one protected demand between Nodes A and D. The working path is routed over A-E-D and the protect path over A-B-C-D. Assume that a request for a second protected demand between Node A and Node D arrives. If the working path of this second demand is also routed over A-E-D, which is the shortest path, it will not be able to share the existing protection capacity along A-B-C-D. If, however, the new working path is routed over path A-F-G-D, which is 800 km longer, the protection capacity along A-B-C-D can be shared by both demands (because the working paths are disjoint).

While selecting the longer working path for the new demand may reduce the overall capacity requirements, it may ultimately result in a higher cost. Assume that the optical reach in Fig. 7.25 is 1,000 km. If the new working path is routed on A-F-G-D and the protect path shares the existing protection capacity along

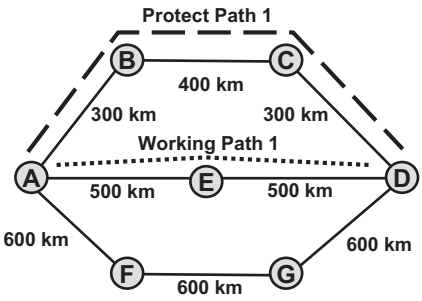


Fig. 7.25 The existing working and protect paths are shown. If another connection between *A* and *D* is added, routing it along *A-F-G-D* allows it to share the existing protection capacity

A-B-C-D, then two additional regenerations are required (on the working path). If the new working path is routed on A-E-D and the protect path is a new wavelength on A-B-C-D, then no additional regenerations are needed, although the additional protect wavelength may incur cost at its endpoints (e.g., a transponder and a switch port, depending on the protection mechanism). It may be less costly to add this extra protection capacity rather than select a working path with two extra regenerations. Thus, simply maximizing protection sharing may not always be the optimal strategy. These types of trade-offs need to be evaluated as part of the design process.

Note that multipath routing may similarly be used as a means to decrease the amount of required protection capacity. A demand is split into multiple lower-rate signals, each of which is routed over a *diverse* working path. A diverse protect path can then be shared by each of the working paths (see Sect. 3.11.2).

The choice of the protect path also affects the amount of attainable sharing. Figure 7.26 shows a network with two existing protected demands. Demand 1 has working path A-H-D and protect path A-B-C-D; Demand 2 has working path D-I-G and protect path D-E-F-G. If a new protected demand is added between Nodes A and G, and the working path is routed over A-J-K-G, then there are several options for its protect path. Assume that the shortest possible protect path for the new demand is A-L-M-G. If this protect path is selected for the new demand, then a new protection wavelength will need to be allocated along the path, with no sharing of existing protection resources. Alternatively, if the protect path is routed over A-B-C-D-E-F-G, then it can share the protection capacity that is already deployed, with no additional spare capacity needed. The second option is more capacity efficient, although it produces a protect path that is twice as long.

Routing the working path and/or the protect path over longer paths, or paths with more hops, generally increases the vulnerability to failure. As was discussed in Chap. 6 relative to grooming, the planning algorithm may need to enforce rules regarding how much excess routing can be tolerated in the working path and protect path in order to attain better sharing (the excess factors for the working and protect paths can be different). This can be based on the desired availability of the demands.

Next, three strategies for routing demands with shared protection are outlined.

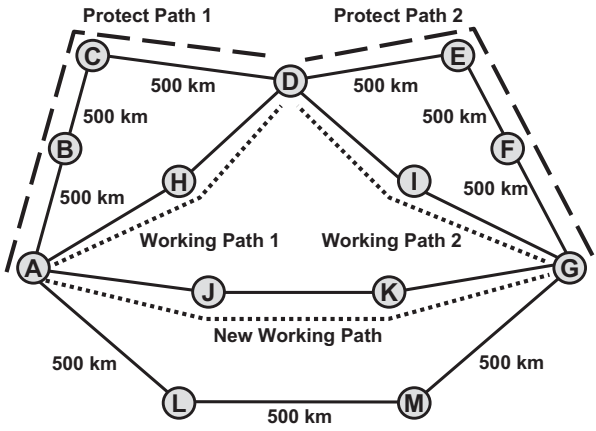


Fig. 7.26 If the protect path for the new demand is selected to be *A-B-C-D-E-F-G,* then the existing protection capacity can be shared. However, this path is twice as long as a protect path along *A-L-M-G*

7.11.2.1 Candidate-Path Strategy

Assume that a request arrives for a protected demand between a particular source and destination. One strategy is to consider each of the candidate path pairs that have been pre-calculated for the source/destination/protected class. (Generating candidate paths for protected demands using an SPDP algorithm was covered in Sect. 3.7. More candidate path pairs should be pre-calculated when the protection mode is shared as opposed to dedicated, e.g., at least five path pairs. Recall that while the paths comprising a pair should be maximally disjoint, the set of candidate path pairs may have some links/nodes in common.) For each candidate path pair, the total amount of capacity that would need to be added to the network to accommodate the working and protect paths is evaluated. Capacity needs to be added along the whole working path, whereas the protect path may be able to share protection capacity that is already deployed. The cost, due to added transponders and regenerators, along with the loading on the path links, is also evaluated for each candidate path pair. Based on the relative weighting of factors such as capacity and cost, one of the candidate path pairs is selected.

If the selected candidate path pair yields little or no sharing of the protection capacity, then other paths can be considered. For example, the candidate paths can be examined again, where the working and protect paths are swapped. Using a candidate protect path as the working path, and vice versa, may allow better sharing, although it likely increases the distance of the working path.

7.11.2.2 Shareability-Metric Strategy

Another commonly used shared protection design strategy is to first select a working path, and then search for a protect path using a relatively simple graph modification that assigns link metrics based on the shareability of the protection capacity (e.g., [BLRC02]). To simplify the discussion, assume that the demands are at the

wavelength level, and that protection capacity is allocated in units of wavelengths. The first step in the graph transformation is to temporarily remove all links from the topology that are included in the working path. (To be more precise, any link that is part of an SRLG that contains a working-path link is removed.) Furthermore, any link that has no free capacity *and* no shared protection capacity is removed. Each remaining link is examined to determine if there is a shared protection wavelength already allocated on the link that could potentially be used to protect the new demand. For a protection wavelength to be shareable with the new demand, the working paths that already use this wavelength for protection must be disjoint from the working path of the new demand. Any link that is found to have at least one shareable protection wavelength is assigned a small cost metric. All other links are assigned their usual cost metric (e.g., link distance).

A shortest-path algorithm is then run on the modified graph to determine the protect path. Assigning a relatively low metric to the shareable links drives the protect path towards those links. Note that if the metric is too low for these links, then the path that is found may be excessively long (e.g., if the metric of the shareable links were zero, then there would be no penalty at all for traversing more links). As investigated in Bouillet et al. [BLRC02], the metric can be adjusted to strike the desired balance of sharing and path length; using a metric that is roughly 50 % of the usual link metric was shown to achieve a good balance.

If this method is successful in finding a path, and the path distance and sharing are acceptable, then this path can be used as the protect path for the new demand. The path is guaranteed to be disjoint from the working path because the working-path links were eliminated from the transformed graph. As detailed in Sect. 3.7, using a two-step process that searches for a protect path after eliminating the links of the working path may be suboptimal; thus, with this shareability-metric technique, it may be desirable to run through this process with different candidate working paths.

For shared protection based on pre-deployed subconnections (Sect. 7.8), where the subconnections are terminated in a transponder at either endpoint, the above algorithm needs to be modified. Rather than looking at the shareability of links, the modified algorithm focuses on the shareability of the existing protection subconnections. If a protection subconnection is potentially shareable with the new demand, a link should be added to the graph that captures the subconnection, and the link should be assigned a relatively low metric. For example, if a shareable subconnection extends between Node X and Node Y, then a link between Nodes X and Y should be added to the graph. Any links in the true topology that do not have a shareable subconnection between them should be assigned their usual link metric. Running a shortest-path algorithm on this modified graph then favors routing the protect path over shareable subconnections.

7.11.2.3 Potential-Backup-Cost Strategy

Note that the previous method focuses on finding a good protect path given a particular working path. However, because of the constraint that two working paths that have a link or intermediate node in common cannot share a protection wavelength,

the selection of the working path itself affects the amount of achievable sharing. There are various shared protection methodologies that take this into account, where the algorithm is proactive in searching for a working path that will likely yield a protect path that can take advantage of sharing.

For example, in Xu et al. [XuQX07], P_i is defined to be the maximum amount of protection capacity required on any link in the network in order to protect against a failure of link i. Let M be the maximum P_i over all links i in the network. Assume that the goal is to minimize the total utilized wavelength-links. Then each link i in the network is assigned a metric equal to $(1+\alpha)$, where α is proportional to P_i/M. The "1" term captures that the working path will occupy one wavelength on the link, whereas the α term represents the potential backup cost. A shortest path algorithm is run with this metric assignment, and the resulting path is taken as the working path. One can then use the methodology described in the previous section, where the working path links are removed and the remaining links are assigned a metric based on their shareability, to find a suitable protect path; this requires a second invocation of the shortest-path algorithm.

7.11.2.4 Combining Strategies

A good overall design strategy is to use a combination of these techniques. The candidate-path strategy can be used initially to select the working and protect paths for a new demand. After the paths are selected, the amount of sharing can be evaluated. For a demand where there is little or no sharing of its protection capacity, the shareability-metric methodology described above (possibly in concert with the potential-backup-cost strategy) can be used to improve the sharing.

By using the candidate-path strategy first, demands are more likely to be routed over a preferred path. Additionally, this typically produces good sharing for the bulk of the demands, such that the other techniques, which are slower, are only needed for a relatively small number of demands.

In the study of Sect. 7.8.1, the candidate-path strategy was used, combined with the shareability-metric strategy for demands not achieving a high degree of sharing. Using the potential-backup-cost strategy, or using the shareability-metric strategy for *all* demands, did not appreciably affect the results.

7.11.2.5 Distributed Shared Protection with Partial Information

Thus far, it has been assumed that the shared protection algorithm has knowledge of all existing working paths and their respective backup paths. The algorithm can use this information to determine whether certain protection capacity is shareable by a new working path. If shared protection is calculated in a distributed environment, updating each node with information on each working path and its respective backup path may be too onerous. Thus, schemes that operate on only partial information have been devised.

For example, Kodialam and Lakshman [KoLa00] assume that only aggregated link information is disseminated indicating how much capacity on link i has been allocated for working paths, how much has been allocated for backup paths, and how much capacity is unallocated; let these parameters be represented by W_i, B_i, and R_i, respectively. For any new working path that is chosen, the W_i of the links on the new path increases. Let $W_{\max}$ be the maximum such W_i on the new working path. If the backup path is carried on link j, then $W_{\max}$ backup capacity is needed on link j to *guarantee* that there are sufficient protection resources for the new working path. Thus, B_j can be compared to $W_{\max}$ to determine how much more protection capacity would need to be allocated on link j if the backup path of the new demand is routed on this link. This is used to assign link metrics; i.e., link j is assigned a metric of $\text{Max}[(W_{\max} - B_j), 0]$; if this value is greater than the residual capacity, R_j, it is instead assigned a metric of infinity. The links of the selected working path are also assigned a metric of infinity. By running a shortest-path algorithm with these metrics, the likelihood of finding a backup path that minimizes the required amount of new capacity is enhanced. Multiple iterations can be run where different working paths are considered. Studies showed that while this scheme requires more capacity as compared to when complete information is provided regarding working and backup paths, it performs significantly better than the case where protection capacity is not shared at all. (To accommodate node failures, each node is replaced by two dummy nodes interconnected by a link; the incoming nodal traffic terminates on one of the dummy nodes, and the outgoing nodal traffic is sourced by the other node. The link connecting the two dummy nodes is then included in the above algorithm.)

In Qiao and Xu [QiXu02], information that is somewhat more detailed is maintained, allowing the scheme to be more efficient in allocating protection capacity. Rather than simply tracking the total working capacity allocated on each link, the scheme tracks P_i^k, the amount of working capacity on link i that is protected by link k, for all links i and k in the network. When working paths are established, the signaling message specifies the protect path as well, so that P_i^k can be updated accordingly. This can then be used to estimate the amount of additional protection bandwidth that would need to be allocated on a link to protect a particular new working path. This estimate is used as the link metric, similar to the scheme above, so that running a shortest-path algorithm minimizes the estimated additional protection bandwidth. More details can be found in Qiao and Xu [QiXu02].

Related schemes that operate on partial information are proposed by Sridharan et al. [SrSS02]. The emphasis in these schemes is on finding a wavelength-continuous backup path. Thus, information regarding whether a given wavelength on a given link is used for a working path, for a protect path, or is unassigned is disseminated.

Another scheme, *distributed path selection with local information*, is designed for *selecting the wavelengths* on the protect path using a distributed protocol such as Generalized Multi-Protocol Label Switching (GMPLS) [AYDA03]. Nodes track the state of the wavelengths on each of their outgoing links, where the three possible states are: available, assigned and non-shareable, and assigned but shareable. For

the last category, the nodes must also track which working paths are being protected by these shareable wavelengths. GMPLS signaling can then be used to favor reusing a shareable wavelength, assuming that the working path of the new connection is disjoint from the working paths already protected by that shareable wavelength. Whenever shared protect resources are reserved on a path, it is necessary to also propagate to the nodes on that path information regarding the associated working path. This scheme is discussed in further detail in Chap. 8, which covers dynamic networking.

All of the methods described above avoid having to disseminate to *all* nodes the details of each working and protect path, and thus are more suitable for distributed computation of shared protection.

7.12 Protection of Subrate Demands

Chapter 6 covered subrate demands, where the bit rate of the demand is less than that of a wavelength. This can be, for example, IP, Optical Transport Network (OTN), or SONET/SDH traffic. As discussed there, the two most common ways of handling subrate demands are with end-to-end multiplexing (Sect. 6.2) or with grooming (Sect. 6.3). With end-to-end multiplexing, subrate demands with the same source and destination are grouped together in a wavelength and carried as a unit from source to destination. With grooming, arbitrary subrate demands can be bundled together to form well-packed wavelengths, where repacking of the wavelengths can occur at intermediate points along the demand paths. While Chap. 6 focused on the multiplexing and grooming aspects, this section specifically addresses protection for subrate demands.

There are generally two approaches to protecting subrate demands. First, there is wavelength-level (i.e., optical-layer) protection, where the subrate demands are bundled into wavelengths, and then the wavelengths are routed with protection. Second, there is subrate-level (i.e., grooming-layer) protection, where the individual subrate demands are routed with protection, and then the working and the protect paths are bundled into wavelengths. Both methods are discussed below in the context of grooming (end-to-end multiplexing can be considered simplified grooming where bundling occurs at only the source and destination nodes).

7.12.1 Wavelength-Level (Optical-Layer) Protection

The network shown in Fig. 7.27 is used to illustrate wavelength-level protection, where the wavelength line rate is assumed to be 40 Gb/s. As shown in the box at the top of the figure, there are three protected 10G demands, A-E, A-F, and G-F, as well as one unprotected 10G demand, A-F. Assume that Nodes A, B, and E are equipped with grooming switches. The following grooming scheme is used (others are possible):

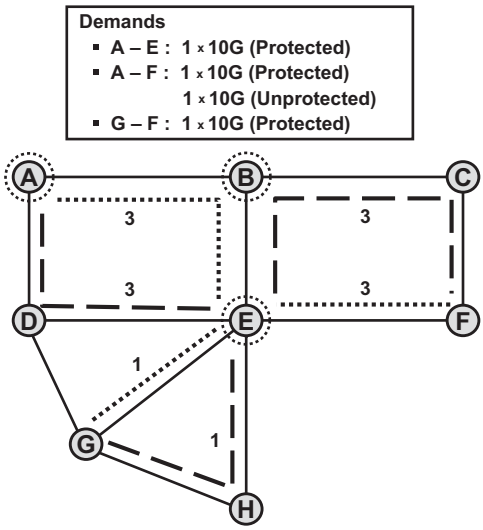


Fig. 7.27 Protection of the specified subrate demands using wavelength-level protection. The *circled nodes* indicate those nodes equipped with grooming switches. The *dotted lines* indicate the working paths of the grooming connections, and the *dashed lines* indicate the corresponding protect paths. The *numbers* indicate how many 10G subrate demands are carried in the grooming connections

- All of the demands from Node A are grouped into a wavelength and routed to Node E.
- The demand from Node G is routed on a wavelength to Node E.
- All demands destined for Node F are carried on a wavelength from Node E to Node F.

Using the terminology of Chap. 6, each of these three groupings is a *grooming connection* (GC). Each GC contains at least one protected subrate demand, thus each GC requires protection. With protection at the wavelength level, the GCs can be treated as three independent wavelength-level demands (A-E, G-E, and E-F) that require protection. Any of the protection schemes discussed in this chapter can be used to protect the GCs. For example, either dedicated or shared protection can be used.

Assume that the working and protect paths for each of the GCs are as shown in Fig. 7.27, with the working paths shown by the dotted lines and the protect paths by the dashed lines. The numbers next to the paths indicate how many 10Gs are carried in the GC. This design requires grooming only at Node E.

Note that the AF demand follows A-B-E-F for the working path and A-D-E-B-C-F for the protect path, with grooming occurring at Node E. Although the end-to-end working and protect paths for this connection are not link diverse, the scheme does provide protection against a single link failure, because the A-E and E-F GCs are protected independently. (This is similar to segment protection.)

Being able to treat the GCs like a wavelength service simplifies the planning and operation of the network; e.g., the system needs to track the protection path for each GC as opposed to each subrate demand. One disadvantage of wavelength-level protection is that the subrate demands are vulnerable to failures at the grooming nodes. For example, if the grooming switch at Node E fails, then all of the traffic in the figure

is brought down. Another drawback is that an unprotected subrate demand that has been groomed with a protected subrate demand will end up being protected. In the example, the AF demand that did not require protection is routed in protected GCs, thereby receiving a higher level of service than the customer requested (and paid for). Another option would have been to route the unprotected subrate demand in a separate, unprotected GC, but this would have utilized more wavelengths in the network and required more transponders.

Capacity-wise, another potential disadvantage of wavelength-level protection is that the fill rate of the GC applies to both its working and protect paths; i.e., the working and protect portions of the GC cannot be filled independently. Thus, in the example, unless additional subrate demands are added to the GC between Nodes G and E, the wavelength routed on G-E and the wavelength routed along G-H-E will remain only 25 % full.

7.12.2 Subrate-Level (Grooming-Layer) Protection

With subrate-level protection, each subrate demand is individually protected. Most grooming switches incorporate some type of subrate-level protection mechanism. For illustration purposes, we assume that path-based protection is employed; however, more general schemes are possible. For example, the MPLS Next-Hop Fast Reroute scheme, commonly used to protect IP traffic, is link based (see Sect. 6.7.3).

Each subrate demand requiring protection is initially routed along disjoint working and protect paths. The individual subrate paths are then groomed together into wavelengths. During this grooming step, the working and protect paths can, for the most part, be treated independently. However, as described in Chap. 6, the grooming algorithm may reroute some of these subrate demands in order to improve the grooming efficiency. When dealing with a protected subrate demand, it is important to check that link/node diversity requirements will not be violated by the new route.

Either dedicated or shared protection can be used for the subrate demands. For example, with shared protection, two subrate demands can share the same subrate protection capacity, as long as the two working paths are link/node disjoint. The shared protection capacity must enter the grooming switch at the "sharing points." The grooming switch is actively involved in the recovery mechanism, to direct the failed subrate demand to the protection capacity.

Due to the large number of subrate demands that may be part of a network design (e.g., tens of thousands of protected subrate demands may be added at one time in a network design exercise), processing each demand individually may not be a scalable option. As discussed in Chap. 6, it is preferable to first group subrate demands that have the same source and destination and required protection, and select an initial working path and protect path for the group. The grooming operations described in Chap. 6 are then implemented to improve the packing of the wavelengths. In performing these operations, individual subrate demands may be shifted to different GCs. With this strategy, the bulk of the routing and grooming is performed on groups of subrate demands, whereas fine-tuning occurs on a per-demand basis.

Because protection is at the subrate demand level, failure recovery requires more system memory and more signaling, as the system needs to track the protection path of each subrate demand and restore each one individually. One could use *protection groups* to reduce the complexity and restoration signaling overhead, where demands with the same working and protect paths are treated as a single unit for recovery [ADHN01].

Consider using subrate-level protection for the example of Fig. 7.27. Figure 7.28a illustrates the working and protect paths that are assumed for each of the subrate demands (the working paths are shown with dotted lines, and protect paths with dashed lines). Note that the working and protect paths for each demand are end-to-end link-and-node disjoint, thereby providing protection against a grooming node failure. The paths are groomed together, taking advantage of the grooming switches at Nodes A, B, and E, to form the GCs shown by the thick lines in Fig. 7.28b (other grooming arrangements are possible). The numbers indicate how many 10Gs are carried in a GC.

In addition to protecting against grooming node failures, subrate-level protection also provides finer control of the protection resources as compared to wavelength-level protection. This can be advantageous in that an unprotected demand does not end up unnecessarily protected simply because it has been groomed together with a protected demand. Furthermore, it allows the protect paths of some demands to be groomed together with the working paths of other demands. For example, in Fig. 7.28, if a new working path were added along AB, it could be groomed together with the three protect paths that are in the GC along this link. Thus, a GC created with subrate-level protection cannot generally be designated as a "working GC" or a "protect GC." However, with shared protection, note that mixing working paths and shared-protect paths in the same GC provides less sharing flexibility as the network evolves. For example, if a wavelength carries both working and shared-protect paths, that wavelength cannot be manipulated (e.g., passed through more grooming points) to provide better sharing of the protection capacity because it would disrupt a live connection. Ultimately, this may lead to lower overall capacity efficiency. Thus, in some scenarios, mixing working and shared-protect paths in the same GC is discouraged [ThSo02, OZZS03]. However, in the case of an IP network, which generally maintains a stable virtual topology, it is standard practice to utilize a particular wavelength to carry both working and protect traffic.

Further discussion on subrate-level protection can be found in Yao and Ramamurthy [YaRa05b].

7.12.3 Wavelength-Level Versus Subrate-Level Protection

The designs of Figs. 7.27 and 7.28 offer insights into the relative performance of wavelength-level and subrate-level protection. These designs can be compared on a few measures. However, direct comparisons of the two schemes are not entirely

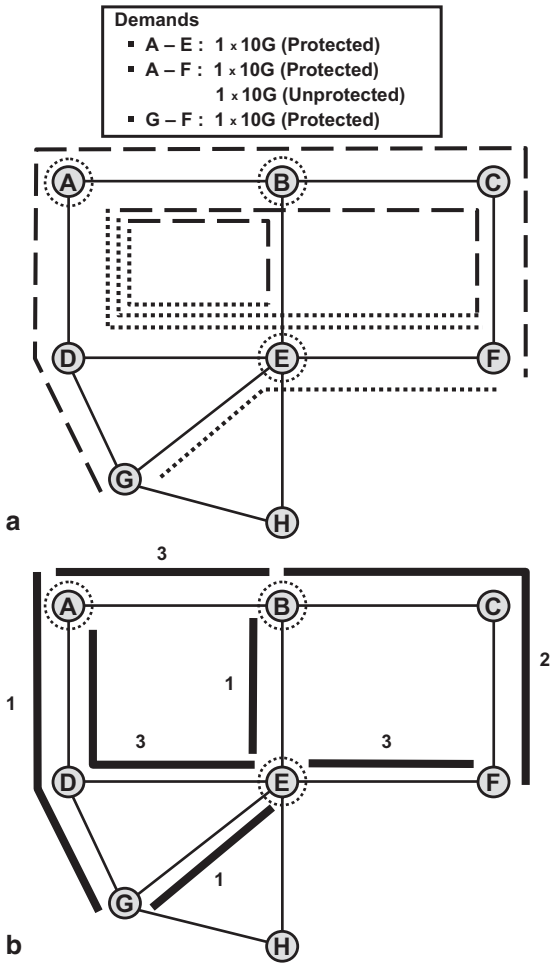


Fig. 7.28 Protection of the specified subrate demands using subrate-level protection. **a** The working paths of the subrate demands are shown by the *dotted lines* and the protect paths by the *dashed lines*. Grooming is performed at Nodes *A, B,* and *E*. **b** The *thick lines* indicate the grooming connections that result from bundling the subrate paths. The *numbers* indicate how many 10G subrate demands are carried in the grooming connections

fair, as the subrate-level protection scheme protects against grooming-node failures and the wavelength-level scheme does not. In terms of capacity, as measured by wavelength-links, the subrate-level design requires 10 wavelength-links, whereas the wavelength-level design requires 11 wavelength-links. If measured by 10G-links, the subrate-level design requires 20 10G-links, whereas the wavelength-level design requires 27 10G-links. In terms of number of GCs, the subrate-level design produces seven GCs, each one terminating in a grooming port at either endpoint. The wavelength-level design generates three protected GCs; the protected GCs utilize either one or two grooming ports at the endpoints, depending on whether protection of the grooming port is desired. Even assuming two grooming ports are used per GC endpoint, the wavelength-level design utilizes fewer grooming switch ports and transponders. These trends tend to hold in more general networks, where subrate-level protection may be more capacity efficient but wavelength-level protection requires less terminating equipment [OZZS03].

The greatest advantage of wavelength-level protection is its speed of restoration. Its greatest disadvantage is that wavelength-level protection alone cannot protect against failures in the grooming layer. Given that each network layer has its own particular strengths and weaknesses, it is natural to consider implementing protection in multiple layers. This is discussed next, in the context of an IP-over-optical architecture.

7.12.4 Multilayer Protection

Consider the example of Fig. 7.29, where an IP connection is established between Nodes A and C, via the IP router at Node B, as shown by the dotted line. From the optical-layer viewpoint, there are two separate connections: one between A and B and one between B and C. If protection is desired in both the IP and optical layers, the question is how to operate the protection across the two layers.

First, consider uncoordinated protection. In Fig. 7.29, assume that the IP connections between A and B and between B and C are established as (shared) protected connections in the optical layer. Assume that the optical layer provides protection for the two connections along A-E-C-B and B-A-E-C, respectively; this protection capacity is shown in the figure by the dashed lines. Additionally, assume that the IP layer plans protection for the AC connection by requesting an unprotected connection from the optical layer between A and C. Assume that the optical layer routes this on the path A-D-C, as shown by the dotted/dashed line.

Assume that Link AB fails. Because it is assumed that there is no coordination among the layers, both protection mechanisms are triggered, such that the optical layer attempts to move the AB connection to A-E-C-B, while at the same time the IP layer attempts to move the AC connection to its alternate path, routed over A-D-C. This protection redundancy is unnecessary. For example, if both layers support preemptible traffic (where such traffic is carried on the protect path as long as the protection resources are not needed for failure recovery), then some of this traffic may be unnecessarily bumped. Furthermore, studies have shown that simultaneous recovery operations in the optical and IP layers can lead to slow convergence in the IP layer [PDCS06].

To avoid triggering simultaneous recovery mechanisms, the protection in the layers can be coordinated, e.g., via a bottom-up escalation strategy. The optical layer protection mechanism is given the opportunity to respond first to a failure. If this layer is not successful in recovering from the failure (e.g., because the failure occurs in the IP router), then the IP layer takes action. Normally, such escalation strategies are controlled by a backoff timer, where the IP layer protection mechanism is not triggered until a certain time after the onset of the failure, to give time for the optical layer to respond. If the failure does require action at the IP layer, the backoff timer causes recovery delays. Proactive signaling schemes have also been considered to avoid the need for a backoff timer [EBRL02].

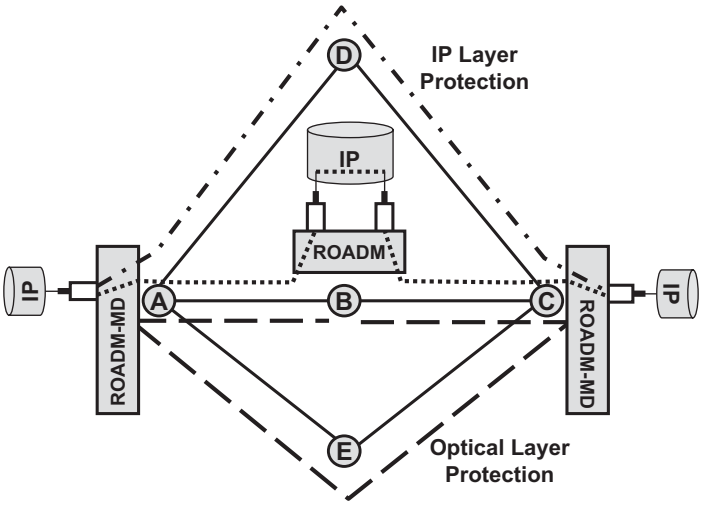


Fig. 7.29 An IP connection is established between Nodes *A* and *C,* via the IP router at Node *B.* With uncoordinated multilayer protection, the optical layer may provide protection capacity along Links *AB, BC, AE,* and *EC,* while the IP layer may protect the connection along *A-D-C*

Further coordination can be implemented between the two layers. Note that in Fig. 7.29, there are two different paths allocated to protect the same connection. If the protection planning is coordinated between the IP and optical layers, then, for example, the backup IP path for the AC connection could be routed over A-E-C. Either layer can then use the same protection capacity along these links, where the escalation strategy is relied upon to ensure both layers do not attempt to grab the capacity at the same time. While more capacity efficient, this scheme requires tighter interaction between the layers.

A scheme that allows sharing of the protection resources for IP services and wavelength services was studied in Simmons [Simm09], in the context of an IP-over-OTN-over-optical architecture. Protection of the subrate demands was handled by the OTN layer, with several nodal architectures proposed to allow the protection capacity to be shared between the OTN and optical layers. An architecture based on an edge switch provided the most versatility.

In Vadrevu et al. [VTWM12], the protection capacity that is allocated for wavelength services can be used to carry IP traffic. An important design requirement is that when a failure occurs and the affected wavelength services are moved to their backup paths, thereby preempting any IP traffic using this capacity, the IP virtual topology must remain intact.

Another alternative is to allow the IP layer to dynamically control the optical layer in order to recover from a failure. For example, if the IP router at Node B in Fig. 7.29 fails, the IP layer could direct the optical layer to tear down the AB and BC connections and create an AC connection along this same path. The communication between the layers would occur via the control plane. Studies have shown this

dynamic protection strategy to be more cost effective than a static protection scheme [PDCS06, ChAN03]. A disadvantage is that tearing down the old path and setting up the new path likely results in a longer restoration time as compared to having a pre-allocated restoration path. However, research on dynamic networking, covered in Chap. 8, indicates that connection-establishment times that are compatible with rapid dynamic protection are possible.

A different type of integrated IP/optical layer protection approach is investigated in Chiu et al. [CCFS11]. The key assumption is that nodes in the IP topology are equipped with two core IP routers, which is standard practice in most large backbone networks. Both routers are directly connected to the ROADM in the node. The failure modes considered in the study were single link failures and single IP router failures. The new twist in the proposed scheme is that the ROADM is reconfigured in response to a failure of one of the two IP routers; i.e., the ROADM rapidly reestablishes the failed IP links (i.e., links in the IP virtual topology) by shifting the endpoint of these links to the surviving router in the node. The IP ports on this router that had been used for intra-nodal communication with the failed router can now be used for internodal IP communication. (Additionally, it is assumed that best-effort traffic can be dropped under a failure condition, such that any ports that had been used for this traffic can be used to recover higher-priority traffic.) Thus, the IP layer takes advantage of rapid reconfigurability *within* the node to avoid rerouting the IP links. Furthermore, it is cost-effective, because the router ports are repurposed at the time of failure. This router-failure recovery scheme was combined with optical-layer recovery for link failures. As compared to using pure IP layer protection for either link or IP router failures, it was shown that the integrated IP/optical layer scheme provided about a 20 % cost savings in a large backbone network.

The integrated IP/optical-layer protection scheme proposed in Autenrieth et al. [ANEG12] also assumed two core routers per node (designated A and B), with corresponding diverse A and B "operating planes" (i.e., topologies). Only link failures were considered. If a link in the A plane fails, then IP-layer restoration is used to restore the high-priority traffic using the B routers. The best-effort traffic is significantly scaled back until optical-layer protection reroutes the failed IP links. As with Chiu et al. [CCFS11], this study found a significant number of IP ports could be saved as compared to using pure IP layer protection.

More general discussions of multilayer recovery can be found in Pickavet et al. [PDCS06], Lee et al. [LeLM11], Schupke [Schu12], and Gerstel et al. [GFTG14].

7.13 Fault Localization and Performance Monitoring

The chapter up to this point has only dealt with failure recovery. Another equally important component of failure management is determining the actual cause and location of the failure. This operation is known as *fault localization*. (*Fault isolation* is often used as a synonym for fault localization. However, it is sometimes used

to specifically refer to the process of isolating the portion of the network that has failed from the remainder of the network, so that traffic avoids it.) As discussed in Sect. 7.5, with fault-dependent protection schemes, the recovery mechanism depends on what has failed, thereby requiring that fault localization occur prior to failure recovery. Failure-independent protection allows fault localization to occur after the failed connections have been recovered, thus providing more time for the fault localization operation. However, in either case, the cause of the failure must ultimately be determined so that it can be fixed.

Fault localization methods are often coupled to the system architecture. In O-E-O-based networks, signals are converted to the electrical domain at each node, allowing various error checks to be performed to determine the health of the signal. This link-by-link monitoring enhances the ability of the network management system to localize the root cause of a failure. In optical-bypass-enabled networks, an optical signal may traverse several links and nodes prior to it being converted back to the electrical domain. Thus, the same type of link-by-link electronic-based error checking that is performed in O-E-O networks is typically not possible. The remainder of this section specifically focuses on fault localization in the presence of optical bypass.

Many failures (e.g., loss of light, or out-of-specification wavelength frequency) trigger a system alarm. Based on the alarms that are received, the network management system can determine what has failed. For example, decision trees can be established, where the particular combination of alarms is used to isolate the failure [MaTo03]. In principle, this appears straightforward; however, alarm correlation is somewhat more challenging in an optical-bypass-enabled network because an initial failure can cause a chain reaction of alarms in the network as the failure (e.g., the loss of light) propagates to other links. It is important for the system to be able to suppress alarms when appropriate, to avoid overwhelming the network management system. In Mas et al. [MaTT05], a methodology for selecting the network locations at which to deploy monitoring equipment is presented, with the goal of minimizing the possible failure events that could cause a given sequence of alarms.

Total link failures (e.g., fiber cuts) are generally easier to isolate in a network as compared to intermittent failures or failures that affect just a single wavelength. One method is to analyze all of the connections that have suddenly failed, and deduce on which link the failure has occurred. A more reliable method makes use of the optical supervisory channel (OSC) that is typically carried in-fiber, but out-of-band (e.g., 1,510 nm), on each link of the network. The OSC is terminated in the electrical domain on each managed device in the network to provide a communications channel for remote management, monitoring, and control. For example, it can be terminated on each nodal network element and possibly on each optical amplifier. (There is usually an out-of-fiber means of communicating with the nodal network elements as well.) The OSC is typically carried at a rate that is much lower than the data line rate so that the associated electronics are relatively low cost. The loss of the OSC channel between two endpoints, or error messages encoded on the OSC, can be used to isolate failures along the link [LiRa97].

7.13.1 Monitoring Structures

Another proposed method of link-failure localization is based on sending monitoring signals over specific paths in the network. A failed link is determined by the monitoring signals that are not successfully received. There are numerous variations of this fault-localization methodology, as summarized below.

In Wen et al. [WeCZ05], a sequence of probes is sent (one at a time) over various network paths, where the path of a probe is selected based on which of the previous probes were not received successfully. More specifically, the probe path is chosen to maximize the network state information that is likely to be provided by the probe. By making such a choice, the number of probes that need to be sent to identify a single link failure is kept to a minimum. An alternative probe-based scheme, proposed in Harvey et al. [HPWY07], relies on sending a predetermined set of probes at one time, where the combined results of the probes (i.e., whether successfully received or not) uniquely determines which link has failed. This method is faster, although it typically requires more probes.

A related scheme utilizes a predetermined set of monitoring cycles and paths, where any single link failure results in a unique combination of failed cycles and paths [AhRK09]. (Cycles start and end at the same node; paths have different endpoints.) In contrast to the probe-based methods where a probe can originate and terminate at any network node, monitoring cycles and paths can originate and terminate only at a fixed set of nodes that have been designated as monitoring locations. It is generally desirable to minimize the number of such monitoring sites to reduce cost. Strategies for selecting the monitoring locations and designing the set of cycles and paths to use are provided in Ahuja et al. [AhRK09]. These strategies are largely based on the network connectivity. For example, for networks with an average nodal degree of about 2.5 (which is realistic for a US backbone network), roughly 50–60 % of the nodes are selected as monitoring locations, and about two to three wavelengths per link, on average, are utilized for carrying the monitoring cycles and paths. For an average nodal degree of about 3.5 (which is more characteristic of European backbone networks), roughly 40 % of the nodes are selected as monitoring locations, with an average of 2.5–3 monitoring wavelengths per link. (Note that nodes of higher degree provide more pathways through the node for the monitoring paths. This makes it easier to find a unique set of monitoring paths to route on each of the incident links, without having to source/terminate a path at the node. Thus, fewer monitoring locations are generally required with denser networks.)

To reduce the amount of monitoring resources required, Stanic and Subramaniam [StSu11] utilize the status of the demand paths that have been established in the network. The demand paths are augmented by a set of monitoring paths to ensure that all relevant failures can be localized. As the number of demands in the network increases, the number of required monitoring paths decreases. Thus, as the network fills with traffic, fewer wavelengths need to be reserved for fault localization purposes.

The notion of a monitoring path was extended to a monitoring trail, where a trail can traverse a node more than once. Furthermore, bidirectional trails allow the monitoring signal to loop back at a node, onto the fiber carrying traffic in the opposite direction. Various monitoring trail schemes can be found in Haddad et al. [HaDG10], Tapolcai et al. [TWHR11], Tapolcai et al. [THRW12], and Tapolcai et al. [TaRH13]; an overview of monitoring cycles and trails is presented in Wu et al. [WHYT11].

As indicated above, this class of monitoring schemes, whether based on probes, cycles, paths, or trails, is targeted at detecting a single link failure. One can represent a successfully received monitoring signal by a "0" and a failed monitoring signal by a "1." It is required that any single failed link produce a unique pattern of 0's and 1's (i.e., a unique alarm code or "syndrome"). Thus, if the network has E links, a lower bound on the number of monitoring structures required to uniquely identity a failed link is $\log_2[E+1]$, where the "+1" term is needed because the all "0" pattern indicates that no links have failed.

7.13.2 Optical Performance Monitoring

The previous section focused on locating link failures. Finer granularity monitoring can be provided by various types of *optical performance monitors* (OPMs) [KBBE04, Will06, WiYW09, PaYW10]. For example, an OPM can tap off a small amount of the power from the WDM signal, and continually scan through each wavelength, checking signal parameters such as power level, wavelength accuracy, and OSNR. This provides at least some feedback on each individual wavelength. However, if an OPM is limited to checking the three aforementioned signal parameters, then some optical impairments, such as an increase in dispersion, will not be detected. More advanced OPMs, e.g., ones that can monitor the "Q-factor" of the signal (the Q-factor is a measure of signal quality that is strongly correlated to the bit error rate), are required to provide better fault detection capabilities [KBBE04]. From the point of view of enhancing the fault management capabilities of the network, one or more OPMs are ideally deployed on each link. However, due to the cost of OPMs, carriers may choose not to deploy them, or to deploy them only sparingly.

A different approach to monitoring the quality of a path is through *network kriging* [ChKC06, PoCR08, SPCV09]. In this methodology, the actual performance metrics of just a small sample of paths are measured. These measurements are then used to calculate the performance of the remaining paths or the expected performance of a path that is being considered for a new demand request. The theory is based on the observation that most routing matrices have an *effective rank* that is relatively small (in a routing matrix, the $[i, j]$ entry is 0 or 1 depending on whether path i is routed on link j). As presented in Chua et al. [ChKC06], the network kriging approach holds only for additive link metrics, such as dispersion and

polarization-mode dispersion (PMD). For example, by measuring the end-to-end PMD of a small set of paths, the PMD of individual links can be derived. From this, the end-to-end PMD of other paths can be calculated (assuming that the PMD is not significantly different across the wavelengths on a fiber). If the desired number of actual measurements cannot be made, then statistical analyses can be used to estimate various link performance metrics.

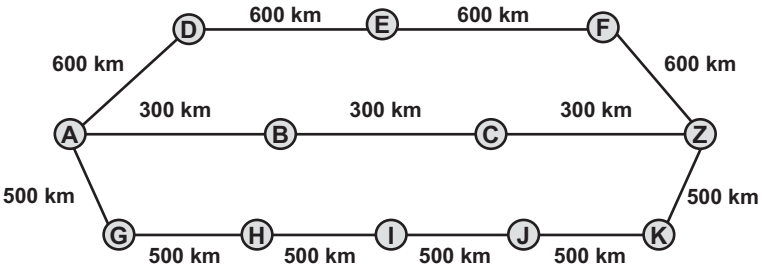
Monitoring the quality of a path potentially enables proactive protection [GLNS08, LFWT12]. As the performance of a connection begins to degrade, a backup path is established. Assuming the failure process occurs over tens of milliseconds, a cutover to the backup path can be effected before the connection completely fails, thereby providing "hitless" protection.

A topic related to fault localization is detecting malicious attacks on a network. An introduction to this topic is provided in Médard et al. [MMBF97], Médard et al. [MeCS98], and Rejeb et al. [ReLG06]. Using performance monitoring methods for this purpose is described in Willner et al. [WiYW09].

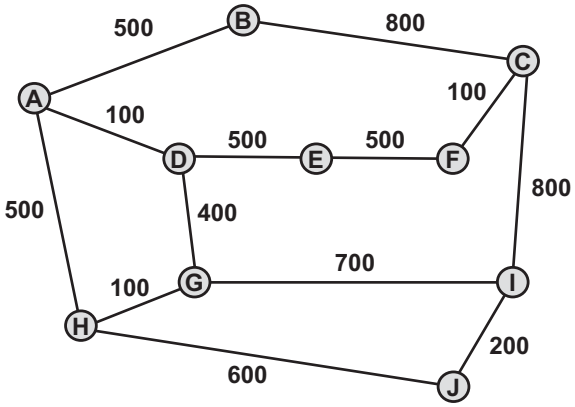
7.14 Exercises

7.1 Consider a network where three diverse paths exist between a given pair of nodes, where each of the paths has *availability R*. Consider a connection between these two nodes. (a) If 1 + 1 protection is used (where the working and backup paths are diverse), what is the availability of the connection? (b) If there is a second connection between these two nodes, and shared protection is used (the two working connections are routed on diverse paths, with a diverse backup path shared between them), what is the availability of the connection? Assume that the two working connections fail independently. (c) Compare the availability of the connection with 1 + 1 protection versus with shared protection, assuming R is 95 %.

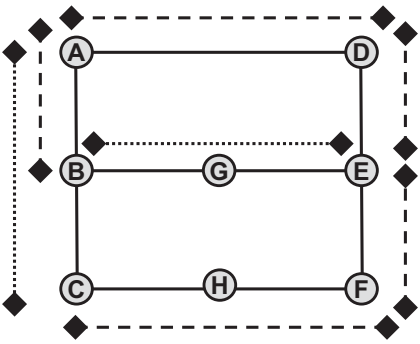
7.2 Assume that three line-rate demands between Nodes A and Z must be routed on the optical-bypass-enabled network shown below. Assume that the optical reach is 1,000 km, and assume that regeneration is implemented with back-to-back transponders. (a) If 1 + 1 protection is required for each demand, how should the working and protect paths be routed to minimize the number of required transponders? (b) If 1 + 1 protection is required for each demand, and one transponder has an equivalent cost of 200 km of bidirectional transmission, how should the working and protect paths be routed to minimize cost? (c) Repeat (a) and (b) for the scenario where either shared protection or 1 + 1 protection can be used for any of the demands. (In any of the scenarios, if multiple designs are tied for the minimum, then select the one that minimizes the total lengths of the working paths.)



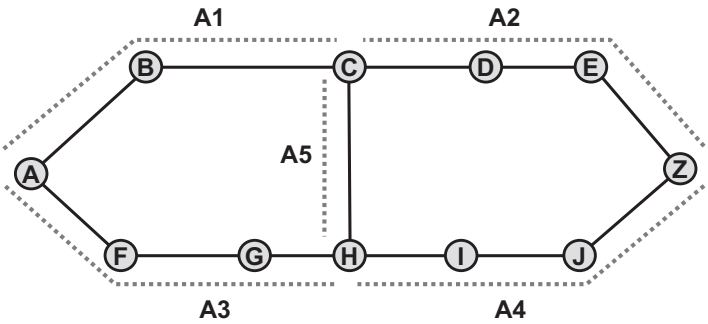
7.3 In the optical-bypass-enabled network shown below, where the links are labeled with their lengths in km, three protected demands need to be routed: AC, AH, and HJ. The demands are at the line rate. Any of these demands can be protected with either 1+1 or shared protection. Assume that any protect paths or segments have transponders at the endpoints (i.e., the transponders of a failed working path are not reused for protection). Assume that no regeneration is required. Assuming that one transponder has the equivalent cost of 200 km of bidirectional transmission, what is the minimum-cost design? (If multiple designs are tied for the minimum cost, then select the one that minimizes the total lengths of the working paths.)



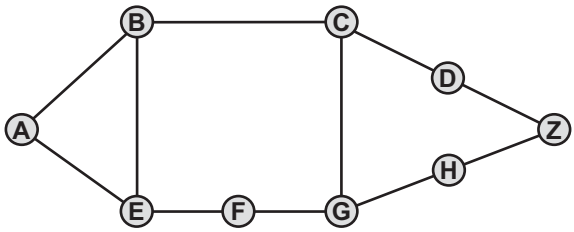
7.4 In the network below, the dotted lines indicate the two working paths (B-G-E and A-B-C) and the dashed lines indicate the protection segments (B-A, A-D, D-E, E-F, and F-H-C). Assume that the two demands are unidirectional. It is assumed that a demand requires recovery from any single link failure or any single node failure (unless the failed node is one of its endpoints). With this requirement, can the BE and AC demands share the protection capacity along Links AD and DE? Are there any issues if they do?



7.5 Consider a connection between Nodes A and Z in the network shown below. Assume that the network can be decomposed into five segments as shown, where Ai indicates the availability of the *i*th segment. Assume that the segments fail independently. (a) What is the formula for the availability of the connection if 1 + 1 protection is provided, where the two paths are A-B-C-D-E-Z and A-F-G-H-I-J-Z? (b) What is the formula for the availability of the connection if the protection is more dynamic, such that the connection can also make use of the link between Nodes C and H? (Assume that the only significant sources of downtime are the five segments; e.g., ignore switching failures at Nodes C and H.) (c) Evaluate the availability for the two protection scenarios, assuming: A1 = 0.999, A2 = 0.9975, A3 = 0.9985, A4 = 0.9965, and A5 = 0.9995. (d) How many minutes of downtime per year are expected for the connection in the two protection scenarios?



7.6 Consider the network shown below, and assume that a protected connection is desired between Nodes A and Z. (a) If 1 + 1 protection is employed, how many different double-link failures can occur where the connection *survives* (i.e., either it recovers from the failures, or one, or both, of the failures does not affect it)? (b) Answer the same question, but assume that a dynamic protection scheme is used, where the protection path can be configured after a failure occurs.



7.7 Consider transponder protection, such as the architectures shown in Fig. 7.4. Assume that transponders fail independently, and that transponders are the only component that can fail. Which of the following two transponder protection schemes do you think results in higher availability for a connection: *M* spares to protect *N* transponders, or *2M* spares to protect *2N* transponders? Why?

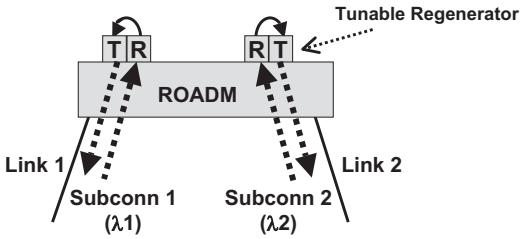
7.8 Figure 7.7 illustrates Node A from Fig. 7.6, where shared ring protection is employed. Assume that, due to lasing issues, the protection ring is regenerated at Node A. With this assumption, draw the configuration for Node A (a) when there are no failures; and (b) when Link AB has failed, and the protection ring is used to restore the demand between Nodes A and B. (c) Are there potentially contention issues on the add/drop ports of the ROADM at Node A?

7.9 Assume that a single-ended protection protocol is used to implement shared protection where the demand source initiates recovery by sending a control message over the protect path, requesting that the necessary cross-connections of protection capacity be performed at certain nodes along the path. Assume that the control message is not forwarded by a node until its own cross-connection is completed (the nodes that do not need to perform a cross-connection forward the control message immediately). Additionally, assume that before the source initiates transmission on the protect path, it waits for a verification message from the destination indicating that the protect path has been established. How much time expires between the onset of recovery and the initial transmission on the protect path? Assume that the one-way end-to-end fiber propagation delay is *D*; there are *C* cross-connects required on the protect path; and the cross-connect switching time is *S*. (Ignore any message processing time; also ignore any switching time that may be required at the demand endpoints.)

7.10 In contrast to the assumptions of Exercise 7.9, assume that pipelining is used, where a node on the protect path immediately forwards the control message without waiting for its cross-connection to be completed. Also, assume that the source begins to transmit on the protect path as soon as possible, without waiting for an acknowledgment that the protect path has been established (i.e., the source waits just long enough that any cross-connects will have been established by the time its transmission reaches the node). How much time expires between the onset of recovery and the initial transmission on

the protect path? Does this time depend on how many cross-connections are required? How do the results here compare to the results of Exercise 7.9?

7.11 Consider a pre-deployed-subconnection protection scheme, where the ROADM is used to concatenate the subconnections at the time of failure. The figure below shows two protection subconnections prior to failure. Assume that they both terminate in tunable regenerator cards. Assume that a failure occurs that requires the two subconnections to be concatenated for recovery. After the concatenation, λ1 must still be used in both directions on Link 1 and λ2 must still be used in both directions on Link 2. (a) Draw two possible configurations that can be used to concatenate the two subconnections. (b) From the perspective of minimizing issues with optical amplifier transients, which of the two configurations is preferred? (c) In either option, does the ROADM need to be directionless? Colorless? What are the implications with respect to contention? (Assume that the process must be automated, with no manual intervention.)



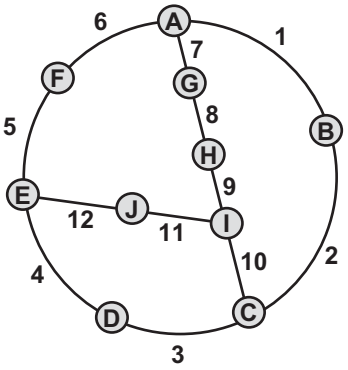
7.12 Consider the network that was shown in Fig. 7.21. Assume that the network is optical-bypass enabled with no required regeneration, links are equipped with a single fiber pair, and the two connections shown are bidirectional (e.g., B to D and D to B). Assume that the p-cycle shown is used for *link-based* protection (not all link failures are covered by this one p-cycle). (a) Assign wavelengths to the two connections and to the p-cycle such that no wavelength conversion is required for failure recovery. (b) Does this wavelength assignment still work if the connection along B-F-G is replaced by a connection along B-F-G-D-C (and the wavelength that was assigned to B-F-G is now assigned to B-F-G-D-C)? (c) What limitations are placed on the p-cycle link-based approach if the cycle has a single chordal link as in Fig. 7.21 and wavelength conversion is not permitted?

7.13 Consider the PXT configuration that was shown in Fig. 7.22. Assume that the PXT is pre-lit at Nodes A and D, and that all intermediate ROADMs along the PXT path are initially configured to allow the PXT wavelength to pass through. Additionally, assume that the connection endpoints reuse the working transponder when switching to the protect path, and that the wavelength

used for the working path is *not* the same as the wavelength used for the PXT. Assume that the ROADMs are directionless and support drop-and-continue. (a) Draw two nodal diagrams of Node I: first, under the no-failure condition, and second, after the connection between Nodes A and I fails, and the PXT is used to recover from the failure. (b) What operations need to be performed at the time of failure and what are the ramifications with respect to optical amplifier transients?

7.14 Assume that catastrophes can be modeled as affecting a disk-shaped area of radius R. Assume that links can be drawn as straight lines. Draw the region of vulnerability for an arbitrary link, i.e., the region in which a catastrophe must be centered in order to affect the link.

7.15 The network below has 12 links, as numbered; assume that no more than one link fails at a time. (a) What is a lower bound on the number of monitoring cycles or paths that are needed to uniquely identify which link has failed? (b) Design a monitoring scheme that meets this bound. Assume that any node can serve as a monitoring location.



7.16 Consider a ring with N nodes, where $N \geq 5$. Prove that exactly CEILING[N/2] monitoring trails are needed to uniquely identify a single link failure on the ring.

7.17 *Research Suggestion:* Monitoring schemes with cycles, paths, and/or trails have been studied, where these structures are all linear. Consider making use of the multicast feature of some ROADMs, such that the monitoring structure has branch points (multiple destinations would report whether the probe sent by one source was received properly). For example, does this allow fewer wavelengths to be dedicated to monitoring, or fewer monitoring locations to be used?

References

[ACLY00] R. Ahlswede, N. Cai, S.-Y. R. Li, R. W. Yeung, Network information flow. IEEE. Trans. Inf. Theory. **46**(4), 1204–1216 (July 2000)

[ADHN01] G. P. Austin, B. T. Doshi, C. J. Hunt, R. Nagarajan, M. A. Qureshi, Fast, scalable, and distributed restoration in general mesh optical networks. Bell. Labs. Tech. J. **6**, 67–81 (Jan-June 2001)

[AEGH10] P. K. Agarwal, A. Efrat, S. K. Ganjugunte, D. Hay, S. Sankararaman, G. Zussman, Network vulnerability to single, multiple, and probabilistic physical attacks. *Proceedings, IEEE Military Communications Conference (MILCOM 2010*), San Jose, CA, 31 Oct–3 Nov 2010, pp. 1824–1829

[AhRK09] S. S. Ahuja, S. Ramasubramanian, M. M. Krunz, Single-link failure detection in all-optical networks using monitoring cycles and paths. IEEE/ACM. Trans. Netw. **17**(4), 1080–1093 (August 2009)

[AlAy99] M. Alanyali, E. Ayanoglu, Provisioning algorithms for WDM optical networks. IEEE/ACM. Trans. Netw. **7**(5), 767–778 (October 1999)

[ANEG12] A. Autenrieth, M. Neugirg, J.-P. Elbers, M. Gunkel, Evaluation of IP-over-DWDM core network architectures with CD-ROADMs using IP protection in combination with optical restoration. *Proceedings, International Conference on Transparent Optical Networks (ICTON'12)*, United Kingdom, 2–5 July 2012 (Paper Tu.A2.1)

[AYDA03] C. Assi, Y. Ye, S. Dixit, M. Ali, Control and management protocols for survivable optical mesh networks. J. Lightwave Tech. **21**(11), 2638–2651 (November 2003)

[BELR07] E. Bouillet, G. Ellinas, J.-F. Labourdette, R. Ramamurthy, in *Path Routing in Mesh Optical Networks*, (John Wiley & Sons Ltd, West Sussex, England, 2007)

[BLRC02] E. Bouillet, J.-F. Labourdette, R. Ramamurthy, S. Chaudhuri, Enhanced algorithm cost model to control tradeoffs in provisioning shared mesh restored lightpaths. *Proceedings, Optical Fiber Communication (OFC'02)*, Anaheim, CA, 17–22 March 2002 (Paper ThW2)

[CCCD12] A. L. Chiu, G. Choudhury, G. Clapp, R. Doverspike, M. Feuer, J. W. Gannett, J. Jackel, G. T. Kim, J. G. Klincewicz, T. J. Kwon, G. Li, P. Magill, J. M. Simmons, R. A. Skoog, J. Strand, A. Von Lehmen, B. J. Wilson, S. L. Woodward, D. Xu, Architectures and protocols for capacity efficient, highly dynamic and highly resilient core networks. J. Opt. Commun. and Netw. **4**(1), 1–14 (January 2012)

[CCFS11] A. L. Chiu, G. Choudhury, M. D. Feuer, J. L. Strand, S. L. Woodward, Integrated restoration for next-generation IP-over-optical networks. J. Lightwave Tech. **29**(6), 916–924, (15 March 2011)

[CDLS09] A. Chiu, R. Doverspike, G. Li, J. Strand, Restoration signaling protocol design for next-generation optical network. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'09)*, San Diego, CA, 22–26 March 2009 (Paper NTuC2)

[ChAN03] C. Chigan, G. W. Atkinson, R. Nagarajan, Cost effectiveness of joint multilayer protection in packet-over-optical networks. J. Lightwave Tech. **21**(11), 2694–2704, (November 2003)

[ChCF04] T. Y. Chow, F. Chudak, A. M. Ffrench, Fast optical layer mesh protection using pre-cross-connected trails. IEEE/ACM. Trans. Netw. **12**(3), 539–548, (June 2004)

[ChKC06] D. B. Chua, E. D. Kolaczyk, M. Crovella, Network kriging. IEEE. J. Sel Areas in Commun. **24**(12), 2263–2272, (December 2006)

[ClGr02] M. Clouqueur, W. D. Grover, Mesh-restorable networks with complete dual failure restorability and with selectively enhanced dual-failure restorability properties. *Proceedings, SPIE OptiComm 2002*: *Optical Networking and Communications*, vol. 4874, Boston, MA, 29 July–2 Aug 2002, pp. 1–12

[Conw11] A. E. Conway, Fast simulation of service availability in mesh networks with dynamic path restoration. IEEE/ACM. Trans. Netw. **19**(1), 92–101 (February 2011)

[DDHH99] B. T. Doshi, S. Dravida, P. Harshavardhana, O. Hauser, Y. Wang, Optical network design and restoration. Bell Labs Tech. J. **4**(1), 58–84 (January–March 1999)

[DSST99] R. Doverspike, G. Sahin, J. Strand, R. Tkach, Fast restoration in a mesh network of optical cross-connects. *Proceedings, Optical Fiber Communication (OFC'99)*, vol. 1, San Diego, CA, 21–26 Feb 1999, pp. 170–172

[DXMK12] O. Diaz, F. Xu, N. Min-Allah, M. Khodeir, M. Peng, S. Khan, N. Ghani, Network survivability for multiple probabilistic failures. IEEE. Commun. Lett. **16**(8), 1320–1323 (August 2012)

[EBRL02] G. Ellinas, E. Bouillet, R. Ramamurthy, J.-F. Labourdette, S. Chaudhuri, K. Bala, Restoration in layered architectures with a WDM mesh optical layer. Int. Eng. Consort. Annu Rev Commun. **55** (June 2002)

[EBRL03] G. Ellinas, E. Bouillet, R. Ramamurthy, J.-F. Labourdette, S. Chaudhuri, K. Bala, Routing and restoration architectures in mesh optical networks. Opt. Netw. Mag. **4**(1), 91–106 (January/February 2003)

[EiLS11] M. I. Eiger, H. Luss, D. F. Shallcross, Network restoration under a single link or node failure using preconfigured virtual cycles. Telecommun. Syst. **46**(1), 17–30 (January 2011)

[Feue05] R. Feuerstein, Interconnecting the cyberinfrastructure. *Cyberinfrastructure 2005*, Lincoln, NE, 15–16 Aug 2005

[GeRa00] O. Gerstel, R. Ramaswami, Optical layer survivability—an implementation perspective. IEEE. J. Sel. Areas Commun. **18**(10), 1885–1899 (October 2000)

[GFTG14] O. Gerstel, C. Filsfils, T. Telkamp, M. Gunkel, M. Horneffer, V. Lopez, A. Mayoral, Multi-layer capacity planning for IP-optical networks. IEEE. Commun. Mag. **52**(1), 44–51 (January 2014)

[GGCS07] A. Grue1, W. D. Grover, M. Clouqueur, D. A. Schupke, J. Doucette, B. Forst, D. Onguetou, D. Baloukov, Comparative study of fully pre-cross-connected protection architectures for transparent optical networks. *Proceedings, 6th International Workshop on Design of Reliable Communication Networks (DRCN'07)*, La Rochelle, France, 7–10 Oct 2007

[GLNS08] O. Gerstel, I. Leung, G. Nicholl, H. Sohel, W. Wakim, K. Wollenweber, Near-hitless protection in IPoDWDM networks. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'08)*, San Diego, CA, 24–28 Feb 2008 (Paper NWD4)

[Grov03] W. Grover, in *Mesh-based Survivable Transport Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking* (Prentice-Hall, Upper Saddle River, 2003)

[GrSt98] W. Grover, D. Stamatelakis, Cycle-oriented distributed preconfiguration: Ring-like speed with mesh-like capacity for self-planning network restoration. *Proceedings, IEEE International Conference on Communications (ICC'98)*, vol. 1, Atlanta, GA, 7–11 June 1998, pp. 537–543

[GuPM03] K. P. Gummadi, M. J. Pradeep, C. S. R. Murthy, An efficient primary-segmented backup scheme for dependable real-time communication in multihop networks. IEEE/ACM. Trans. Netw. **11**(1), 81–94 (February 2003)

[HaDG10] A. Haddad, E. A. Doumith, M. Gagnaire, A meta-heuristic approach for monitoring trail assignment in WDM optical networks. *Proceedings, International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT'10)*, Moscow, Russia, 18–20 Oct 2010, pp. 601–607

[HoMo02] P. H. Ho, H. T. Mouftah, A framework for service-guaranteed shared protection in WDM mesh networks. IEEE. Commun. Mag. **40**(2), 97–103 (February 2002)

[HPWY07] N. J. A. Harvey, M. Patrascu, Y. Wen, S. Yekhanin, V. W. S. Chan, Non-adaptive fault diagnosis for all-optical networks via combinatorial group testing on graphs. *Proceedings, IEEE INFOCOM 2007*, Anchorage, AK, 6–12 May 2007, pp. 697–705

[HTDM13] M. F. Habib, M. Tornatore, F. Dikbiyik, B. Mukherjee, Disaster survivability in optical communication networks. Comput Commun. **36**(6), 630–644 (15 March 2013)

[IrMG98] R. R. Iraschko, M. H. MacGregor, W. D. Grover, Optimal capacity placement for path restoration in STM or ATM mesh-survivable networks. IEEE/ACM. Trans. Netw. **6**(3), 325–336 (June 1998)

[KaAr04] E. Karasan, M. Arisoylu, Design of translucent optical networks: Partitioning and restoration. Photonic Netw. Commun. **8**(2), 209–221 (March 2004)

[Kama08] A. E. Kamal, A generalized strategy for 1 + N protection, *IEEE International Conference on Communications (ICC'08)*, Beijing, China, 19–23 May 2008, pp. 5155–5159

[KBBE04] D. C. Kilper, R. Bach, D. J. Blumenthal, D. Einstein, T. Landolsi, L. Ostar, M. Preiss, A. E. Willner, Optical performance monitoring. J Lightwave Tech. **22**(1), 294–304 (January 2004)

[KiAJ09] M. S. Kiaei, C. Assi, B. Jaumard, A survey on the p-cycle protection method. IEEE. Commun. Surv. & Tutor. **11**(3), 53–70 (Third Quarter 2009)

[KiLu03] S. Kim, S. Lumetta, Evaluation of protection reconfiguration for multiple failures in WDM mesh networks. *Proceedings, Optical Fiber Communication (OFC'03)*, Atlanta, GA, 23–28 Mar 2003 (Paper TuI7)

[KiMO09] M. Kim, M. Médard, U.-M. O'Reilly, Network coding and its implications on optical networking. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'09)*, San Diego, CA, 22–26 Mar 2009 (Paper OThO3)

[KNSZ04] P. M. Krummrich, R. E. Neuhauser, H.-J. Schmidtke, H. Zech, M. Birk, Compensation of Raman transients in optical networks. *Proceedings, Optical Fiber Communication (OFC'04)*, Los Angeles, CA, 22–27 Feb 2004 (Paper MF82)

[KoGr05] A. Kodian, W. D. Grover, Failure-independent path-protecting p-cycles: Efficient and simple fully preconnected optical-path protection. J. Lightwave Tech. **23**(10), 3241–3259 (October 2005)

[KoLa00] M. Kodialam, T. V. Lakshman, Dynamic routing of bandwidth guaranteed tunnels with restoration. *Proceedings, IEEE INFOCOM 2000*, vol. 2, Tel-Aviv, Israel, 26–30 Mar 2000, pp. 902–911

[KoMe03] R. Koetter, M. Médard, An algebraic approach to network coding. IEEE/ACM. Trans. Netw. **11**(5), 782–795 (October 2003)

[KSCA11] M. S. Kiaei, S. Sebbah, A. Cerny, H. Alazemi, C. Assi, Efficient network protection design models using pre-cross-connected trails. IEEE. Trans. Commun. **59**(11), 3102–3110 (November 2011)

[LeML10] H.-W. Lee, E. Modiano, K. Lee, Diverse routing in networks with probabilistic failures. IEEE/ACM. Trans. Netw. **18**(6), 1895–1907 (December 2010)

[LeLM11] K. Lee, H.-W. Lee, E. Modiano, Reliability in layered networks with random link failures. IEEE/ACM. Trans. Netw. **19**(6), 1835–1848 (December 2011)

[LFWT12] C. P. Lai, F. Fidler, P. J. Winzer, M. K. Thottan, K. Bergman, Cross-layer proactive packet protection switching. J. Optic. Commun. & Netw. 4(10), 847–857 (October 2012)

[LiCS05] G. Li, A. L. Chiu, J. Strand, Failure recovery in all-optical ULH networks. *Proceedings, 5th International Workshop on Design of Reliable Communication Networks (DRCN'05)*, Island of Ischia, Italy, 16–19 Oct 2005

[LiRa97] C.-S. Li, R. Ramaswami, Automatic fault detection, isolation, and recovery in transparent all-optical networks. J. Lightwave Tech. **15**(10), 1784–1793 (October 1997)

[LLLC12] Z. Liu, M. Li, L. Lu, C.-K. Chan, S.-C. Liew, L.-K. Chen, Optical physical-layer network coding. IEEE. Photonics Tech. Lett. **24**(16), 1424–1427 (15 August 2012)

[LSTN05] M.-J. Li, M. J. Soulliere, D. J. Tebben, L. Nederlof, M. D. Vaughn, R. E. Wagner, Transparent optical protection ring architectures and applications. J. Lightwave Tech. **23**(10), 3388–3403 (October 2005)

[MaLe03] B. Manseur, J. Leung, Comparative analysis of network reliability and optical reach. *National Fiber Optic Engineers Conference (NFOEC'03)*, Orlando, FL, 7–11 Sept 2003

[MaTo03] C. M. Machuca, I. Tomkos, Failure detection for secure optical networks. *Proceedings, International Conference on Transparent Optical Networks (ICTON'03)*, Warsaw, Poland, 29 June—3 July 2003, pp. 70–75

[MaTT05] C. Mas, I. Tomkos, O. K. Tonguz, Failure location algorithm for transparent optical networks. IEEE. J. Sel. Areas Commun. **23**(8), 1508–1519, (August 2005)

[MDXA10] E. D. Manley, J. S. Deogun, L. Xu, D. R. Alexander, All-optical network coding. J. Optic. Commun. & Netw. **2**(4), 175–191 (April 2010)

[MeCS98] M. Médard, S. R. Chinn, P. Saengudomlert, Attack detection in all-optical networks. *Proceedings, Optical Fiber Communication (OFC'98)*, San Jose, CA, 22–27 Feb 1998 (Paper ThD4)

[MeGa08] R. C. Menendez, J. W. Gannett, Efficient, fault-tolerant all-optical multicast networks via network coding. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'08)*, San Diego, CA, 24–28 Feb 2008 (Paper JThA82)

[MMBF97] M. Médard, D. Marquis, R. A. Barry, S. G. Finn, Security issues in all-optical networks. IEEE. Netw. **11**(3), 42–48 (May/June 1997)

[MoLS02] R. Monnard, H. K. Lee, A. Srivastava, Suppressing amplifier transients in lightwave systems. *Proceedings, IEEE/LEOS Summer Topicals*, Mont Tremblant, Quebec, 15–17 July 2002 (Paper WE3)

[NZCM11] S. Neumayer, G. Zussman, R. Cohen, E. Modiano, Assessing the vulnerability of the fiber infrastructure to disasters. IEEE/ACM. Trans. Netw. **19**(6), 1610–1623 (December 2011)

[OuMu05] C. Ou, B. Mukherjee, in *Survivable Optical WDM Networks* (Springer, New York, NY, 2005)

[OZSZ04] C. Ou, H. Zang, N. K. Singhal, K. Zhu, L. H. Sahasrabuddhe, R. A. MacDonald, B. Mukherjee, Subpath protection for scalability and fast recovery in optical WDM mesh networks. IEEE. J. Sel. Areas Commun. **22**(9), 1859–1875 (November 2004)

[OZZS03] C. Ou, K. Zhu, H. Zang, L. H. Sahasrabuddhe, B. Mukherjee, Traffic grooming for survivable WDM networks—shared protection. IEEE. J. Sel. Areas Commun. **21**(9), 1367–1383 (November 2003)

[PaYW10] Z. Pan, C. Yu, A. E. Willner, Optical performance monitoring for the next generation optical communication networks. Optic Fiber Tech. **16**(1), 20–45 (January 2010)

[PDCS06] M. Pickavet, P. Demeester, D. Colle, D. Staessens, B. Puype, L. Depré, I. Lievens, Recovery in multilayer optical networks. J. Lightwave Tech. **24**(1), 122–134 (January 2006)

[PoCR08] Y. Pointurier, M. Coates, M. Rabbat, Active monitoring of all-optical networks. *Proceedings, International Conference on Transparent Optical Networks (ICTON'08)*, Athens, Greece, 22–26 June 2008

[QiXu02] C. Qiao, D. Xu, Distributed partial information management (DPIM) schemes for survivable networks—Part I. *Proceedings, IEEE INFOCOM 2002*, vol. 1, New York, NY, 23–27 June 2002, pp. 302–311

[ReLG06] R. Rejeb, M. S. Leeson, R. J. Green, Fault and attack management in all-optical networks. IEEE. Commun. Mag. **44**(11), 79–86 (November 2006)

[RPAG11] M. Rahnamay-Naeini, J. E. Pezoa, G. Azar, N. Ghani, M. M. Hayat, Modeling stochastic correlated failures and their effects on network reliability. *Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN 2011)*, Maui, HI, 31 July—4 Aug 2011

[ScAF01] D. A. Schupke, A. Autenrieth, T. Fischer, Survivability of multiple fiber duct failures. *Proceedings, Third International Workshop on the Design of Reliable Communication Networks (DRCN'01)*, Budapest, Hungary, 7–10 October 2001, pp. 213–219

[ScGC04] D. A. Schupke, W. D. Grover, M. Clouqueur, Strategies for enhanced dual failure restorability with static or reconfigurable p-cycle networks. *Proceedings, IEEE International Conference on Communications (ICC'04)*, Paris, France, 20–24 June 2004, pp. 1628–1633

[Schu12] D. A. Schupke, Multilayer and multidomain resilience in optical networks. Proc. IEEE. **100**(5), 1140–1148 (May 2012)

[SeJa12] S. Sebbah, B. Jaumard, Differentiated quality-of-recovery in survivable optical mesh networks using *p*–structures. IEEE/ACM. Trans. Netw. **20**(3), 798–810 (June 2012)

[SHCJ10] J. P. G. Sterbenz, D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, P. Smith, Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines. Comp. Netw. **54**(8), 1245–1265 (1 June 2010)

[ShGr04] G. Shen, W. D. Grover, Segment-based approaches to survivable translucent network design under various ultra-long-haul system reach capabilities. J. Optic. Netw. **3**(1), 1–24 (January 2004)

[Simm99] J. M. Simmons, Hierarchical restoration in a backbone network. *Proceedings, Optical Fiber Communication (OFC'99)*, San Diego, CA, 21–26 February 1999 (Paper TuL2)

[Simm07] J. M. Simmons, Cost vs. capacity tradeoff with shared mesh protection in optical-bypass-enabled backbone networks. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'07)*, Anaheim, CA, 25–29 March 2007 (Paper NThC2)

[Simm09] J. M. Simmons, Nodal architectures for shared mesh restoration of IP and wavelength services. IEEE. Photonics Tech. Lett. **21** (22), 1677–1679 (15 November 2009)

[Simm12] J. M. Simmons, Catastrophic failures in a backbone network. IEEE. Commun. Letts. **16**(8), 1328–1331 (August 2012)

[SiSB01] J. M. Simmons, A. A. M. Saleh, L. Benmohamed, Extending Generalized Multi-Protocol Label Switching to configurable all-optical networks. *Proceedings, National Fiber Optic Engineers Conference (NFOEC'01)*, Baltimore, MD, 8–12 July 2001, pp. 14–23

[SkCW10] N. Skorin-Kapov, J. Chen, L. Wosinska, A new approach to optical networks security: Attack-aware routing and wavelength assignment. IEEE/ACM. Trans. Netw. **18**(3), 750–760 (June 2010)

[SPCV09] N. Sambo, Y. Pointurier, F. Cugini, L. Valcarenghi, P. Castoldi, I. Tomkos, Lightpath establishment in distributed transparent dynamic optical networks using network kriging. *Proceedings, European Conference on Optical Communication (ECOC'09)*, Vienna, Austria, 20–24 Sept 2009 (Paper 1.5.3)

[SrSS02] M. Sridharan, R. Srinivasan, A. K. Somani, Dynamic routing with partial information in mesh-restorable optical networks. *Proceedings, Sixth Working Conference on Optical Networks Design and Modelling (ONDM'02)*, Torino, Italy, 4–6 Feb 2002

[StSu11] S. Stanic, S. Subramaniam, Fault localization in all-optical networks with user and supervisory lightpaths. *IEEE International Conference on Communications (ICC'11)*, Kyoto, Japan, 5–9 June 2011

[TaRH13] J. Tapolcai, L. Rónyai, P.-H. Ho, Link fault localization using bi-directional m-trails in all-optical mesh networks. IEEE. Trans. Commun. **61**(1), 291–300 (January 2013)

[THRW12] J. Tapolcai, P.-H. Ho, L. Rónyai, B. Wu, Network-wide local unambiguous failure localization (NWL-UFL) via monitoring trails. IEEE/ACM. Trans. Netw. **20**(6), 1762–1773 (December 2012)

[ThSo02] S. Thiagarajan, A. K. Somani, Traffic grooming for survivable WDM mesh networks. Optic. Netw. Mag. **3**(3), 88–98 (May/June 2002)

[TWHR11] J. Tapolcai, B. Wu, P.-H. Ho, L. Rónyai, A novel approach for failure localization in all-optical mesh networks. IEEE/ACM. Trans. Netw. **19**(1), 275–285 (February 2011)

[VaPD04] J. Vasseur, M. Pickavet, P. Demeester, in *Network Recovery: Protection and Restoration of Optical, SONET-SDH, IP, and MPLS*, (Morgan Kaufmann, San Francisco, CA, 2004)

[VTWM12] C. S. K. Vadrevu, M. Tornatore, R. Wang, B. Mukherjee, Integrated design for backup capacity sharing between IP and wavelength services in IP-over-WDM networks. J. Optic. Commun. & Netw. **4**(1), 53–65 (January 2012)

[WaQY11] J. Wang, C. Qiao, H. Yu, On progressive network recovery after a major disruption. *Proceedings, IEEE INFOCOM 2011*, Shanghai, China, 10–15 April 2011, pp. 1925–1933

[WeCZ05] Y. Wen, V. W. S. Chan, L. Zheng, Efficient fault-diagnosis algorithms for all-optical WDM networks with probabilistic link failures. J. Lightwave Tech. **23**(10), pp. 3358–3371 (October 2005)

[WHYT11] B. Wu, P.-H. Ho, K. L. Yeung, J. Tapolcai, H. T. Mouftah, Optical layer monitoring schemes for fast link failure localization in all-optical networks. IEEE. Commun. Surv. & Tutor. **13**(1), 114–125 (First Quarter 2011)

[Will06] A. E. Willner, The optical network of the future: Can optical performance monitoring enable automated, intelligent and robust systems? *Optics and Photonics News*, pp. 30–35, (March 2006)

[WiYW09] A. E. Willner, J. Y. Yang, X. Wu, Optical performance monitoring to enable robust and reconfigurable optical high-capacity networks. *Proceedings, IEEE Military Communications Conference (MILCOM 2009)*, Boston, MA, 18–21 Oct 2009

[WLYK04] D. Wang, G. Li, J. Yates, C. Kalmanek, Efficient segment-by-segment restoration. *Proceedings, Optical Fiber Communication (OFC'04)*, Los Angeles, CA, 22–27 Feb 2004 (Paper TuP2)

[XuQX07] D. Xu, C. Qiao, Y. Xiong, Ultrafast potential-backup-cost (PBC)-based shared path protection schemes. J. Lightwave Tech. **25**(8), 2251–2259 (August 2007)

[XuXQ03] D. Xu, Y. Xiong, C. Qiao, Novel algorithms for shared segment protection. IEEE. J. Sel. Areas Commun. **21**(8), 1320–1331 (October 2003)

[YaRa05b] W. Yao, B. Ramamurthy, Survivable traffic grooming with path protection at the connection level in WDM mesh networks. J. Lightwave Tech. **23**(10), 2846–2853 (October 2005)

[ZhFB07] X. Zhou, M. Feuer, M. Birk, Fast control of inter-channel SRS and residual EDFA transients using a multiple-wavelength forward-pumped discrete Raman amplifier. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'07)*, Anaheim, CA, 25–29 March 2007 (Paper OMN4)

[ZhZM06] J. Zhang, K. Zhu, B. Mukherjee, Backup reprovisioning to remedy the effect of multiple link failures in WDM mesh networks. IEEE. J. Sel. Areas Commun. **24**(8), 57–67 (August 2006)

Chapter 8
Dynamic Optical Networking

8.1 Introduction

Transport optical networks today are typically quasi-static, with connections often remaining established for months or years. The process of provisioning a new wavelength has historically been slow, requiring much up-front planning and manual intervention at multiple sites in the network. In addition to the time and cost involved with any network modification, the manual nature of the process leaves it vulnerable to operator error. The situation has improved with optical-bypass-enabled networks, where the amount of equipment required to support a new connection is significantly reduced; however, traffic provisioning is still heavily reliant on manual intervention. While this mode of operation may have sufficed for relatively static and predictable traffic patterns (i.e., when voice was the dominant traffic type), as services have gravitated towards more variable data connections, the need for greater optical-layer agility has grown as well.

As an initial transition from this relatively fixed environment, transport optical networks are becoming *configurable*. In the configurable model, operations personnel generally initiate the provisioning process, e.g., through the use of a planning tool. The connections are established remotely through software control, assuming that the required equipment is already deployed in the network. This eliminates the time and cost involved with sending personnel to sites along the new path. Configurable networks take advantage of network elements such as reconfigurable optical add/drop multiplexers (ROADMs), which can be remotely reconfigured to add, drop, or bypass any wavelength without affecting existing network traffic, and tunable transponders, which can be tuned to any of the wavelengths supported on a fiber.

The next step in this evolution is *dynamic* networking, where connections (i.e., circuits) can be rapidly established and torn down without the involvement of operations personnel. In the dynamic model, not only is the provisioning process automated, but it is also completely under software control. The higher layers of the network automatically request bandwidth from the optical layer, which is then reconfigured accordingly. Connections may be provisioned and brought down in seconds, or possibly sub-seconds.

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4_8,

It is envisioned that the need for dynamic services will burgeon over the next 5–10 years. This growth will likely be a push–pull evolution, where the need for on-demand services by applications such as cloud computing drives the implementation of dynamic networks, and a dynamic network infrastructure fuels the development of more services that can take advantage of a rapidly responsive network.

Section 8.2 examines the motivation for dynamic optical networking from both the carrier and customer perspectives. This section considers the cost and capacity benefits, as well as additional revenue opportunities, that are engendered by a dynamic optical layer. It enumerates an array of applications that become realizable, depending on the achievable connection setup time.

The remainder of the chapter presents the implementation details regarding dynamic optical networking. To fully appreciate some of the design decisions, it is helpful to have an understanding of how various network tasks are apportioned. As noted in Sect. 1.5, networks are generally composed of a data plane, a management plane, and a control plane. The data plane is directly responsible for the forwarding of packets and bit streams, whereas the management and control planes are responsible for network operations. Historically, optical network configuration has been performed via a centralized network management system (NMS). However, with dynamic optical networking, this function is accomplished via the control plane, which is typically more distributed and more autonomous in nature. For example, a request for a new wavelength potentially can be received from a higher layer at any of the network nodes, e.g., through a user network interface (UNI; see Sect. 1.5). The optical-layer control plane responds by performing any required computations, assigning resources, and issuing commands to the various network elements to actually provision the new connection.

One of the most important system design dichotomies involves where the control-plane "machinery" (i.e., the processing power, memory, algorithms, etc.) resides to accomplish configuration management. It can be centralized, either at a designated network node or at an adjunct location, or it can be distributed among the network nodes. This has major ramifications for network optimality, latency, and processing, memory, and signaling requirements. The centralized model is covered in Sect. 8.3, and the distributed model is covered in Sect. 8.4. Section 8.5 looks at solutions that combine aspects of both models, where the information required for a particular aspect of connection setup, and the rate of change of that information, largely drives the design choices.

We then consider three particular aspects of dynamic networking in more detail. Section 8.6 focuses on protection, with various options for setting up diverse paths for a connection. Section 8.7 examines how physical-layer impairments and regeneration can be handled in a dynamic optical network. In the past, this aspect of dynamic networking was largely glossed over, with the simplifying assumption that optical-bypass-enabled networks were truly all-optical, with no regeneration. Clearly, this is not the case in many networks, especially those of large geographic extent. Development of a rapid and accurate methodology for selecting connection regeneration sites due to optical reach constraints is one of the major hurdles to be overcome before a dynamic optical layer can be realized. The third area of focus is

that of multi-domain networks, where no single entity has a view of the network as a whole, such that it is necessary to stitch together partial solutions in order to accomplish end-to-end provisioning. The challenge is finding near-optimal solutions while not violating the security and administrative boundaries of the various entities that are involved. The dynamic multi-domain environment is covered in Sect. 8.8.

As noted above, in order for networks to be reconfigured remotely through software, the required equipment must already be deployed in the network. This largely corresponds to deploying an appropriate number of transponders at each of the network nodes. If too few transponders are deployed, blocking will result, whereas deploying too many transponders is not cost effective. Section 8.9 presents numerous techniques for estimating the number of transponders to deploy at a node, where the particular technique to use may depend upon the certitude regarding the traffic forecast. (These techniques are not limited to deploying equipment for purposes of dynamic traffic. Equipment is also periodically added to a network to accommodate typical network growth.)

One dynamic service that is somewhat less challenging to accommodate is scheduled traffic, where the customer requests bandwidth well in advance of the time it is actually required. This affords the network operator with time to more optimally plan how such traffic should be provisioned. Various aspects of scheduled traffic are covered in Sect. 8.10.

Finally, Sect. 8.11 covers *Software-Defined Networking* (SDN), a relatively new networking paradigm that has gained traction in data and telecommunications networks. One major tenet of SDN is the separation of the data and control planes. While the realm of SDN is broader than just traffic provisioning, its areas of largest impact are likely routing and network configuration. Additionally, while SDN is more of a transformational proposal for layers such as Internet Protocol (IP) and Ethernet, where the data and control planes are intricately intertwined, it is envisioned as a unifying architecture that would be implemented across networking layers, including the optical layer. Thus, it is worthwhile to examine the ramifications and relevancy of SDN on dynamic optical networking.

8.2 Motivation for Dynamic Optical Networking

Dynamic networking is advantageous for both network carriers and their customers because it delivers bandwidth where and when it is needed. From the carrier perspective, a dynamic infrastructure provides two major benefits. First, it allows the carrier to adapt its network to sudden changes in network traffic. These changes may be operational in nature; for example, an adjustment of the peering points between Internet service providers (ISPs) can result in major swings of traffic to the new border nodes. Alternatively, the traffic changes may reflect unanticipated exogenous events, especially with regard to the Internet. The Internet is growing in its role as: a major component of disaster recovery for corporations and large entities; a means of sharing computing and data resources across an enterprise; a prime source

of information during breaking new events; and a distribution channel for video and huge data sets. This trend will likely result in network traffic that exhibits more "discontinuities," where a sudden flux of traffic in various areas of the network occurs. Rapidly adding more capacity to the stressed portion of the network allows the carrier to continue delivering satisfactory network performance as traffic spikes. In contrast, achieving this performance in a static environment requires overengineering the network to account for a confluence of worst-case traffic scenarios.

In addition to dynamic networking being a reactive strategy to cost effectively deal with traffic fluctuations, it can also be used as a proactive sales tool to derive more value from the network. Once a dynamic infrastructure is in place, a variety of network services become viable, which adds to the revenue opportunities for the carrier; some of these are enumerated in Sect. 8.2.2. There is also the possibility of reaping operational cost benefits due to, for example, redirecting traffic based on power consumption considerations (see Sect. 6.10.1).

Overall, dynamic networking effectively decreases the network bandwidth requirements and allows a carrier to maximize the revenues derived from a given level of deployed network capacity.

Network customers benefit from dynamic service offerings as well. Traffic within an individual enterprise may exhibit bursty tendencies (in fact, it is likely to be burstier than the traffic that is aggregated in the network as a whole). Bandwidth-on-demand allows the user to establish, tear down, and adjust connections between the various remote locations of an enterprise as needed. Even though there is a cost premium relative to the *average* amount of bandwidth used, a dynamic service is still more cost effective than nailing up maximum-sized circuits between each of the locations.

Although it may seem paradoxical that dynamic services potentially allow customers to reduce their networking expenses while network providers maximize their revenues, the twin benefits are derived from circuit-based *statistical multiplexing*. Statistical multiplexing gains arise when the sum of the maximum supported data rates is more than the sum of the allocated capacity, with the expectation that not all customers will utilize their maximum rates at the same time. The multiplexing gains typically increase with the number of dynamic customers. Such statistical analyses have been incorporated into the dimensioning of higher-layer networks (e.g., Frame Relay) for many years. Dynamic optical networking allows these same gains to be realized in the optical layer as well. (To be clear, we are referring to dynamic optical *circuits,* not connectionless architectures such as optical burst switching.) In a simplified view, the same optical capacity is being "sold" to multiple customers, while still meeting the performance requirements of each customer.

8.2.1 Capacity Benefits of Dynamic Optical Networking

To quantify the benefits of dynamic optical networking, a study was performed in Saleh and Simmons [SaSi11] using Reference Network 2, with realistic traffic patterns. Demands were modeled as *on/off services,* with an average on-time of 10%.

Connections were established and torn down as the demands toggled between the *on* and *off* states. This was compared to a scheme where connections were maintained for the duration of the service regardless of whether the service was active or not.

The more traffic that can take advantage of dynamism, the greater the capacity benefits of a dynamic network. This particular study was focused on networks in the 2025 time frame, where it was projected that 25 % of the connections in the optical layer would be dynamic. The study showed that dynamic networking reduces the capacity required *for these services* by a factor of 5, where capacity was measured in total bandwidth-km. Looking at a more near-term scenario, the study was repeated but with 10 % of the connections in the optical layer assumed to be dynamic. With this more conservative assumption, dynamic optical networking reduces the capacity required *for these services* by a factor of 4.

Note that when calculating the bandwidth required in the dynamic scenario, the links were sized to the maximum wavelengths needed over time. If one is willing to accept a small blocking probability, then the capacity benefits can be further improved.

8.2.2 Applications Enabled by a Dynamic Optical Network

Once a dynamic infrastructure is in place, it is natural to offer customers bandwidth-on-demand services. Some telecommunications carriers already offer such services, although these offerings are limited to sub-wavelength rates (e.g., 2.5 Gb/s or less) with setup times on the order of a minute [LiCh07, ATT10]. The customer typically pays for an access pipe into the network, with a prescribed maximum data rate. The customer can establish connections as desired, subject to the maximum aggregate rate. This allows for the implementation of, for example, fluid virtual private networks. Expanding this flexibility to wavelength services would be desirable.

A more opportunistic service could also be offered, where the carrier signals that it has available capacity for a given period of time, and customers grab the capacity as needed. This is suitable for applications where the customer temporarily desires more bandwidth but does not have stringent time requirements. Such an ad hoc service would be commensurately priced to make it attractive to customers. From the carrier perspective, revenue is brought in for capacity that would otherwise sit idle.

Dynamic optical networks can also be promoted as an enabler of high-performance cloud computing. In the cloud model, an enterprise uses the resources of provider data centers distributed across the network, instead of the local resources of their own offices, for tasks such as application hosting, backup and storage, content delivery, web hosting, and large-scale simulations [VaMa12]. Latency, and how it compares to the response time of performing the task locally, is one of the most conspicuous measures used by the customer to evaluate a cloud service. Enabling higher layers to automatically request more capacity from the optical layer when needed should improve the response time of cloud services.

A somewhat related application is grid computing, which is used as a means of sharing distributed processing and data resources that are not under centralized

control. Grid computing is used to support research in "e-science" areas such as high-energy physics, genomics, and astrophysics, where the requirements are expected to grow to exabyte data sets and petaflop computation [SaSi06]. For example, in some high-energy physics experiments, multi-terabyte data files need to be disseminated to multiple locations in a very short period of time. Such applications require on the order of terabit per second capacity, but for just minutes to hours. Dynamically allocating wavelengths is the most cost-effective means of supporting such traffic patterns. (Grid computing is discussed further in Sect. 8.10, in the context of *scheduled* dynamic traffic.)

For the applications discussed thus far, connection setup times on the order of seconds to minutes would generally suffice. However, there is ongoing research to provide setup times on the order of 100 ms in the optical layer; e.g., Saleh [Sale06], Baldine et al. [BJJL11], and Chiu et al. [CCCD12]. While clearly not all services have such a stringent connection setup requirement, some applications do require a very rapid network response time.

For example, with highly interactive visualization and data fusion, a user may pull together large chunks of data from numerous global locations to form a comprehensive situational awareness. One approach is to set up permanent connections between all locations that may participate in the process. However, given the numerous locations that may be involved and the relatively small proportion of the time that any one connection is needed, establishing permanent connections can be prohibitively expensive. Dynamic networking is an attractive alternative, where the connection setup time must be on the order of 100 ms to meet the human tolerance for delay with interactive applications.

Similarly, global-scale distributed computing can benefit from dynamic networking, where the connections must be established and torn down very rapidly in order to realize significant savings in required capacity. The study in Saleh [Sale07] showed that a connection setup time of 100 ms as opposed to 1 s can reduce the bandwidth requirements for distributed computing by a factor of 2.

Another application that becomes feasible with very fast service setup is "route hopping" for purposes of security, where a connection is rapidly moved to a new path to avoid eavesdropping. This is similar to the idea of "frequency hopping" used in some military radio systems. A path would likely need to remain in place for at least a couple of seconds for this to be practical; otherwise the overhead in establishing the paths, even with 100 ms setup times, would be too large.

Setup times on the order of 100 ms also allow for the possibility of restoring a failed connection by issuing another setup request. This was discussed in Sect. 7.6.4. This restoration method can be very beneficial, especially when there are connections that must be able to survive multiple concurrent failures in the network. It is unlikely, however, that it would be suitable as the primary means of restoration for all demands. The number of demands brought down by a failure could be quite large, such that the system would be flooded with simultaneous setup requests.

Whether or not 100 ms connection setup can be achieved in practice depends on the protocols that are developed, the capabilities of the network equipment

(e.g., laser tuning times, ROADM reconfiguration times), and the ability to manage physical-layer issues such as optical amplifier transients. Furthermore, support for dynamic applications, regardless of the setup time requirements, goes beyond rapid and automatic reconfiguration of the optical layer. It is necessary that the edge network and the customer premises equipment be compatible with a dynamic paradigm as well.

8.2.3 IP over a Dynamic Optical Layer

While the benefits of a dynamic optical layer have been elucidated above, it is important to consider the ramifications for the higher electronic layers, especially in an IP-over-optical environment [BSBS08]. First, existing capacity between IP routers can be increased through the provisioning of additional wavelengths that are routed over the same path as the existing capacity (i.e., another wavelength is added to the trunk between two routers). This has minimal effect on the IP layer as the router adjacencies remain intact. Second, the wavelengths between two adjacent routers may be shifted to a new path; this can affect parameters such as latency, which would need to be advertised in the IP layer. A more disruptive operation is when the optical layer is reconfigured such that IP router adjacencies are either created or deleted. This affects the IP-layer topology, which can lead to convergence issues; thus, such changes would need to be performed judiciously. Note that the SDN paradigm, discussed in Sect. 8.11, advocates unified, logically centralized control across networking layers, such that instability issues related to changes in the IP virtual topology would be minimized.

8.3 Centralized Path Computation and Resource Allocation

To implement dynamic optical networking, the optical-layer control plane is responsible for: knowledge of the underlying system parameters (e.g., optical reach), the network topology, and the current state of network resources; path computation; resource selection; and control of the network elements. We first consider a centralized model, where the path computation and resource allocation functions are performed at a single location in the network. For example, the Internet Engineering Task Force (IETF) has defined a *Path Computation Element* (PCE)-based architecture, where the PCE is a powerful computing platform that is capable of performing constraint-based routing [FaVA06, PCGS13]. In the centralized model, the PCE is located at a single node or server, and all routing requests are directed to it. There may be multiple PCEs for purposes of reliability; however, only one is active at any given time.

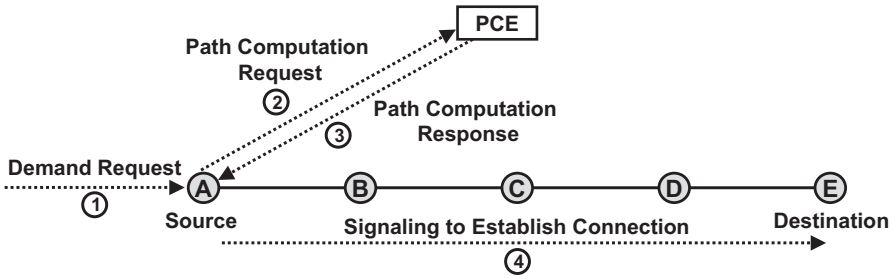


Fig. 8.1 *Centralized Operation:* A request arrives for a demand from Node *A* to Node *E*. The source node, *A,* requests a path from the Path Computation Element (*PCE*), and the *PCE* responds with the route and the resources to use. Node *A* uses signaling along the calculated route to establish the new connection

8.3.1 PCE-Based Operation

In the centralized single-PCE architecture, it is assumed that the PCE can have full visibility into the state of the network. (This implicitly assumes that the PCE is operating in the "stateful" mode as opposed to the "stateless" mode [FaVA06].) There are various means that the PCE can employ to maintain its *traffic engineering database* (TED). System parameters, such as optical reach or line rate, are likely to change very infrequently, if at all, and thus can be manually configured on the PCE, e.g., through the NMS. The physical network topology may change, on a relatively slow scale, as links fail and are repaired. The PCE can track the current network topology by having it receive the *link-state advertisements* (LSAs) that are periodically flooded by the control plane, e.g., via a routing protocol such as Open Shortest Path First (OSPF-TE) [KaKY03]. (Other methods of tracking the topology can be found in Paolucci et al. [PCGS13].) The state of the network resources, e.g., the wavelengths and transponders, may change on a rapid timescale depending on the level of dynamism in the network. However, because the single PCE has sole ownership of assigning these resources, it is capable of having full knowledge of their state. (There may be brief scenarios where the PCE believes a resource is in the *Assigned* state, when it actually is still available, due to a problem with, for example, a switch carrying out a connection setup request. However, this should be quickly remedied after the failure notification is received by the PCE. This type of inconsistency should not be problematic.)

The operation of the single-PCE-based architecture is shown in Fig. 8.1 (this operation follows the specifications of Farrel et al. [FaVA06]). The process is triggered when a new demand request (e.g., from Node A to Node E) is received by the source node. The source node, which acts as a *Path Computation Client* (PCC), sends a path request to the PCE, along with any special requirements, such as which nodes to avoid in the path, which metric to optimize in the path computation, the level of diversity (for a protected path), etc. The *PCE Communication Protocol* (PCEP) is used for all communication between the PCC and PCE [VaLe09]. After

receiving the request, the PCE performs the necessary routing calculations and resource assignments and returns the result (assuming one can be found) to the source node. The source node then uses a signaling protocol to communicate with the nodes along the resultant path to establish the new connection (i.e., the lasers are tuned and the ROADMs are configured, as calculated by the PCE).

8.3.2 Advantages of Centralized Operation

The biggest advantage of centralizing all path computation and resource allocation in one entity is the potential for optimality. By having full knowledge of the network state, the PCE can calculate the "best" path, regeneration locations, and wavelength(s) for a new connection. Furthermore, the PCE can operate in a batch mode, where the calculations are performed across a set of new demand requests rather than processing each request one by one (this is referred to as "synchronized path computation"). This potentially improves the quality of the solution as analyzed in Ahmed et al. [ACMW12].

A second advantage of this model is the ability to avoid resource contention during the connection setup phase. The prime resource that is assigned during provisioning is the specific wavelength(s) that will carry the connection. Because the PCE can track all wavelengths that have previously been assigned, it can avoid assigning the same wavelength to multiple connections routed on the same fiber.

A third advantage to centralizing the computation is that the processing and memory resources are deployed in one element as opposed to being required at multiple network sites. Depending on the complexity of the network design algorithms, the processing requirements can be significant.

8.3.3 Disadvantages of Centralized Operation

Conversely, centralized computation is often flagged as a disadvantage because the PCE could be overwhelmed with new demand requests. However, with processors becoming ever more powerful (e.g., due to multi-core technology and hardware-based accelerators) and with the price of memory continually trending downward, it would be expected that, under most circumstances, the latency due to queuing of requests at the PCE can be kept to a minimum. Additionally, prioritization can be enforced within the PCE, where route calculations are preferentially performed for the most urgent requests, based on the priority specified by the PCC.

Furthermore, if a multistep approach to optical network design has been adopted, where routing and wavelength assignment are treated as separate problems, it is possible to apportion these tasks to separate processing units. For example, there could be a routing PCE, which calculates the path, and a wavelength-assignment PCE, which then selects the wavelength for each link of the path [LBMT13]. Due to the separability of the tasks in the multistep design approach, the outcome should

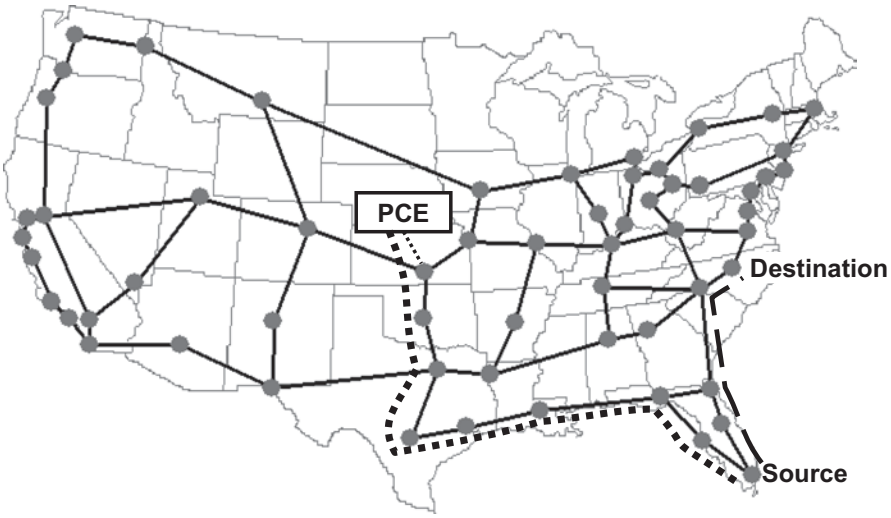


Fig. 8.2 In the centralized Path Computation Element (*PCE*) model, the round-trip propagation delay between the source node and the *PCE* can be tens of milliseconds (or hundreds of milliseconds in a global network), adding to the overall connection setup latency

be the same as one PCE sequentially performing each of the steps. Thus, while multiple PCEs are involved, this model falls under the centralized architecture. The trade-off is the benefit of having dedicated processors versus the additional latency due to inter-PCE communication.

Thus, processing bottlenecks are not likely to be the most significant impediment (although managing delays will still be a challenge). Rather, the biggest potential drawback of the centralized PCE approach is the latency due to propagation delays between the source node (i.e., the PCC) and the PCE. Consider a single-carrier single-domain continental-scale network. In the centralized model, one PCE would be designated for the whole network; presumably, it would be centrally located within the network, as illustrated in Fig. 8.2. At a minimum, the connection setup time includes the round-trip propagation delay between the source node and the PCE (to determine the new path information), and the propagation delay from the source to the destination (to establish the new path). Added to the propagation delay are the computation time at the PCE and the time to configure the equipment (e.g., lasers, ROADMs) along the new path. While the resulting total setup time is sufficient for many applications, it may be too slow for applications with stringent connection-time requirements (e.g., less than 100 ms). Given that the bulk of the delay is due to propagation, which is dictated by the speed of light in fiber, system enhancements, such as having a more powerful PCE, do not solve the problem.

Furthermore, the delay may be even greater, depending on the mechanism for determining when the source node can safely begin transmission. In an aggressive approach, the source node estimates the time it will take for its setup message to reach the destination node, and for all of the nodes along the path to configure their switches, tune their transmitters, etc. More precisely, the source needs to wait long

enough such that when the transmission reaches a given node, that node will be appropriately configured. Thus, the source node begins transmission without receiving any positive confirmation that the path has actually been established properly. This is a potential security risk; for example, a misconfigured switch could direct the transmitted data to the wrong destination. (Note, however, that in schemes such as optical burst switching, transmission occurs based on timing, without receiving any confirmation of path setup.) Alternatively, in a more conservative approach, the source node waits for an acknowledgment from the destination node that indicates all of the relevant network equipment along the path has been properly configured. While potentially "safer," it adds the propagation delay from destination node to source node to the connection setup time, further exacerbating the delay problem.

It may be possible to improve the setup time in some scenarios (see Exercises 8.1 and 8.2) if the PCE communicates the setup commands directly to each of the nodes along the new path, rather than have the source node send out a configuration message [BJJL11]. However, this mode of operation is not explicitly supported by PCEP, where PCE path replies are sent only to the node that requests the path. (Note that *OpenFlow*™, to be discussed in Sect. 8.11.1, does support this parallelized operation mode.) Even with this variation, it is unlikely that the single-PCE architecture can meet the most demanding connection setup times, especially in a network of large geographic extent. Furthermore, one could consider scenarios that are more extreme than what is shown in Fig. 8.2. For example, in a global network with one PCE located in the USA, a demand request between two cities in Europe would need to be directed to this PCE.

8.3.4 Multiple PCEs

A logical alternative is to deploy multiple PCEs throughout the network, such that the propagation delay from any network node to a PCE is below a certain threshold. This potentially reduces the setup time by tens of milliseconds, depending on the geographic extent of the network. Note that the most stringent connection setup time requirements drive this architecture, even though the corresponding applications may represent just a small percentage of the dynamic demand requests.

Multiple PCEs are often deployed when it is necessary to find paths that span multiple domains or carriers, where it is not possible for a single PCE to have full knowledge of the network. In that scenario, each PCE has a set of resources for which it is responsible, and the end-to-end path is formed by stitching together the partial paths calculated by each PCE (see Sect. 8.8). The multi-PCE solution that we discuss here, however, is of a very different nature and is strictly motivated by the desire to minimize the setup time.

It is assumed that each of the PCEs has visibility into the whole network, just as the single PCE did. Thus, each PCE is capable of calculating an end-to-end path without having to communicate with other PCEs. (If communication with other PCEs that are spread across the network were required in order to compute a path, it would defeat the purpose of trying to minimize the delay.)

From the point of view of an individual PCC, the process remains centralized; however, from the network's perspective, it is not. As soon as multiple PCEs are involved in the resource allocation process, the potential for contention arises. Even though a PCE can communicate the results of all of its path computations to the other PCEs (via PCEP), there will be a delay in propagating the information. Thus, for some amount of time, the other PCEs will be unaware, for example, that a particular wavelength on a link has already been assigned. Another PCE may select that same wavelength for one of its demand requests, thereby causing contention.

One means to address contention is to partition the network resources among the PCEs. For example, each of the PCEs could be responsible for a designated set of wavelengths throughout the network. When assigning a wavelength to a new demand, it must select a wavelength from that set. However, a fixed partitioning of resources is generally not optimal, and would likely lead to excessive blocking of new demand requests.

To summarize, a centralized single-PCE-based solution is likely satisfactory if the setup time requirements are on the order of 1 s or more, but is not adequate if the requirements are on the order of 100 ms. A quasi-centralized multi-PCE solution may meet the timing requirements in the latter scenario, but it introduces other drawbacks, most importantly, contention.

8.4 Distributed Path Computation and Resource Allocation

The previous section considered a centralized approach to path computation and resource allocation, with the major impediment likely being delay. In contrast, this section discusses a purely distributed approach. For illustrative purposes, we use the Generalized Multi-Protocol Label Switching (GMPLS) architecture [Mann04], with the Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) signaling protocol [ABGL01]. GMPLS has a number of options that can be implemented; the operation that is presented here should be considered just one variation. For example, for simplicity, we describe the setup of a unidirectional path from one source to one destination; however, bidirectional paths can be established with similar procedures. When describing GMPLS-based operation, it is assumed that regeneration is not required due to optical reach constraints; optical reach is not adequately addressed in GMPLS, as discussed in Sect. 8.7. However, it is assumed that regeneration may occur for purposes of wavelength conversion, using the procedure described in Lee et al. [LeBI11].

8.4.1 GMPLS-Based Operation

It is assumed that a demand request is received at the source node corresponding to that demand. In a fully distributed implementation, each of the nodes is capable

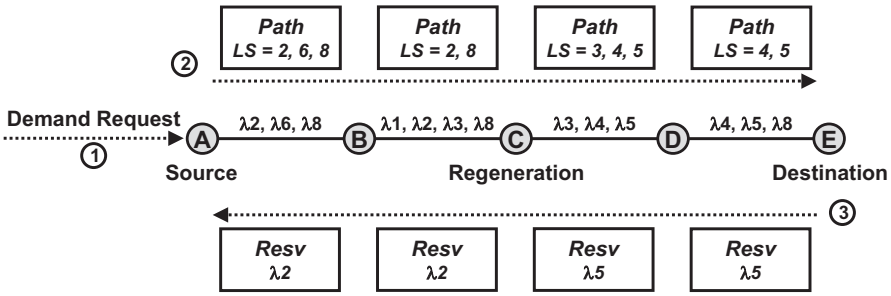


Fig. 8.3 *Distributed operation:* A *request* arrives for a demand from Node *A* to Node *E.* Node *A* initiates a *Path* message with a Label Set (*LS*) containing the wavelengths that are free on its outgoing link (as shown by the link labels in the figure). Node *C* designates itself as a *regeneration* site for purposes of wavelength conversion. At the destination, Node *E* selects λ_5 from the *LS* and initiates a *Resv* message; λ_5 is reserved on Links *DE* and *CD.* Node *C* performs wavelength conversion by selecting λ_2 for the remainder of the path

of path computation without consulting external entities. For example, information regarding the network topology can be disseminated to the nodes via OSPF-TE LSAs.

Once the path is computed by the source node, a two-pass signaling process is initiated, as illustrated in Fig. 8.3. The wavelengths that are assumed to be available on each link are shown in the figure. Using RSVP-TE signaling, the source transmits a *Path* message along the route that it just calculated. An important field in this message is the *Label Set,* which is initialized to a list of wavelengths that are available on the outgoing link from the source node (all possible available wavelengths do not have to be included in the list). The next node in the path removes from the Label Set any wavelengths that are not available on its outgoing link. This process continues at each node in the path, such that at any node, the Label Set indicates a set of wavelengths that are free along each of the links from the source node to that node.

If none of the wavelengths in the Label Set are free on the outgoing link of a node, then regeneration must occur for purposes of wavelength conversion. In the figure, this occurs at Node C. That node selects one of the wavelengths from the incoming Label Set (which will become the wavelength used on the path up to that point), designates itself as a regeneration location for the connection, and then resets the Label Set field with a list of wavelengths that are free on its outgoing link (i.e., as if it were the source node). (A node could also be more proactive in performing wavelength conversion; e.g., if the number of wavelengths in the Label Set is very small, it can choose to regenerate.) Assuming that there is at least one wavelength available on each link (and transponders available for regeneration, if needed), the *Path* message is eventually received at the destination. If a link is encountered where there are *no* wavelengths available, then the *Path* message is dropped, a failure message is sent back to the source node, and the demand request is blocked.

Assuming that the *Path* message is successfully received at the destination, a *Resv* message is sent in response. The destination selects one of the wavelengths

from the Label Set, and includes that wavelength in the *Resv* message. For example, in Fig. 8.3, it is assumed that the destination selects λ_5. As the *Resv* message travels back to the source node, the specified wavelength is reserved (on the links in the source to destination direction; e.g., Link DE in Fig. 8.3) and the ROADMs are configured accordingly. If a node was designated as a regeneration site on the first pass, then that node needs to update the wavelength field of the *Resv* message. As the *Resv* continues on its way to the source, this new wavelength will be the one that is reserved. For example, in the figure, Node C places λ_2 in the *Resv* message, thereby accomplishing wavelength conversion.

Eventually, the *Resv* message is received at the source node. After the equipment at the source node is configured, transmission can begin.

8.4.2 Advantages of Decentralized Operation

The most important advantage of the distributed approach is that the setup time is potentially very fast. The delay primarily consists of the round-trip propagation time between the source node and the destination node, and the time to configure the switches.

There are generally two options regarding switch configuration that may be supported in the reservation phase of a distributed setup protocol; the methodology employed has a significant impact on setup delay. RSVP-TE specifies that a node *should* configure its switch prior to forwarding the *Resv* message to the upstream node [ABGL01, ShFa11]. The drawback of this mode of operation is that each required switch configuration contributes to the overall delay. This method is typically implemented so that the source node is assured that the path is properly established prior to commencing transmission. (However, because the terminology "should" is used in Awduche et al. [ABGL01], as opposed to "must," it is not an absolute requirement that RSVP-TE be implemented in this manner.)

In the second option, often referred to as *pipelining,* the reservation is sent to the upstream node without waiting for the switch to be configured at the current node. This is faster, as it allows switch setup to occur in parallel at the nodes, and would clearly be preferred for connections with stringent setup time requirements. It may appear that this method necessitates that the source estimate when transmission can safely begin, without receiving assurance of path setup. However, it is possible to send an additional message from the destination to the source that verifies that the path has been properly established. Assuming that the switch configuration times are approximately the same at each node (and assuming that the control-plane topology coincides with the data-plane topology), the verification message does not add to the connection setup time (except for a small processing delay). The timing of the verification message is explored in Exercise 8.6.

To be able to achieve very rapid setup times, it is necessary to pipeline the *Resv* message. It is assumed here that this is a feasible implementation.

8.4.3 Disadvantages of Decentralized Operation

Among the disadvantages of the distributed scheme are the need for potentially powerful processing resources at each of the nodes, and the loss of optimality in performing route calculations. However, the most significant drawback is the amount of resource contention that can potentially occur. A key property of RSVP-TE signaling is that there is no reservation of any wavelengths in the forward direction, from source to destination. The Label Set only indicates the wavelengths that are free at the time the *Path* message is processed at a node. It is not until the backward direction, from destination to source, that the wavelengths are actually reserved. This is known as a *destination-initiated reservation* (DIR) scheme.

Some amount of time transpires between the arrival of the *Path* message and the arrival of the *Resv* message at any given node. During that time, wavelengths that had been available (and which were included in the Label Set) may have been assigned to other connections that were also in the process of being established. Thus, by the time the *Resv* message is received at an intermediate node, the wavelength selected by the destination (or by a wavelength-conversion site) may no longer be available. This is called *backward blocking* because it occurs in the backward direction, from destination to source. When it occurs, the *Resv* message is dropped, a failure message is sent both to the source and the destination, and any successful reservations that already occurred for this connection are released.

The source node may choose to initiate another setup, which adds to the delay, or it may drop the request. Thus, backward blocking may cause a demand request to be blocked even though feasible paths exist between the source and destination. Studies have shown that under light load or when the network is highly dynamic, backward blocking is the predominant cause of blocking [GSCA09].

8.4.4 Schemes to Minimize Contention

Backward blocking occurs because of the delay between the probing message and the reservation message. It is natural to consider reserving the wavelength during the forward pass, as soon as the *Path* message is received. This is known as *source-initiated reservation* (SIR). The difficulty is in selecting which wavelength to reserve. A particular node does not know what wavelengths will be available on downstream links (assuming wavelength-state information is not flooded), which is why several potential wavelengths are typically included in the Label Set. Thus, in an SIR scheme, multiple wavelengths would likely need to be reserved in the forward direction. This ties up a set of resources until a message is received from the destination indicating which wavelength will actually be used for the connection, at which point the other reserved wavelengths can be released. During this time, however, other demand requests may be blocked due to the number of wavelengths that have been reserved. Studies have shown that SIR schemes result in greater overall blocking than DIR schemes [YuMG99].

GMPLS does allow for a "quasi-SIR" operation [Berg03]. On the forward pass, a node may designate one wavelength in the Label Set as the *Suggested Label* and may begin configuring its switch based on this choice. This Suggested Label is passed downstream in the *Path* message. However, there is no guarantee that this will be the wavelength that is ultimately selected by the destination. If a different wavelength is specified in the *Resv* message, it overrules the Suggested Label. Thus, backward blocking may still occur.

One factor that affects the likelihood of contention is the methodology that the destination (or a wavelength-conversion site) uses to select a wavelength from the Label Set. Randomly selecting the wavelength generally reduces backward blocking as compared to a scheme such as First-Fit, where the lowest indexed wavelength in the Label Set is selected [LiWM07, GSCA09]. This is expected, because if other connections that are simultaneously being established select the lowest indexed wavelength in their respective Label Set, there is a greater likelihood that the same wavelength will be chosen on a particular link. However, under heavy load, random wavelength assignment ultimately makes it more difficult to find a wavelength that is free along an entire path, thereby requiring more regeneration for purposes of wavelength conversion.

A variety of other mechanisms for minimizing contention in a DIR scheme have been proposed, e.g., Ozugur et al. [OzPJ03], Lin et al. [LiWM07], and Giorgetti et al. [GSCA09]. As a generalization, most of these schemes try to predict which wavelengths are likely to be selected by other setup requests, and then try to avoid selecting these same wavelengths for the current request. For example, each node may be required to tally how many times a wavelength has been placed in the Label Set or Suggested Label fields for recently processed *Path* messages (i.e., *Path* messages for which a corresponding *Resv* message has not yet been received). Mechanisms are then implemented to reduce the probability that a wavelength that has been placed in a Label field many times will be selected by the destination in response to a new *Path* message.

A different approach to deal with contention was proposed in the *3-Way Handshake* (3WHS) signaling protocol of Skoog and Neidhardt [SkNe09] and Chiu et al. [CCCD12]. There are numerous enhancements to GMPLS that are included in 3WHS; the focus here is on the methodology to reduce backward blocking. After receiving the *Path* message, the destination selects a wavelength from the Label Set as the primary wavelength to use for the connection. In addition, it picks a secondary wavelength to be used in case backward blocking of the primary wavelength occurs. Both wavelengths are included in the *Resv* message, and nodes along the path reserve and configure their switches for both of them. The source selects one of the wavelengths that was successfully reserved along the whole path, with preference given to the primary wavelength; i.e., the secondary wavelength is not selected unless the primary one was blocked. (If wavelength conversion is needed, then there will be a primary wavelength and a secondary wavelength for each "subconnection." Either all primary wavelengths or all secondary wavelengths are selected.) The source sends a message back towards the destination indicating its choice of wavelength so that any node that also reserved the other wavelength can release it. The source can begin transmission after its own switch is configured; i.e., the reservation release process does not add to the setup delay. If neither the primary nor the secondary wavelength was reserved successfully, then the setup fails.

Simulations on Reference Network 1 showed that by reserving a secondary wavelength, the rate of backward blocking can be significantly reduced [SCGN12]. The trade-off is that, for a brief time, an extra wavelength may be reserved such that it cannot be used by other concurrent demand requests; however, it was shown that this had little negative effect. (The effect is smaller as compared to an SIR scheme that reserves extra wavelengths. In an SIR scheme, several extra wavelengths may need to be reserved, due to the uncertainty regarding which wavelengths may be free on downstream links.) A similar technique was considered in Yuan et al. [YuMG99] as one of numerous possible reservation schemes; this study also found that a DIR scheme that reserved one to two extra wavelengths performed the best.

8.4.5 Subconnection as the Label

Establishing a new connection in an optical-bypass-enabled network typically involves tuning one or more lasers and configuring a number of ROADMs. As discussed in relation to protection, these actions, when performed rapidly, potentially result in optical amplifier transients. At least in the near term, while transients are still an issue in most systems, this could preclude extremely fast connection setup. One of the solutions proposed in relation to shared restoration was to deploy a set of pre-lit subconnections. When failure recovery is necessary, a sequence of subconnections are concatenated together *in the electrical domain* to establish an end-to-end recovery path (see Sect. 7.8).

As outlined in Simmons et al. [SiSB01], the same type of solution also can be employed for rapid connection setup. Each pre-lit subconnection would be assigned an ID number. The GMPLS Label Set field would then contain the IDs and link sequences of available subconnections that lie along the desired new path, rather than available wavelengths. The destination selects a set of these subconnections to form the new path, if possible. The concatenation of the subconnections would occur on the second pass. This methodology avoids any issues with transients. The disadvantage is that, at times, wavelengths are lit without being used to carry traffic, thereby "burning" capacity. However, the same set of pre-lit subconnections could be used for dynamic setup as for restoration. Furthermore, it would only be those demands with the most stringent setup requirements that would require this method, thus moderating the number of required pre-lit subconnections.

8.5 Combining Centralized and Distributed Path Computation and Resource Allocation

As described above, both the centralized and distributed architectures have strengths and weaknesses. This section considers combining the two architectures, using a GMPLS/PCE model [LeBI11], to capitalize on their respective advantages. The major strengths of the single-PCE architecture are centralized processing

requirements and improved optimality of the solution; the main disadvantage is the potential connection setup delay. While the delay can be addressed using a multiple-PCE approach, this may lead to excessive contention when assigning wavelengths (due to stale state information). The major strength of the distributed GMPLS architecture is the minimization of delay, assuming backward blocking is adequately addressed; the main disadvantages are non-optimality and the required processing at each node.

It is assumed here that a two-step approach to routing and wavelength assignment yields good performance, such that these tasks can be separated. When deciding where a particular task should reside, the important factors to consider are the processing requirements, the required state information, and how quickly that state information changes.

Route calculation in the optical layer is characterized by high processing requirements, especially if optical impairments need to be considered. Thus, from a processing perspective, a PCE-based routing implementation is favored (so that powerful processors are not needed at every node).

Next, consider the state information needed for routing. To calculate a *valid* route, the network topology needs to be known. The topology changes on a relatively slow timescale; thus, disseminating topology information, whether to the PCEs or to the network nodes, should not be challenging. To calculate a *good* route, parameters such as link length and link load may be used, where link load is typically the most important time-varying routing metric. For an algorithm such as Least-Loaded routing (see Sect. 3.5.2), it should be sufficient to know the approximate load on each link, e.g., within a few wavelengths of the actual load. (Knowledge of the exact link load is only important if every wavelength has been assigned on a link, such that the link can be eliminated from the route calculation.) Furthermore, it is not necessary to know the particular wavelengths that have been assigned, but simply *how many* (or approximately how many) have been assigned.

Thus, routes that have been calculated based on somewhat stale load information are likely still valid. This has two implications. First, it implies that routes can be calculated periodically rather than in response to each demand request. Thus, delegating the routing function to the PCE does not imply that the PCE must be "consulted" whenever a new demand request arrives. Second, calculating routes in a multiple-PCE architecture should not be problematic, despite the state-synchronization delays.

As compared to routing, wavelength assignment has an opposite characterization. Given the route, wavelength assignment is typically a relatively simple procedure (e.g., First-Fit), such that powerful processors are not required. (This assumes that it is not necessary to take into account the nonlinear effects of the adjacent wavelengths, as described in Sect. 5.9.)

Conversely, efficient wavelength assignment does require up-to-date state information regarding the status of the wavelengths. Otherwise, the ensuing contention, where the same wavelength is assigned to multiple concurrent demand requests, could lead to a high rate of blocking.

This implies that if the task of wavelength assignment is delegated to the PCE, then the PCE would need to be consulted for the proper wavelength(s) to be used,

on a per-demand basis. This would require that communication with the PCE be a part of the setup process for each demand; as discussed previously, the resulting delay may be prohibitively long for some applications. Conversely, GMPLS can readily determine the available wavelengths on each link of a path at the time of each demand request, via the *Path* message. Furthermore, using techniques as described above for the 3WHS protocol, backward blocking can be kept to a minimum. This favors a GMPLS-based approach to wavelength assignment.

Thus, we conclude that in a combined centralized/distributed architecture, routing should be PCE based and wavelength assignment should be GMPLS based. One mode of operation is that the PCE periodically calculates a route for each source/destination pair, and distributes the resulting path for each pair to the source node. The optimal route is likely to remain the same for a relatively long period of time (i.e., relative to the connection setup times), such that frequent updates are not required [CCCD12]. When a demand request is received at the source node, it uses the prescribed path, without further consultation with the PCE. It then initiates wavelength assignment using a distributed signaling scheme, as described earlier. This combined GMPLS/PCE mode of operation, for the most part, takes advantage of the strengths of both architectures. Note that this architecture is amenable to one or more active PCEs.

One potential weakness is with respect to optimality, where some of the benefits of having the entire design process reside in a single PCE are lost. However, the 3WHS signaling protocol adds two features to improve optimality with respect to standard GMPLS signaling [CCCD12]. First, it allows multiple candidate paths (as calculated periodically by a PCE) to be simultaneously probed for a new demand request, as opposed to probing just a single path as in GMPLS. The destination can then select the best path based on the resource information that was collected by each of the *Path* messages. It is desirable that the distances of the candidate paths not be very different so that the latency in receiving the *Path* messages is not excessive (or, a destination can potentially select a route without waiting for all of the *Path* messages to be received).

A second feature of 3WHS that improves optimality is that all resource decisions are made by the destination. With the GMPLS-based procedure described in Lee et al. [LeBI11], as the *Path* message propagates over the calculated route, a node may choose to perform wavelength conversion, where these nodes are then responsible for selecting the new wavelength. Having the destination make all of these decisions likely improves the performance. This is especially true when regeneration due to optical reach is required, as considered in Sect. 8.7.

8.6 Dynamic Protected Connections

There are a variety of methods that can be used to establish protected connections in a dynamic environment. First, consider the PCE-based architecture of Sect. 8.3. A PCC can request that N paths between a source and destination be calculated in a "synchronized fashion," such that the paths demonstrate link diversity, node

diversity, and/or shared risk link group (SRLG) diversity. The PCE should be able to calculate paths that are suitable for either dedicated or shared protection, where the latter requires more computational complexity.

With a *single PCE,* all working paths, along with their respective protect paths, are known. This allows reserving resources for shared protection to be optimized. With *multiple PCEs,* allocation of shared restoration resources may not be as efficient due to out-of-sync state information among the PCEs. For example, two protected demand requests may simultaneously arrive at different PCEs. The resultant working paths may be diverse such that they could potentially share restoration capacity. The PCEs would initially not be aware of this, which may result in excess reserved shared capacity. This should be a relatively minor problem, however. Shared mesh resources can be periodically re-optimized, with the endpoints of a connection informed of any change. Because the restoration path is only used during a failure, the path to use can be updated without disrupting live traffic. The bigger problem with the multi-PCE architecture is the potential for resource contention, as discussed earlier. For example, a PCE may designate wavelengths for restoration that are simultaneously being selected by another PCE for a working path.

Next, consider the combined GMPLS/PCE approach, described in Sect. 8.5, where the PCE calculates the routes and GMPLS signaling is used to select the wavelengths. In a protection scheme such as 1 + 1, this is straightforward, as GMPLS signaling would probe both the working and protect paths, and use the Label Set to track the available wavelengths. In a shared restoration scheme, ideally the shared resources that have already been reserved are reused if possible, which requires storing more state information at the nodes. One scheme specifically designed for this purpose is *distributed path selection with local information* [AYDA03]. (This scheme was also noted in Sect. 7.11.2.5 for its suitability for a distributed environment.) In this scheme, a node classifies each wavelength on its outgoing links into one of three states: available, assigned and non-sharable, and assigned but sharable. For the last class, the node also tracks which working path links (and possibly which nodes and SRLGs) are protected by that shareable wavelength. This can be accomplished, for example, by having any *Resv* message sent on a protect path indicate the corresponding working path information. When a new request arrives for a demand that requires shared restoration, the *Path* message that is sent to probe the protect path must specify the corresponding working path that is being probed as well. As this *Path* message propagates, each node checks whether the wavelengths in the Label Set are shareable and viable for the new demand (i.e., the node checks whether the potentially new working path is disjoint from all of the working paths already protected by the shared wavelength). If so, this would be recorded in a Label Set attribute, so that these wavelengths can be preferentially reserved for the protect path.

For a protocol such as 3WHS, where multiple working paths are probed for a new demand, it may be desirable to first establish only the working path (the working paths that are probed should be paths that potentially have a diverse protect path to avoid issues with "trap topologies," as described in Sect. 3.7). For connections with the most stringent setup time requirements, transmission may start after the working path is established. Concurrently, the source node requests from the PCE

a protect path for the selected working path. Assuming that the PCE is able to find a suitable protect path, the source node then proceeds with the resource notification along that path. Until that task is complete, the connection is unprotected. If the PCE is unable to find a protect path, then the connection remains unprotected. However, connections that require very rapid establishment are often short lived, such that the lack of a backup path may not pose a big risk. For connections with a setup time requirement of a second or greater, the source node can delay the start of transmission until it receives notification from the PCE of the protect path. If a protect path cannot be found, there is time to restart the setup process.

Finally, with the ability to establish connections in less than 100 ms, pre-calculating a protect path may be unnecessary. A failed connection can be recovered by issuing a new connection request. As has been mentioned several times, relying on such a methodology as the primary recovery scheme for all connections is unlikely to be viable, as the burden on the control plane would be too great. This was explored experimentally in Perelló et al. [PSAA12]. As expected, restoration time grew with the number of demands restored.

8.7 Physical-Layer Impairments and Regeneration in a Dynamic Environment

Capturing the relevant physical-layer impairments and selecting the sites at which a connection should be regenerated can be challenging even in a non-dynamic environment. It can be more difficult in a dynamic environment where full resource state information may not be known. Many of the protocols utilized for dynamic networking were developed prior to the widespread deployment of optical-bypass-enabled networks. Thus, to a large extent, adding the proper support for these networks has been through a series of patchwork additions to the existing protocols. Some aspects have been adequately addressed; for example, the GMPLS Label Set can be used to enforce the wavelength continuity constraint. However, other aspects, such as proper treatment of physical-layer impairments, still require that additional support be added to the protocols. Some preliminary IETF proposals to remedy this can be found in Agraz et al. [AgYH10], Martinelli and Zanardi [MaZa10], and Lee et al. [LBLM12].

There is also much ongoing research into how best to enforce *quality of transmission* (QoT) in a dynamic environment; e.g., Martínez et al. [MPCA06], Cugini et al. [CSAG08], Sambo et al. [SPLC09], Sambo et al. [SGCA09], and Angelou et al. [Ange12]. Some of this research makes the simplifying assumption that all paths must be purely all-optical, with no regeneration. With this assumption, the emphasis is on verifying that a calculated end-to-end path meets the QoT threshold, either through analytic means or by sending a probe to test the actual path. However, it is preferable to develop solutions that support dynamic optical networking in more general settings, where regeneration may be required along a path due to optical reach constraints. Various options are discussed in Sects. 8.7.1 and 8.7.2, all of which require extensions to current signaling protocols.

Note that using analytic methods to calculate QoT, whether it be for determining if an end-to-end path meets the desired QoT threshold or determining where regeneration is required along a path, may be time consuming depending on the complexity of the approach. While tools have been developed for such calculations, e.g., Azodolmolky et al. [Azod11], the run times are typically on the order of seconds (using hardware accelerators). Furthermore, multiple such calculations may need to be performed as part of the evaluation process for a potential new connection (e.g., to compare different routes). This process may be too slow to meet the setup time requirements for some dynamic applications; faster methods are desired. For example, as discussed in Chap. 4, QoT can be estimated by mapping the various impairments to optical signal-to-noise ratio (OSNR) penalties, and comparing the overall effective OSNR to a desired threshold. Such a calculation can be performed very rapidly, and it presents a reasonable trade-off between run time and accuracy.

Another proposed methodology is to make use of a cognitive QoT estimator [JADD13]. In this strategy, a database is maintained for a set of paths for which the QoT is known (through analysis, experimentation, and/or performance monitoring in the network). There are likely to be thousands of paths in this database; the number of entries scales with the size of the network. Each of these paths is characterized by a set of metrics (e.g., path distance, wavelength, dispersion, etc.). When a potential new connection, or subconnection, is being evaluated, the paths in the database that are most similar to it are used to determine whether the QoT will meet the system threshold (interpolation may be necessary). The key to this technique is maintaining the database. If there are too many entries, the run time will be slow; if there are too few entries, the comparisons will not be very accurate. Ideally, the database is periodically updated based on the monitoring of established connections. This technique produced correct decisions with respect to the QoT threshold in roughly 98% of the cases tested in Jiménez et al. [JADD13], while running in tens of milliseconds.

The strategy used to determine QoT ultimately depends on the setup time requirements, the complexity of the underlying system, and the desired accuracy.

8.7.1 Regeneration in a PCE-Based Implementation

We first consider selecting regeneration sites for a connection in the single-PCE centralized architecture of Sect. 8.3. It is assumed that the PCE would be provided with the impairment information on each of the links; e.g., OSNR, dispersion, PMD, and with the system rules detailing how impairments should be taken into account. Then, assuming that the PCE has full knowledge of each of the wavelengths that have been allocated already on a link, and full knowledge regarding the usage of transponders (or regenerator cards) at each node, the PCE should be able to calculate the proper sites at which to regenerate a connection.

The result of the regeneration calculations would be sent back to the source node (i.e., the PCC), along with the calculated route and wavelength assignments. The source node would then include this information in the signaling message that

actually establishes the new path (extensions to RSVP-TE are needed to encode the regeneration information). For example, the source node would need to issue an order that an intermediate node interconnect transponders *A* and *B* (e.g., using an edge cross connect) and configure its ROADM such that the incoming path is received on transponder *A* on wavelength *i* and the outgoing path is transmitted on transponder *B* on wavelength *j*.

Furthermore, programmable transponders, which will be discussed in Sect. 9.9, allow properties such as modulation format and *forward error correction* (FEC) to be set via software for each transponder. If such transponders are present in the network, then the setup message would need to specify these parameters as well.

In a *multi-PCE* architecture, the possibility of out-of-sync information regarding the wavelengths and transponders poses at least two problems. First, a PCE may calculate that a regeneration occur at a particular node along the path, using a particular pair of transponders. However, another PCE, handling a simultaneous request, may have already allocated these transponders, thereby resulting in contention. Resource contention can also occur with regard to assigning wavelengths, as described previously.

The more serious problem with regard to stale wavelength-state information, however, is that it potentially results in erroneous regeneration site selection, resulting in a path that is infeasible or that does not have enough system margin. The root-cause of this particular problem is that the performance of a connection may depend on the properties of the connections that are co-propagating on adjacent, or nearby, wavelengths. As discussed in Chap. 4, carriers typically implement conservative system rules such that provisioning a new connection does not cause any previously established connections to become infeasible. This allows new connections to be added without having to analyze the neighboring connections. While this is an expedient approach, there are some scenarios that make it difficult to implement. For example, as described in Sect. 4.2.6 with respect to mixing line rates on a single fiber, the presence of a 10G OOK wavelength can be detrimental to the performance of an adjacent 40G DP-QPSK wavelength, due to cross-phase modulation. If one PCE were to calculate the regeneration sites for a new 40G connection without knowledge that another PCE was simultaneously establishing a new 10G connection on an adjacent wavelength, the 40G connection may not have satisfactory performance. (Even if the PCEs implement a "soft" partitioning scheme to minimize the likelihood of adjacent 10G and 40G connections on a fiber, the situation is still likely to arise periodically.) To avoid this problem, the PCEs could always assume a worst-case mixed line-rate scenario; however, this would be too extreme of a measure that would result in excessive regeneration. Alternatively, after a short delay to receive updated status messages from other PCEs, a PCE could verify the QoT of a new connection.

8.7.2 Regeneration in a GMPLS-Based Implementation

Regeneration could also be determined as part of the distributed signaling protocol, e.g., GMPLS [SGCA09]. The *Path* message would collect information on the state

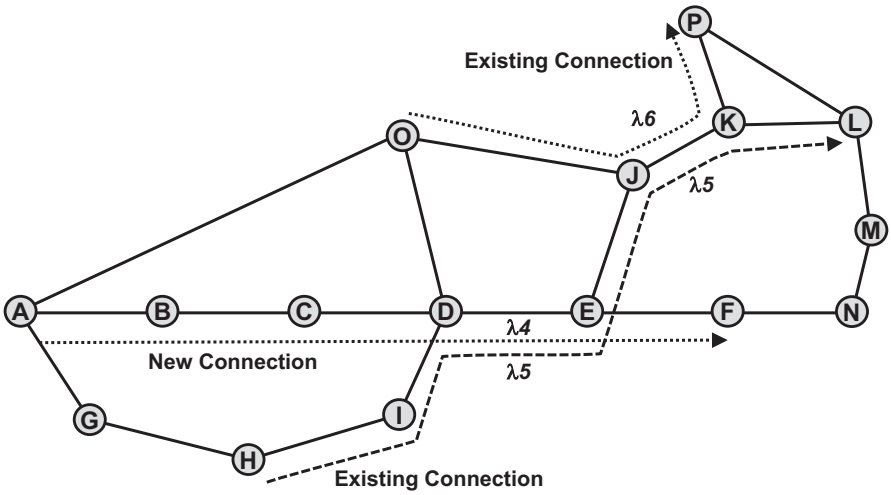


Fig. 8.4 Assume that aggressive system rules have been adopted, such that whether or not a connection meets the QoT threshold depends on the co-propagating connections assigned to nearby wavelengths. If a *new connection* is established from *A* to *F,* using λ_4 on Link *DE,* then the *existing connection* from *H* to *L,* using λ_5 on Link *DE,* is potentially affected. To determine if the *HL* connection still meets its desired QoT would require knowledge of the connection from *O* to *P* (due to the *OP* connection using an adjacent wavelength on Link *JK*)

of the transponders at each traversed node and the state of the wavelengths on each traversed link. Metrics such as link OSNR change relatively slowly, such that they would not need to be tracked in the *Path* message. (Alternatively, the wavelength- and transponder-state information could be flooded in the network using OSPF-TE [MCMT10]. However, the required flooding would consume significant signaling resources if the network is highly dynamic [SkNe09].)

As noted in Sect. 8.5, it is recommended that all regeneration decisions, and hence all wavelength-assignment decisions, be made by the destination node. This would likely produce more optimal results as compared to regeneration decisions being made on a node-by-node basis as the *Path* message propagates. For example, consider the setup of path A-B-C-D-E-F in Fig. 8.4. Assume that one regeneration is required along this path, and that it can occur at Node B, C, or D. If regeneration calculations are performed node-by-node, then the "furthest" node would typically be selected as the regeneration site, i.e., Node D in this example. However, Node D may have few available transponders, making it a poor choice for regeneration. Or, it may have no free transponders, making it an invalid choice for regeneration. Nodes B and C would not have knowledge of this, unless this detailed resource information were to be disseminated network-wide. (Alternatively, if an initial set-up attempt fails, the error message may include the reason for the failure, thereby allowing a better design choice to be made if reattempts are permitted [SGCA09].) In contrast, the *Path* message will have collected full resource information when it arrives at the destination, allowing the destination to make more strategic decisions. It is also faster to have optical-reach calculations performed just once in the setup

process, instead of performing the calculation at each node in the path. Furthermore, with programmable transponders (see Sect. 9.9), the destination can make more optimal decisions regarding factors such as the modulation format and FEC format to use (both of which affect the optical reach). (The simulations in Sambo et al. [SGCA09] also showed that making all regeneration decisions at the destination node, as opposed to on a node-by-node basis, delivered the best performance.)

When selecting the regeneration sites for a connection, *conservative* system rules should be adopted with regard to adjacent connections (with the exception of the mixed line-rate scenario, which is addressed below). Otherwise, the setup process would not only need to calculate regeneration sites for the new connection, but would also need to verify that the new connection does not cause any existing connections assigned to a nearby wavelength to fall below an acceptable QoT threshold. This could be a time-consuming process, depending on the complexity of the analytic QoT model and on the amount of information stored at the nodes, as the following example demonstrates.

Consider Fig. 8.4 again, and assume that *aggressive* system rules have been implemented, such that whether or not a connection meets its QoT threshold is dependent on the state of its neighboring wavelengths. The new path being established, from Node A to Node F, is A-B-C-D-E-F. Assume that the wavelength selected on link DE is λ_4. Also, assume that there is an existing connection, from Node H to Node L, routed on link DE, which has been assigned to λ_5. Assume that Node F is responsible for verifying that the addition of the AF connection will not invalidate any existing connections.

To perform this QoT calculation, Node F would need to have the full routing details of the HL connection, plus the information regarding any neighboring wavelengths of the HL connection (e.g., it would need to know that a connection has been assigned to λ_6 on Link JK). Essentially, information regarding each wavelength on each link in the network would need to be disseminated to all nodes; as noted above, this could lead to a potentially large signaling burden. An alternative is for Node F to communicate with Node L, and request that Node L perform the QoT verification for the HL connection [Azod11]. The communication between Nodes F and L would add to the delay and complexity of the setup process.

Thus, to reiterate, *conservative* rules regarding the presence of neighboring connections should be implemented *when possible*.

If the system is such that the QoT of a connection may be *significantly* affected by the existing neighboring connections (e.g., the mixed 10G/40G line-rate scenario), then more information would need to be included in the *Path* message. For example, for each wavelength in the Label Set, the *Path* message would need to track the bit rate and modulation format of any nearby assigned wavelengths. (Each node would be required to store this information for any path routed on any of its outgoing links.) Note that the distributed scheme is susceptible to stale information, just as the multi-PCE architecture is. The same scenario as described earlier, where simultaneous setup requests result in a 40G and a 10G connection being assigned to adjacent wavelengths, can occur. The nodes would need to check for this type of conflict during the reservation phase; the connection whose *Resv* message arrives later would be blocked. To minimize this form of backward blocking, a secondary

wavelength (or even a tertiary wavelength) should be reserved by the *Resv* message, as proposed in the 3WHS protocol (refer back to Sect. 8.4.4).

8.8 Multi-Domain Dynamic Networking

To this point, the dynamic network operations described in this chapter have assumed a single-domain network, where it is possible for the PCEs to gather detailed knowledge of the state of the whole network or where probe messages can collect complete resource information along an end-to-end path. (For our purposes here, a domain can be defined as "a collection of network elements within a common sphere of address management or path computational responsibility" [VZBL09].) The ability to fully share information enhances the quality of any routing and resource allocation decisions. Multi-domain environments, which typically limit the information shared between domains, are more challenging. One can consider both single-carrier multi-domain networks, as well as multiple-carrier multi-domain networks. In the former, a carrier chooses to partition its network into multiple domains to demarcate the boundaries between administrative entities or different equipment vendors, or to limit the size of the routing area in which state information must be flooded. In multiple-carrier multi-domain networks, end-to-end paths may need to traverse the networks of more than one carrier; for example, multiple Internet service providers. The challenge is for a domain to advertise enough information to allow for (close to) optimal routing decisions to be made, without exposing proprietary information, creating security vulnerabilities, or causing scalability problems.

A PCE-based approach is well suited for multi-domain routing; indeed, numerous PCE-based solutions have been proposed, as summarized in Paolucci et al. [PCGS13]. Typically, each domain has one or more PCEs, with each PCE having detailed knowledge of just that one domain, including any inter-domain links that extend from the domain. Communication among the PCEs is permitted, but the information that can be exchanged is limited, especially in a multiple-carrier environment. The PCEs need to agree on how various traffic engineering metrics should be interpreted (e.g., distance, load, reliability) so that there is a consistency in calculating each segment of an end-to-end path.

The multi-domain network shown in Fig. 8.5 is used here for illustration purposes. It includes five domains, labeled A through E. The shaded nodes represent *boundary nodes* (BNs), which lie on either end of an *inter-domain* link (e.g., Nodes A1 and D1) or which lie at the boundary of two domains (e.g., Node C4). All other nodes and links are considered "interior" to a domain. One PCE per domain is shown. Assume that a demand request arrives at the source node shown in Domain A; the request is forwarded to that domain's PCE. Based on the destination node address, the PCE determines that the destination lies in Domain C, such that a multi-domain route is required. (If addressing is not sufficient to determine the destination domain, then PCE_A may need to query the other PCEs.)

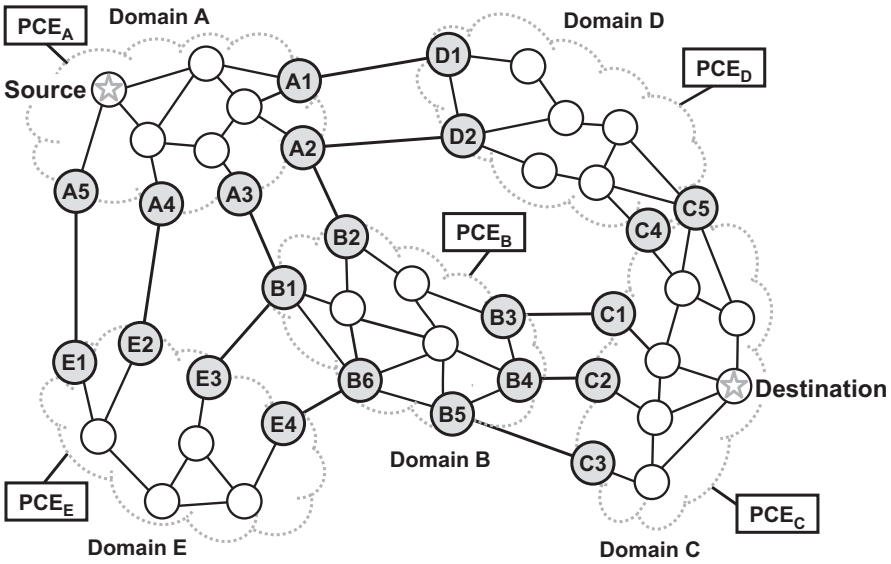


Fig. 8.5 A multi-domain network composed of five domains, *A* through *E*. One Path Computation Element (*PCE*) is assigned per domain. The boundary nodes are *shaded.* The *source* node is located in *Domain A,* and the *destination* node in *Domain C*

There are two major components of multi-domain routing. The first is selecting the sequence of domains that should be traversed from source node to destination node; the second is determining the actual path through each of the domains. Using the example of Fig. 8.5, we outline two particular multi-domain routing strategies that have been introduced in the IETF: Backward-Recursive PCE-Based Computation (BRPC) [VZBL09] and Hierarchical PCEs [KiFa11]. This is followed by a discussion of the multi-domain connection setup process in Sect. 8.8.3.

8.8.1 Backward-Recursive PCE-Based Computation

In BRPC, the sequence of domains to follow between the source node and destination node is assumed to be known, e.g., via administrative configuration. Alternatively, route calculations can be performed over a number of domain sequences, with the best result taken as the final route. Assume that in the example of Fig. 8.5, the domain sequence to be used by BRPC is A-B-C. The demand request will be forwarded from PCE_A to PCE_B to PCE_C.

PCE_C initiates the routing process by calculating a tree from its set of BNs that border Domain B (i.e., C1, C2, and C3) to the destination node. The resulting tree is shown by the dotted lines extending from the destination node in Fig. 8.6. PCE_C has full knowledge of Domain C, and thus is able to calculate the optimal path for each of the three branches of this tree, where "optimal" is with respect to whatever

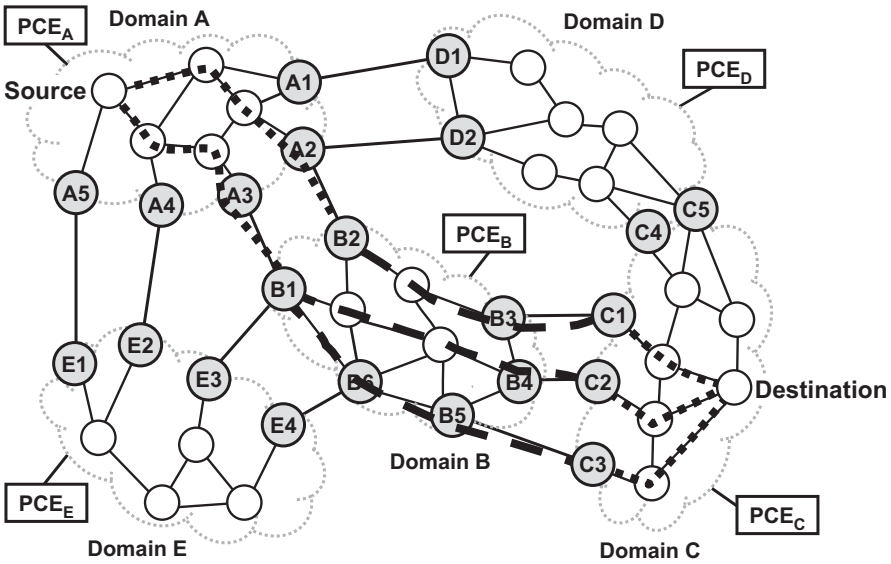


Fig. 8.6 The *virtual shortest path tree* created in the BRPC scheme is shown on the multi-domain network of Fig. 8.5, assuming a domain sequence of A-B-C. Each domain calculates the minimum-cost paths from each of its entry BNs to the destination node, using the cost information provided by the downstream domain and the detailed knowledge of its own domain.

metric was specified in the demand request. The tree that is calculated is referred to as the *Virtual Shortest Path Tree* (VSPT).

PCE_C calculates the total cost of each branch of the tree and passes this information to PCE_B. In a single-carrier environment, PCE_C may optionally send the path details as well. Alternatively, PCE_C stores the path details for each branch, assigns the path an ID number, and simply passes the ID number to PCE_B [BrVF09]. PCE_B proceeds to calculate the optimal paths from its BNs that border Domain A (i.e., B1 and B2) to the destination node, using the VSPT cost information provided by PCE_C, combined with its own detailed knowledge of Domain B (which includes knowledge of the inter-domain links). The resulting extensions of the tree branches are shown by the dashed lines in Fig. 8.6. Thus, the tree branches now extend from B1 and B2 to the destination node. PCE_B forwards the total cost of each branch to PCE_A. (It is necessary to forward the information of only the lower cost of the two paths that extend from B1 to the destination node.)

PCE_A proceeds to calculate an optimal tree from the source node to its two BNs that border Domain B (i.e., A2 and A3), using its detailed knowledge of Domain A. This is shown in Fig. 8.6 by the dotted lines extending from the source node. Combining the costs of these tree branches with the cost information forwarded by PCE_B, PCE_A can then determine the minimum-cost path from source node to destination node, given the domain sequence A-B-C.

Various experiments indicate that BRPC performs well with respect to resource allocation and setup time [PCGS13], although congestion can develop if the same domain sequence is always selected between a pair of source/destination domains.

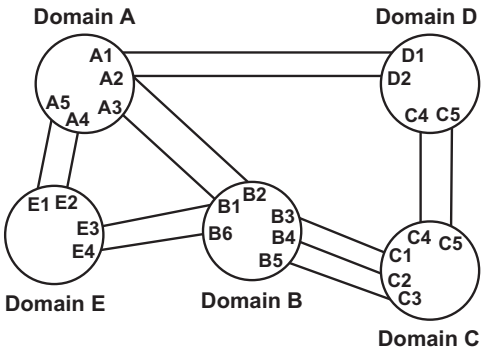


Fig. 8.7 The abstract topology of the network shown in Fig. 8.5, as seen by the parent PCE. Each of the nodes represents a domain, with the links being the inter-domain links. "Dummy" inter-domain links are added between *Domains C* and *D,* where Nodes *C4* and *C5* lie in both domains.

8.8.2 Hierarchical PCEs

One drawback to BRPC is that it does not include an explicit mechanism for determining the sequence of domains to be followed from source node to destination node. In contrast, the Hierarchical PCE scheme both determines the sequence of domains and finds a route from the source node to the destination node. In this architecture, each PCE associated with a domain is considered a child PCE. There is an additional PCE, known as the parent PCE (pPCE), which can communicate with each of the children PCEs. The pPCE has knowledge of how the domains are interconnected, including knowledge of any inter-domain links, but does not have visibility into any of the domains themselves. (The information regarding inter-domain connectivity can be administratively configured or it can be provided by the children PCEs.) From the point of view of the pPCE, the network topology of Fig. 8.5 is abstracted to that shown in Fig. 8.7. Note that links with a metric of 0 need to be added between Domains C and D. (Nodes C4 and C5 lie in both Domains C and D; there is no true inter-domain link. This configuration typically occurs only when Domains C and D are owned by the same carrier.)

When an inter-domain route is requested, the request is forwarded to the pPCE. The pPCE runs a routing algorithm on its abstracted topology to determine several candidate domain sequences. Using the same example as earlier, it may calculate three candidate sequences: A-B-C (via inter-domain links A2-B2 and B4-C2), A-B-C (via inter-domain links A3-B1 and B4-C2), and A-D-C (via inter-domain link A1-D1 and dummy link "C4-C4"). The pPCE generally selects the candidate sequences based on the traffic engineering properties of the inter-domain links; e.g., link load. The pPCE then requests that each of the domains that appear in a candidate sequence provide cost information for their portion of the desired path. For example, for the first domain sequence, the pPCE requests cost information from: PCE_A, for the path from the source node to A2; PCE_B, for the path from B2 to B4; and PCE_C, for the path between C2 and the destination node. After collecting all of the cost information, and adding in the costs of the inter-domain links, the pPCE determines the solution that produces the least-cost path. It is expected that the children PCEs do not provide the path details, just the costs (as noted earlier, in a single-carrier environment, the path details may be provided; otherwise just path IDs are provided).

Note that the path produced by this method is not necessarily optimal, due to the pPCE relying on only its knowledge of the *inter-domain* links to select the candidate domain sequences. First, the set of candidate domain sequences that are selected may not contain the optimal sequence. Second, the inter-domain links that the pPCE selects to use for a particular domain sequence may not be optimal. For example, selecting B5-C3 as the link between Domains B and C, as opposed to B4-C2, may produce a lower cost solution. This latter limitation could be remedied by having the pPCE request cost information for all possible paths between the entry and exit BNs of a domain; however, this may not be scalable, depending on the number of BNs (see Exercise 8.11).

Note that developing better methods for selecting the domain sequence to use in multi-domain routing is an area of active research.

8.8.3 Establishing Multi-Domain Connections

The previous two sections outlined two schemes for determining a route from a source node to a destination node that lie in different domains. The next step in the provisioning process is signaling this route to the relevant network elements so that the path can actually be established. As with single domain routing, this can be performed with either centralized or distributed mechanisms.

One option is centralized path setup within a domain, where selection of resources is controlled by the domain's PCE. As outlined in the two schemes discussed above, each PCE has knowledge of the path details for the portion of the end-to-end path that passes through its domain. After the end-to-end route calculation has been successfully completed, the relevant PCEs would be notified that the provisioning process should proceed. For example, in BRPC, PCE_A could signal PCE_B, which could then signal PCE_C. The signaling message would specify the IDs of the paths that should be established in each domain. With Hierarchical PCEs, the pPCE could signal all relevant PCEs in parallel to commence provisioning; again, the path IDs would be specified in the signaling message. Each PCE would proceed to establish its portion of the end-to-end path, just as it would in a single-domain network. Additionally, the resources on each inter-domain link of the path need to be assigned. It is assumed that one of the PCEs associated with the two interconnected domains would be responsible for all resource decisions on the link.

Alternatively, the end-to-end path can be provisioned using a distributed approach within each domain. RSVP-TE signaling can be used to establish the path segment that has been calculated by a domain's PCE. In the source domain, this signaling process is initiated by the source node. In all other domains, the signaling is initiated by the entry BN. The RSVP-TE signaling could be extended to the inter-domain link, or resource allocation on the inter-domain link may be performed by one of the PCEs.

The various provisioned segments would be stitched together to form an end-to-end path. Note that it is unlikely that all-optical segments would extend across domain boundaries, especially if the domains represent multiple vendors or multiple carriers.

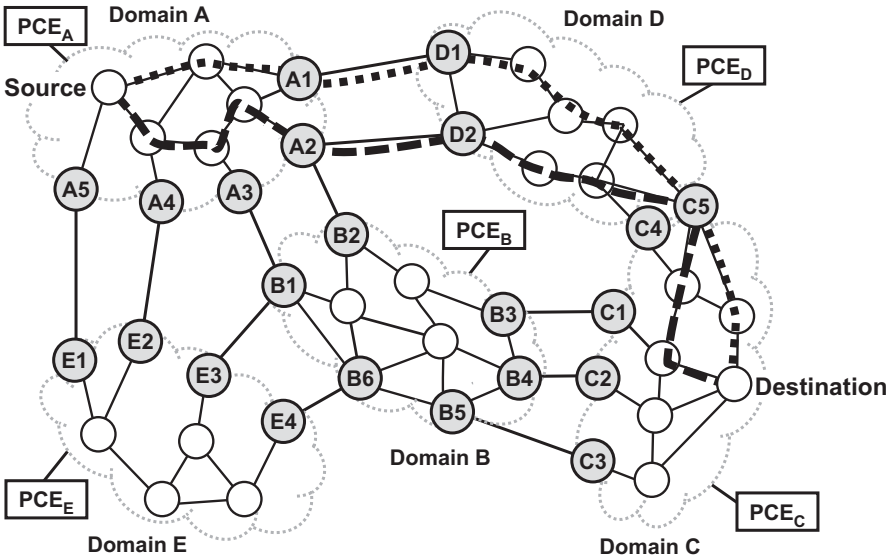


Fig. 8.8 Link-diverse paths from *source* to *destination.* 1-to-2 diverse routing is used in *Domain A,* 2-to-1 diverse routing is used in *Domain D,* and standard 1-to-1 diverse routing is used in *Domain C*

Additionally, note that meeting stringent setup time requirements in a multi-domain environment may be difficult, due to the inter-domain communication that is involved.

8.8.4 Protected Multi-Domain Connections

It may be desirable to establish *diverse* paths between a source node and a destination node that lie in different domains. There are various levels of diversity that can be enforced: link, node, and domain (also SRLG diversity, however, we do not specifically address that option here). For illustration purposes, we consider the Hierarchical PCE scheme, as applied to Fig. 8.5; however, similar techniques can be applied to other multi-domain routing schemes, including BRPC.

First, consider link diversity, where common nodes and domains are permitted. Assume that the pPCE has calculated A-D-C as one of the candidate domain sequences, as indicated in Fig. 8.8. In this scenario, the pPCE would request cost information from PCE_A for link-diverse paths from the source node to both A1 and A2. This is an example of where "1-to-2" diverse routing, with one source and two destinations, is required (algorithms for this scenario were covered in Sect. 3.7.3). Given that node diversity is not required, the pPCE may choose to use only boundary node C5 to transit from Domain D to Domain C. If so, it would request cost information from PCE_D for link-diverse paths from D1 and

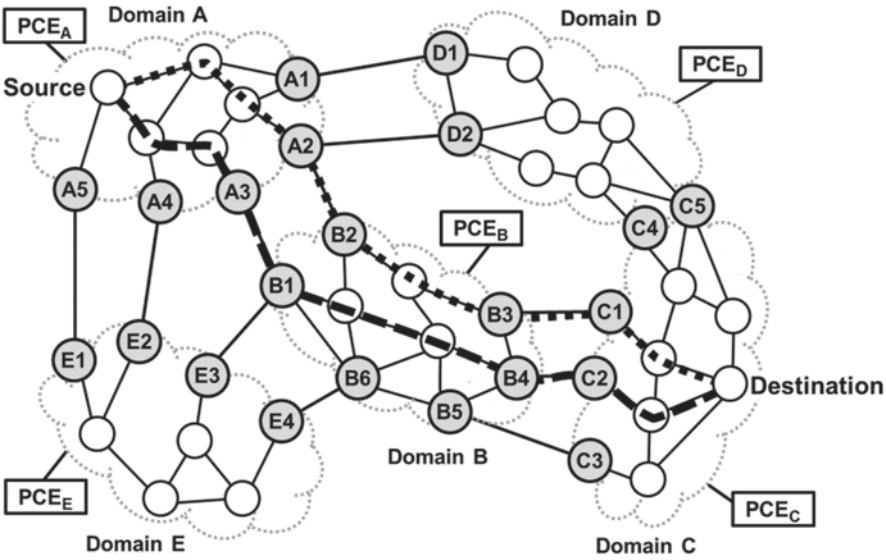


Fig. 8.9 Node-diverse paths from *source* to *destination.* 1-to-2 diverse routing is used in *Domain A,* 2-to-2 diverse routing is used in *Domain B,* and 2-to-1 diverse routing is used in *Domain C*

D2 to C5. This is an example of "2-to-1" diverse routing. Finally, the pPCE would request cost information from PCE_C for link-diverse paths between C5 and the destination node.

Next, assume that node diversity is also required. This is illustrated in Fig. 8.9, using the domain sequence A-B-C. The pPCE requests the cost information from PCE_A for link-and-node-diverse paths from the source node to A2 and A3. Assuming that the pPCE has selected B3 and B4 as the exiting BNs for Domain B, it would request cost information from PCE_B for link-and-node-diverse paths from B1 and B2 to B3 and B4. This is an example of "2-to-2" diverse routing, also covered in Sect. 3.7.3. (It may be desirable to explore additional combinations of exiting BNs. Thus, the pPCE could also request the costing of diverse paths from B1 and B2 to, for example, B4 and B5.) Finally, the pPCE would request information for link-and-node-diverse paths between C1 and C2 and the destination node.

In the third scenario, domain diversity is also required (clearly, the source and destination domains will be the same for the two paths). Thus, when the pPCE performs its initial routing calculation on the abstract topology shown in Fig. 8.7, it looks for "node"-diverse paths between Domains A and C, where the nodes represent domains in the abstract topology. Assume that the pPCE selects diverse domain sequences A-B-C and A-D-C, with inter-domain links A3-B1, B5-C3, and A2-D2, and inter-domain node C4; see Fig. 8.10. 1-to-2 diverse routing would be utilized in Domain A, 2-to-1 diverse routing would be utilized in Domain C, and unprotected routing would be used in the transit domains, B and D.

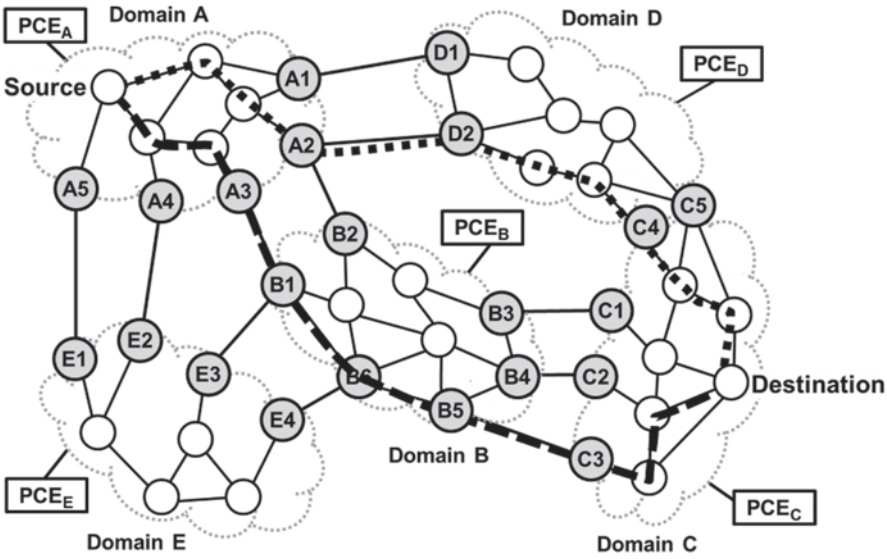


Fig. 8.10 Domain-diverse paths from *source* to *destination*. 1-to-2 diverse routing is used in *Domain A,* 2-to-1 diverse routing is used in *Domain C,* and unprotected routing is used in *Domains B* and *D*

8.9 Pre-deployment of Equipment

The previous sections covered routing, regeneration, and wavelength assignment in various dynamic settings. Another essential aspect of dynamic networking is pre-deploying equipment, where equipment is placed in the network in anticipation of future demands. To satisfy connection setup times on the order of seconds, or less, any equipment required by a connection must already be installed. Selecting how much equipment to pre-deploy, as well as where to place the equipment, is strategically very important. Pre-deploying too little equipment leads to suboptimal routing (e.g., a poor route may be selected if that is the only one with the required equipment) or excessive blocking of demand requests. Pre-deploying too much equipment is an unnecessary expense.

Pre-deployed equipment in large part refers to the transponders (or regenerator cards). Clearly, the number of transponders to install depends on the underlying transport system; e.g., one would expect to require more transponders in an optical–electrical–optical (O-E-O) network than in an optical-bypass-enabled network. Furthermore, the degree of configurability provided by a particular network element mandates different levels of accuracy in the estimation process [GeRa04]. For example, in an O-E-O architecture, one must calculate the number of transponders to pre-deploy on each optical terminal (i.e., a per-link estimate). Similarly, in an optical-bypass-enabled architecture with *non-directionless* ROADMs (and no other means of edge configurability), where the add/drop ports are tied to a particular

network link, a per-link estimation of required transponders is required. However, with *directionless* ROADMs, where the add/drop ports can access any network link, it is necessary to estimate only the total number of transponders to pre-deploy at each node (although contention on the add/drop ports does need to be considered). A per-node estimate is likely to be more accurate than multiple per-link estimates. (Exercise 2.10, in Chap. 2, explored the benefits of a directionless ROADM with respect to equipment pre-deployment.)

Several strategies exist for assessing the amount of equipment to pre-deploy. A common strategy is to run simulations based on traffic forecasts, where the simulation results can be used to estimate the amount of equipment that should be pre-deployed at each node in order to reduce the blocking probability below a given threshold. If the desired blocking probability (due to transponder unavailability) is very low, this method may require extensive simulations.

One particular simulation study took this method a step further, and fit curves to model the required transponder pool size distribution [SkWi10, CCCD12, SCGN12]. The simulations were run on a 100-node global network; the continental US portion of the network corresponds to Reference Network 1. All nodes were equipped with directionless ROADMs. All traffic was at the wavelength level, with a relatively high level of dynamism. Only 40 of the 100 nodes generated traffic; an additional 13 of the nodes needed to be equipped with transponders for regeneration purposes. (The remaining 47 nodes of the network were assumed to generate subrate services only; these services were not included in the simulations.) Only the transponders required for the working traffic, whether at a demand endpoint or at a regeneration site, were tracked. Initially, simulations were run in which the nodes had an unlimited transponder pool. Every 30 min of "network time," the number of transponders actually in use at a node was recorded, with roughly 2,500 sample points taken during the simulations. The generated transponder-usage histograms at each of the 53 nodes were found to closely follow a chi-squared distribution. (The chi-squared distribution is a family of curves, characterized by one parameter, referred to as the *degree of freedom*.)

For the 40 nodes that generated traffic, the histogram data was best modeled by a chi-squared distribution with 1–15 degrees of freedom; for the 13 nodes used for regeneration only, a chi-squared distribution with one degree of freedom best fit the data. An example of one of the histograms for a traffic-generating node is shown in Fig. 8.11. For this particular node, the best-fit chi-squared curve has five degrees of freedom. Given the mean, standard deviation, and degrees of freedom of a particular nodal histogram, one can determine the number of required transponders to deploy at the node to meet a particular blocking probability (i.e., the blocking probability due to no available transponders at a node; [SCGN12]).

Thus, while this methodology requires a simulation to generate the histograms, once these are produced, it is straightforward to size the pre-deployed transponder pool based on the target transponder blocking probability. (Note that deploying enough transponders to eliminate *all* blocking is not a good strategy. For example, in the simulation described above, about 30% more transponders are needed to achieve no blocking as opposed to roughly 10^{-4} blocking.) More research is needed

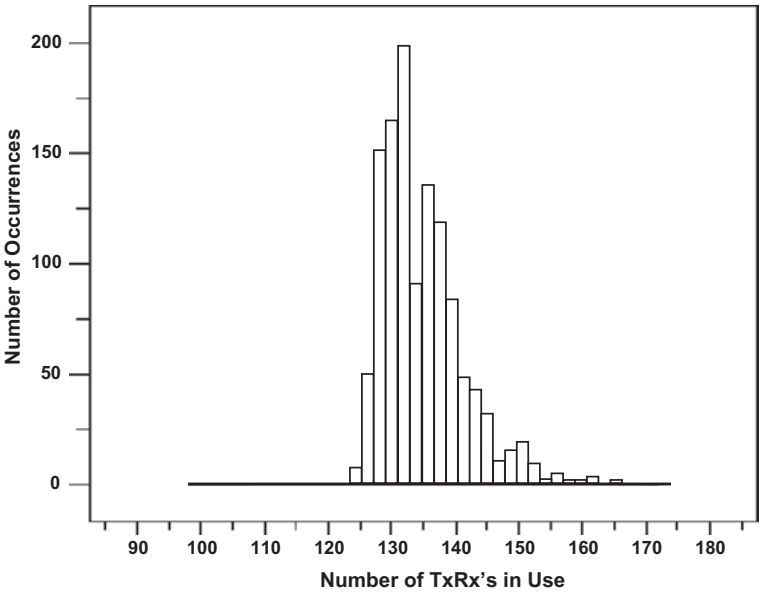


Fig. 8.11 An example of a histogram of the number of required transponders (*TxRxs*) at given points in time for a traffic-generating node in the simulations of Skoog and Wilson [SkWi10]. This particular histogram correlates well with a chi-squared distribution with five degrees of freedom. (© 2010 OSA)

to determine how general these results are with respect to network size, topology, traffic, and level of dynamism.

Alternatively, one can use queuing theory to estimate how much equipment to pre-deploy, where each source/destination demand pair is associated with requiring equipment at particular sites in the network. This assumes that one path is used for a given source/destination pair, so that the associated regeneration points are known; i.e., each arrival of a particular source/destination demand requires a transponder at the two endpoints and two transponders at each regeneration site along the path. (If alternative-path routing is used instead, where, at the time of the demand request, a path is selected from the pre-calculated candidate path set, then the probability that a particular candidate path will be selected would need to be estimated.) The arrival and departure processes of the demands can be modeled to estimate the required equipment at each node to reduce the blocking probability below the desired threshold (e.g., Mokhtar et al. [MoBB04]).

Another strategy for optical-bypass-enabled networks, proposed in Barakat and Leon-Garcia [BaLe02], involves estimating for each node the probability that any new demand will require regeneration at that node. The probabilities are determined based on the nodal position within the network and the lengths of the links feeding into the node, where being closer to the center of the network and being an endpoint of a long link increases the likelihood of regeneration at a node. This analysis can be used to assist in determining the amount of regeneration equipment to pre-deploy at each node.

The pre-deployment strategy to use may depend on the level of detail in the traffic forecast. If the forecast is very specific, then running a simulation is probably the most accurate strategy to determine where to pre-deploy equipment. If there are only approximate models of the demand arrivals and departures, then queuing analysis can be used. If only a forecast of the total number of demands in the network is available, and not the specific source/destination pairs, then the method based on nodal regeneration probabilities can be used.

A different perspective on the equipment pre-deployment problem was considered in Woodward et al. [WFKP12]. This work assumed that connections are added to the network at random times, and never removed (i.e., the network continues to grow in size). It was required that the equipment for any new connection already be installed, to accelerate the provisioning process. It was assumed that wavelength conversion was *not* permitted when a connection was regenerated. This implied that the two transponders used for a regeneration must lie on different add/drop ports of the ROADM (it was assumed that wavelength contention can occur on the add/drop ports). The various pre-deployment strategies that were considered focused on deploying enough transponders across the add/drop ports relative to the number of contiguous wavelength paths that could pass through the node. Rather than pre-deploying numerous transponders up-front, the approach taken was to periodically install a small number of transponders, where the goal was to strike a balance between the number of required "truck-rolls" and the number of idle transponders. It is anticipated that this gradual deployment strategy is the approach a carrier would follow as a first step towards a dynamic network.

8.10 Scheduled or Advance Reservation Traffic

The discussion regarding dynamic traffic thus far has assumed that demand requests need to be served as soon as they are received; i.e., the demands require *immediate reservation* (IR). With scheduled, or *advance reservation* (AR), traffic, the demand request arrives in advance of when the connection is actually required. (The terms "scheduled" and "AR" are used interchangeably in this section.) The time between the request arrival and the desired start time of the connection is referred to as the *book-ahead time*. In addition to specifying the start time, scheduled traffic typically specifies the holding time as well. The advanced notification, combined with knowledge of the holding time, allows the network to more optimally allocate resources to scheduled demands. Of course, not all traffic can be scheduled, such that a network must be able to accommodate a mix of both AR and IR traffic.

Scheduled traffic has grown in importance in the optical layer [ZhMo02], especially with the advent of grid networks, where an array of computing and storage resources is shared among a community of geographically distributed users. Because of the large transfers of data that are often required, the wavelengths to establish the required connectivity have become an additional resource that needs to be scheduled, giving rise to the concept of a *lambda grid* [DDMN03, SFPF05, Take06, Batt07, ZENS08].

In much of the research regarding scheduled services, it is assumed that there is a fixed set of AR demands that needs to be accommodated, e.g., Kuri et al. [KPGD03]. The emphasis of much of this work is partitioning the set of AR demands into time-independent groups; resources can then be shared among the groups. Standard RWA techniques are often applied; e.g., alternative-path routing with a set of candidate paths (Sect. 3.5.2) combined with First-Fit wavelength assignment (Sect. 5.5.1). Here, we are more interested in the dynamic aspect, where requests for scheduled services continue to arrive over time. The goal of this section is to highlight some of the more important design decisions.

A comprehensive survey paper covering many aspects of scheduled traffic, including a number of variations with respect to the start- and end-time specifications, can be found in Charbonneau and Vokkarane [ChVo12]. One variant is where a time range, or "window," is specified for the connection start time, thereby providing more flexibility in scheduling the request [WLLF05, AYTM09]. This scenario arises with applications such as offsite backup; the actual start time is not critical, as long as the backup is completed each evening. Other types of AR demands require an exact start time. For example, with grid computing, the establishment of a wavelength path needs to coincide with the time that a user has been scheduled to use particular computing resources.

An important design decision is whether the scheduling function should be centralized or distributed. There are potentially three major aspects of the design process: selecting a start time (for those demands with flexible start time), selecting a route, and assigning wavelengths. The first two of these operations are tightly coupled because it is necessary to ensure that each link in the path will have the required resources available for the entire holding time of the AR demand. Furthermore, these processes are well suited to centralized operation, where batch scheduling can be performed to improve optimality (assuming the AR requests do not require an immediate response as to whether they have been accepted; [BoSt04]), or where the reserved resources can be re-optimized for scheduled connections that are not yet in service [SYTR07]. Scheduling is more challenging in a distributed environment because it requires knowledge of the resources that have already been reserved to ensure that enough resources will be available to accommodate a new AR demand. This would necessitate flooding information regarding all accepted scheduled demands to each of the nodes. This would place an additional burden on the signaling channel and on each node, without providing significant tangible benefits. (One of the main benefits of distributed implementation is minimization of delay. However, latency should not be a critical factor in responding to an AR request.) Considering all of these factors, the processes of selecting a start time and selecting a route are preferably handled in a centralized approach, e.g., a PCE-based architecture. If the arrival rate of scheduled demand requests is very high, such that the computational burden is too high for a single PCE, then multiple PCEs can be utilized. Synchronization delays among the PCEs should be relatively small compared to the book-ahead time, such that conflicting reservations can be minimized.

The timing with regard to wavelength assignment is a separate design decision; i.e., the wavelength(s) to use can be selected at the time the AR request is accepted, or the assignment process can be postponed until the connection start time. Much

of the research on scheduled demands has assumed purely all-optical networks, where the same wavelength must be utilized end-to-end for a connection. With this assumption, wavelength assignment in advance is preferable, to better ensure the end-to-end wavelength continuity constraint can be met. This mode of operation is likely more compatible with centralized wavelength assignment; thus, the selection of the start time, route, and wavelength would all be performed centrally. (Issues such as the propagation delay with PCC-to-PCE or PCE-to-PCE communication should not be problematic; as indicated above, latency should not be critical in the AR acceptance process.)

In more general architectures, where regeneration can be performed to accomplish wavelength conversion, reserving specific wavelengths ahead of time is not crucial. Delaying assignment may provide more flexibility to the overall network design process, where both AR and IR demands need to be accommodated. It is sufficient to know that there will be an available wavelength on each link of the path, not necessarily *which* wavelength. If wavelengths are not assigned until the connection setup time, then this portion of the AR design process is similar to that of IR demands. To minimize blocking due to stale information, distributed wavelength assignment is preferred (as discussed in Sect. 8.5).

Thus, for networks where some wavelength conversion is possible, the following is an effective design strategy. Selection of the start time and routing are performed in a PCE, with the PCE informing the source node of the results. At the time of connection setup, the source node initiates RSVP-TE signaling to select the resources and establish the desired path.

One possible downside of delaying wavelength assignment is with regard to transponders (or regenerator cards). If regeneration sites, whether required due to optical reach or due to wavelength conversion, are not selected until wavelength assignment is performed, then it is not known in advance at which nodes the AR demand will utilize transponders. This precludes reserving transponders for the scheduled traffic, which may ultimately result in an AR demand being blocked at its connection start time. There are some mitigating factors, however. First, there is often a choice as to where regeneration can occur in a path; if there are no available transponders at one node, there may be some at a neighboring node. Second, enough transponders should be pre-deployed such that blocking is below a desired threshold. Finally, while not desirable, it is possible to "bump" a (low priority) IR demand in order to free up a transponder for an AR demand.

As this last point illustrates, there may be contention for resources among the AR and IR demands; this is analyzed in Greenberg et al. [GrSW99] and Triay et al. [TrCV13]. It is important to ensure that the bulk of the resources do not end up being reserved for the AR demands such that the IR demands are "starved." Various resource-assignment strategies and/or traffic admission-control policies can be used to reduce the contention between the two traffic types. One possibility is to partition the resources between the AR and IR demands; however, inflexible partitioning typically results in excess blocking. Another possibility is to assign wavelengths at one end of the spectrum for AR demands and from the other end for IR demands [ESCJ08]. This type of solution is especially beneficial in pure all-optical systems,

where the same wavelength must be assigned along the entire path. In more general systems, where the wavelength can change with regeneration, simulations on Reference Network 1 showed that limiting the number of wavelengths that can be reserved on each link for AR demands is an effective admission control policy [CCCD12]; i.e., at any given time, no more than S wavelengths on a link can be scheduled. The key is that it can be any S wavelengths, not a specific set of S wavelengths. This is similar to the admission-control policy proposed in Greenberg et al. [GrSW99].

Additionally, if the end times of the IR services are not known, then it is possible that so many IR demands are still in the system at a given time that there are not enough resources available for the demands that are scheduled to start at that time. If this occurs, either some of the IR demands are preempted or some of the AR demand reservations are "betrayed" (i.e., an accepted reservation cannot be honored), depending on the priorities of the demands. To reduce the probability of this scenario to an acceptable level, it may be necessary to implement a *blackout* period, B, on the IR demands [CCCD12]. For example, if R wavelengths on link L are required for scheduled services at time T, then during the time $[T–B, T]$ IR demand requests are blocked if they would result in there being fewer than R wavelengths available on link L. The length of the blackout period needs to be chosen to strike an appropriate balance between IR demand blocking during the blackout period and the number of IR preemptions/AR betrayals.

8.11 Software-Defined Networking

To wrap up this chapter on dynamic networking, we examine a relatively new paradigm known as SDN. The major concepts behind SDN are straightforward to state: The network control plane and data plane should be decoupled, and network control should be *logically* centralized [ONF12]. To better understand the implications of SDN, it is instructive to consider the antitheses of this model, namely the IP and Ethernet layers, where the control plane is both coupled to the data plane and distributed. In an IP network, the routers make all control decisions regarding traffic forwarding; e.g., each router runs a distributed routing protocol, such as OSPF, to populate its routing tables. Furthermore, each router operates autonomously, without network-wide coordination. Implementing new control policies typically requires interacting with each physical router. In addition to running the control software, the routers perform all packet forwarding (e.g., the routing table look-up). In some sense, both "the brains and the brawn" are encompassed in the routers. Coupling of the control and data planes exists in Ethernet as well. If an Ethernet switch receives a frame for which it does not recognize the destination address, it broadcasts the frame to the other switches. Based on the return traffic, it adds an entry in its switch table for this new address. Thus, again, each switch operates without centralized control, and is responsible for both populating the forwarding table *and* directing traffic based on the table.

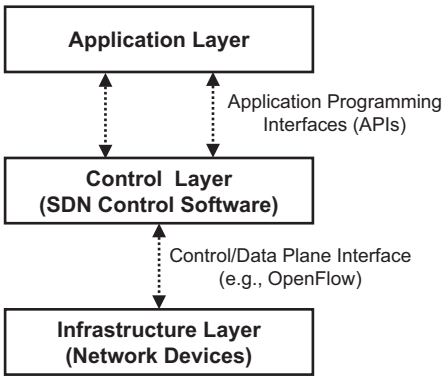


Fig. 8.12 Three-layer SDN architectural model

In the SDN model, the IP routers and Ethernet switches would be controlled via software that runs in a decoupled control plane. The SDN control software (i.e., the SDN "controller") is essentially a network-wide operating system that performs tasks such as topology discovery, routing, traffic engineering, and recovery. It is capable of populating the forwarding tables of the routers and switches, allowing it to exercise very fine granularity control of the network flow.

The SDN model can be conceptualized as a three-layered architectural stack, as shown in Fig. 8.12 [ONF12]. At the bottom is the infrastructure layer, composed of the network devices, e.g., the switches and routers. At the top are the network applications. The SDN control layer sits in the middle, responsible for presenting a vendor-independent interface and a network-wide view to the applications, and translating requests/requirements of the applications into instructions for the infrastructure. Clearly, buy-in from the equipment vendors is needed to support this vision, as the individual boxes must be able to act in accordance with the commands from the SDN controllers.

There are several potential benefits that can be realized by decoupling the control and data planes. First, it allows the control software to be more easily modified or customized, without having to adjust each individual router or switch or wait for a new software release from the vendor. This provides carriers and enterprises with greater control of their network, allowing them to more easily introduce new services and innovations. Network management should be simpler and more automated, with less chance for manual configuration errors. Another potential benefit of SDN, which has been debated, is that pulling the control functionality out of the IP router may significantly lower the router cost.

Furthermore, the SDN paradigm is amenable to *network virtualization,*[1] where a single physical network, i.e., the communication, computing, and storage resources,

[1] Network virtualization is different from, though related to, the concept of *network functions virtualization* (NFV). NFV is an initiative to instantiate networking functions as software applications or virtual machines running on commercial off-the-shelf (COTS) servers rather than employing an array of proprietary hardware. NFV is similar to SDN in that it represents a move away from proprietary solutions, and provides carriers and enterprises with greater control of their networks.

can be partitioned into multiple logically isolated networks (alternatively, multiple physical networks can be consolidated into one virtual network). Each virtual network, or network "slice," can be customized for, or even by, the end user, depending on their requirements for services and resources. This capability supports the infrastructure-as-a-service (IAAS) model of cloud computing [BGPV12]. The fine-granularity control afforded by SDN allows attributes such as the networking protocols and the virtual network topology to be tailored to the customer's needs.

Additionally, SDN is an enabler of dynamic networking. One of the biggest drivers for SDN is the need for flexible "QoS-on-demand" cloud-computing backbones. While the SDN concept originated out of enterprises wanting more control of the Ethernet and IP layers, SDN is envisioned as a unifying multi-layer control-plane architecture. The SDN controllers would be capable of provisioning across layers, vendors, and domains. In this vision, SDN would extend to the optical transport layer as well. Note that the optical layer already decouples the control and data planes. For example, in a PCE-based architecture, the PCE performs all of the routing calculations, and specifies how the network elements (e.g., ROADMs) should be configured. The ROADM itself is responsible only for directing wavelengths between the proper input and output ports.

Having a single control plane across network layers is advantageous from an operational and optimization perspective. For example, provisioning a new service would be seamless across the IP, Optical Transport Network (OTN), and optical layers, where the multi-layer network is abstracted to a single-layer, flat representation. The details of the underlying network layers are hidden from the network applications. If new bandwidth is required between two points, the SDN controller can automatically determine which layers should be involved; e.g., it can decide whether a particular IP router should be bypassed or whether grooming in an OTN switch is required. SDN proponents consider this global view as preferable to the GMPLS or ASON (overlay) model, where isolated instances of various control protocols are deployed in each layer and a combination of UNIs and E-NNIs are used to stitch together a cross-layer connection [DaPM12]. A single network-wide control plane could also enable efficient, coordinated multi-layer restoration.

Furthermore, it is envisioned that centralizing the control process will lead to greater network stability. For example, as noted in Sect. 8.2.3, one of the concerns with a dynamic optical layer is that it may disrupt IP adjacencies. Because of the delays inherent in propagating the new adjacency information, the potential for routing instabilities arises. By centralizing topology discovery and routing, SDN is designed to avoid such problems.

The challenge is in scaling centralized control across an entire network. Some of the issues are considered in Yeganeh et al. [YeTG13]. Note that the SDN control layer is *logically* centralized, but would likely be implemented in a distributed computing environment, for improved responsiveness and reliability. Thus, the scalability issues are not so much related to processing power or memory (though these may be challenging), but to the need to manage grossly different network-layer characteristics. For example, the IP layer involves tracking thousands of flows and addresses, performing table look-ups, managing queues, etc. In contrast,

provisioning in the optical layer involves functions such as wavelength assignment, impairment analysis, and managing optical amplifier transients. It is not clear if a single control plane can meet the disparate needs of the various layers without growing unwieldy. This is especially challenging with respect to the optical layer, where network operation is still very coupled to the particular equipment vendor. A unified control plane would require exposing the details of the transport layer (e.g., power levels, impact of nonlinear impairments). Some optical-layer equipment vendors, while expressing support for the SDN model, envision a more hybrid approach, where SDN exists in the higher layers (down to the edge of the optical network), but where GMPLS is used in the optical layer. Some early experimental work has taken this approach as well [ANEJ11, LiTM12]. A hybrid centralized/distributed model is also espoused by some network operators [GrBX13].

One protocol defined for the SDN control plane/data plane interface that has garnered a lot of attention is *OpenFlow*™ [MABP08].[2] We present a brief overview of this protocol to further illustrate the SDN concept.

8.11.1 OpenFlow

The OpenFlow protocol operates between the control and infrastructure layers of Fig. 8.12. OpenFlow allows direct access to the data plane (or forwarding plane) of the various network elements, thereby allowing the software controllers to "program" these elements. This enables the software to adjust traffic flow based on factors such as usage pattern, application, required resources, or business policy.

OpenFlow is much further along in its development in the IP and Ethernet layers than in the optical layer. We describe its operation with respect to IP (it operates similarly with Ethernet). For each IP flow, the OpenFlow controller can specify an action in the forwarding table that controls the processing of the packets in that flow. The actions can be programmed proactively, before the flow of packets begins, or reactively, after a packet is received at a router for which there is no corresponding table entry. For example, the controller can specify that the flow be forwarded to a particular output port of the router, that the packets of a particular flow be dropped due to congestion or security reasons, or that a particular field in all packets of a flow be modified. For greater scalability, OpenFlow supports "groups" of flows, where an action in the flow table can be associated with the entire group, for coarser granularity control. Additionally, it can optionally indicate that a flow should be handled via normal processing, to allow for gradual adoption of the SDN paradigm in the network.

Experiments with regard to OpenFlow in the optical layer are still in a relatively nascent stage [GDSP10, Liu13]. In some aspects, an OpenFlow-based control plane is similar to one based on a PCE. In discussing the drawbacks of the centralized PCE model in Sect. 8.3.3, the potential for excessive latency was noted, due to the

[2] OpenFlow is a registered trademark of the Open Networking Foundation.

need for a source node to communicate with a PCE that may not be geographically close. This type of effect was noted in the multi-layer OpenFlow field trial reported on in Liu et al. [Liu13]. In this trial, when the first packet of a new flow was received by the edge IP router, it was forwarded to the remotely located OpenFlow controller. The controller computed the end-to-end path, and issued instructions back to the IP router, as well as all of the network elements along the path, to configure their forwarding tables for the new connection. It was determined that the propagation delay in communicating with the controller was the key contributor to the path setup latency in the control plane. In an actual deployment, the OpenFlow controller is likely to be implemented in a distributed-computing fashion over several locations, which potentially reduces the propagation time. However, this introduces the problem of out-of-sync state information at the various locations due to the inherent delay of propagating updates across a network (similar to deploying multiple PCEs). Thus, as discussed for the PCE model, meeting the most stringent time requirements for connection setup (e.g., 100 ms) may be very challenging in the SDN/OpenFlow model. Furthermore, propagation delays also could limit the ability to use OpenFlow to orchestrate restoration from a failure. Such delays were noted in the restoration experiments of Liu et al. [Liu13], where link failures were reported to the controller, which in turn reprogrammed the forwarding tables of all network elements involved in rerouting the failed demands.

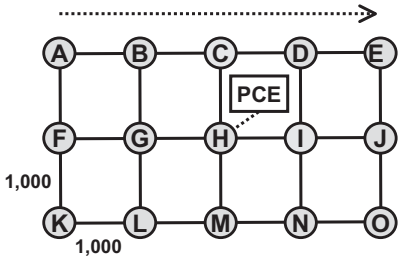
Whether OpenFlow, or more broadly SDN, takes hold in carrier networks depends largely on the use models that can take advantage of it, and their associated business case. Its fate may be tied to other new paradigms that are seen as major drivers, such as cloud computing and network virtualization, or possibly even more forward-looking concepts, such as cognitive networking [DeMi13], where the network continues to "learn" how certain conditions affect metrics such as resource usage and user quality-of-experience, and autonomously adjusts its behavior (e.g., routing decisions) accordingly.

8.12 Exercises

In the exercises below regarding transmission start-time calculations, consider only fiber propagation delays and switch configuration times (i.e., ignore processing delays). Take the speed of light in fiber to be 2×10^8 m/s. When a path is being set up, switches need to be configured at all intermediate nodes, as well as at the source and destination. If verification of path setup is not required, then the transmission start time at the source node is determined by the requirement that each switch in the path be configured by the time the initial transmission reaches it.

When an exercise specifies that verification of path setup is required, assume that the verification message is initiated at the destination node upon completion of its own switch configuration. The verification message is sent to the source node; the intermediate nodes in the path do not forward the message until their own switch is configured.

8.1 Consider the network shown below, with all link distances 1,000 km, and with one PCE located at Node H. Assume that the control plane uses the same topology as the data plane, and that shortest distance routing is used for both data-plane and control-plane communications. Assume that switches can be configured in 15 ms. A new demand request, from Node A to Node E, arrives at Node A, which then directs the request to the PCE. Assume that verification of path setup is *not* required before the source can begin transmission. (a) If the PCE can only communicate with the source node, how long does it take from the receipt of the demand request to the time the source can begin transmission? Assume that the setup message from the source to the other nodes in the path is pipelined. (b) Repeat part (a), except assume that the PCE is allowed to directly communicate a setup message to each of the nodes in the path. (c) Repeat parts (a) and (b), except assume that Node B requires 30 ms to configure its switch, instead of 15 ms.



8.2 Repeat Exercise 8.1, parts (a) and (b), except assume that a verification message must be received by the source node indicating that the path has been properly configured.

8.3 Repeat Exercise 8.1, parts (a) and (b), except assume that the grid topology shown in the figure holds only for the data plane. Assume that the control plane has a different topology such that the propagation delay between any two nodes or between a node and the PCE is 20% longer than in the data plane. In this example, which switch configuration scheme is more adversely affected by the extra delay in the control plane (i.e., the scheme of Exercise 8.1a or 8.1b)?

8.4 Consider the grid network shown in Exercise 8.1, and assume that a demand is requested from Node F to Node L. The shortest path from F to L can be established along links F-G-L or F-K-L. Assume that the switches at Nodes F and L can be configured in 15 ms, and the switches at Nodes G and K can be configured in 50 ms. Assume that a path setup verification message is *not* required before transmission can begin. (a) If the PCE can directly communicate the setup message to each of the nodes in the path, how much sooner can the source initiate transmission if the F-G-L path is selected instead of the F-K-L path? (b) If the switch configuration times were 15 ms at all nodes, does the selected path affect the transmission start time?

8.5 Consider a GMPLS-based implementation for establishing the AE demand shown in Exercise 8.1. Consider two scenarios, one where pipelining of the *Resv* message is not permitted, and one where it is permitted. Assume that a

path setup verification message is *not* required before transmission can begin. (a) If the switches at each node can be configured in 15 ms, how much sooner can the source begin transmission when pipelining is used? (b) If Node B requires 30 ms to configure its switch, instead of 15 ms, how much sooner can the source begin transmission when pipelining is used? (c) Express the transmission start time as a function of the switch configuration time at Node B for the non-pipelined and pipelined scenarios (take Time = 0 to be when the *Path* message is received by the destination).

8.6 Consider a GMPLS-based implementation for establishing the AE demand shown in Exercise 8.1, where the *Resv* message is pipelined. Assume that a path setup verification message must be received by the source node before it can start transmission. (a) If the switches at each node can be configured in 15 ms, by how much is the initial transmission time delayed due to the source node awaiting the verification message? (b) If Node B requires 30 ms to configure its switch, instead of 15 ms, by how much is the initial transmission time delayed due to the source node awaiting the verification message?

8.7 Consider a path from Amsterdam to Paris with an intermediate node of Brussels. Assume that the distance between Amsterdam and Brussels is 200 km, and the distance between Brussels and Paris is 300 km. (a) Compare the transmission start times for a PCE-based architecture and a distributed GMPLS-based architecture. Assume that in both scenarios: all switch times are 15 ms; a path setup verification message is *not* required before transmission can begin; and any signaling messages can be pipelined. For the PCE scenario, assume that the PCE is located in Athens, which is assumed to be at a distance of 2,500 km from each of the three nodes in the path. Assume that the PCE can send configuration messages directly to each node in the path. (b) Repeat part (a), except assume that a path setup verification message must be received by the source node before it can start transmission (in either architecture).

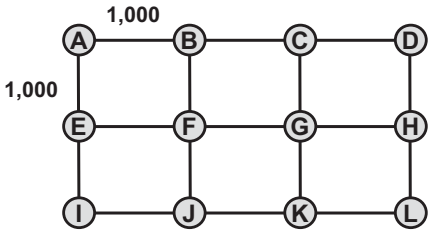
8.8 Consider a path A-B-C-D-E from Node A to Node E, where each link in the path is 1,000 km. Assume that a distributed GMPLS-based architecture is used to establish the connection. In a variation from typical operation, assume that the demand request is sent to *both* the source node and destination node (and assume that the request is received at approximately the same time at both nodes). Assume that *both* endpoint nodes can initiate path setup; i.e., both Node A and Node E send *Path* messages (e.g., to collect information on the free wavelengths on each link). The node where the *Path* messages meet (assumed to be Node C) is treated as a "destination node" for both *Path* messages. Node C sends *Resv* messages to both of the endpoints. (a) How much faster can transmission begin at Node A in this two-ended scheme, as compared to typical GMPLS operation where Node A sends the *Path* message and Node E sends the *Resv* message? Assume that all switch configuration times are 15 ms, no path setup verification message is required, and *Resv* messages are pipelined. (b) Repeat part (a), except assume that a path setup verification message (from Node E to Node A) is required. (c) For simplicity, it was assumed in parts (a) and (b) that the two *Path* messages arrive at the same time at Node C. Describe how a two-ended approach could work if the two *Path* messages are not received at the same time at one of the intermediate nodes.

8.9 Some switches are limited to performing a certain number of reconfigurations within a given time period. How might this limitation affect the dynamic connection setup process?

8.10 Assume that a network supports 20 Tb/s of dynamic traffic. Half of this traffic has an average holding time of 30 s; the other half has an average holding time of 5 min. All of this traffic is at the line rate, which is assumed to be 100 Gb/s. On average, each connection occupies five hops. If link state advertisements (LSAs) need to be broadcast every time a wavelength on a link is either assigned or released, how many LSAs are sent per second on average? (Assume that the traffic is bidirectional, with the same wavelengths used in both directions, such that one LSA is sent per bidirectional link. Assume that any blocking of the traffic is negligible.)

8.11 Assume that a new demand must be routed over a sequence of five domains. Assume that four links interconnect each pair of adjacent domains in the sequence. Assume that the Hierarchical PCE scheme is used for routing. (a) If an unprotected path is desired, how many possible paths would the parent PCE see in its high-level abstracted view of the domains? (b) How about if two link-diverse paths are desired?

8.12 Consider the 12-node optical-bypass-enabled grid network shown below. Assume that all links are 1,000 km in length, the optical reach is 3,000 km, and shortest path routing is used. Assume that demand requests arrive to the network according to a Poisson process of 60 Erlangs (all demand requests are bidirectional and at the line rate). Assume that 75 % of the traffic is between adjacent nodes; this traffic is split randomly among the adjacent node pairs, with any pair equally likely. Assume that 25 % of the traffic is between the corner nodes (A and L; D and I); this traffic is split randomly, with traffic between either node pair equally likely. (a) For the inter-corner traffic, assume that regeneration must occur in Node G, and that regenerator cards are used. How many transponders should be pre-deployed at each node, and how many regenerator cards should be pre-deployed at Node G, such that the probability of blocking (due to no available transponders or regenerators) for any demand pair is less than 10^{-4}? (b) Assume that regeneration for the inter-corner traffic is split randomly between Nodes F and G, with either site equally likely to be selected for regeneration. How does this change the result from part (a)? (c) If the goal is to minimize the number of pre-deployed regenerator cards, is it better to perform all regeneration at one node, or split the regeneration between two nodes?



8.13 Assume that time is partitioned into 1-min time slots. Consider scheduled demands that have flexible start and end times, where the triplet [s, e, d] is used to indicate the earliest acceptable starting time slot *s*, the latest acceptable ending time slot *e*, and the duration of the demand *d* (in units of time slots). Assume that the following five demands need to be scheduled on one wavelength: (1) [12, 15, 2], (2) [2, 30, 5], (3) [4, 16, 10], (4) [3, 19, 3], and (5) [21, 26, 4]. Assume that no more than one demand can be scheduled in a time slot. (a) Prepare a schedule for the demands to yield the earliest time at which all services have completed. What is the latest time slot occupied by any of the demands? (b) Repeat part (a), except assume that the demands can be served by noncontiguous time-slots, as long as the total service time equals the required duration of the demand.

8.14 *Research Suggestion:* Investigate why a chi-squared distribution may be suitable for modeling the histogram of the required number of transponders required over time at a node with dynamic traffic (Sect. 8.9). How general is this result? Can the chi-squared "degrees of freedom" parameter for a particular node be related to the node's position in the network (e.g., its *betweenness*), its level of traffic, its optical bypass level, etc.?

References

[ABGL01] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow, RSVP-TE: Extensions to RSVP for LSP Tunnels. (Internet Engineering Task Force, Request for Comments (RFC) 3209, Dec 2001)

[ACMW12] J. Ahmed, C. Cavdar, P. Monti, L. Wosinska, A dynamic bulk provisioning framework for concurrent optimization in PCE-based WDM networks. J. Lightwave Technol. **30**(14), 2229–2239, 15 Jul 2012

[AgYH10] F. Agraz, Y. Ye, J. Han, RSVP-TE extensions in support of impairment aware routing and wavelength assignment in wavelength switched optical networks (WSONs), draft-agraz-ccamp-wson-impairment-rsvp-00. (Internet Engineering Task Force, Work In Progress, Oct 2010)

[ANEJ11] S. Azodolmolky, R. Nejabati, E. Escalona, R. Jayakumar, N. Efstathiou, D. Simeonidou, Integrated OpenFlow–GMPLS control plane: An overlay model for software defined packet over optical networks, *Proceedings, European Conference on Optical Communication (ECOC'11)*, Paper Tu.5.K.5, Geneva, Switzerland, 18–22 Sept 2011

[Ange12] M. Angelou et al., Benefits of implementing a dynamic impairment-aware optical network: Results of EU project DICONET. IEEE Commun. Mag. **50**(8), 79–88, Aug 2012

[ATT10] AT&T Optical Mesh Service – OMS. (AT&T Product Brief 1 Jul 2010), www.business.att.com/binary/content/productbrochures/PB_OMS_20676.pdf

[AYDA03] C. Assi, Y. Ye, S. Dixit, M. Ali, Control and management protocols for survivable optical mesh networks. J.Lightwave Technol. **21**(11), 2638–2651, Nov 2003

[AYTM09] D. Andrei, H.-H. Yen, M. Tornatore, C. U. Martel, B. Mukherjee, Integrated provisioning of sliding scheduled services over WDM optical networks. J. Opt. Commun. Netw. **1**(2), A94–A105, Jul 2009

[Azod11] S. Azodolmolky et al., Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks. J. Lightwave Technol. **29**(4), 439–448, 15 Feb 2011

[BaLe02] N. Barakat, A. Leon-Garcia, An analytic model for predicting the locations and frequencies of 3R regenerations in all-optical wavelength-routed WDM networks. *Proceedings, IEEE International Conference on Communications (ICC'02)*, New York, 28 Apr–2 May 2002. vol. 5, pp. 2812–2816

[Batt07] L. Battestilli et al., EnLIGHTened computing: An architecture for co-allocating network, compute, and other grid resources for high-end applications, *International Symposium on High Capacity Optical Networks and Enabling Technologies* (HONET 2007), Dubai, United Arab Emirates, 18–20 Nov 2007, pp. 1–8

[Berg03] L. Berger, Editor, Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description. (Internet Engineering Task Force, Request for Comments (RFC) 3471, Jan 2003)

[BGPV12] L. Badger, T. Grance, R. Patt-Corner, J. Voas, Cloud Computing Synopsis and Recommendations. (National Institute of Standards and Technology (NIST), Special Publication 800–146, May 2012)

[BJJL11] I. Baldine, A. W. Jackson, J. Jacob, W. E. Leland, J. H. Lowry, W. C. Milliken, P. P. Pal, S. Ramanathan, K. Rauschenbach, C. A. Santivanez, D. M. Wood, PHAROS: An architecture for next-generation core optical networks, in Next-Generation Internet: Architectures and Protocols, ed. by B. Ramamurthy, G. N. Rouskas, K. M. Sivalingam, (Cambridge University Press, 2011), pp. 154–178

[BoSt04] C. Bouras, K. Stamos, An adaptive admission control algorithm for bandwidth brokers. *Proceedings, Third IEEE International Symposium on Network Computing and Applications (NCA 2004)*, Cambridge, MA, 30 Aug–1 Sep 2004

[BrVF09] R. Bradford, J. P. Vasseur, A. Farrel, Preserving topology confidentiality in inter-domain path computation using a path-key-based mechanism. (Internet Engineering Task Force, Request for Comments (RFC) 5520, Apr 2009)

[BSBS08] J. Berthold, A. A. M. Saleh, L. Blair, J. M. Simmons, Optical networking: Past, present, and future. J. Lightwave Technol. **26**(9), 1104–1118, 1 May 2008

[CCCD12] A. L. Chiu, G. Choudhury, G. Clapp, R. Doverspike, M. Feuer, J. W. Gannett, J. Jackel, G. T. Kim, J. G. Klincewicz, T. J. Kwon, G. Li, P. Magill, J. M. Simmons, R. A. Skoog, J. Strand, A. Von Lehmen, B. J. Wilson, S. L. Woodward, D. Xu, Architectures and protocols for capacity efficient, highly dynamic and highly resilient core networks. J. Opt. Commun. Netw. **4**(1), 1–14, Jan 2012

[ChVo12] N. Charbonneau, V. M. Vokkarane, A survey of advance reservation routing and wavelength assignment in wavelength-routed WDM networks. IEEE Commun. Surv. Tutor. **14**(4), 1037–1064, Fourth Quarter, 2012

[CSAG08] F. Cugini, N. Sambo, N. Andriolli, A. Giorgetti, L. Valcarenghi, P. Castoldi, E. Le Rouzic, J. Poirrier, Enhancing GMPLS signaling protocol for encompassing quality of transmission (QoT) in all-optical networks. J. Lightwave Technol. **26**(19), 3318–3328, 1 Oct 2008

[DaPM12] S. Das, G. Parulkar, N. McKeown, Why OpenFlow/SDN can succeed where GMPLS failed. *Proceedings, European Conference on Optical Communication (ECOC'12)*, Paper Tu.1.D.1, Amsterdam, The Netherlands, 16–20 Sep 2012

[DDMN03] T. DeFanti, C. de Laat, J. Mambretti, K. Neggers, B. St. Arnaud, Translight: A global-scale lambdagrid for e-science. Commun. ACM, **46**(11), 34–41, Nov 2003

[DeMi13] I. de Miguel, et al., Cognitive dynamic optical networks. J. Opt. Commun. Netw. **5**(10), A107–A118, Oct 2013

[ESCJ08] E. Escalona, S. Spadaro, J. Comellas, G. Junyent, Advance reservations for service-aware GMPLS-based optical networks. Computer Netw. **52**(10), 1938–1950, Jul 2008

[FaVA06] A. Farrel, J.-P. Vasseur, J. Ash, A Path Computation Element (PCE)-Based Architecture. (Internet Engineering Task Force, Request for Comments (RFC) 4655, Aug 2006)

[GDSP10] V. Gudla, S. Das, A. Shastri, G. Parulkar, N. McKeown, L. Kazovsky, S. Yamashita, Experimental demonstration of OpenFlow control of packet and circuit switches. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'10)*, Paper OTuG2, San Diego, CA, 21–25 Mar 2010

[GeRa04] O. Gerstel, H. Raza, Predeployment of resources in agile photonic networks. J.Lightwave Technol. **22**(10), 2236–2244, Oct 2004

[GrBX13] S. Gringeri, N. Bitar, T. J. Xia, Extending Software Defined Network principles to include optical transport. IEEE Commun. Mag. **51**(3), 32–40, Mar 2013

[GrSW99] A. G. Greenberg, R. Srikant, W. Whitt, Resource sharing for book-ahead and instantaneous-request calls. IEEE/ACM Trans. Netw. **7**(1), 10–22, Feb 1999

[GSCA09] A. Giorgetti, N. Sambo, I. Cerutti, N. Andriolli, P. Castoldi, Label preference schemes for lightpath provisioning and restoration in distributed GMPLS networks. J. Lightwave Technol. **27**(6), 688–697, 15 Mar 2009

[JADD13] T. Jiménez, J. C. Aguado, I. de Miguel, R. J. Durán, M. Angelou, N. Merayo, P. Fernández, R. M. Lorenzo, I. Tomkos, E. J. Abril, A cognitive quality of transmission estimator for core optical networks. J. Lightwave Technol. **31**(6), 942–951, 15 Mar 2013

[KaKY03] D. Katz, K. Kompella, D. Yeung, Traffic Engineering (TE) Extensions to OSPF Version 2. (Internet Engineering Task Force, Request for Comments (RFC) 3630, Sep 2003)

[KiFa11] D. King, A. Farrel, The application of the Path Computation Element architecture to the determination of a sequence of domains in MPLS and GMPLS, draft-king-pce-hierarchy-fwk-06. (Internet Engineering Task Force, Work In Progress, Apr 2011)

[KPGD03] J. Kuri, N. Puech, M. Gagnaire, E. Dotaro, R. Douville, Routing and wavelength assignment of scheduled lightpath demands. IEEE J. Sel. Areas Commun. **21**(8), 1231–1240, Oct 2003

[LBLM12] Y. Lee, G. Bernstein, D. Li, G. Martinelli, A Framework for the Control of Wavelength Switched Optical Networks (WSONs) with Impairments, (Internet Engineering Task Force, Request for Comments (RFC) 6566, Mar 2012)

[LBMT13] Y. Lee, G. Bernstein, J. Martensson, T. Takeda, T. Tsuritani, O. G. de Dios, PCEP requirements for WSON routing and wavelength assignment, draft-ietf-pce-wson-routing-wavelength-09. (Internet Engineering Task Force, Work In Progress, June 2013)

[LeBI11] Y. Lee, G. Bernstein, W. Imajuku, Framework for GMPLS and Path Computation Element (PCE) Control of Wavelength Switched Optical Networks (WSONs). (Internet Engineering Task Force, Request for Comments (RFC) 6163, Apr 2011)

[LiCh07] S. Liu, L. Chen, Deployment of carrier-grade bandwidth-on-demand services over optical transport networks: A Verizon experience. *Proceedings, Optical Fiber Communication/ National Fiber Optic Engineers Conference (OFC/NFOEC'07)*, Paper NThC3, Anaheim, CA, 25–29 Mar 2007

[LiTM12] L. Liu, T. Tsuritani, I. Morita, Experimental demonstration of OpenFlow/GMPLS interworking control plane for IP/DWDM multi-layer optical networks. *Proceedings, International Conference on Transparent Optical Networks (ICTON'12)*, Paper Tu.A2.5, United Kingdom, 2–5 Jul 2012

[Liu13] L. Liu, et al., Field trial of an OpenFlow-based unified control plane for multilayer multigranularity optical switching networks. J. Lightwave Technol. **31**(4), 506–514, 15 Feb 2013

[LiWM07] W. Lin, R. S. Wolff, B. Mumey, A Markov-based reservation algorithm for wavelength assignment in all-optical networks. J. Lightwave Technol. **25**(7), 1676–1683, Jul 2007

[MABP08] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, J. Turner, OpenFlow: Enabling innovation in campus networks. White Paper, ACM SIGCOMM Comput. Commun. Rev. **38**(2), 69–74, Apr 2008

[Mann04] E. Mannie, Editor, Generalized Multi-Protocol Label Switching (GMPLS) Architecture. (Internet Engineering Task Force, Request for Comments (RFC) 3945, Oct 2004)

[MaZa10] G. Martinelli, A. Zanardi, GMPLS signaling extensions for optical impairment aware lightpath setup, draft-martinelli-ccamp-optical-imp-signaling-03. (Internet Engineering Task Force, Work In Progress, Oct 2010)

[MCMT10] R. Martínez, R. Casellas, R. Muñoz, T. Tsuritani, Experimental translucent-oriented routing for dynamic lightpath provisioning in GMPLS-enabled wavelength switched optical networks. J. Lightwave Technol. **28**(8), 1241–1255, 15 Apr 2010

[MoBB04] A. Mokhtar, L. Benmohamed, M. Bortz, OXC port dimensioning strategies in optical networks – a nodal perspective. IEEE Commun. Lett. **8**(5), 283–285, May 2004

[MPCA06] R. Martínez, C. Pinart, F. Cugini, N. Andriolli, L. Valcarenghi, P. Castoldi, L. Wosinska, J. Comellas, G. Junyent, 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, Dec 2006

[ONF12] Open Networking Foundation, "Software-Defined Networking: The new norm for networks," ONF White Paper, 13 Apr 2012

[OzPJ03] T. Ozugur, M.-A. Park, J. P. Jue, Label prioritization in GMPLS-centric all-optical networks. *Proceedings, IEEE International Conference on Communications (ICC'03)*, Anchorage, AK, 11–15 May 2003, vol. 2, pp. 1283–1287

[PCGS13] F. Paolucci, F. Cugini, A. Giorgetti, N. Sambo, P. Castoldi, A survey on the path computation element (PCE) architecture. IEEE Commun. Surv. Tutor. **15**(4), 1819–1841, Fourth Quarter, 2013

[PSAA12] J. Perelló, S. Spadaro, F. Agraz, M. Angelou, S. Azodolmolky, Y. Qin, R. Nejabati, D. Simeonidou, P. Kokkinos, E. Varvarigos, I. Tomkos, Experimental demonstration of a GMPLS-enabled impairment-aware lightpath restoration scheme. J. Opt. Commun. Netw. **4**(5), 344–355, May 2012

[Sale06] A. A. M. Saleh, Program Manager, "Dynamic Multi-Terabit Core Optical Networks: Architecture, Protocols, Control And Management (CORONET)," Defense Advanced Research Projects Agency (DARPA) Strategic Technology Office (STO), BAA 06–29, Proposer Information Pamphlet (PIP), Aug 2006

[Sale07] A. A. M. Saleh, Technologies, architecture and services for the next-generation core optical networks. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'07)*, Workshop on the Future of Optical Networking, Anaheim, CA, 25–29 Mar 2007

[SaSi06] A. A. M. Saleh, J. M. Simmons, Evolution toward the next-generation core optical network. J. Lightwave Technol. **24**(9), 3303–3321, Sept 2006

[SaSi11] A. A. M. Saleh, J. M. Simmons, Technology and architecture to enable the explosive growth of the Internet. IEEE Commun. Mag. **49**(1), 126–132, Jan 2011

[SCGN12] R. Skoog, G. Clapp, J. Gannett, A. Neidhardt, A. Von Lehman, B. Wilson, Architectures, protocols and design for highly dynamic optical networks. Opt. Switch. Netw. **9**(3), 240–251, Jul 2012

[SFPF05] L. Smarr, J. Ford, P. Papadopoulos, S. Fainman, T. DeFanti, M. Brown, J. Leigh, The OptIPuter, Quartzite, and Starlight Projects: A campus to global-scale testbed for optical technologies enabling LambdaGrid computing. *Proceedings, Optical Fiber Communication (OFC'05)*, Paper OWG7, Anaheim, CA, 6–11 Mar 2005

[SGCA09] N. Sambo, A. Giorgetti, F. Cugini, N. Andriolli, L. Valcarenghi, P. Castoldi, Accounting for shared regenerators in GMPLS-controlled translucent optical networks. J. Lightwave Technol. **27**(19), 4338–4347, 1 Oct 2009

[ShFa11] K. Shiomoto, A. Farrel, Advice on When it is Safe to Start Sending Data on Label Switched Paths Established Using RSVP-TE. (Internet Engineering Task Force, Request for Comments (RFC) 6383, Sep 2011)

[SiSB01] J. M. Simmons, A. A. M. Saleh, L. Benmohamed, Extending Generalized Multi-Protocol Label Switching to configurable all-optical networks. *Proceedings, National Fiber Optic Engineers Conference (NFOEC'01)*, Baltimore, MD, 8–12 Jul 2001, pp. 14–23

[SkNe09] R. A. Skoog, A. L. Neidhardt, A fast, robust signaling protocol for enabling highly dynamic optical networks. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'09)*, Paper NTuB5, San Diego, CA, 22–26 Mar 2009

[SkWi10] R. A. Skoog, B. J. Wilson, Transponder pool sizing in highly dynamic translucent WDM optical networks. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'10)*, Paper NTuA3, San Diego, CA, 21–25 Mar 2010

[SPLC09] N. Sambo, C. Pinart, E. Le Rouzic, F. Cugini, L. Valcarenghi, P. Castoldi, Signaling and multi-layer probe-based schemes for guaranteeing QoT in GMPLS transparent networks. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'09)*, Paper OWI5, San Diego, CA, 22–26 Mar 2009

[SYTR07] L. Shen, X. Yang, A. Todimala, B. Ramamurthy, A two-phase approach for dynamic lightpath scheduling in WDM optical networks. *Proceedings, IEEE International Conference on Communications (ICC'07)*, Glasgow, Scotland, 24–28 Jun 2007, pp. 2412–2417

[Take06] A. Takefusa, et al., G-lambda: Coordination of a grid scheduler and lambda path service over GMPLS. Futur. Gener. Comp. Sy. **22**(8), 868–875, Oct 2006

[TrCV13] J. Triay, C. Cervello-Pastor, V. M. Vokkarane, Analytical blocking probability model for hybrid immediate and advance reservations in optical WDM networks. IEEE/ACM Trans. Netw. **21**(6), 1890–1903, Dec 2013

[VaLe09] J. P. Vasseur, J. L. Le Roux, Path Computation Element (PCE) Communication Protocol (PCEP). (Internet Engineering Task Force, Request for Comments (RFC) 5440, Mar 2009)

[VaMa12] J. Varia, S. Mathew, "Overview of Amazon Web Services," White Paper, Oct 2012

[VZBL09] J. P. Vasseur, R. Zhang, N. Bitar, J. L. Le Roux, A Backward-Recursive PCE-Based Computation (BRPC) Procedure To Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths. (Internet Engineering Task Force, Request for Comments (RFC) 5441, Apr 2009)

[WFKP12] S. L. Woodward, M. D. Feuer, I. Kim, P. Palacharla, X. Wang, D. Bihon, Service velocity: Rapid provisioning strategies in optical ROADM networks. J. Opt. Commun. Netw. **4**(2), 92–98, Feb 2012

[WLLF05] B. Wang, T. Li, X. Luo, Y. Fan, C. Xin, On service provisioning under a scheduled traffic model in reconfigurable WDM optical networks. *Proceedings, IEEE 2nd International Conference on Broadband Networks (BroadNets 2005)*, vol. 1, Boston, MA, 3–7 Oct 2005, pp. 13–22

[YeTG13] S. H. Yeganeh, A. Tootoonchian, Y. Ganjali, On scalability of Software-Defined Networking. IEEE Commun. Mag. **51**(2), 136–141, Feb 2013

[YuMG99] X. Yuan, R. Melhem, R. Gupta, Distributed path reservation algorithms for multiplexed all-optical interconnection networks. IEEE Trans. Comput. **48**(12), 1–9, Dec 1999

[ZENS08] G. Zervas, E. Escalona, R. Nejabati, D. Simeonidou, G. Carrozzo, N. Ciulli, B. Belter, A. Binczewski, M. Poznan, A. Tzanakaki, G. Markidis. PHOSPHORUS grid-enabled GMPLS control plane (G^2MPLS): Architectures, services, and interfaces. IEEE Commun. Mag. **46**(6),128–137, Jun 2008

[ZhMo02] J. Zheng, H. T. Mouftah, Routing and wavelength assignment for advance reservation in wavelength-routed WDM optical networks. *Proceedings, IEEE International Conference on Communications (ICC'02)*, vol. 5, New York, NY, 28 Apr–2 May 2002, pp. 2722–2726

Chapter 9
Flexible Optical Networks

9.1 Introduction

The previous chapter examined dynamic optical networking, where the virtual topology of the network is flexibly reconfigured to deliver bandwidth where and when it is needed. In contrast, this chapter considers flexibility with respect to the underlying technology. Historically, network technology has largely been "one size fits all"; for example, a transponder card generates a single transmission rate with a specified optical reach. The flexible approaches discussed in this chapter allow telecommunications carriers to tune the technology to better match the characteristics of the current network traffic.

As with most shifts in networking paradigms, one of the major drivers behind the trend towards greater flexibility is ultimately cost. However, the more imminent impetus is the desire to use fiber capacity more efficiently. For decades, the capacity of a fiber has been so much greater than the carried bandwidth, that fiber has been viewed as an almost infinite-capacity medium. However, two decades of explosive traffic growth[1] have brought networks close to the capacity limit of conventional fiber. With the amount of network traffic doubling approximately every 30 months [Cisc13, Koro13], this may necessitate lighting multiple fiber pairs per link in the near future. While this is certainly a feasible solution for addressing network growth, instantiating what is essentially multiple copies of a network does not deliver the economies of scale on which network operators rely. (As was noted in Chap. 4, large carriers often light up a new fiber pair when introducing a new wavelength line rate, where they take advantage of the generally superior economics that accompany an increase in line rate. However, if current traffic growth rates continue, new fibers will need to be lit much more often, likely using the same generation of technology.)

The ramifications of the impending fiber capacity limit are twofold. First, there has been a groundswell of research on innovative techniques that can *cost-effectively*

[1] Internet traffic from 2000 to 2005 exhibited exponential growth. However, over a much longer period, the *compound annual growth rate* (CAGR) of Internet traffic is better represented by a hyperbolically decreasing function [Koro13].

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4_9,

increase fiber capacity, to postpone the need for multiple fiber pairs on a link. The methods largely fall under the category of *space division multiplexing* (SDM), where spatially diverse light paths are utilized. (We are most interested in SDM solutions where the spatially diverse light paths in a given direction on a link are carried on a single fiber. However, note that deploying multiple fiber pairs on a link is a form of SDM as well.) While potentially increasing the fiber capacity by more than an order of magnitude, such solutions are likely several years away from practical implementation. Second, a greater premium has been placed on using capacity more efficiently, which ultimately translates to deriving greater flexibility from the underlying network technology. This is likely a nearer-term solution, with standards bodies already taking steps to incorporate more flexibility in their specifications.

As fiber capacity limits are so intimately tied to the need for greater flexibility, Sect. 9.2 examines this topic in more detail, including an overview of the technological approaches that are being pursued in an attempt to markedly increase the capacity of a fiber. The focus of this section is more on technology than on architecture (this section can be skipped without affecting the readability of the remainder of the chapter). However, it is important that network architects understand the implications of some of these proposals. For example, in contrast to the theme of the rest of the chapter, some of these schemes may actually *reduce* the flexibility of the network (e.g., by constraining the granularity of the reconfigurable optical add/drop multiplexer, ROADM).

The remainder of the chapter examines mining greater utilization from current systems by using bandwidth more efficiently. One major line of attack is allowing more flexible usage of the fiber spectrum. For many years, networks have used a fixed grid plan: Wavelengths are typically spaced 50 GHz apart in a backbone network and 100 GHz apart in a metro-core network. Section 9.3 looks at relatively minor modifications of this plan, where finer granularity spacing is utilized and a mix of wavelength spacings can be efficiently accommodated on one fiber. Throughout the chapter, we refer to this as the *flexible-grid* architecture.

Section 9.4 discusses a much more extreme proposal, where the spectrum can be partitioned almost arbitrarily. To distinguish this from the flexible-grid architecture, we use the term *gridless* architecture. The greater spectral flexibility allows the network bandwidth to be more efficiently allocated to match the data rates of customer traffic. While improvement in bandwidth efficiency is one motivation for this approach, another benefit is its potential to reduce the need for electronic grooming (in fact, this was the original motivation for the scheme). This would lessen the burden on Internet Protocol (IP) routers, thereby delivering savings in both cost and power consumption as well.

The flexible spectral partitioning approach of the gridless architecture brings numerous operational challenges. The *routing and wavelength assignment* (RWA) problem of conventional optical-bypass-enabled networks is replaced by the more complex *routing and spectrum assignment* (RSA) problem, as covered in Sect. 9.5. In addition to complexities in assigning spectrum to new connections and in tracking spectral usage across the network, there will likely be “mismatches” in how the spectrum is partitioned on each link. In an optical-bypass-enabled network, this will

ultimately inhibit the ability to route traffic all-optically from one link to another. Defragmentation will be required to improve the alignment of the available spectrum across links and regain "stranded" bandwidth. Defragmentation strategies are discussed in Sect. 9.6.

In addition to the operational challenges, new technology is needed to enable the greater spectral flexibility, as discussed in Sect. 9.7. First, the per-wavelength filtering inherent in most bypass-capable network elements needs to be as flexible as the grid plan. Flexible, or "gridless," ROADMs were discussed in Sect. 2.9.6; we revisit this topic in Sect. 9.7. Note that flexible ROADMs are required for both the flexible-grid and gridless architectures, although the degree of required flexibility is greater in the latter. Furthermore, to support the gridless architecture, with arbitrary connection bandwidths, new transmission formats are needed. An overview of two such formats is included in Sect. 9.7. Virtual transponders, which would likely be an important cost-reducing technology for gridless networks, are discussed in this section as well.

Section 9.8 compares the flexible-grid and gridless architectures at a high level. The discussion is focused on bandwidth utilization and on the role of electronic grooming. This comparison is supplemented by a network study in Chap. 10, which explores the potential savings in cost and capacity that may be provided by a gridless architecture.

Even without changes in the underlying grid plan, there was already a push for programmable transponder technology. This enables the signal characteristics generated by a transponder to be adjusted, via software, to better meet the requirements of the traffic being carried and the conditions of the network. For example, it may be desirable to trade off data rate for optical reach. Programmable transponders, and the ramifications for network design, are covered in Sect. 9.9.

As optical networks grow more flexible, they are coming closer to the vision of a future-proof network that can accommodate any new service or transmission innovation. It must be emphasized, however, that while some of the flexible approaches covered in this chapter may ultimately be implemented, much of the chapter is largely speculative.

9.2 Fiber Capacity Limits

The ultimate capacity of a fiber-based network depends on the system characteristics. Today's networks typically employ single-core, single-mode fibers. (The fiber core is the portion of the fiber through which the light is guided; the core typically has a circular cross-section and is surrounded by a cladding that confines most of the light to the core. A mode is a spatial phase and amplitude distribution that propagates unchanged in a waveguide [MARW12]. Single-mode fibers have a small-diameter core such that only one mode is supported.) In addition to having these fiber characteristics, most optical networks operate in only the C-band portion of the spectrum (refer to Sect. 1.6 for a discussion of the spectral bands). Furthermore,

most networks support optical bypass; in a continental-scale network, this translates to an optical-reach requirement of about 2,000–2,500 km.

With these system characteristics, it is estimated that the traffic carried by backbone networks in the 2015 time frame was within a factor of 25 of the maximum bandwidth supportable on a single-fiber-pair system. (Recall that fibers are deployed in pairs, one fiber for each direction of traffic.) This estimate, which is expounded upon in Sect. 9.2.1, is based on an analysis of the maximum *spectral efficiency* of a fiber. Spectral efficiency is defined as the ratio of the information bit rate to the total bandwidth consumed. For example, if a 100-Gb/s wavelength consumes 50 GHz of spectrum, the spectral efficiency is 2 bits/s/Hz.

Given the capacity limits of conventional networks, we consider increasing the fiber capacity by challenging the underlying system assumptions; i.e., expanding the transmission band, increasing the number of fiber cores, and/or increasing the number of fiber modes. These strategies are covered at a high level in Sect. 9.2.2 through 9.2.4, respectively. Any of these solutions pose many implementation challenges, especially the latter two.

Note that simply increasing the capacity is not the end game; the capacity increase must be achieved *cost-effectively,* such that economies of scale are realized. Otherwise, increased capacity can be accomplished by deploying multiple fiber pairs on a link. With N fiber pairs per link, the capacity increases by a factor of N, but the number of deployed optical amplifiers and the number of ROADM ports increase by N also. Thus, while relatively straightforward to implement, the multi-fiber-pair solution does not provide benefits in cost per bit/s, and equally important, power per bit/s (i.e., energy per bit).

Another means of forestalling the need to deploy multiple fiber pairs per link is to use capacity more efficiently. This section wraps up with a discussion on various near-term architectural approaches that can improve the utilization of network capacity. In contrast to the bulk of this chapter, these approaches can be supported with today's technologies, although more sophisticated network management may be needed.

Much of the discussion in the remainder of this section follows that of Saleh and Simmons [SaSi11], which looked 20–25 years out, and speculated how a 1,000-fold growth in network traffic could ultimately be handled.

9.2.1 Spectral Efficiency

Historically, transmission systems have kept pace with traffic growth by both increasing the number of wavelengths supported on a fiber and increasing the bit rate of each wavelength. In the mid to late 1990s, state-of-the-art transmission systems supported 16 wavelengths of 2.5 Gb/s each, in the ~4,000 GHz of spectrum in the C-band, representing a spectral efficiency of 0.01 bits/s/Hz. In the 2010 time frame, the prototypical backbone network supported 80 wavelengths of 40 Gb/s each, corresponding to a spectral efficiency of 0.8 bits/s/Hz. Deployment of 80×100 Gb/s

systems, with a spectral efficiency of 2.0 bits/s/Hz, began in 2012. As noted in Chap. 4, this technological progress has been attained through more complex modulation schemes and more advanced electronic signal processing.

Increased capacity through increased spectral efficiency has provided favorable economies of scale. For example, the 10-Gb/s transponder cost is approximately twice that of a 2.5-Gb/s transponder, resulting in a halving of the cost per bit/s. Similarly, the power consumption and size of a 10-Gb/s transponder is less than that of four 2.5-Gb/s transponders, providing benefits in power and space per bit/s. It is expected that similar benefits will eventually apply to higher line-rate equipment as these technologies mature; e.g., it is estimated that in 2020, a 100–Gb/s 2,000-km-reach transponder will consume roughly 15% more power than a 40-Gb/s 2,500-km-reach transponder [Gree13].

Continuing the trend of increased spectral efficiency, however, will become increasingly more difficult. The analysis of Essiambre et al. [EKWF10] indicates that for an optical reach of 2,000 km, the theoretical limit on spectral efficiency is about 6–7 bits/s/Hz per polarization. This analysis assumed single-mode fiber (SMF) (Sect. 4.2.4), ideal distributed Raman amplification (Sect. 4.2.3), modulation formats based on ring constellations [EKWF10], and took into account various optical impairments.

A few comments regarding this analysis are in order. First, the results are highly dependent upon the assumed optical reach. For example, for an optical reach of 500 km, the spectral efficiency limit was calculated to be about 9 bits/s/Hz per polarization. While yielding about 30% more capacity, a reach of 500 km would increase the number of required regenerations in a backbone network by a factor of almost 10 as compared to a reach of 2,000 km (e.g., see the study of Sect. 10.3); the cost and power consumption per unit of capacity would increase significantly. The analysis of Sect. 10.4 shows that, from a cost perspective, the optimal optical reach for a continental-scale network is 2,000–2,500 km; it is assumed that this will remain the targeted reach.

Second, the results stated above are per polarization. State-of-the-art modulation formats, e.g., dual-polarization quadrature phase-shift keying (DP-QPSK), use two polarizations (see Sect. 4.2.3); it is expected that future systems will as well. With two polarizations, the overall spectral efficiency limit at 2,000-km reach is, at best, 12–14 bits/s/Hz.

Third, while the theoretical limit may be approached in "hero experiments," it is unlikely to be attainable in a practical system. For the assumptions made by Essiambre et al. [EKWF10], it is reasonable to consider the realizable spectral efficiency limit to be more on the order of 10 bits/s/Hz. Thus, 100×400 Gb/s or 40×1 Tb/s systems are likely feasible in the C-band (though still challenging).

As compared with 80×40 Gb/s systems, 10 bits/s/Hz represents a factor of 12.5 increase in system capacity. In the 2015 time frame, carrier networks based on 80×40 Gb/s technology were on the order of 50% filled (this is clearly a rough estimate, which can vary from one carrier to another). Overall, this analysis indicates that networks, circa 2015, were within a factor of 25 of reaching the capacity of *conventional* single-fiber-pair systems.

9.2.2 Expanded Transmission Band

Most optical systems are accommodated in approximately 32 nm (i.e., ~4,000 GHz) of spectrum in the C-band. However, expansion into other bands can be used to increase system capacity. For example, the L-band provides low fiber loss comparable to the C-band, making it the most likely choice for expansion. It is important that an expanded system require only a single amplifier across the spectrum, to avoid the cost of deploying multiple band amplifiers. Furthermore, the tunable transponders ideally should tune across the whole utilized spectrum.

One commercially available system supports 54 nm across the C- and L-bands with a single amplifier [FiTV06]. Additionally, hero experiments have been performed with ultra-wideband Raman amplifiers that can amplify approximately 90 nm of spectrum in the C- and extended L-bands (however, separate C-band and L-band erbium-doped fiber amplifiers (EDFAs) were also required; additionally, the reach for the 224-wavelength experimental system was just 240 km) [SKYM12].

It is reasonable to expect that, in practice, a single amplifier could cover ~65 nm of spectrum across the C- and L-bands, thereby yielding a factor of 2 increase in system capacity as compared to C-band-only systems.

9.2.3 Multicore Fiber

While the fiber plant of carrier networks is typically composed of single-core fiber, there have been recent advances in *multicore fiber* (MCF) [MARW12; HaSS12; ZFYL12]. MCF is one example of SDM. Each core is capable of supporting other forms of multiplexing (e.g., wavelength-division multiplexing (WDM) and polarization muxing), to achieve a multiplicative effect. Ideally, the total fiber capacity increases in proportion to the number of cores; however, this will be more difficult to achieve as the number of cores increases.

In order for the MCF solution to be effective, it must continue to demonstrate the benefits of a conventional single-fiber, single-core solution. For example, ideally, a single optical amplifier would be capable of amplifying each of the fiber cores, rather than requiring one amplifier per core. Operationally, it is desirable that a single connector be capable of interconnecting all of the cores, as opposed to requiring one connector per core.

MCF presents many challenges, most notably crosstalk between the cores [FTZY10]. If the amount of crosstalk is too large, then electronic multiple-input multiple-output (MIMO) digital signal processing is needed for mitigation [Winz13]. This has important ramifications. First, MIMO processing is likely to consume a significant amount of power. Furthermore, it will preclude a ROADM from being able to drop one wavelength frequency from a single core; rather, all wavelengths at that frequency across the cores will need to be dropped/bypassed at a ROADM as a single unit, as demonstrated by Feuer et al. [Feue13].

MCF also requires that new fiber plant be deployed.

Note that many of the multicore experiments where the total carried bandwidth was 100 Tb/s or higher covered distances of 100 km or less [Saka13]. The number of cores per fiber in these experiments was typically 7 or 19, as these numbers are compatible with hexagonal packing. To attain a reasonable optical reach, the number of cores in a practical system may need to be smaller to allow for greater inter-core distance within the fiber and reduced crosstalk.

9.2.4 Multimode Fiber

Another SDM proposal for increasing fiber capacity is *multimode fiber* (MMF) [MARW12; SLAC12]. MMF has a larger core diameter than SMF, which allows multiple modes to propagate; i.e., a given frequency (or wavelength) of light has different modes along the fiber, where each mode can potentially be exploited as a channel. Some small "edge" networks already use MMF due to its lower cost compared to SMF (it is also easier to splice due to the larger core). However, standard MMF fiber, which can support more than 100 modes, is more suitable when there is just a single wavelength transmitted and the transmission distance is relatively short. Supporting a WDM signal with many modes per wavelength would be very difficult to process. Thus, for purposes of increasing fiber capacity in a backbone network, it is envisioned that the fiber would support a much more modest number of modes, e.g., six. To distinguish this type of fiber from standard MMF, it is frequently referred to as *few-mode fiber* (FMF).

Ideally, the capacity of the fiber increases in proportion to the number of modes; this is dependent on the power per mode and would be difficult to realize in practice as the number of modes increases [EsMe12; SLAC12]. Similar to the discussion regarding MCF, amplifiers and components that can operate on all of the modes are highly desirable to derive cost and power benefits. (It is expected that an FMF amplifier would be more power efficient than an MCF amplifier [Krum12].) As with the MCF solution, a new fiber plant would need to be deployed.

Due to the coupling between the modes, FMF requires MIMO processing. It is expected that a ROADM would not be capable of dropping an individual mode; i.e., all of the modes corresponding to a particular frequency would be handled as one unit by the ROADM [CLYA13].

Note that it is possible to combine SDM approaches; e.g., multiple modes supported within multiple cores.

9.2.5 Architectural Approaches for Improved Capacity Utilization

Several architectural proposals to more efficiently use network capacity were outlined by Saleh and Simmons [SaSi11]. These include taking advantage of multicasting, distributed caching, traffic asymmetry, dynamic networking, and improved IP bundling. We briefly address the potential capacity benefits of each one of these strategies.

With multicast traffic, one source communicates with multiple destinations. Provisioning a single multicast connection rather than separate unicast connections between the source and each destination potentially saves a factor of 3 in capacity, depending on the number of destinations (see Sect. 3.10). Multicasting will likely grow in importance with the proliferation of video distribution.

Distributed caching, as implemented by *content distribution networks* (CDNs), stores content on multiple servers in the network to be closer to the consumers of the content. As the Internet is used increasingly as a repository of data and video, the proportion of traffic that will be distributed via CDNs is likely to grow.[2] Furthermore, caching algorithms are improving, which increases the probability that the desired data are stored on a nearby server [SLBN13]. One study reported a factor of 3 benefit in using CDNs to reduce capacity requirements, as compared with content distribution via a centralized server [GeDo11].

Asymmetric traffic demands arise when the data rates required in the two directions of a bidirectional connection are not equal. For simplicity, such demands are typically provisioned symmetrically, with the greater of the two data rates used for both directions of the connection. Establishing asymmetric connections, to better match the data rates that are actually required, will reduce the capacity requirements, though the factor of reduction is at most 2. For example, one study of a carrier IP network found that by allowing asymmetric IP links, the capacity requirements could potentially be reduced by a factor of roughly 1.3 [WZBC13]. (There is also the potential to save equipment if unidirectional transponders are utilized; i.e., if the transponder has just a transmitter or just a receiver, rather than both. For example, if three wavelengths of traffic are sent from A to Z, but only one wavelength from Z to A, then A requires three transmitters but only one receiver, and Z requires three receivers but only one transmitter.)

Chapter 8 examined dynamic networking in detail. As reported in Sect. 8.2.1, a dynamic environment can reduce the bandwidth requirements for bursty services by a factor of 5, depending on the traffic characteristics.

Finally, packing IP traffic onto wavelengths has historically been very inefficient, due to the "headroom" needed to accommodate bursty traffic. For example, in 2005, the average fill-rate of IP-carrying wavelengths in the US Internet was about 25 % [Robe05]. However, as noted in Sect. 6.10, 40-Gb/s (and higher) wavelengths carry so many individual IP flows that there is a smoothing effect on the aggregate traffic, which allows the wavelengths to be more tightly filled. Under conditions of no failures, wavelength fill-rates of 65 % or more are feasible. The bulk of the 35 % headroom is to allow for rerouting during failure conditions (i.e., if failure recovery were not required, wavelengths could be as much as 95 % filled).

In addition to these approaches, capacity can be used more efficiently if the fiber spectrum is partitioned to better align with transmission requirements and with customer traffic. This is the subject of the next several sections.

[2] About 1/3 of the global Internet traffic crossed CDNs in 2012; this is forecast to grow to 50 % in 2017 [Cisc13].

9.3 Flexible-Grid Architectures

Networks have historically used a fixed grid plan. Wavelengths of line rate 2.5, 10, 40, and 100 Gb/s have all been accommodated with 50-GHz spacing in backbone networks;[3] i.e., the nominal wavelength center frequencies are spaced 50 GHz apart (this corresponds to ~80 wavelengths in the C-band). Spacing 400-Gb/s wavelengths every 50 GHz is theoretically possible (as this would correspond to a spectral efficiency of 8 bits/s/Hz); however, it is likely that such spacing would be technically challenging to achieve for initial rollouts of 400-Gb/s technology. Rather, it is expected that a bandwidth of 62.5 or 75 GHz will be required. Thus, if a grid granularity of 50 GHz were maintained, it would necessitate allocating 100 GHz for each 400-Gb/s wavelength, thereby wasting 25–37.5 % of the spectral capacity.

A more efficient solution is to utilize a finer grid granularity, either 12.5 or 25 GHz, to better match the 400-Gb/s requirements. These finer granularities have been supported by the International Telecommunication Union (ITU) grid-plan since 2002 [ITU02]. Furthermore, in 2012, the ITU added a "flexible-grid" option to its recommendation to better support a mix of wavelength spacings on one fiber [ITU12b]. This permits any combination of wavelength spacing, as long as each wavelength aligns with a 6.25-GHz grid and the bandwidth assigned to each wavelength is an integral multiple of 12.5 GHz. There have been corresponding proposals in the Internet Engineering Task Force (IETF) to add fields to Generalized Multi-Protocol Label Switching (GMPLS) signaling in support of this flexible-grid option [KFLZ12].

Figure 9.1 illustrates how the flexible-grid option potentially results in more efficient spectral utilization. Consider allocating two adjacent wavelengths, one requiring a bandwidth of 62.5 GHz and the other requiring 50 GHz. If the center frequencies need to be aligned on a 12.5-GHz grid, as specified in the 2002 ITU recommendation, then a 6.25-GHz gap would be required (representing wasted spectrum), as shown in Fig. 9.1(a). The flexible-grid option allows alignment on a 6.25-GHz grid, such that no gap results, as shown in Fig. 9.1(b).

The flexible-grid option may appear to alleviate any concerns regarding the support of heterogeneous wavelength bandwidths on one fiber. However, challenges may still arise in the presence of optical bypass, as illustrated in Fig. 9.2. Assume that it is desirable to route a new connection all-optically through the degree-two ROADM shown in the figure, where the new connection requires 62.5 GHz of bandwidth. Assume that the evolution of connections on the two links incident on the ROADM has resulted in the spectral fill pattern that is shown in the figure, where the shaded blocks indicate assigned spectrum and the unshaded blocks indicate available spectrum. Even though there is an available block of 62.5 GHz on both of the links, the new connection cannot be routed all-optically, due to the misalignment of the available spectrum on the two links.

[3] As noted in Sect. 2.9.6, early generations of 2.5-Gb/s technology required more than 50 GHz of bandwidth.

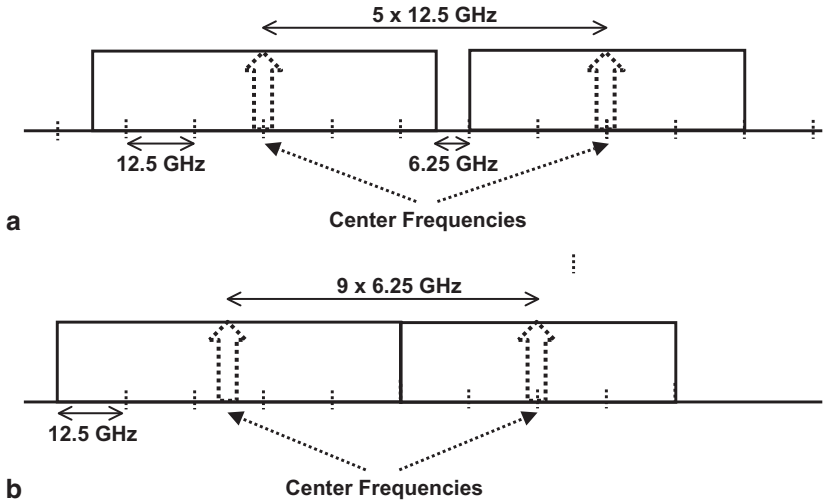


Fig. 9.1 **a** With alignment on a 12.5-GHz grid, allocating adjacent wavelengths of 62.5-GHz and 50-GHz bandwidths leads to a gap of 6.25 GHz. **b** No gap results if alignment on a 6.25-GHz grid is permitted

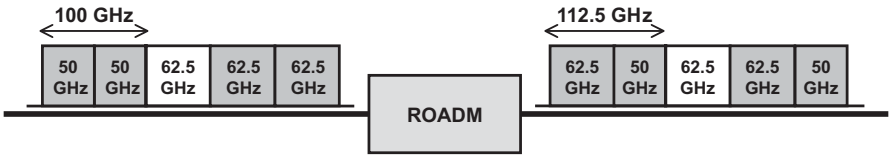


Fig. 9.2 The *shaded blocks* indicate assigned spectrum, the *unshaded blocks* indicate available spectrum. It is not possible to route a new 62.5-GHz connection all-optically through the reconfigurable optical add/drop multiplexer (*ROADM*), because the available spectra on the two links do not align

Of course, this type of inefficiency occurs in optical-bypass-enabled networks even when all wavelengths are aligned on a 50-GHz grid. There may be one wavelength free on two adjacent links, but if the wavelengths are not the same, all-optical routing between the links is not possible. However, heterogeneous wavelength bandwidths exacerbate the problem.

To investigate the effect of mixing different bandwidths, a degree-three node was simulated where all of the traffic at the node was modeled as dynamic with Poisson arrivals and exponential holding times. Fifty percent of the traffic was assumed to be sourced/sunk at the node, such that it occupied just one of the nodal links (any of the three links equally likely). The remaining 50% of the traffic was bypass traffic (any of the three bypass paths through the node equally likely). Three scenarios were considered: (1) all of the demand requests required 50-GHz bandwidth (80 wavelengths per fiber); (2) all of the demand requests required 62.5-GHz bandwidth (64 wavelengths per fiber); and (3) demand requests equally split between 50 GHz

and 62.5 GHz (on average, about 71 wavelengths per fiber). The demand arrival rates were scaled such that the spectral request rates were the same in all three scenarios. First-Fit wavelength assignment (WA) was used. For the rates chosen in the study, the blocking probabilities for the three scenarios were roughly 0.15%, 0.30%, and 1.2%, respectively. The relatively large blocking probability in the third scenario indicates that mixing bandwidths has an adverse effect on blocking. (When the wavelength bandwidth is uniform, the wavelength-assignment blocking probability increases with the size of the bandwidth because there are fewer wavelengths on a fiber. If the bandwidth of the wavelengths is *uniformly* 56.25 GHz, i.e., halfway between 50 and 62.5 GHz, the blocking probability in the simulation above is approximately 0.22%. The fact that mixing 50- and 62.5-GHz bandwidths resulted in 1.2% blocking is indicative of the detrimental effect of mixing bandwidths.)

It may be desirable to implement a *soft* partitioning of the spectrum, where, for example, the connections requiring 50 GHz are assigned wavelengths starting at the low end of the spectrum, whereas the connections requiring 62.5 GHz are assigned wavelengths starting at the upper end. (This is similar to the soft partitioning discussed for mixed line-rate systems to quasi-segregate the 10-Gb/s on-off-keying (OOK) wavelengths and the 40-Gb/s or 100-Gb/s DP-QPSK wavelengths co-propagating on a fiber; see Sect. 5.9.1.) Using this WA scheme with the mixed-bandwidth 50/62.5-GHz scenario studied above reduces the blocking probability from 1.2% to 0.45%, mitigating most of the penalty due to mixed bandwidths.

9.4 Gridless Architectures and Elastic Networks

While a bandwidth granularity of 12.5 GHz provides a lot of flexibility in the line rates that could possibly be supported, it is envisioned that wavelength line rates will continue to scale up by a factor of 4 or 2.5 (i.e., 40 Gb/s, 100 Gb/s, 400 Gb/s, 1 Tb/s). However, this vision has been challenged by the proposed *Spectrum-sLICed Elastic optical path* (SLICE) gridless architecture [Jinn08; Jinn09; GJLY12]. In SLICE, there is no notion of a wavelength line rate or a bandwidth grid. Rather, each demand is allocated the amount of spectrum to best meet its required rate; i.e., the spectrum on each fiber is *sliced* arbitrarily, based on the current traffic. The spectral allotment is sometimes referred to as an *optical corridor,* which is the terminology adopted here.

The SLICE scheme was primarily motivated by its ability to address the mismatch between the wavelength line rate and the customer traffic rate that exists in current networks and is expected to worsen. For example, as the wavelength line rate ultimately increases to 1 Tb/s, it is expected that roughly 90% of the client services will still require a data rate of 10 Gb/s or lower [Infi12]. The present mode of operation is to electronically groom the subrate traffic, e.g., using IP routers and/or Optical Transport Network (OTN) switches, to achieve a high wavelength fill-rate. As noted in Chap. 6, electronic grooming is a costly and power-consuming operation. By allocating spectrum to better match the customer traffic rate, SLICE has the

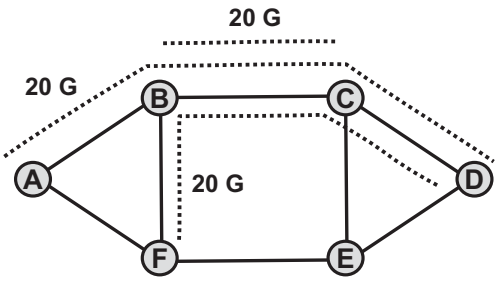


Fig. 9.3 Three 20-G demands are routed in the network. In a conventional network, with 100-Gb/s line rate, the three demands would likely be electronically groomed into one wavelength at Node *B* and sent to Node *C*. In a gridless scheme, the demands are allocated spectrum to carry exactly 20 G, such that no electronic grooming is needed

potential to reduce the amount of electronic grooming that is required. In addition to the cost and operational benefits, SLICE potentially allows the network capacity to be used more efficiently.

For example, consider the network shown in Fig. 9.3, which shows three connections routed on the network, where each connection requires 20 Gb/s. First, consider a conventional network, with a wavelength line rate of 100 Gb/s. An electronic grooming switch at Node B would likely be used to bundle the three connections together onto a wavelength that is sent to Node C. Only 60% of the bandwidth of this wavelength is used. After dropping one of the connections at Node C, the wavelength sent to Node D is only 40% filled.

In contrast, with SLICE, each connection is assigned to a 20-Gb/s optical corridor along the length of its path, such that no electronic grooming is necessary (e.g., if the spectral efficiency is 2 bits/s/Hz, then 10 GHz of spectrum is allocated to each connection). Theoretically, there is no "wasted" bandwidth. However, practically speaking, this is not realizable, as guardbands are required between the optical corridors, as will be discussed below.

SLICE is essentially performing grooming in the optical layer. In contrast to most optical-layer grooming schemes, SLICE grooms in the frequency domain rather than in the time domain. This eliminates the challenge of dealing with contention in the time domain, which has plagued most optical-layer grooming proposals; e.g., optical packet switching and optical burst switching. (Time-domain optical-layer grooming schemes often require the use of optical buffers or complex scheduling; see Sect. 6.10.)

In addition to supporting gridless optical grooming, SLICE also provides *elasticity,* where the spectrum allocated to a demand is allowed to grow or shrink as needed (growth, of course, is contingent on there being available contiguous spectrum). This adds a degree of dynamism to the network, to deliver bandwidth where it is needed.

Realistically, the fiber spectrum cannot be sliced into arbitrarily fine portions. The granularity must be feasible for the underlying technology, namely the ROADM

filters and the transmission modulation format (both are discussed in Sect. 9.7). A spectral granularity as fine as 5–10 GHz *may* be feasible, although a 12.5-GHz granularity is probably more realistic. With 12.5-GHz granularity and a spectral efficiency of 2 bits/s/Hz, the minimum-sized optical corridor would be 25 Gb/s. This level of granularity is too coarse to efficiently carry a single low-rate demand. Thus, some amount of multiplexing or grooming would still be needed, to fill the optical corridors more efficiently (see Sects. 9.8 and 10.6).

Additionally, some fraction of the spectrum will be unusable because guardbands are required between the "slices" of spectrum; i.e., two optical corridors cannot be immediately adjacent in the frequency domain [KTYY09]. The guardband serves two purposes. First, it reduces the crosstalk between neighboring optical corridors. Second, it minimizes the spectral clipping that occurs when a signal passes through a ROADM, due to imperfect filter characteristics (see Sect. 2.9.1). The bandwidth of the required guardband is in part determined by the quality of the filtering technology. Furthermore, the bandwidth of the guardband may not be uniformly assigned between all optical corridors. For example, a corridor that optically bypasses several consecutive ROADMs may require larger guardbands to prevent excessive spectral clipping.

The need for guardbands could be very detrimental to the system efficiency. For example, if each optical corridor occupies B GHz of bandwidth, followed by a guardband of B GHz, 50% of the system bandwidth will be lost. Thus, optical corridors of wider bandwidth are desirable, as long as they can be efficiently packed with traffic. The presence of guardbands ultimately limits the spectral benefits that can be attained in the gridless architecture. This point is revisited in Sect. 9.8.

Nevertheless, a gridless architecture such as SLICE does offer other ancillary benefits, which are discussed next. This is followed by a presentation of some of the implementation challenges, in Sects. 9.5 and 9.6.

9.4.1 Superchannels

While much of the emphasis of SLICE is on finely partitioning the spectrum to better meet low-rate traffic demands, the same technology that can be used to support a gridless architecture can also be used to construct *superchannels* [CLZP09], which have a capacity that is *larger* than that of a wavelength in a conventional network. For example, in the experiment reported by Jinno et al. [Jinn08], optical corridor capacity could be allocated in the range from 40 to 440 Gb/s, in 10-Gb/s increments. The same technique should be extensible for rates up to 1 Tb/s (or higher). Conventional networks currently use inverse multiplexing to carry demands that require more than one wavelength, which is likely less spectrally efficient than a gridless approach (see Sect. 9.7.2).

However, it should be noted that superchannels are not restricted to a gridless architecture. It is largely fortuitous that some of the technologies being investigated for superchannels happen to be well suited for an architecture such as SLICE. These same technologies could be used, for example, to support 1-Tb/s superchannels in a conventional network or a flexible-grid network as well.

9.4.2 Multipath Routing

In multipath routing, discussed in Sect. 3.11, the aggregate capacity of a demand is divided up into multiple lower-rate signals, with each signal potentially routed over a different path. The destination must be capable of re-aggregating the original signal. Multipath routing is especially advantageous when a network is heavily loaded and there is not sufficient available bandwidth along any one path to carry a new demand. Splitting the demand into lower-rate signals and using multiple paths, despite the added complexity, may be preferable to blocking the demand. The likelihood of finding a set of feasible paths is clearly improved with finer granularity signals.

A gridless network is well suited to support multipath routing. Demands can be partitioned into arbitrarily-sized bandwidth signals that fit into the available blocks of spectrum [ZLZA13]. For example, assume that the system granularity is 12.5 GHz, and assume that a particular link has only two blocks of available spectrum, both of which can accommodate a demand requiring up to 37.5 GHz of bandwidth. Assume that a new demand requiring a total of 50 GHz of bandwidth is ideally routed on this link. The new demand can be split into two lower-rate signals (e.g., one requiring 37.5 GHz of bandwidth and the other requiring 12.5 GHz of bandwidth) and be carried within the two blocks of available spectrum. Note, however, that it would not be desirable to partition a demand into too many lower-rate signals because guardbands are required between each signal (assuming that each signal is carried in a separate optical corridor).

Of course, multipath routing also can be supported in a conventional wavelength-grid network, but it may not be as straightforward. For example, consider the scenario where a new demand request requires a full wavelength, and no path exists with sufficient bandwidth to carry it. Multipath routing implies that the new demand will be partitioned into multiple subrate connections, each of which may require electronic grooming in a conventional network. In contrast, in a gridless network, ideally no such grooming is required for the lower-rate signals, depending on the granularity of the optical corridors.

In addition to providing greater provisioning flexibility, multipath routing can also be employed to reduce the amount of required protection resources (see Sect. 3.11.2). This usually requires that the paths over which a demand is split be diverse. The study by Ruan and Xiao [RuXi13] specifically focused on the protection-capacity benefits of multipath routing in the context of a gridless network. In theory, the benefits increase with the number of diverse paths utilized for a demand, assuming the lengths of the paths are not excessively long. However, as demonstrated in this study, the benefits level off in a gridless network due to the need for guardbands for each optical corridor over which the demand rides.

9.4.3 Bandwidth Squeezing Restoration

The elasticity property of SLICE, where the bandwidth of a connection can be dynamically modified up or down, also provides more restoration flexibility. One

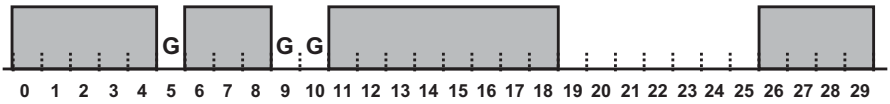


Fig. 9.4 As shown, the spectrum is divided into 30 spectral slots. Four optical corridors are assigned on the fiber, as indicated by the *shaded boxes*. *G* indicates a guardband slot

restoration scheme that takes advantage of this flexibility is referred to as *bandwidth squeezing* [SWIT09]. This scheme assumes that a network customer specifies both a minimum rate and a maximum rate for each demand. Under no failures, the maximum rate is in force. Upon failure of the demand, the rate allocated to the demand on the recovery path is only the minimum rate. Furthermore, non-failed demands that are routed along the recovery path can be scaled back to their minimum rate as well, to free up bandwidth for the restored demand. The flexibility to scale down the traffic potentially allows a significant reduction in the amount of protection capacity that needs to be allocated. Perhaps the greatest benefits can be obtained under scenarios of multiple concurrent failures. The bandwidth that remains intact can be allocated more judiciously among the affected demands, to allow at least some connectivity between demand endpoints.

9.5 Routing and Spectrum Assignment

As noted above, despite being called a "gridless" architecture, there is inherently a grid in the SLICE architecture, albeit a fine granularity one, due to the limitations of the underlying technology (probably somewhere between 5 and 12.5 GHz). The granularity of this grid represents one *spectral slot* (or one *frequency slot*). Each optical corridor that is created can be identified by its starting spectral slot (i.e., its lowest numbered slot) and the number of slots that it occupies. If the guardband size is not uniform, then it would also be necessary to track how many slots on either side of the optical corridor need to be reserved as guardbands.

An example of a spectrum partitioning on one fiber is shown in Fig. 9.4. Four optical corridors have been allocated, as indicated by the shaded blocks. The first is assigned to slots 0–4. One slot is used as a guardband to separate this corridor from the corridor assigned to slots 6–8. A third optical corridor, extending over slots 11–18, is assumed to require a two-slot guardband. A fourth corridor extends over slots 26–29. Slots 19–25 are not assigned, although some of these slots would be needed as guardbands if another corridor is added.

Such detailed slot information needs to be stored and disseminated by the network management system, thereby requiring changes in the protocols that are used. Clearly, there is more complexity than in simply tracking wavelength usage in a conventional network. With a spectral slot size of 5 GHz, the number of slots to manage on each fiber is on the order of 800, as opposed to 80 wavelengths on a fiber in a conventional network. Thus, algorithm scalability becomes a growing concern.

In conventional optical-bypass-enabled networks, some of the most essential algorithms are for RWA. In the gridless architecture, the analog is routing and *spectrum* assignment (RSA, also called routing and spectrum allocation). Some of the spectrum assignment (SA) constraints exactly parallel those of WA. If two optical corridors are routed over the same fiber, then they must be assigned non-overlapping spectrum (this applies to both optical-electrical-optical (O-E-O) and optical-bypass-enabled networks). When a corridor is all-optically routed through a ROADM, the spectrum that is used for that corridor when entering the ROADM is the same as the spectrum that is used when exiting the ROADM. This *spectral continuity constraint* is analogous to the *wavelength continuity constraint*. While all-optical "spectral converters" are technologically possible, they are unlikely to be commercially viable for quite some time (just as with all-optical wavelength converters).

There is an additional constraint that must be observed with SA, namely *contiguousness*. The slots that are assigned to one optical corridor must be adjacent. (With multipath routing, a demand is partitioned into multiple lower-rate signals, with each one potentially routed on different paths. The spectrum that is assigned to each lower-rate signal does not need to be contiguous in this scenario because each signal can be carried in a completely separate corridor.)

9.5.1 Routing

Many of the algorithms discussed in Chaps. 3–5 are directly extensible to the RSA problem. It is assumed here that routing, regeneration, and spectral assignment can be treated as three separate steps and still produce efficient results. One-step algorithms tend to be computing intensive; scalability issues will be even more problematic with RSA. (As with RWA, one-step RSA algorithms may be more desirable when the network is heavily loaded. Multistep methods do not perform as well under heavy load; furthermore, under heavy load, so few of the resources are available that the size of the problem becomes more tractable.)

With respect to routing, the same options exist as for RWA, i.e., fixed-path routing, alternative-path routing, and dynamic routing (see Sect. 3.5). As noted in Chap. 3, fixed-path routing is undesirable due to the resulting load imbalances. Also, as discussed in Chap. 3, one of the drawbacks of dynamic routing is that it typically leads to several different paths being chosen between a given source and destination, which makes WA more challenging. This negative effect is magnified with spectral assignment, where greater freedom in selecting paths is likely to lead to more spectral fragmentation; i.e., it is preferable to assign spectrum along the same link sequences so that contiguous blocks of spectrum remain free on the links.

Thus, alternative-path routing is favored for RSA, as it is for RWA. A set of candidate paths is calculated for each relevant source/destination pair, where the paths in the set provide diversity with respect to the expected "bottleneck" links of a network. When a demand request arrives, the criterion for selecting a particular candidate path can be based on, for example, minimizing the resulting load on the path links, where the link load is determined by the number of slots in use (including guardband slots).

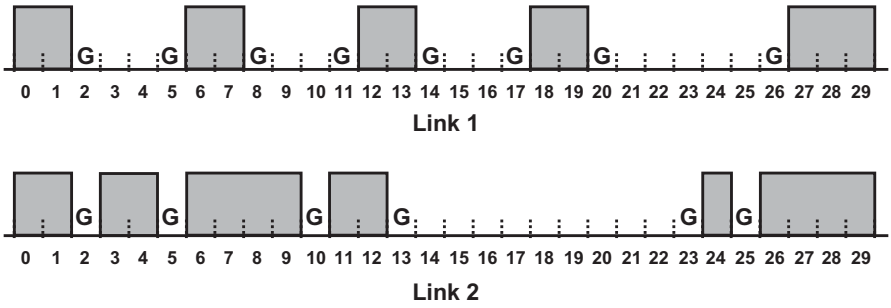


Fig. 9.5 *Link 1* is less loaded than *Link 2,* but its spectrum is more fragmented. Thus, routing strictly based on minimizing the maximum link load may not be optimal

However, load alone may not give a complete picture of how "full" a link is. It may be desirable to also consider the fragmentation of the available spectrum on a link [SHKJ11]. For example, consider the spectral state of the two links shown in Fig. 9.5, where the guardbands are explicitly shown for each corridor. Assume that a new optical corridor can be routed on either link. Link 1 is less loaded than Link 2 (19 vs. 21 used slots, including guardbands); however, the available slots on Link 1 are much more fragmented. If the new optical corridor requires only two slots (not counting guardbands), then Link 1 may be preferred because it has several two-slot gaps. However, if the new corridor requires four slots, Link 2 is likely preferred, as it has more ways of accommodating a four-slot corridor as compared to Link 1. Routing on Link 2 would leave the network in a better position to accept another four-slot corridor in the future. One can develop a variety of metrics to capture these fragmentation considerations.

9.5.2 Spectrum Assignment

The notion of optical reach extends to an optical corridor, where after the quality of the optical signal has degraded below the acceptable threshold, the corridor must be regenerated. Regeneration partitions the end-to-end path into multiple all-optical *subconnections*. (The "subconnection" terminology was introduced in Chap. 4; it refers to the portions of a connection that fall between two regeneration points or between an endpoint and a regeneration point.) It is assumed that regeneration occurs in the electrical domain, using back-to-back transponders. After undergoing regeneration, the corridor can be assigned to a different set of contiguous slots. Thus, the spectral continuity constraint holds only along each subconnection, similar to wavelength continuity in a conventional network.

As has been noted several times, guardbands are required between the optical corridors, which needs to be considered when assigning spectral slots. If the size of the guardband is uniform, i.e., G slots on either side of all corridors, then it is straightforward to incorporate this in the SA algorithm. If the optical corridor requires S slots, then SA is performed with a size requirement of $S + G$ slots, where

the G slots are positioned immediately after the end of the corridor; if the number of possible slots on the fiber is F, then SA proceeds as if there were actually $F + G$ slots on the fiber (because the highest-positioned corridor does not require a guardband above it). (Note that each corridor size is not increased by 2 G slots, as that would double-count the guardbands; i.e., the upper guardband of one corridor is the lower guardband of the adjacent corridor.)

In the algorithm descriptions below, it is assumed that the optical corridor being assigned spectrum requires S slots, and that G is one slot. The corridor is treated as extending over $S + 1$ slots.

Similar to WA, there have been numerous algorithms proposed for SA, e.g., by Christodoulopoulos et al. [ChTV11], Sone et al. [SHKJ11], and Duran et al. [Dura12]. Some of these schemes are straightforward extensions of the WA algorithms described in Sect. 5.5.

The simplest scheme to consider is First-Fit, where a subconnection is assigned the lowest-numbered feasible slot; i.e., the new subconnection is assigned to start at slot j, where j is the lowest slot number such that slots $j, j + 1, \ldots, j + S$ are available on each link of the subconnection.

In a variation of this, the candidate *path* is selected that has the lowest-numbered available slot sequence along the path [ChTV11]. This is actually a one-step RSA algorithm; i.e., it selects the route and spectral assignment together. This scheme is more appropriate for networks that do not support regeneration.

Another WA algorithm that can be readily extended to the SA problem is Most-Used. For each feasible slot assignment, $j, j + 1, \ldots, j + S$ for the subconnection, the sum

$$\sum_{k=j}^{j+S} \text{Number of Times Slot k is Assigned in Network}$$

is calculated. The slot sequence that yields the largest sum is assigned to the subconnection. Despite the additional information that is considered in Most-Used, First-Fit yielded a slightly lower blocking probability in the studies of Sone et al. [SHKJ11] and Durán et al. [Dura12].

Other SA schemes specifically take into account fragmentation. For example the Best-Fit scheme considers all spectral "gaps" of size at least $S + 1$ slots, where all slots in the gap are available on each link of the subconnection. Best-Fit selects the smallest such gap and assigns slots starting at the low end of the gap. The idea is to "fill in the holes" as best as possible. This is illustrated in Fig. 9.6, where the spectral states of the three links on which a new corridor is all-optically routed are shown; the solid shaded portions indicate the spectrum that has previously been assigned. $S + 1$ is assumed to be four. There are three gaps but only Gaps 1 and 3 are large enough to accommodate the new corridor. Of these, Gap 1 is chosen for the new corridor because its gap size is smallest. (In Fig. 9.6, the new corridor is shown assigned to Gap 1.)

Despite the intuitiveness of this algorithm, simulations that were run on Reference Network 2 yielded a slightly higher blocking probability with Best-Fit as compared to First-Fit. One of the deficiencies of Best-Fit is that, after a new corridor

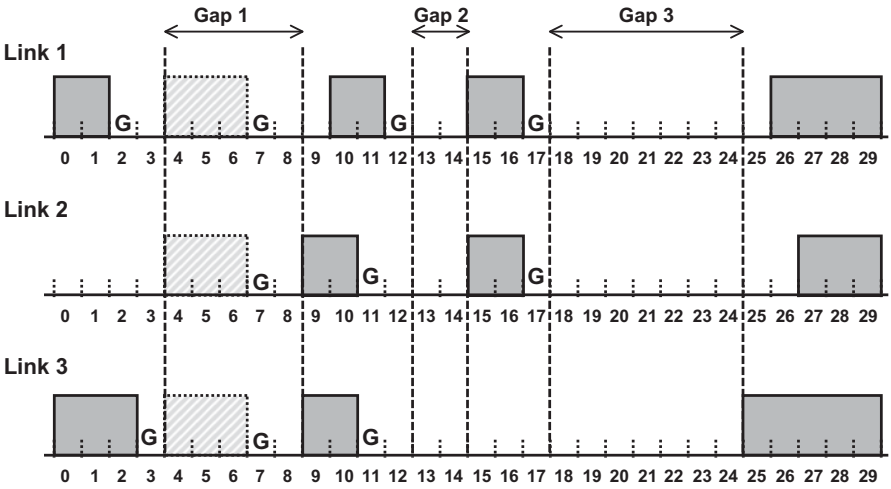


Fig. 9.6 A new corridor, requiring a total of four slots, is routed all-optically on three links. The *solid shaded slots* indicate spectrum that has previously been assigned. Of the three existing gaps, the best-fitting one is *Gap 1*. The new corridor, represented by the *hatched slots,* is shown assigned to this gap. This strands bandwidth on *Link 1* (slot *3*) and on *Links 2* and *3* (slot *8*)

is assigned to a gap, the number of free slots that remain in that gap may be very small, such that bandwidth is stranded. In the example of Fig. 9.6, after assigning the new corridor as shown, bandwidth is stranded on slot 3 of Link 1, and on slot 8 of both Links 2 and 3; i.e., no new corridors can be assigned to these slots. Thus, as suggested by Durán et al. [Dura12], it may be desirable to devise SA schemes that consider the number of stranded slots that would result from a particular slot assignment. This would have favored selecting Gap 3, as this would not strand any bandwidth. (Note, however, that if corridors are allowed to grow and shrink, then the notion of a slot being stranded is more "transient.")

In addition to the SA algorithm that is used, the order in which subconnections are assigned spectrum may affect the resulting spectral fragmentation and the ultimate blocking in the network. This comes into play with long-term network planning exercises, where many demands are added at once to the network. With RSA, typical orderings are to assign spectral slots either starting with the subconnections that require the most slots or starting with the subconnections that are routed over the most hops. Improvements (i.e., lower blocking levels) were reported by Patel et al. [PJJW12], using simulated annealing. In each annealing iteration, the order in which two subconnections were considered in the assignment algorithm was swapped. As was mentioned in Sect. 3.8, the effectiveness of such ordering techniques may depend on how many subconnections are being assigned spectrum at one time, which determines whether enough of the solution space can be explored in a reasonable time.

The graph-coloring techniques of Sect. 5.6.1 can also be extended to the SA problem. As in Sect. 5.6.1, a conflict graph is constructed where each vertex corresponds to a subconnection, and two vertices are connected by an edge if the corresponding

subconnections have at least one network link in common. Additionally, for SA, each vertex is weighted by the number of slots that are required by the optical corridor to which the subconnection belongs. The graph is colored using a *weighted* graph-coloring algorithm. (This is equivalent to unweighted graph coloring, with a vertex of weight w transformed into w fully connected vertices.) In a study by Popescu et al. [PCSC13], a methodology is presented for mapping any solution to the weighted graph-coloring problem to one where the slots assigned to each vertex are contiguous. This is necessary to enforce the spectral contiguousness constraint that is required for gridless networks.

9.5.3 Spectral Elasticity

Another aspect of the elastic gridless architecture that needs to be considered during SA is the ability of an optical corridor to grow or shrink as the connection bandwidth requirements of the client layer change [ChTV13; KRVC13]. It may be desirable to allocate extra slots to a corridor to give it the ability to grow [ZhMM13]. When the slots are not needed by the corridor, low-priority traffic can be assigned to them, where this traffic is bumped if the corridor is extended. One could also consider schemes where two corridors with opposite temporal tendencies are assigned near each other; i.e., one corridor may need more bandwidth in the morning, whereas the other needs more bandwidth in the evening. The slots at the boundary of the two corridors could potentially be time-shared, thereby providing statistical multiplexing benefits on a coarse time scale.

Two spectral adjustment schemes were considered by Christodoulopoulos et al. [ChTV13]. In both schemes, described below, assume that each optical corridor is assigned a nominal "center" slot.

In one scheme, requests for an increase in bandwidth are accommodated by first adding spectral slots to the upper end of the corridor. Once expansion at the upper end is no longer possible (due to meeting up with the corridor above it in the spectrum), spectral slots are added to the lower end of the corridor. When the bandwidth requirements decrease, spectral slots are first removed from the lower end of the corridor (until the "center" slot is reached), and then from the upper end.

In a second scheme, every attempt is made to grow and shrink the corridor symmetrically about the "center" slot; i.e., the preference is to alternate adding (or removing) slots at the upper and lower ends of the corridor. If the corridor becomes asymmetric with respect to the "center" slot, future growth and shrinkage are favored in one direction until symmetry is restored.

The results of Christodoulopoulos et al. [ChTV13] indicate that the symmetric scheme (i.e., the second scheme) results in a lower blocking probability for requests for additional slots.

Another possible growth/contraction scheme is to preferentially add slots at the lower end of the corridor, whereas contraction is performed from the upper end of the corridor. This would have a tendency to shift the corridors to the lower end

of the spectrum; i.e., the "center" slot may shift downwards. By packing the corridors more closely together, it improves the likelihood of finding spectrum that is free on all links of a subconnection, thereby reducing spectrum contention issues. However, the closer packing also makes future expansion of an existing corridor more difficult.

9.6 Spectral Defragmentation

The gridless approach has two sources of spectral inefficiency (in addition to the guardbands). First, gaps may arise in the spectrum of a particular link, where the gaps are not large enough to accommodate new optical corridors and thus represent stranded bandwidth. Second, inefficiencies arise from the spectrum being sliced differently on the various links, thereby impeding all-optical routing. These effects are likely to grow worse over time, especially if the traffic is dynamic.

Thus, it is expected that gridless networks would require defragmentation, where a number of optical corridors are shifted to different spectral slots and/or rerouted, to better pack the used spectrum and to better align the spectrum that is available on adjacent links. To be more precise, it is the optical-corridor subconnections that are shifted/rerouted. (This assumes that spectrum-conversion can accompany regeneration, such that the whole optical corridor is not necessarily shifted/rerouted.) Any corridor subconnection that is assigned new spectrum must be assigned that same new spectrum on each link of the subconnection. Furthermore, the switching configuration of the intermediate ROADMs along the subconnection, and the spectrum used by the transponders at the two subconnection endpoints, must be updated accordingly. Thus, defragmentation is potentially a complex operation.

Defragmentation can be performed on a periodic basis, as part of network maintenance. Refer back to the spectrum of the two links shown in Fig. 9.5 and assume that these are adjacent links. The goal of defragmentation might be to better align the available spectrum on these two links. For example, to create a section of eight contiguous slots that are free on both Links 1 and 2, the corridor on slots 18 and 19 of Link 1 could be shifted to slots 3 and 4 of that link. Alternatively, in a more aggressive approach, the goal might be to pack the used slots on each link at one end of the spectrum. In this scenario, three corridors on Link 1 and two corridors on Link 2 would need to be moved. This would create a section of nine contiguous slots that are free on both links. However, the small additional benefit (one extra contiguous free slot as compared with the less aggressive strategy) may not justify the amount of required disruptions.

It is also possible to adopt a more reactive approach, where defragmentation occurs only when a new demand request cannot be accommodated [THSS11]. The selection of the route for the new demand can, at least in part, be based on minimizing the number of conflicts with existing corridors. These conflicting corridors are either moved to different slots on the same link, or they are rerouted.

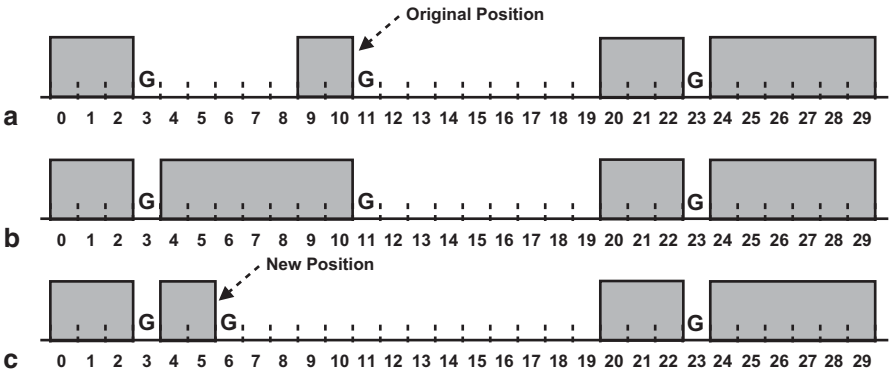


Fig. 9.7 **a** The original spectral partitioning. Assume that it is desired to shift the optical corridor on slots *9* and *10* to slots *4* and *5* on that same link. **b** In the first phase, the corridor is extended to encompass the original slots, the new slots, and any slots in between. **c** In the second phase, the corridor is contracted to encompass just the new slots, *4* and *5*

As defragmentation requires adjusting live traffic, it is important that it be performed in a "hitless" manner. Typically, this is accomplished by using a "make-before-break" mechanism, where the new optical corridor is established prior to the original one being removed.

A hitless defragmentation mechanism that takes advantage of the elasticity of a corridor was proposed by Gerstel [Gers10] and Cugini et al. [CPMB13]. This mechanism, referred to as Push–Pull by Cugini et al. [CPMB13], is illustrated in Fig. 9.7. Assume that the corridor to be moved is initially assigned to slots 9 and 10, as shown in Fig. 9.7(a). The goal is to move the corridor to slots 4 and 5. To accomplish this, the corridor is first expanded to encompass both the original slots and the new slots, and all slots in between, as shown in Fig. 9.7(b). The corridor is then contracted to occupy only the desired new slots as shown in Fig. 9.7(c). Note that these expansion and contraction operations must be performed concurrently on each ROADM along the path of the subconnection, as well as at the transponders at either end of the subconnection, so that end-to-end connectivity is never lost. The experiments of Cugini et al. [CPMB13] demonstrated that this method of defragmentation can be performed without any traffic disruption.

Push–Pull provides support for only a limited amount of defragmentation; i.e., spectrum can only be shifted along the same fiber, and all slots between the old and new spectral assignments must be unassigned. For example, looking back at Fig. 9.5, it would not be possible to shift the optical corridor on slots 18 and 19 of Link 1 to slots 3 and 4, because of the intervening existing corridors. Nevertheless, the Push–Pull method can provide a notable improvement in performance. For example, in the study by Wang et al. [WaMu13], this method was shown to reduce the level of blocking by one to two orders of magnitude when used reactively. The scheme involves shifting a subset of the existing optical corridors on a candidate path to create a contiguous block of free spectrum that can accommodate a new demand that would otherwise be blocked.

Some form of defragmentation would almost certainly be necessary to maintain capacity efficiency in a gridless architecture such as SLICE. Fragmentation may also arise in the wavelength-based flexible-grid architecture, due to the possibly disparate allocation of wavelength bandwidths on each link. This was illustrated in Fig. 9.2, where the misalignment of the available spectrum on the two links prevented a new demand from being routed all-optically through the ROADM. Due to the relatively small number of possible bandwidths that are likely to be supported on one fiber (e.g., perhaps 50, 62.5, and 75 GHz), however, the level of fragmentation and contention should not be as severe as in a gridless architecture. Nevertheless, it is possible that defragmentation could be warranted. The defragmentation heuristic proposed by Patel et al. [PJJW11a] moves connections to the lowest numbered wavelengths possible. This approximates a First-Fit type of WA. However, as described in Sect. 9.3, a soft partitioning of the spectrum, where an attempt is made to segregate connections based on their required bandwidths, outperforms First-Fit in terms of blocking probability. Incorporating soft partitioning into the defragmentation scheme for a flexible-grid architecture might be beneficial as well.

9.7 Technologies for Flexible-Grid and Gridless Networks

Implementing a gridless architecture, and to a lesser extent a flexible-grid architecture, requires new, flexible technology. The most important components are covered below.

9.7.1 Gridless ROADMs

For more than a decade, wavelengths in a backbone network have typically been aligned on a 50-GHz grid, even though the line rate has increased from 2.5 to 100 Gb/s (wavelengths in a metro-core network are typically aligned on a 100-GHz grid). This homogeneity carried over to the design of the ROADM, where the switch technology has been limited to this same fixed wavelength spacing. However, switching the individual wavelengths of a flexible-grid network or the optical corridors of a gridless architecture requires that the ROADM incorporate a commensurate amount of flexibility. This has given rise to the "gridless" ROADM, where both the channel spacing and the filter passband can be adjusted through software to match the spectral state of the fibers feeding into the ROADM. (Gridless ROADMs were discussed at a high level in Sect. 2.9.6.)

The "switching engine" of the ROADM must have fine enough granularity to be able to add/drop/switch the individual optical signals. Additionally, the filters must be sharp enough so as not to cause excessive spectral clipping. To match the granularity of the flexible-grid option specified by the ITU, a bandwidth granularity of 12.5 GHz is required. A gridless architecture could potentially take advantage of a bandwidth granularity as fine as 5 GHz; however, it may not be possible to build

filters with sharp enough "skirts" at this granularity. Without sharp skirts, the guard-band would need to be increased, such that no net capacity benefit may be achieved as compared to utilizing a coarser granularity. Thus, 12.5 GHz may necessarily end up being the required granularity for the gridless architecture as well.

One ROADM technology that is well suited to provide the necessary flexibility is *liquid crystal on silicon* (LCoS) [BFAZ06; Fris07; CoCo11; MaSi12]. In a wavelength-selective switch (WSS)-based ROADM architecture using LCoS technology, there is a two-dimensional array of small, densely packed liquid-crystal pixels, where the phase of light can be programmed for each pixel; the particular phase image of the pixels determines to which output port the incident light is steered. The set of pixels that are assigned to steer a particular wavelength can be modified to conform to the desired channel spacing and modulation format. Furthermore, the set of pixels can be configured independently for each wavelength, thereby providing the requisite flexibility across the spectral band. (Rather than using liquid crystal technology, an alternative is a two-dimensional array of Digital Light Processor (DLP®) micro-mirrors.)

The pixelated array is combined with a dispersive element that spreads an incoming WDM signal into its different frequencies across the array. To meet the stringent requirements of a fine-granularity flexible ROADM, a high-resolution dispersive element is required. Using current technology, a 3-GHz resolution is possible [MaSi12]; experimental systems demonstrate a resolution as fine as 0.5–1 GHz (this corresponds to the precision of the device; it does not imply spectral slots of this bandwidth can be supported).

Overall, ROADM technology should not be a bottleneck in realizing flexible-grid and gridless architectures, although it does prevent using an arbitrarily small slot size in the gridless architecture.

9.7.2 Flexible Transmission for Gridless Networks

While a flexible ROADM is required for both the flexible-grid and gridless architectures, flexible transmission is a necessity only for the latter. Specifically, the transmission must support fine-granularity bandwidths and must demonstrate elasticity, where the bandwidth can grow or shrink. Furthermore, the transmission must remain compatible with an optical-bypass-enabled network of extended optical reach. Two multi-carrier solutions have emerged as the favored transmission techniques for the gridless architecture: the optical analog of *orthogonal frequency-division multiplexing* (OFDM) [DjVa06; ShAt06; ShBT08], and *Nyquist-WDM* (N-WDM) [BCCP10; Gavi10].

There are numerous possible implementations of OFDM in the optical domain, with the major differences being the methods of signal synthesis and detection [ZDMM13]. The terminology to distinguish the different OFDM variants is used inconsistently; thus, we will simply use the generic term, *optical OFDM* or *O-OFDM*.

With O-OFDM, the optical signal is carried on a number of low-rate, orthogonal carriers. By adding or deleting more carriers, the aggregate optical signal increases or decreases in bandwidth. The orthogonality of the carriers allows them to partially overlap in the frequency domain, thereby providing a spectrally efficient means of carrying a high data-rate signal. This is in contrast to conventional inverse multiplexing, where a connection requiring bandwidth greater than the line rate of one wavelength is carried across multiple wavelengths that are spaced according to the underlying grid. Another advantage of O-OFDM is that, due to the lower rate of the constituent carriers, the speed of the underlying electronics can be lower, making it less technically challenging to attain very high bit rates (e.g., 1 Tb/s). In addition, there is a greater tolerance to impairments such as chromatic dispersion and polarization-mode dispersion (PMD). One of the original motivations for O-OFDM was its use as a "superchannel" technology in more conventional architectures, to go beyond line rates of 100 Gb/s [CLZP09]. The flexibility and fine bandwidth granularity of O-OFDM make it well suited to the requirements of an elastic, gridless architecture as well.

N-WDM is an alternative spectrally efficient multi-carrier technology. The carriers are conventional WDM channels that have been spectrally shaped at the transmitter to produce close to rectangular pulses in the frequency domain, which allows the carriers to be tightly packed. The bandwidth of each carrier is slightly larger than the symbol rate. As with O-OFDM, N-WDM was initially targeted as a superchannel technology; however, it also is compatible with the requirements of an elastic, gridless network.

There are implementation challenges with either method, although they are not likely to be insurmountable.

9.7.3 Virtual Transponders

One aspect of the gridless scheme that has not received enough attention is the number of transponders that are potentially required, and the attendant cost impact. If discrete transponders are required at both endpoints and at all regeneration points for each optical corridor, then simulations have shown that the expected number of transponders in the network may increase dramatically as compared to a conventional network (see Sect. 10.6). This clearly depends on the granularity of the gridless scheme and on the traffic profile. Furthermore, the transponders must be compatible with the elastic architecture, such that they are capable of supporting bit rates from, for example, 10–100 Gb/s (or higher). These "bandwidth variable" transponders (BVTs) are likely to be *more* expensive than a conventional 100-Gb/s transponder. Given that transponders are already responsible for the bulk of the cost (as well as failures) in the optical layer, the overall network cost could actually increase if a large number of BVTs are deployed. The net impact on cost would depend on how much electronic grooming can be eliminated with the gridless architecture and how much more efficiently the fiber capacity is used.

One proposal to address this is a BVT that can be "sliced" into several "virtual transponders," each of which serves one optical corridor [Gers10]. For example, a BVT capable of supporting a maximum of 100 Gb/s could be used for two optical corridors of size 25 Gb/s and a third corridor of size 20 Gb/s, with guardbands required between the corridors. Each of these optical corridors can be routed independently in the network.

As a step in this direction, a coarse granularity "multi-flow" transmitter was demonstrated by Takara et al. [TGSY11]. The data rate of each flow was a multiple of 100 Gb/s. Multiple light sources were required in the transmitter. A Nyquist-WDM-based multi-flow transponder, composed of ten 100-Gb/s subtransponders, is described by Jinno et al. [JTYH13]. Multiple optical corridors can be supported on one transponder, with each corridor assigned an appropriate number of subtransponders. The same principle can be applied to produce elastic regenerators that are capable of regenerating multiple variable-sized optical corridors. In such multi-flow devices, photonic integration would be highly desirable to drive down the cost (and size).

More research in sharing transponders across optical corridors is needed, to maximize the cost benefits of a gridless architecture.

9.8 Flexible-Grid Versus Gridless Architectures

A discussion of the merits of the gridless architecture versus a wavelength-based flexible-grid architecture is warranted. In order to be a viable architecture, the benefits of the gridless scheme have to outweigh its added complexity, its need for new technology, and its need for new network management protocols and algorithms. The focus in this section is on the two chief motivations for the gridless architecture: more efficient use of fiber capacity and a reduction in the amount of electronic grooming.

First, consider capacity utilization. The initial problem with wavelength-based architectures was the mismatch of 400-Gb/s and possibly 1-Tb/s line rates with a 50-GHz grid, which would result in a significant amount of wasted bandwidth (e.g., assigning 100-GHz bandwidth to a wavelength that requires only 62.5 or 75 GHz). The second problem was that mixing certain bandwidths on one fiber would lead to unusable spectral gaps. However, the modified grid plan of the ITU, by allowing wavelengths to have a bandwidth granularity as fine as 12.5 GHz, with alignment on 6.25 GHz spacing, largely removes these issues. Furthermore, the high-level analysis in Sect. 9.3 showed that a soft partitioning of the spectrum is likely to be effective at removing much of the spectral fragmentation resulting from mixing a small number of bandwidths on one fiber. (Most carriers tend not to mix more than two different line rates on a fiber; large carriers may not mix line rates at all. However, the advent of programmable transponders, where the data rate and/or bandwidth of a signal can be adjusted via software, may result in an increased occurrence of mixed bandwidths on one fiber; see Sect. 9.9.)

Another source of inefficiency with conventional *single-carrier* wavelengths is the likely need for inverse multiplexing in order to support data rates as high as 1 Tb/s (due to the speed limitations of the underlying electronics). However, a more spectrally efficient, *multi-carrier* technique, such as O-OFDM or N-WDM, can be used in a conventional architecture; i.e., employing 1-Tb/s superchannels does not depend on adopting a gridless architecture.

There have been several studies that have shown the large amount of spectrum that can be saved by using a gridless architecture as compared to a wavelength-based architecture. However, in many of these studies, it is assumed that electronic grooming is not used in either scenario. For example, consider how a 10-Gb/s demand is treated in many studies: in the conventional wavelength architecture, it is carried with a 10-Gb/s wavelength requiring 50 GHz of bandwidth; in the gridless architecture, it is assigned to, say, one 12.5-GHz optical corridor. Clearly, this type of comparison will favor a gridless scheme with respect to capacity efficiency.

However, it is unlikely that a carrier would utilize a 10-Gb/s wavelength, with 50-GHz bandwidth, when more spectrally efficient methods exist. Rather, the carrier would use electronic grooming (e.g., IP or OTN) and carry the demand in a more spectrally efficient wavelength, e.g., 100 Gb/s. Electronic grooming is costly, but wasting capacity can be costly as well. To capture this trade-off, one major carrier uses a cost equivalence of one network-side IP router port to 770 wavelength-km of transport [CCCD12]. This particular cost equivalence was used in studies with 40-Gb/s line rates. We would expect the cost equivalence to increase at 100-Gb/s line rates, given that transport costs scale better than router costs (due to the better scaling of optical technology as compared to electronics). Let us assume that at 100-Gb/s line rates, the cost equivalence is one IP router port to 1,500 wavelength-km of transport. Furthermore, from the results of the grooming study of Sect. 6.9, we know that electronic grooming can produce fill-rates on the order of 80–90%, depending on the level of traffic. With this, consider a scenario where eight 10-Gb/s demands need to be transported between two IP routers that are 1,000 km apart. One option is to carry each demand, without any grooming, in a separate 10-Gb/s wavelength, each with a bandwidth of 50 GHz. A second option is to groom all eight of the demands onto one 100-Gb/s wavelength, with a bandwidth of 50 GHz. The grooming option requires two network-side IP ports, but frees up enough capacity to carry another seven 100-Gb/s wavelengths on the 1,000-km path. (The grooming option also requires client-side IP ports, but their cost is much less than that of the network-side ports.) The cost versus capacity trade-off in this scenario is thus two network-side IP ports versus 7,000 wavelength-km of transport, at 100-Gb/s line rates. With the assumption that one router port is cost-equivalent to 1,500 wavelength-km of transport, grooming is the preferred option in this simple example. The cost benefits of grooming are borne out in actual networks as well. Additionally, if an IP-over-OTN-over-optical architecture is adopted, where much of the grooming is moved to the OTN layer, the cost benefits of grooming will be higher due to the lower cost of an OTN switch port as compared to an IP router port. Thus, although electronic grooming is costly, it does provide a significant bandwidth benefit. (Of course, operational factors should be considered as well. Gridless

networks, by reducing the need for electronic grooming, may significantly reduce power consumption; this is highly desirable. However, this benefit may be offset by the increased operational complexity.)

In the wavelength-based architecture, the major source of capacity inefficiency is the fill-rate of the groomed wavelengths. Using the results of Sect. 6.9, we conservatively estimate the fill-rate to be about 80 % on average, such that 20 % of the network capacity is wasted. In the gridless scheme, the major source of capacity inefficiency is the need for guardbands, and the stranding of bandwidth that may occur due to imperfect spectrum assignment. It is not unreasonable to assume that guardbands and stranded bandwidth take up 20 % of the capacity. Thus, in terms of capacity utilization, a wavelength-based scheme with electronic grooming is probably no worse than that of a gridless scheme. Even if the gridless scheme is slightly more efficient, it is not clear that carriers would want to add so much complexity to their network just to delay a capacity upgrade to their network by a few months or a year. (In fact, the study of Sect. 10.6 indicates that the presence of guardbands may actually result in the gridless architecture being *less* capacity efficient.)

Given that electronic grooming can be beneficial, the next logical point of comparison is the desired amount of grooming in the two schemes. Many studies on the gridless architecture assume that the lowest-rate demand matches the data rate of one spectral slot, such that no electronic grooming is required. However, traffic forecasts indicate that a large percentage of demands will continue to be at relatively low rates (e.g., 1.25 Gb/s) [Infi12]. It would be undesirable to allocate a 12.5-GHz spectral slot for 1.25 Gb/s of traffic; thus, in all likelihood, grooming would be utilized in the gridless scheme as well. This has the additional benefit of reducing the number of guardbands needed because grooming packs multiple demands into one optical corridor rather than having separate corridors per demand [ZZLH11; PJJW11b; ZhMM13]. (There have been optical-layer grooming proposals where it is assumed that adding/dropping/switching of individual 1.25 Gb/s demands can be performed efficiently, e.g., by manipulating the subcarriers of an O-OFDM signal. However, most technologists do not think this would be possible, due to limitations on filtering.)

One would still expect the amount of electronic grooming in the gridless scheme to be lower than that of a wavelength-based scheme. However, the savings is largely dependent on the traffic assumptions. Thus, the impact of a gridless architecture on network cost and power consumption is difficult to state with any certainty and may vary widely from one network to another.

Differences in cost between the two architectures will also depend on the cost of a bandwidth variable transponder relative to a conventional transponder. Another important cost factor is whether the "virtual" transponder, discussed in Sect. 9.7.3, can be realized. Switching costs are not likely to be a major differentiator between the two architectures. First, both the gridless architecture and the wavelength-based flexible-grid architecture would need some form of flexible ROADM. Second, the optical network elements typically represent a small percentage of the overall network cost.

Overall, while there is much momentum driving research into the gridless architecture, it is not clear that it will deliver on all of its promised benefits and justify the additional architectural and technological complexities.

9.9 Programmable (or Adaptable) Transponders

Transponders have historically been fixed-performance devices, characterized by a specific optical reach, bandwidth, and data rate. The one tunable aspect has been the wavelength. However, as the wavelength line rate has increased, transponder technology has become more complex, necessitating that digital signal processors (DSPs) be incorporated in the transmitter and/or receiver design. In addition to providing the processing power needed for advanced modulation and detection schemes, the presence of DSP chips can be harnessed to enable transponder operational agility, where the signal properties are modified via software.

9.9.1 Data Rate Versus Optical Reach

One possible performance trade-off is data rate versus optical reach [BSRK09; KRMD10; RiVM11], where a transponder that nominally transports a data rate of D with an optical reach of R can be programmed to transport a larger D but with a smaller R. This can be accomplished, for example, by keeping the symbol rate fixed but using a modulation scheme with a greater number of bits per symbol. (It may be desirable to also employ variable-rate forward error correction (FEC) codes with soft-decision decoding in order to optimize the reach for each modulation scheme [GhKa12].) Another option to increase D while reducing R is to increase the symbol rate (but not beyond what is compatible with 50-GHz wavelength spacing, if using a fixed grid), with the bits per symbol kept fixed [KRMD10].

The D versus R trade-off allows short connections that do not require an extended optical reach to take advantage of a higher data rate. Such a trade-off is well suited to a network with a strongly distance-dependent traffic profile, where geographically proximate nodes exchange the most traffic. Additionally, given that some fraction of the system margin calculation that goes into determining optical reach is allocated to the impairments suffered from passing through multiple ROADMs, a connection that travels a relatively long distance (though less than the nominal R) but traverses very few ROADMs could take advantage of the higher data rate as well.

One complication that results from programming the data rate by adjusting the modulation scheme is that there will likely be wavelengths with a mix of modulation formats and data rates co-propagating on one fiber. As was discussed in Sect. 5.9, when adjacent wavelengths employ certain combinations of modulation formats, the performance of one or both of the wavelengths may degrade. This necessitates adding guardbands between these wavelengths or reducing their reach, either of which negates some of the benefits expected from programmable transponders. The study of Rival et al. [RiVM11] examined this effect further, using transponders capable of three different bit rates: 25 Gb/s (using binary phase-shift keying (BPSK)), 50 Gb/s (using dual-polarization BPSK (DP-BPSK)), and 100 Gb/s (using DP-QPSK). Depending on the traffic profile, the results indicated that the benefits of deploying programmable transponders may be fairly modest.

9.9.2 Bandwidth Versus Optical Reach

Another possible trade-off is bandwidth versus optical reach, where the transponder supports the same data rate but occupies a smaller bandwidth [JKTW10; ChTV11], thereby saving capacity. For example, a 100-Gb/s transponder could be capable of either 3,000-km reach with 75-GHz bandwidth, or 2,000-km reach with 50-GHz bandwidth. This option is somewhat more difficult to implement from a network management and equipment perspective because it requires that the network support wavelengths of different bandwidths. Thus, it requires support for a flexible grid and the accompanying flexible ROADMs. One method for trading off reach to enable a smaller required bandwidth is to reduce the symbol rate but increase the bits per symbol. OFDM provides another alternative, where fewer carriers are used, but the bits per symbol on each carrier are increased. For example, four carriers could be modulated with BPSK (1 bit per symbol), two carriers could be modulated with QPSK (2 bits per symbol), or one carrier could be modulated with 16-quadrature amplitude modulation (16-QAM, 4 bits per symbol). As the bits per symbol increases, the optical reach decreases. (An estimate of the reach of various combinations of modulation formats, bandwidths, and number of carriers can be found in the study by Teipen et al. [TeGE12].) Again, note that wavelengths with a mixture of modulation formats will end up co-propagating on one fiber; thus, as described above, some of the benefits of flexibility may be curbed. Nyquist-WDM offers another means of trading off reach and bandwidth by varying the spacing of the carriers.

The network economics of a programmable transponder capable of trading off data rate versus reach as compared to one that is capable of trading off bandwidth versus reach was considered by Zhou et al. [ZhNM13]. The former results in fewer transponders, whereas the latter results in less utilized capacity. Thus, from a cost perspective, the preferred programmability depends on the relative cost of a transponder compared to a wavelength-km of capacity. For reasonable cost assumptions, it is expected that trading off data rate versus reach will likely produce the lower-cost network. Furthermore, this type of programmable transponder has the additional benefit of not requiring a flex-grid architecture.

9.9.3 Utility of Programmable Transponders

Of course, the same reach, rate, and bandwidth performance points discussed in Sects. 9.9.1 and 9.9.2 can be attained using different types of transponders. However, having one transponder that is capable of a range of operational modes is beneficial with respect to inventory and sparing, and becomes especially advantageous in a dynamic network, where a given transponder may be used to carry different connections as demands enter and leave the network. Even if a transponder is assigned long-term to a single connection, the flexibility of a bandwidth-programmable transponder allows it to adjust to diurnal changes of that connection to optimize the use of the network capacity.

Programmable transponders also offer the possibility of saving energy [MRBL13]. For example, using a simpler modulation scheme or a lower symbol rate, or lighting up fewer carriers in OFDM, may reduce the power requirements. Thus, if the required data rate of a connection is reduced (e.g., due to time of day), the transponder can be programmed to support a lower rate with lower power consumption.

Another advantage afforded by programmable transponders is the ability to react to the conditions of the network [GPYS11]. Some physical-layer impairments are time varying, affected by factors such as the temperature. If the quality of transmission (QoT) of a connection falls below a threshold, the connection can be transmitted with a less aggressive modulation format in order to improve its performance. For example, when this technique is used in a gridless, elastic network, the bandwidth allocated to the connection (i.e., the optical corridor) can be increased, such that the connection data rate can remain unchanged. For this operation to be successful, there must be available spectrum on each of the path links to allow for the bandwidth expansion. This may require that other connections on these links be shifted to different spectral slots to provide room for expansion. It is also necessary to adjust the filter settings of the ROADMs through which any of the affected connections pass. While not a simple process, it may be preferable to tearing down the poorly performing connection and reestablishing it on a new route with better metrics. If the link conditions later improve, the modulation scheme can be upgraded, thereby requiring less bandwidth and freeing up spectrum.

Note that a programmable transponder could be used for QoT improvement in a conventional fixed-grid network as well. In this scenario, the bandwidth would be maintained and the data rate would be decreased; e.g., the data rate could be reduced from 100 to 40 Gb/s, while maintaining a 50-GHz spacing. This is a less desirable means to improve performance, as it decreases the data rate, which may be unacceptable to the client.

The flexibility of the transponder provides another degree of freedom when performing routing, regeneration, and wavelength/spectrum assignment. For example, assume that a new connection is to be routed on a 3,000-km path with a data rate of D. Assume that transponders have two modes of operation: 2,500-km reach at 50-GHz bandwidth or 1,500-km reach at 25-GHz bandwidth (with a data rate of D in either case). At either optical-reach setting, one regeneration is required. From a capacity perspective, it is preferable to carry the connection with the 1,500-km reach setting.

Just as regeneration provides the opportunity to change the wavelength or spectrum assigned to a connection, it also allows the transmission properties to be changed. This is more applicable when the trade-off is reach versus bandwidth, as opposed to reach versus data rate (i.e., the data rates of all transponders carrying an end-to-end connection would be expected to be the same). Consider the example above, except assume that the new connection has a distance of 4,000 km. The connection can be routed for the first 2,500 km, with 50-GHz bandwidth; after regeneration, it can be transmitted with the 1,500 km, 25-GHz setting, thereby saving capacity on the final portion of the route. From a capacity perspective, this is preferable to using the 2,500 km setting for both subconnections; from a cost perspective,

it is likely preferable to using only the 1,500-km setting, which would require two regenerations instead of one (although this depends on the cost tradeoff between transponders and capacity).

Optimizing the network with respect to modulation format (or other transmission setting) is thus another facet to consider when designing a network. In fact, terminology such as *routing, modulation level, and spectrum assignment* (RMLSA) has emerged to capture the additional step in the design process [ChTV11].

Of course, one needs to consider the cost of a programmable transponder. In general, a transponder that operates over a set of operational modes will cost somewhat more than a fixed transponder that operates in the most challenging of these modes. However, given the DSP technology that is now incorporated in many transponders, it is likely that the price premium will be fairly small.

As with most of the schemes discussed in this chapter, the additional flexibility of a programmable transponder brings additional algorithmic and network management complexities. Unless a carrier is able to reap significant capacity benefits, such that a network upgrade can be postponed for a meaningful period of time, it is not clear that such methods will be implemented.

9.10 Exercises

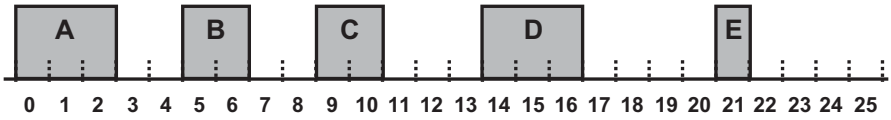
9.1. The ITU has specified a FlexGrid option that allows wavelengths to be aligned on a 6.25-GHz grid. (a) If all wavelength bandwidths are an even multiple of 12.5 GHz, does the FlexGrid option provide any benefit, as compared to wavelengths being aligned on a 12.5-GHz grid? (b) How about if all wavelength bandwidths are an odd multiple of 12.5 GHz? (c) If wavelengths needed to be aligned on a 12.5-GHz grid rather than on a 6.25-GHz grid, and wavelengths with 50- and 62.5-GHz bandwidths are mixed on one fiber, and up to 4,000 GHz of spectrum is available, in the worst case, how much bandwidth would be wasted due to unused gaps?

9.2. In Sect. 9.3, simulations showed that mixing 50- and 62.5-GHz bandwidths on a single fiber leads to excess blocking as compared to supporting just a single bandwidth. *Relative* to mixing 50- and 62.5-GHz bandwidths on a fiber, would you expect the "mixing effect" with 50- and 75-GHz bandwidths on a fiber to be better or worse, assuming First-Fit wavelength assignment is used? Why?

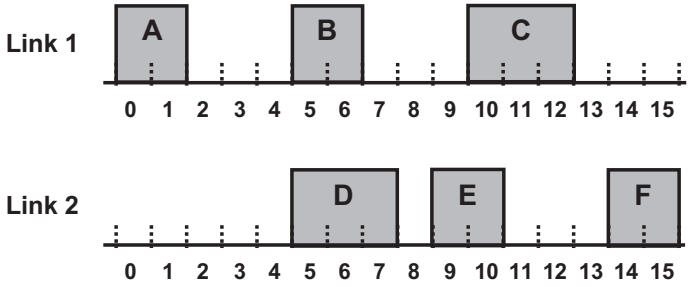
9.3. Assume that a network is 1/3 full, and that the traffic in the network doubles every 30 months. (a) How many months before the network is filled? (b) If the remainder of the network is deployed with a strategy that is 20% more efficient in using capacity (and the existing traffic is not modified), how many months before the network is filled? How does this compare to the result from part (a)?

In Exercises 9.4 through 9.8, assume that the entire spectrum on the links is shown in the figures (i.e., the spectrum does not extend further to the right). Guardbands are not explicitly shown in the figures.

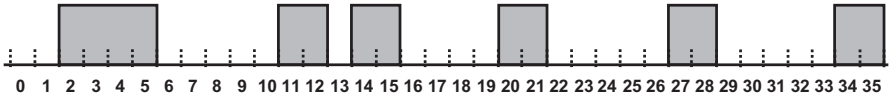
9.4. Assume that the spectrum on a link has been allocated to five optical corridors (labeled A through E) as shown below. Assume that one slot of guardband is needed between any two corridors. Assume that it is desired to defragment the spectrum, such that all of the allocated spectrum is contiguously packed (with appropriate guardbands) at either the low end or the high end. (a) What strategy (i.e., what sequence of corridor movements) minimizes the number of corridor moves to accomplish this? (If the same corridor is moved twice, this counts as two moves.) (b) If the *Push–Pull* defragmentation scheme is used, how many corridors need to be moved to accomplish this?



9.5. Assume that the spectrum on two adjacent links has been allocated to six optical corridors (labeled A through F) as shown below. Assume that one slot of guardband is needed between any two corridors. Assume that it is desired to create the largest contiguous block of spectrum that is available for a new corridor on both links. What strategy minimizes the number of corridors that need to be moved to accomplish this?

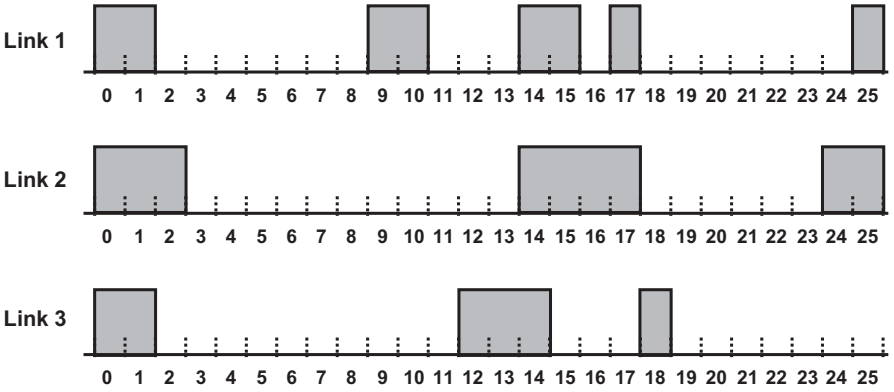


9.6. Assume that all optical corridors require an *even* number of slots, and that one slot of guardband is needed between any two corridors. Assume that the spectrum on a link is allocated as shown below. (a) At a minimum, how many slots are stranded (due to the even number of slots per corridor)? (b) What is the best strategy to defragment the link if only two corridors are permitted to be moved?

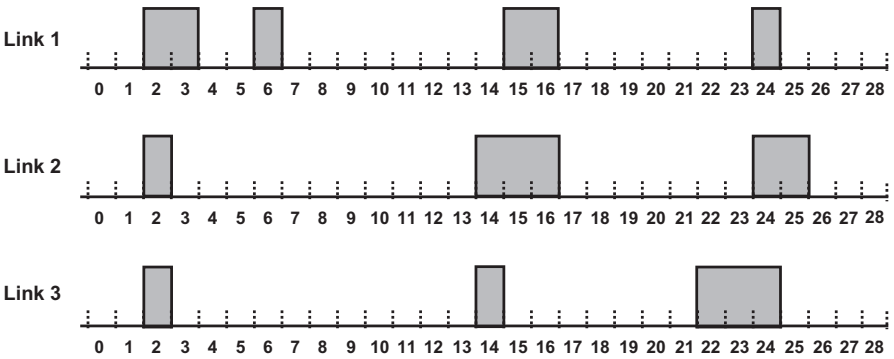


9.7. Assume that a new optical corridor that requires three slots of bandwidth is routed all-optically on three consecutive links, where the spectral state on these three links is shown below. Assume that one slot of guardband is required

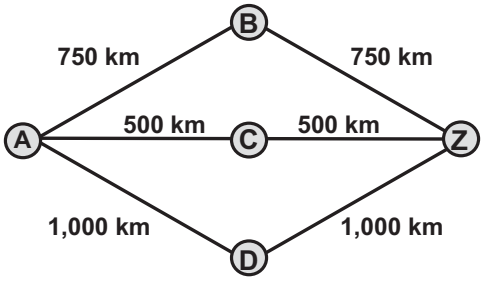
between any two corridors. (a) Where is the new corridor assigned if First-Fit assignment is used? (b) Where is it assigned if it is desired that the number of stranded slots be minimized? (c) Where is it assigned if the goal is to maximize the largest contiguous block of spectrum that is available on all three links, after assignment?



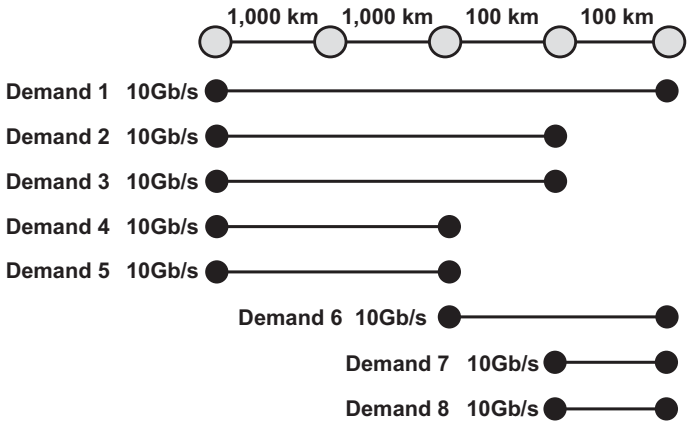
9.8. Assume that a new four-slot optical corridor is routed on three consecutive links (Link 1, Link 2, Link 3, in that order), where the spectral state on these three links is shown below. Assume that one regeneration is required, either after Link 1 or after Link 2. Assume that one slot of guardband is needed between any two corridors. (a) Where should the regeneration occur and what spectral slots should be assigned to the resulting subconnections if the goal is to maximize the largest contiguous block of spectrum that is still available on all three links after assignment? (b) If multipath routing is permitted, which allows the optical corridor to be broken into multiple "thinner" corridors, what might a better slot assignment strategy be? (One regeneration is required for each thinner corridor; the location of the regenerations can be selected independently for each of these corridors.)



9.9. In the gridless network shown below, there is one 50-Gb/s protected demand request between Nodes A and Z. The optical corridor slot granularity is 12.5 GHz and the system spectral efficiency is 2 bits/s/Hz. Consider two architectural options: multipath routing, where a demand is split across multiple paths; and *bandwidth squeezing* restoration, where the bandwidth of the demand can be reduced by up to 50% under failure conditions. Assume that shared protection can be used, and protection against only single failures is required. For parts (a) through (d), answer the following: What design minimizes the utilized capacity, and how much capacity, in *GHz*-km, is utilized? Ignore the capacity that may be required for guardbands. (a) Assume that neither multipath routing nor bandwidth squeezing is permitted. (b) Assume that multipath routing is permitted, but not bandwidth squeezing. (c) Assume that bandwidth squeezing is permitted, but not multipath routing. (d) Assume that both bandwidth squeezing and multipath routing are permitted.



9.10. The topology and demands for a gridless network are shown below. Assume that: the slot granularity is 12.5 GHz, optical corridors can occupy any number of slots, one-slot guardbands are required between spectrally adjacent corridors, the system spectral efficiency is 2 bits/s/Hz, and First-Fit slot assignment is used. Assume that one electronic grooming port is the cost equivalent of 34,000 *GHz*-km of capacity. (Ignore any other costs, including the cost of any transponders that may be utilized.) (a) If no grooming or multiplexing of demands is used (i.e., one demand per corridor), how much *GHz*-km of capacity is utilized, including the capacity occupied by the guardbands? (b) If electronic grooming is used to *minimize* the capacity consumed, how much *GHz*-km of capacity is utilized, and how many grooming ports are required (only count the network-side grooming ports, which feed into the WDM system)? How does the cost compare to that of part (a)? (c) What grooming strategy minimizes the cost? How much *GHz*-km of capacity is utilized (including the capacity occupied by the guardbands), and how many network-side grooming ports are required in this design?



9.11. Consider a network of 12 nodes arranged in a 3×4 grid. Assume that the distance of each link is 1,000 km, and assume that shortest-path routing is used. Assume that there are two types of transponders: a fixed transponder of capacity 10 Gb/s and reach 4,000 km and a programmable transponder that can support either 10 Gb/s with a reach of 4,000 km or 40 Gb/s with a reach of 2,000 km. Assume that the traffic is all-to-all, and distance dependent, such that nodes that are one link apart exchange 50 Gb/s, nodes that are two links apart exchange 40 Gb/s, nodes that are three links apart exchange 30 Gb/s, etc. (a) How many fixed transponders are required to carry the traffic? (b) How many programmable transponders are required to carry the traffic? (c) If the programmable transponder costs 20% more than the fixed transponder, does using programmable transponders reduce the network cost? (d) Repeat (a) through (c), but assume that the traffic is uniform and all-to-all, where 10 Gb/s is exchanged between all node pairs.

9.12. Assume that a carrier is evaluating which system to deploy in its network, where System 1 has a spectral efficiency that is twice that of System 2. The lifetime of the network is 6 years. *System 1:* Capacity of C. Infrastructure cost (i.e., amps, ROADMs) of P, all of which is incurred at the start of Year 1. Transponder cost of P/3 incurred at the start of each year (i.e., a total of 2P over six years). *System 2:* Capacity of C/2; Infrastructure cost of 0.8P, all of which is incurred at the start of Year 1. Due to the reduced capacity, a second system must be installed at the start of Year 4, again at a cost of 0.8P. The transponder cost incurred at the start of each year is only 3/4 of that in System 1 (a longer optical reach is achievable at lower spectral efficiency, such that fewer regenerations are required). Let P be US$ 100 million. Assume that the annual rate of return on investment is 17%. (a) How do the net present costs of the two systems compare? (b) How about if the rate of return on investment is only 2%? (c) What other factors might the carrier consider when comparing the two systems?

References

[BCCP10] G. Bosco, A. Carena, V. Curri, P. Poggiolini, F. Forghieri, Performance limits of Nyquist-WDM and CO-OFDM in high-speed PM-QPSK systems. IEEE Photonics Technol. Lett. **22**(15), 1129–1131 (1 August 2010)

[BFAZ06] G. Baxter, S. Frisken, D. Abakoumov, H. Zhou, I. Clarke, A. Bartos, S. Poole, Highly programmable wavelength selective switch based on liquid crystal on silicon switching elements. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'06)*, Anaheim, CA, 5–10 March 2006, Paper OTuF2

[BSRK09] A. Bocoi, M. Schuster, F. Rambach, M. Kiese, C.-A. Bunge, B. Spinnler, Reach-dependent capacity in optical networks enabled by OFDM. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'09)*, San Diego, CA, 22–26 March 2009, Paper OMQ4

[CCCD12] A.L. Chiu, G. Choudhury, G. Clapp, R. Doverspike, M. Feuer, J.W. Gannett, J. Jackel, G.T. Kim, J.G. Klincewicz, T.J. Kwon, G. Li, P. Magill, J.M. Simmons, R.A. Skoog, J. Strand, A. Von Lehmen, B.J. Wilson, S.L. Woodward, D. Xu, Architectures and protocols for capacity efficient, highly dynamic and highly resilient core networks. J. Opt. Commun. Netw. **4**(1), 1–14 (January 2012)

[ChTV11] K. Christodoulopoulos, I. Tomkos, E.A. Varvarigos, Elastic bandwidth allocation in flexible OFDM-based optical networks. J. Lightwave Technol. **29**(9), 1354–1366 (1 May 2011)

[ChTV13] K. Christodoulopoulos, I. Tomkos, E. Varvarigos, Time-varying spectrum allocation policies and blocking analysis in flexible optical networks. IEEE J. Sel. Areas Commun. **30**(1), 1–13 (January 2013)

[Cisc13] Cisco Visual Networking Index: Forecast and Methodology, 2012–2017, White Paper, 29 May 2013

[CLYA13] X. Chen, A. Li, J. Ye, A. Al Amin, W. Shieh, Demonstration of few-mode compatible optical add/drop multiplexer for mode-division multiplexed superchannel. J. Lightwave Technol. **31**(4), 641–647 (15 February 2013)

[CLZP09] S. Chandrasekhar, X. Liu, B. Zhu, D.W. Peckham, Transmission of a 1.2-Tb/s 24-carrier no-guard-interval coherent OFDM superchannel over 7200-km of ultra-large-area fiber. *Proceedings, European Conference on Optical Communication (ECOC'09)*, Vienna, Austria, 20–24 September 2009, Paper PD2.6

[CoCo11] P. Colbourne, B. Collings, ROADM switching technologies. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'11)*, Los Angeles, CA, 6–10 March 2011, Paper OTuD1

[CPMB13] F. Cugini, F. Paolucci, G. Meloni, G. Berrettini, M. Secondini, F. Fresi, N. Sambo, L. Potì, P. Castoldi, Push-pull defragmentation without traffic disruption in flexible grid optical networks. J. Lightwave Technol. **31**(1), 125–133 (1 January 2013)

[DjVa06] I.B. Djordjevic, B. Vasic, Orthogonal frequency division multiplexing for high-speed optical transmission. Opt. Express. **14**(9), 3767–3775 (1 May 2006)

[Dura12] R.J. Durán et al., Performance comparison of methods to solve the routing and spectrum allocation problem. *Proceedings, International Conference on Transparent Optical Networks (ICTON'12)*, United Kingdom, 2–5 July 2012, Paper Mo.C2.4

[EKWF10] R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel, Capacity limits of optical fiber networks. J. Lightwave Technol. **28**(4), 662–701 (15 February 2010)

[EsMe12] R.-J. Essiambre, A. Mecozzi, Capacity limits in single-mode fiber and scaling for spatial multiplexing. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'12)*, Los Angeles, CA, 4–8 March 2012, Paper OW3D.1

[Feue13] M.D. Feuer et al., ROADM system for space division multiplexing with spatial superchannels. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'13)*, Anaheim, CA, 17–21 March 2013, Paper PDP5B.8

[FiTV06] D.A. Fishman, W.A. Thompson, L. Vallone, LambdaXtreme® transport system: R&D of a high capacity system for low cost, ultra long haul DWDM transport. Bell Lab. Tech. J. **11**(2), 27–53 (Summer 2006)

[Fris07] S. Frisken, Advances in liquid crystal on silicon wavelength selective switching. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'07)*, Anaheim, CA, 25–29 March 2007, Paper OWV4

[FTZY10] J.M. Fini, T. Taunay, B. Zhu, M. Yan, Low cross-talk design of multi-core fibers. *Proceedings, Conference on Lasers and Electro-Optics (CLEO'10)*, San Jose, CA, 16–21 May 2010, Paper CTuAA3

[Gavi10] G. Gavioli et al., Investigation of the impact of ultra-narrow carrier spacing on the transmission of a 10-carrier 1 Tb/s superchannel. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'10)*, San Diego, CA, 21–25 March 2010, Paper OThD3

[GeDo11] A. Gerber, R. Doverspike, Traffic types and growth in backbone networks. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'11)*, Los Angeles, CA, 6–10 March 2011, Paper OTuR1

[Gers10] O. Gerstel, Flexible use of spectrum and photonic grooming. *International Conference on Photonics in Switching,* Monterey, CA, 25–28 July 2010, Paper PMD3

[GhKa12] G.-H. Gho, J.M. Kahn, Rate-adaptive modulation and low-density parity-check coding for optical fiber transmission systems. J. Opt. Commun. Netw. **4**(10), 760–768 (October 2012)

[GJLY12] O. Gerstel, M. Jinno, A. Lord, S.J.B. Yoo, Elastic optical networking: A new dawn for the optical layer? IEEE Commun. Mag. **50**(2), S12–S20 (February 2012)

[GPYS11] D.J. Geisler, R. Proietti, Y. Yin, R.P. Scott, X. Cai, N.K. Fontaine, L. Paraschis, O. Gerstel, S.J.B. Yoo, Experimental demonstration of flexible bandwidth networking with real-time impairment awareness. Opt. Express. **19**(26), B736–B745 (12 December 2011)

[Gree13] GreenTouch, GreenTouch Green Meter Research Study: Reducing the Net Energy Consumption in Communications Networks by up to 90% by 2020. GreenTouch White Paper, Version 1.0, 26 June 2013

[HaSS12] T. Hayashi, T. Sasaki, E. Sasaoka, Multi-core fibers for high capacity transmission. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'12)*, Los Angeles, CA, 4–8 March 2012, Paper OTu1D.4

[Infi12] Infinera, Network efficiency quotient. White Paper WP-EQ-06-2012, 2012

[ITU02] International Telecommunication Union, Spectral Grids for WDM Applications: DWDM Frequency Grid, ITU-T Rec. G.694.1, 1st edn, June 2002

[ITU12b] International Telecommunication Union, Spectral Grids for WDM Applications: DWDM Frequency Grid, ITU-T Rec. G.694.1, 2nd edn, February 2012

[Jinn08] M. Jinno et al., Demonstration of novel spectrum-efficient elastic optical path network with per-channel variable capacity of 40 Gb/s to over 400 Gb/s. *Proceedings, European Conference on Optical Communication (ECOC'08)*, Brussels, Belgium, 21–25 September 2008, Paper Th.3.F.6

[Jinn09] M. Jinno et al., Spectrum-efficient and scalable elastic optical path network: Architecture, benefits, and enabling technologies. IEEE Commun. Mag. **47**(11), 66–73 (November 2009)

[JKTW10] M. Jinno, B. Kozicki, H. Takara, A. Watanabe, Y. Sone, T. Tanaka, A. Hirano, Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network. IEEE Commun. Mag. **48**(8), 138–145 (August 2010)

[JTYH13] M. Jinno, H. Takara, K.Yonenaga, A. Hirano, Virtualization in optical networks from network level to hardware level. J. Opt. Commun. Netw. **5**(10), A46–A56 (October 2013)

[Koro13] S.K. Korotky, Semi-empirical description and projections of Internet traffic trends using a hyperbolic compound annual growth rate. Bell Lab. Tech. J. **18**(3), 5–21 (December 2013)

[KFLZ12] D. King, A. Farrel, Y. Li, F. Zhang, R. Casellas, Generalized labels for the flexi-grid in lambda-switch-capable (LSC) label switching routers. draft-farrkingel-ccamp-flexigrid-lambda-label-04, Internet Engineering Task Force, Work In Progress, October 2012

[KRMD10] A. Klekamp, O. Rival, A. Morea, R. Dischler, F. Buchali, Transparent WDM network with bitrate tunable optical OFDM transponders. *Proceedings, Optical Fiber Communication/ National Fiber Optic Engineers Conference (OFC/NFOEC'10)*, San Diego, CA, 21–25 March 2010, Paper NTuB5

[Krum12] P.M. Krummrich, Optical amplifiers for multi mode/ multi core transmission. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/ NFOEC'12)*, Los Angeles, CA, 4–8 March 2012, Paper OW1D.1

[KRVC13] M. Klinkowski, M. Ruiz, L. Velasco, D. Careglio, V. Lopez, J. Comellas, Elastic spectrum allocation for time-varying traffic in flexgrid optical networks. IEEE J. Sel. Area. Commun. **31**(1), 26–38 (January 2013)

[KTYY09] B. Kozicki, H. Takara, T. Yoshimatsu, K. Yonenaga, M. Jinno, Filtering characteristics of highly-spectrum efficient spectrum-sliced elastic optical path (SLICE) network. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/ NFOEC'09)*, San Diego, CA, 22–26 March 2009, Paper JWA43

[MARW12] T. Morioka, Y. Awaji, R. Ryf, P. Winzer, D. Richardson, F. Poletti, Enhancing optical communications with brand new fibers. IEEE Commun. Mag. **50**(2), S31–S42 (February 2012)

[MaSi12] D.M. Marom, D. Sinefeld, Beyond wavelength-selective channel switches: Trends in support of flexible/elastic optical networks. *Proceedings, International Conference on Transparent Optical Networks (ICTON'12)*, United Kingdom, 2–5 July 2012, Paper Mo.B1.4

[MRBL13] A. Morea, O. Rival, N. Brochier, E. Le Rouzic, Datarate adaptation for night-time energy savings in core networks. J. Lightwave Technol. **31**(5), 779–785 (1 March 2013)

[PCSC13] I. Popescu, I. Cerutti, N. Sambo, P. Castoldi, On the optimal design of a spectrum-switched optical network with multiple modulation formats and rates. J. Opt. Commun. Netw. **5**(11), 1275–1284 (November 2013)

[PJJW11a] A.N. Patel, P.N. Ji, J.P. Jue, T. Wang, Defragmentation of transparent flexible optical WDM (FWDM) networks. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'11)*, Los Angeles, CA, 6–10 March 2011, Paper OTuI8

[PJJW11b] A.N. Patel, P.N. Ji, J.P. Jue, T. Wang, Traffic grooming in flexible optical WDM (FWDM) networks. *Proceedings, 16th Opto-Electronics and Communications Conference (OECC 2011)*, Kaohsiung, Taiwan, 4–8 July 2011

[PJJW12] A.N. Patel, P.N. Ji, J.P. Jue, T. Wang, A naturally-inspired algorithm for routing, wavelength assignment, and spectrum allocation in flexible grid WDM Networks. *Proceedings, IEEE Global Communications Conference* (*GLOBECOM'12)*, Anaheim, CA, 3–7 December 2012, pp. 340–345

[RiVM11] O. Rival, G. Villares, A. Morea, Impact of inter-channel nonlinearities on the planning of 25–100 Gb/s elastic optical networks. J. Lightwave Technol. **29**(9), 1326–1334 (1 May 2011)

[Robe05] L. Roberts, Enabling data-intensive iGrid applications with advanced network technology. *iGrid 2005*, San Diego, CA, 26–29 September 2005

[RuXi13] L. Ruan, N. Xiao, Survivable multipath routing and spectrum allocation in OFDM-based flexible optical networks. J. Opt. Commun. Netw. **5**(3), 172–182 (March 2013)

[Saka13] J. Sakaguchi et al., 305 Tb/s space division multiplexed transmission using homogeneous 19-core fiber. J. Lightwave Technol. **31**(4), 554–562 (15 February 2013)

[SaSi11] A.A.M. Saleh, J.M. Simmons, Technology and architecture to enable the explosive growth of the Internet. IEEE Commun. Mag. **49**(1), 126–132 (January 2011)

[ShAt06] W. Shieh, C. Athaudage, Coherent optical orthogonal frequency division multiplexing. Electron. Lett. **42**(10), 587–589 (11 May 2006)

[ShBT08] W. Shieh, H. Bao, Y. Tang, Coherent optical OFDM: Theory and design. Opt. Express. **16**(2), 841–859 (21 January 2008)

[SHKJ11] Y. Sone, A. Hirano, A. Kadohata, M. Jinno, O. Ishida, Routing and spectrum assignment algorithm maximizes spectrum utilization in optical networks. *Proceedings, European Conference on Optical Communication (ECOC'11)*, Geneva, Switzerland, 18–22 September 2011, Paper Mo.1.K.3

[SKYM12] A. Sano, T. Kobayashi, S. Yamanaka, A. Matsuura, H. Kawakami, Y. Miyamoto, K. Ishihara, H. Masuda, 102.3-Tb/s (224 × 548-Gb/s) C- and extended L-band all-Raman transmission over 240 km using PDM-64QAM single carrier FDM with digital pilot tone. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'12)*, Los Angeles, CA, 4–8 March 2012, Paper PDP5C.3

[SLAC12] W. Shieh, A. Li, A. Al Amin, X. Chen, Space-division multiplexing for optical communications. IEEE Photon. Soc. Newsl. **26**(5), 4–8 (October 2012)

[SLBN13] S. Spagna, M. Liebsch, R. Baldessari, S. Niccolini, S. Schmid, R. Garroppo, K. Ozawa, J. Awano, Design principles of an operator-owned highly distributed content delivery network. IEEE Commun. Mag. 51(4), 132–140 (April 2013)

[SWIT09] Y. Sone, A. Watanabe, W. Imajuku, Y. Tsukishima, B. Kozicki, H. Takara, M. Jinno, Highly survivable restoration scheme employing optical bandwidth squeezing in spectrum-sliced elastic optical path (SLICE) network. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'09)*, San Diego, CA, 22–26 March 2009, Paper OThO2

[TeGE12] B.T. Teipen, H. Griesser, M.H. Eiselt, Flexible bandwidth and bit-rate programmability in future optical networks. *Proceedings, International Conference on Transparent Optical Networks (ICTON'12)*, United Kingdom, 2–5 July 2012, Paper Tu.C2.1

[TGSY11] H. Takara, T. Goh, K. Shibahara, K. Yonenaga, S. Kawai, M. Jinno, Experimental demonstration of 400 Gb/s multi-flow, multirate, multi-reach optical transmitter for efficient elastic spectral routing. *Proceedings, European Conference on Optical Communication (ECOC'11)*, Geneva, Switzerland, 18–22 September 2011, Paper Tu.5.A.4

[THSS11] T. Takagi, H. Hasegawa, K. Sato, Y. Sone, A. Hirano, M. Jinno, Disruption minimized spectrum defragmentation in elastic optical path networks that adopt distance adaptive modulation. *Proceedings, European Conference on Optical Communication (ECOC'11)*, Geneva, Switzerland, 18–22 September 2011, Paper Mo.2.K.3

[WaMu13] R. Wang, B. Mukherjee, Provisioning in elastic optical networks with non-disruptive defragmentation. J. Lightwave Technol. **31**(15), 2491–2500 (1 August 2013)

[Winz13] P.J. Winzer, Spatial multiplexing: The next frontier in network capacity scaling. *Proceedings, European Conference on Optical Communication (ECOC'13)*, London, UK, 22–26 September 2013, Paper We.1.D.1

[WZBC13] S.L. Woodward, W. Zhang, B.G. Bathula, G. Choudhury, R.K. Sinha, M.D. Feuer, J. Strand, A. L. Chiu, Asymmetric optical connections for improved network efficiency. J. Opt. Commun. Netw. **5**(11), 1195–1201 (November 2013)

[ZDMM13] G. Zhang, M. De Leenheer, A. Morea, B. Mukherjee, A survey on OFDM-based elastic core optical networking. IEEE Commun. Surv. Tutor. **15**(1), 65–87 (First Quarter 2013)

[ZFYL12] B. Zhu, J. M. Fini, M. F. Yan, X. Liu, S. Chandrasekhar, T. F. Taunay, M. Fishteyn, E. M. Monberg, F. V. Dimarcello, High-capacity space-division-multiplexed DWDM transmissions using multicore fiber. J. Lightwave Technol. **30**(4), 486–492 (15 February 2012).

[ZhMM13] S. Zhang, C. Martel, B. Mukherjee, Dynamic traffic grooming in elastic optical networks. IEEE J. Sel. Areas Commun. **31**(1), 4–12 (January 2013).

[ZhNM13] X. Zhou, L. E. Nelson, P. Magill, Rate-adaptable optics for next generation long-haul transport networks. IEEE Commun. Mag. **51**(3), 41–49 (March 2013).

[ZLZA13] Z. Zhu, W. Lu, L. Zhang, N. Ansari, Dynamic service provisioning in elastic optical networks with hybrid single-/multi-path routing. J. Lightwave Technol. **31**(1), 15–22 (1 January 2013).

[ZZLH11] Y. Zhang, X. Zheng, Q. Li, N. Hua, Y. Li, H. Zhang, Traffic grooming in spectrum-elastic optical path networks. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'11)*, Los Angeles, CA, 6–10 March 2011, Paper OTuI1.

Chapter 10
Economic Studies

10.1 Introduction

As a departure from previous chapters, which focused on the algorithmic aspects of optical networking, this chapter addresses network economics. A range of network studies are presented that investigate various properties of optical networks, especially with regard to the economics of optical bypass. The studies are intended to provide guidance on how best to evolve a network as traffic levels continue to grow, and also to shed light on some of the desirable properties for a system vendor's portfolio.

The results of any economic study depend on the topology of the network, the traffic set, and, of course, the cost assumptions used in the study. To simplify the presentation, the results of most of the studies are shown only for Reference Network 2 (a backbone network, first introduced in Sect. 1.10), using a traffic set that is representative of realistic carrier networks. However, the studies have been performed on a range of networks and traffic sets to probe dependencies on these factors. If warranted, results from other topologies are also presented. References to previously published network studies, based on different topologies and traffic, are also provided where appropriate. The topological properties of Reference Network 2, the statistics of the associated traffic set, and the assumptions for various equipment costs are presented in Sect. 10.2.

The first economic study arises from the fact that extended-reach transmission and optical-bypass-enabling network elements come with a cost premium. For example, Raman-based amplifiers are generally needed for extended optical reach as opposed to lower-cost erbium-based amplifiers. The associated transponders are more costly as well, due to the need for more complex modulation schemes, more precise lasers, and more powerful error-correcting coding. The philosophy is that the extra cost of the optical-bypass-enabled infrastructure is more than compensated for by the elimination of a large percentage of the regenerations. However, this implies that whether optical-bypass technology is cost effective in a network depends on the traffic level (i.e., the greater the level of traffic, the greater the number of regenerations that potentially can be removed) and the traffic pattern. This relation is investigated in the study presented in Sect. 10.3.

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4_10,

The cost of optical-bypass-enabling technology also implies that there is a limit to the benefits that can be achieved from increasing the optical reach. After some point, the extra cost of extending the reach further is not offset by the additional reduction in the amount of regeneration. The optimal optical reach for a network, from a cost perspective, is explored in Sect. 10.4. The effect of optical reach on the amount of traffic that needs to add/drop at a node is also investigated in this section. This offers insights into the efficacy of optical network elements that limit the amount of add/drop (see Sect. 2.6.1).

One of the benefits of optical bypass is that it provides more flexibility in choosing the topology of the network for a given set of nodes. In Sect. 10.5, the effect of topology on the network cost is compared for an optical-bypass-enabled system and an optical-electrical-optical (O-E-O)-based system. The comparison is performed in the context of varying the number of links in a metro-core mesh network.

The final two sections are more forward looking. Section 10.6 investigates the gridless architecture of Chap. 9 in more detail. The role of electronic grooming is an important factor in determining the cost and capacity benefits of this architecture. Finally, Sect. 10.7 examines an architecture where the bulk of the grooming occurs exterior to the network core, possibly in the optical domain. This is one possible evolution direction for future networks.

10.2 Assumptions

10.2.1 Reference Network Topology

Reference Network 2 is used in most of the studies in this chapter. Originally presented in Sect. 1.10, it is shown again in Fig. 10.1. This network is a realistic representation of a fairly large US backbone network, with three diverse paths across the country; the topological statistics are presented in Table 10.1. The emphasis of most of the studies in this chapter is on backbone networks due to extended optical reach having more of an impact on such networks. However, where appropriate, the studies are related to regional and metro-core networks as well.

10.2.2 Reference Traffic Set

The reference traffic set that accompanies Reference Network 2 represents realistic carrier traffic for a backbone network. Because the cost of grooming can be highly variable, e.g., due to large differences in cost between IP and Optical Transport Network (OTN) equipment, the baseline traffic set is restricted to line-rate traffic. However, in Sect. 10.6, where grooming is an integral part of the study, appropriate subrate traffic is used. (Subrate traffic is also used in the metro-core study of Sect. 10.5.)

Fig. 10.1 Reference Network 2 is used in most of the network studies in this chapter

Table 10.1 Statistics of Reference Network 2

Number of nodes	60
Number of links	77
Average nodal degree	2.6
Number of nodes with degree 2	34
Largest nodal degree	5
Average link length	450 km
Longest link length	1,200 km
Optical amplifier spacing	80 km

Table 10.2 Statistics of the reference traffic model

Number of demands	400
Percentage of demands requiring protection	50%
Average shortest path distance of a working path	1,800 km
Average shortest path distance of a protect path	3,300 km

The statistics of the baseline traffic set are shown in Table 10.2. When routed on Reference Network 2, the most heavily loaded links are at or near a utilization of 80 wavelengths. This is assumed to represent a relatively full network.

The traffic set is far from uniform all-to-all traffic. If one were to designate the largest 20% of the nodes (based on traffic generated) as *Large,* the next largest 30% of the nodes as *Medium,* with the remaining nodes designated as *Small,* then the traffic breakdown among node pairs is approximately: *Large/Large,* 30%; *Large/Medium,* 30%; *Large/Small,* 15%; *Medium/Medium,* 10%; *Medium/Small,* 10%; and *Small/Small,* 5%.

Table 10.3 Relative costs for an 80-wavelength, 10-Gb/s, 2,500-km optical-reach system

Element	Relative cost
Tunable transponder	1X
Tunable regenerator	1.4X
Bidirectional in-line optical amplifier	4X
Optical terminal	5X
ROADM	14X
Degree-3 ROADM-MD	21X
Degree-4 ROADM-MD	28X
Degree-5 ROADM-MD	35X
OTN grooming switch port	1.5X

10.2.3 Cost Assumptions

Network economics encompasses two major classes of costs: the capital cost of the equipment and the operating cost to run the network. These are often referred to as *CapEx* (i.e., capital expenditures) and *OpEx* (i.e., operational expenditures). Capital cost is generally more straightforward to calculate for a given network design and is the focus of most of the network studies in this chapter. Furthermore, carriers tend to evaluate system proposals based on the capital costs, so it is reasonable to focus on this aspect.

For the studies in this chapter, the relative costs assumed for an 80-wavelength, 2,500-km optical-reach system with 10-Gb/s line rate are shown in Table 10.3. All costs in the table are relative to the cost of a tunable 10-Gb/s transponder with 2,500-km reach. The following assumptions were made: The cost of the shelves for the transponders is amortized in the cost of the transponder; the cost of any equalization or fiber-based dispersion compensation is amortized in the cost of the in-line optical amplifier; the cost of the nodal network elements includes the cost of the pre- and post-nodal amplifiers; and the cost of the grooming-switch fabric is amortized in the cost of the switch ports.

The costs shown in Table 10.3 should be treated as rough estimates, as costs vary across vendors and carriers negotiate varying levels of discounts. (More detailed cost models can be found in Rambach et al. [RKDG13].) The goal of this chapter is to investigate trends and concepts, rather than derive absolute network costs. Furthermore, when networks are full, the capital cost of the *optical layer* is dominated by the cost of the transponders. Thus, for some architectural comparisons, the ratio of the required number of transponders is a first-order estimate of the optical-layer cost differences.

In many of the studies, several optical-reach distances are considered. To capture the fact that extended-reach transmission requires more advanced technology, it was assumed that the cost of the amplifiers, transponders, and regenerator cards increases by a factor of F for every doubling of the reach. Note that the network elements are equipped with nodal amplifiers, such that this portion of the nodal equipment was affected by this assumption as well. Here, the parameter F is referred to as the *cost increase factor.* Based on anecdotal evidence from vendors and carriers, F

was assumed to be 25% in the studies in this chapter (sensitivity to this assumption was probed in several of the studies). Thus, relative to the costs in Table 10.3 for a 2,500-km reach system, the cost of amplification and transmission in a system with an optical reach of R km was scaled by a factor of

$$1.25\hat{}(\log_2(R/2{,}500)) \tag{10.1}$$

For example, using Formula 10.1, the amplification and transmission costs of a 600-km optical-reach system would be scaled down by approximately 35%.

Operating costs are notoriously more difficult to capture as compared to calculating capital costs, because they encompass a wide range of factors, e.g., the cost of electricity to run the equipment, leasing costs for central-office space, and labor costs to install and maintain the equipment. These costs are enumerated in more detail in Verbrugge et al. [Verb05], where it is reported that carriers estimate the ratio of total operating costs to capital costs to be on the order of 1.3–4.0. This wide range is, in part, indicative of the difficulty in tracking operational costs; it also reflects the dependence of these costs on the underlying networking technology. For example, one of the major selling points of optical-bypass technology is that the removal of the bulk of the regenerations from the network results in a reduction in power costs, required rack space, and installation and maintenance costs. This is further borne out by a study in Batchellor and Gerstel [BaGe06] that included an analysis of the cost of operating a network based on different technologies. An optical-bypass-enabled network was shown to have lower operating costs as compared to various O-E-O architectures. As reinforced in Batchellor and Gerstel [BaGe06], the absolute operational cost figures may be difficult to nail down; however, the relative trends should hold. Just as the number of transponders (and regenerator cards) deployed in the network can be used as a first-order estimate of optical-layer *capital* costs, they also can be used as a rough indicator of relative optical-layer *operational* costs.

10.3 Prove-In Point for Optical-Bypass Technology

As described in the introduction, the various components of an optical-bypass system are more costly than those of an O-E-O-based system. In order to attain extended reach, the amplifiers and transponders require more advanced technology. The elements to provide optical bypass in a node are also more costly; e.g., a degree-two reconfigurable optical add/drop multiplexer (ROADM) is more costly than two optical terminals. Thus, the "first-deployed cost" of an optical-bypass-enabled network, prior to any traffic being added, is generally higher than that of an O-E-O network. However, the marginal cost of adding a demand to an O-E-O network is typically significantly higher due to the required amount of regeneration. As the traffic level increases, an optical-bypass-enabled network eventually becomes the lower-cost option, with the cost savings increasing as more traffic is added. The relative costs of optical bypass and O-E-O technologies for different levels of traffic are investigated further in this section.

Reference Network 2 was used for the study. Two architectures were considered: an optical-bypass architecture with 2,500-km optical reach and an O-E-O system with 600-km optical reach. As indicated in Sect. 10.2.3, the cost of amplification and transmission in the 600-km reach system was assumed to be a factor of ~35% lower than that in the 2,500-km reach system (i.e., using the 25% cost increase factor of Formula 10.1).

It was assumed that any regeneration was implemented with regenerator cards as opposed to more costly back-to-back transponders. (The O-E-O architecture, which has significantly more regeneration, benefits more from this assumption.) In the O-E-O architecture, there was a need for dedicated regeneration sites along the links that exceeded 600 km in length. Furthermore, in the O-E-O network, it was assumed that the nodes were equipped with patch panels as opposed to automated switches (i.e., the nodal architecture of Fig. 2.5 was assumed, not that of Fig. 2.6). Thus, this was a *non-configurable* O-E-O architecture. Conversely, the optical-bypass architecture employed ROADMs and multi-degree ROADMs (ROADM-MDs), which provide core configurability. This difference in agility in the two architectures is discussed further in Sect. 10.3.1.

The traffic level on the network was increased from "empty" to "full" to study the impact on the network cost. The reference traffic set discussed in Sect. 10.2.2 was used to represent a "full" network; the traffic set was reduced to probe the costs at lower traffic levels. Note that all demands in this traffic set were at the line rate so that grooming costs did not play a role in the analysis. The 50% of the demands that required protection were assumed to employ 1+1 dedicated client-side protection. The traffic pattern of the demand set was characteristic of real carrier traffic. Much of the traffic was distance dependent, with nodes that were geographically closer exchanging more traffic. Additionally, there was an overlay of traffic between certain nodes on the East and West coasts of the network.

The designs were performed to minimize capital costs in the two architectures. The resulting number of regenerations and optical-layer capital costs for the two architectures are shown in Table 10.4, where the network costs are normalized to 1.0 for the optical-bypass-enabled network (the final column in the table is discussed below). The capital costs include the costs of the amplifiers, transponders, regenerator cards, and network elements. As shown in the table, with no traffic in the network, the cost of the O-E-O network was roughly 33% lower than that of the optical-bypass-enabled network. However, as the traffic level increased, optical bypass gradually became the more cost-effective option. When the network was approximately full, with 80 wavelengths routed on the most heavily loaded link, optical bypass provided almost a 35% cost savings. These results are similar to those produced with a range of network topologies and traffic sets; see Simmons [Simm04].

This study indicates that whether optical-bypass technology is favored, at least from a cost perspective, depends on the level of traffic. For a network with little traffic, O-E-O may be the more cost-effective option, whereas optical bypass is more favored as the traffic level increases.

Table 10.4 Optical-bypass versus O-E-O architecture using Reference Network 2 (60 nodes)

			2,500 km, optical bypass, configurable		600 km, O-E-O, non-configurable (25% cost increase factor)		600 km, O-E-O, non-configurable (15% cost increase factor)
# of demds.	~Max. # of λs on any link	~Avg. # of λs on a link	# of regens.	Normal. cost	# of regens.	Normal. cost	Normalized cost
0	0	0	0	1.00	0	0.67	0.75
100	20	10	54	1.00	754	0.88	1.00
200	40	20	146	1.00	1,757	1.10	1.26
400	80	40	272	1.00	3,373	1.34	1.55

The cost differences shown in Table 10.4 are of course dependent on the cost assumptions that were made. For example, if the cost increase factor were 15% for every doubling of the optical reach, rather than 25%, the normalized capital costs for the O-E-O network would be as shown in the final column of Table 10.4. (This should be thought of as the 600-km O-E-O-system costs remaining the same, and the 2,500-km optical-bypass-system costs increasing by a smaller factor. The costs shown in the final column of Table 10.4 are again relative to the costs of the optical-bypass-enabled system.) As expected, the benefits of optical bypass are realized sooner with this assumption.

The amount of regeneration is primarily a function of the system reach. As shown in the table, optical bypass with a 2,500-km optical reach consistently eliminated more than 90% of the regenerations as compared to an O-E-O system with a 600-km reach. With O-E-O technology, the average unprotected demand required 4.4 regenerations and the average protected demand required 13.6 regenerations. With optical-bypass technology, these numbers were 0.3 and 1.2, respectively. Thus, the marginal cost of adding traffic in the optical-bypass-enabled network was significantly lower.

10.3.1 Comments on Comparing Costs

The numbers in Table 10.4 are indicative of the cost trends in the O-E-O and optical-bypass-enabled architectures. There are numerous factors that make direct cost comparisons difficult.

First, the level of configurability is very different in the two architectures that were considered. It was assumed that the O-E-O nodes were not equipped with switches, such that the nodes had no means of automated configurability. In the optical-bypass-enabled network, the ROADMs and ROADM-MDs provided configurability; i.e., traffic could be configured as add/drop or through, or traffic could be routed in different directions through the node without requiring manual intervention. Thus, the optical-bypass solution was more configurable. Since static traffic was used in the study, these differences in configurability were not reflected in the costs shown in Table 10.4.

Second, the operating costs are likely to be significantly lower with optical-bypass technology. As discussed in Sect. 10.2.3, the number of required transponder and regenerator cards can be used as a rough measure of relative operational costs. With optical bypass, the amount of regeneration was reduced by over 90%, which would be accompanied by large savings in deployment costs and space and power requirements. Furthermore, carriers should be able to provision demands more quickly using optical-bypass technology so that revenues can be generated sooner.

Another consideration is that the "effective capacities" of the two networks are slightly different. As investigated in Sect. 5.11, optical bypass results in roughly 5% less network efficiency because of wavelength contention.

The time value of money also needs to be considered. The up-front costs of an optical-bypass system are higher; the savings afforded by optical bypass are realized over time. Thus, the cost savings are somewhat mitigated depending on the rate of traffic growth in the network and the rate of return on investment capital.

10.3.2 O-E-O Technology with Extended Optical Reach

Another architectural option is to deploy technology that supports an extended optical reach, but continue to use O-E-O technology at the nodes to avoid any issues with wavelength contention. For example, consider the combination of 1,500-km optical reach and O-E-O nodes (a 1,500-km reach is readily attainable with today's erbium amplifiers). In Reference Network 2, where the average link length is 450 km, this combination provided no cost benefit. The nodes are too densely packed to derive much advantage from increasing the reach of the O-E-O architecture from 600 to 1,500 km.

However, this architecture can provide some benefits in a less dense network. The study was repeated for Reference Network 3, with a corresponding line-rate traffic set. This network was first presented in Sect. 1.10, and is shown again in Fig. 10.2. The network has only 30 nodes, with an average link length of 700 km and a maximum link length of 1,450 km. A 600-km O-E-O architecture, a 1,500-km O-E-O architecture, as well as a 2,500-km optical-bypass architecture were considered. It was assumed that the amplification and transmission costs increase by 25% for every doubling of the reach.

The results are shown in Table 10.5. The comparison of the 600-km O-E-O architecture to that of the 2,500-km optical-bypass architecture is similar to that for Reference Network 2. However, in Reference Network 3, maintaining the O-E-O architecture but increasing the reach from 600 to 1,500 km did provide a *small* cost benefit as the traffic level increased. For example, when the network was full, the cost of the 1,500-km design was approximately 5% lower than the cost of the 600-km design. The biggest benefit is that approximately 40% of the regenerations were removed by increasing the reach from 600 to 1,500 km, which would result in operational-cost savings as well.

If the cost increase factor is 15% rather than 25%, then, when the network is full, the cost of the 1,500-km O-E-O architecture is 15% lower than the cost of the

Fig. 10.2 Reference Network 3. With a relatively low nodal density, increasing the reach from 600 to 1,500 km can provide a cost benefit even in a network based on O-E-O technology

Table 10.5 Optical-bypass versus O-E-O architecture using Reference Network 3 (30 nodes)

			2,500 km, optical bypass, configurable		600 km, O-E-O, non-configurable (25% cost increase factor)		1,500 km, O-E-O, non-configurable (25% cost increase factor)	
# of demds.	~Max. # of λs on any link	~Avg. # of λs on a link	# of regen.	Normal. cost	# of regen.	Normal. cost	# of regen.	Normal. cost
0	0	0	0	1.00	0	0.72	0	0.79
60	20	12	78	1.00	653	0.98	392	1.00
120	40	25	138	1.00	1,163	1.12	684	1.11
250	80	50	268	1.00	2,277	1.35	1,331	1.28

600-km O-E-O architecture, but 35% higher than the cost of the optical-bypass architecture. (These results are not shown in the table.)

The caveats mentioned above with respect to the comparison of O-E-O and optical-bypass architectures are relevant here as well.

10.4 Optimal Optical Reach

Given that extended optical reach, combined with optical-bypass elements, provides a significant reduction in the number of required regenerations, it is tempting to assume that the longer the reach, the greater the cost savings afforded by the system. However, increasing the reach typically requires amplifiers with more pumps or higher-powered pumps, transponders with more complex modulation formats

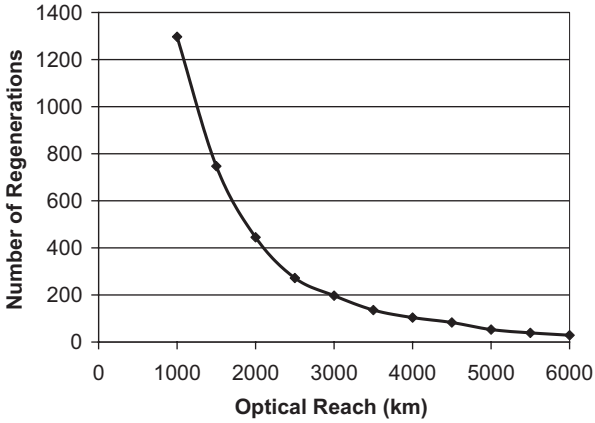


Fig. 10.3 Number of regenerations as a function of the optical reach, in Reference Network 2. Most of the regenerations are eliminated with 2,500-km optical reach

and stronger error-correcting capabilities, and components with stricter tolerances. After some point, the cost of increasing the reach is not fully offset by the savings due to further reductions in regeneration, leading to a concave "cost versus reach" curve, as investigated further in this section.

Reference Network 2 was used for the study, along with the baseline traffic set described in Sect. 10.2.2. Dedicated 1 + 1 protection was used for the demands that required protection (shared mesh restoration is also considered below). Optical-bypass-enabled designs were performed for the network, where the optical reach was increased from 1,000 to 6,000 km at 500-km intervals. Cost-optimized designs were performed for each reach setting.

The cost assumptions provided in Table 10.3 were used for the 2,500-km-reach design. As specified in Sect. 10.2.3, it was assumed that the cost of amplification and transmission increased by 25 % for every doubling of the optical reach; i.e., Formula 10.1 was used as the cost adjustment for a reach of R km.

As expected, the number of regenerations decreased with increasing optical reach, as shown in Fig. 10.3. Initially, the decrease is fairly steep, with roughly 80 % of the regenerations removed by increasing the reach from 1,000 to 2,500 km. After 2,500 km, the curve begins to level off, indicating a diminishing "rate of return" for increasing the reach even further. Note that even with a 6,000-km reach, not all regenerations were eliminated. For this particular network, a reach of 8,500 km would be required to remove all regenerations from both the working and protect paths.

The solid curve in Fig. 10.4 plots the normalized network capital cost as a function of the optical reach. The minimum cost was achieved with an optical reach in the range of 2,000–2,500 km. As this graph illustrates, continuing to increase the reach beyond this point resulted in a more costly network. The minimum-cost point is clearly dependent on the assumption that the cost increase factor is 25 % for every doubling of the reach. As extended-reach technology matures, the cost premiums will continue to decrease, which will shift the minimum-cost point to the right. For example, a cost increase factor of 15 % produces the cost versus reach curve shown

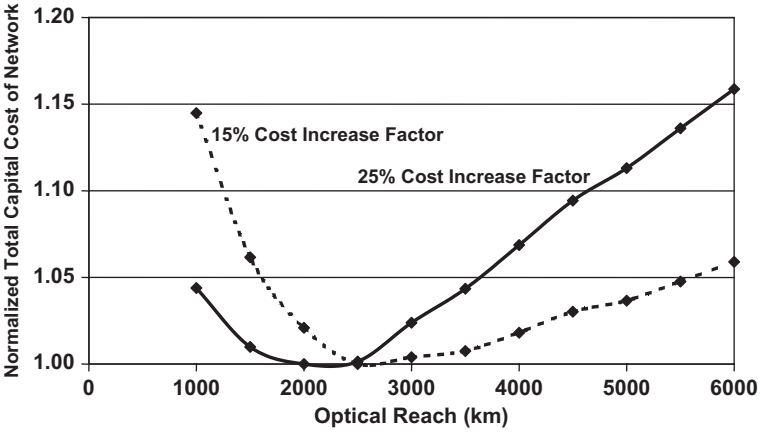


Fig. 10.4 Normalized network capital costs as a function of the optical reach. After some point, increasing the optical reach leads to a more costly network. The cost increase factor is the percentage increase in the cost of amplification and transmission for every doubling of the reach

by the dashed line in Fig. 10.4. With this assumption, the minimum cost was attained with a reach in the range of 2,500–3,000 km.

However, note that if connections need to be brought into the electrical domain for reasons other than regeneration, e.g., for grooming and/or shared protection, the optimal reach shifts to the left. To test this, another set of designs was performed where shared mesh protection was used instead of dedicated protection (more specifically, subconnection-based shared protection, with 20% of the nodes selected as protection hubs, was used; see Sect. 7.8 for the details of this protection scheme). The associated cost versus reach curve, assuming a 25% cost increase factor, is shown by the dashed line in Fig. 10.5 (the solid-line curve in this figure is the same as that in Fig. 10.4). With shared protection, the minimum-cost optical reach was reduced to 1,500 km.

As demonstrated by the results of Sect. 10.3, the economics of optical-bypass technology are dependent on the amount of traffic. While Fig. 10.4 shows the cost for a "full" network, Fig. 10.6 includes additional cost curves corresponding to lower amounts of traffic (assuming a 25% cost increase factor and dedicated protection). As expected, with lower traffic levels, the optimal reach from a cost perspective decreased. For example, with a level of traffic that is 50% smaller than the baseline traffic set, the lowest cost was achieved with an optical reach in the range of 1,000–1,500 km. (Each curve shown in Fig. 10.6 is independently normalized to 1.0 at its minimum value.)

In Simmons [Simm05], a similar study was performed for four other network topologies, ranging from a 16-node network to a 55-node network. The results for these other networks are similar to those shown here. Additionally, similar concave cost versus reach results were presented in Sardesai et al. [SaSR05].

As has been emphasized previously, the costs that are plotted are only the capital costs (of the optical layer), not the operational costs. Operational costs are related

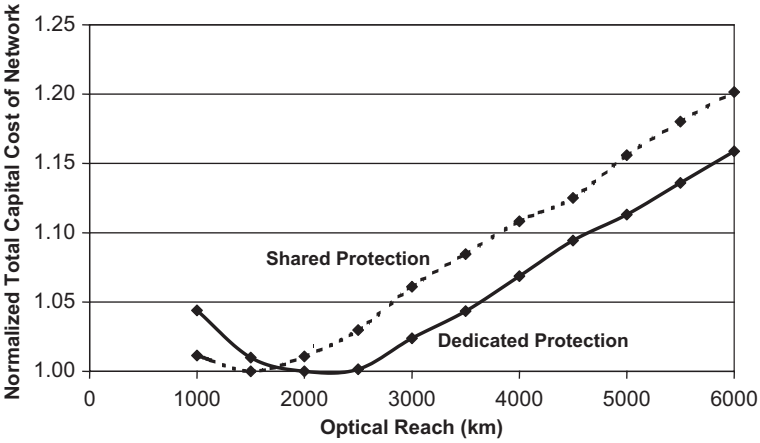


Fig. 10.5 Normalized network capital costs as a function of the optical reach, for dedicated protection and shared protection. The minimum-cost point shifts to the *left* with shared protection due to the protection capacity entering the electrical domain at the protection hubs. (Both *curves* assume a cost increase factor of 25%.)

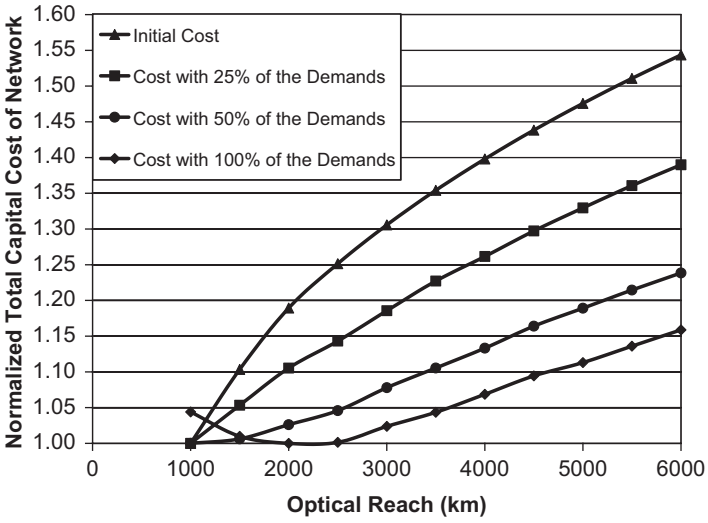


Fig. 10.6 Normalized network capital costs as a function of the optical reach, for different levels of traffic. With less traffic, shorter optical reach is more cost-effective. Note that each *curve* is independently normalized to 1.0 at its minimum value

to the number of regenerations; thus, the optimal reach may shift to the right when these costs are considered. However, given that the bulk of the regenerations have been removed with a reach of 2,500 km, including operational costs in the total cost is likely to have only a small effect on the overall minimum-cost point.

Similar studies can be performed with respect to power consumption versus optical reach. For example, to increase the optical reach of a dual polarization quadrature phase-shift keying (DP-QPSK) 100-Gb/s transponder, more complex operations are required in the digital signal processor. This reduces the amount of regeneration, but translates to higher power consumption. This particular trade-off was investigated in Rizzelli et al. [RMTP13] for an IP-over-optical network that included subrate demands and electronic grooming. Again, a convex curve was demonstrated, with the region of lowest power consumption occurring at a reach of approximately 1,000–1,300 km. This is in reasonable consonance with the *cost* versus reach study performed in Simmons [Simm05] for a network with subrate demands and electronic grooming, where the optimal reach was found to be 1,500 km. As noted earlier, the presence of electronic grooming, or shared protection, tends to shift the optimal reach towards shorter distances.

10.4.1 Add/Drop Percentage as a Function of Optical Reach

As noted in Sect. 2.6.1, some commercially available ROADMs and ROADM-MDs limit the number of wavelengths that can be added or dropped. With a *non-directionless* ROADM (or ROADM-MD), the limit is generally specified on a per-fiber basis; i.e., no more than P% of the fiber wavelengths can be added/dropped to/from each fiber entering the ROADM. With a *directionless* ROADM, the limit is generally specified on a per-node basis, which provides more flexibility; i.e., no more than P% of the total wavelengths entering a ROADM can be added/dropped at the node. A typical value used commercially for P is 50%.

The network designs that were performed to investigate the effect of optical reach on network cost are used here to determine how the add/drop percentage varies with reach and whether 50% add/drop is sufficient to meet the needs of a typical network. For these purposes, we focus on the scenario where 100% of the baseline demands were added to the network, so that the network was essentially full, and where demands requiring protection employed 1 + 1 dedicated client-side protection. The fiber capacity was assumed to be 80 wavelengths.

First, consider the case where the ROADMs and ROADM-MDs are *non-directionless*; the important figure of merit in this scenario is the add/drop percentage for each fiber. This percentage includes those wavelengths that add/drop for purposes of regeneration.

Figure 10.7 shows a histogram of the required fiber add/drop percentage when the optical reach was 2,500 km (there was one fiber pair per link; thus, the 77 links correspond to 154 fibers). *The add/drop percentages are relative to 80 wavelengths per fiber.* For example, the first bar in this histogram indicates that there were 81 fibers from which 10% or less add/drop was required (i.e., eight or fewer add/drop wavelengths). There were four fibers from which more than 50% add/drop was required. (No node needed more than 60% add/drop.) This indicates that a limit of 50% add/drop per fiber would not have been sufficient for all nodes. If the system

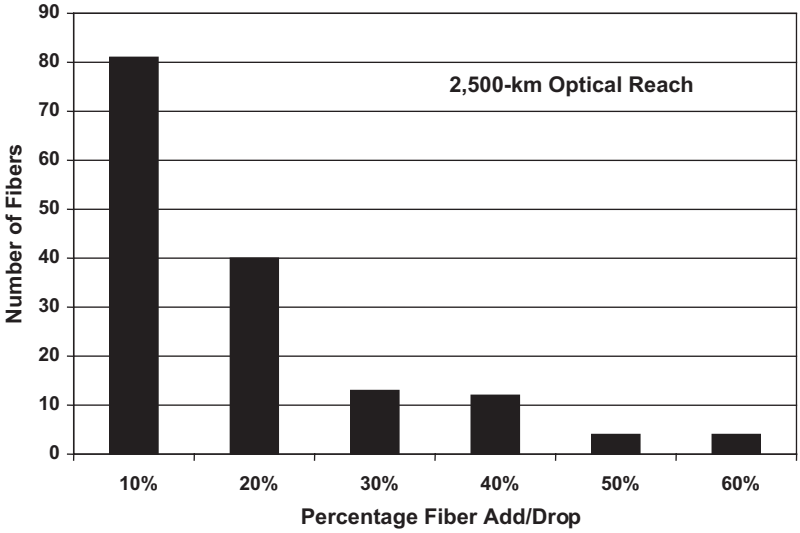


Fig. 10.7 A histogram of the required fiber add/drop percentage for the network design with dedicated protection, 2,500-km optical reach, and non-directionless ROADMs. For example, 81 of the fibers had 10% or less add/drop, whereas four fibers had between 50 and 60% add/drop. The add/drop percentages are relative to 80 wavelengths per fiber

had such a limit, then optical terminals would need to be deployed at the sites where the threshold was violated, or traffic would need to be routed differently to reduce the number of add/drop wavelengths from these particular four fibers.

With a 1,500-km optical reach, the 50% add/drop threshold was even more limiting, as shown by the histogram in Fig. 10.8. Here, 14 of the fibers required an add/drop percentage greater than 50% (i.e., more than 40 add/drop wavelengths).

If *directionless* ROADMs had been used, where the add/drop limit is on a per-node (rather than per-fiber) basis, then none of the nodes violated the 50% limit in the scenario with 2,500-km optical reach. However, with 1,500-km optical reach, two of the nodes required close to 60% add/drop and thus violated the 50% limit.

It is interesting to also consider the effect of shared protection on the per-fiber add/drop percentage. (With shared protection, the average utilization in the network decreased by 20%; however, the maximum link utilization decreased by less than 10%. Thus, with a fiber capacity of 80 wavelengths, the network was still close to full.) With 2,500-km reach and shared protection, the average per-fiber add/drop percentage increased, due to breaking the protection capacity into smaller segments to enable sharing. However, the *maximum* per-fiber add/drop percentage remained approximately the same so that there were still four fibers that required more than 50% add/drop. With 1,500-km reach and shared protection, the number of fibers requiring more than 50% add/drop was only 6, as opposed to 14 with dedicated protection. This decrease was due to sharing of the regenerations on the protection capacity, which resulted in fewer add/drop wavelengths.

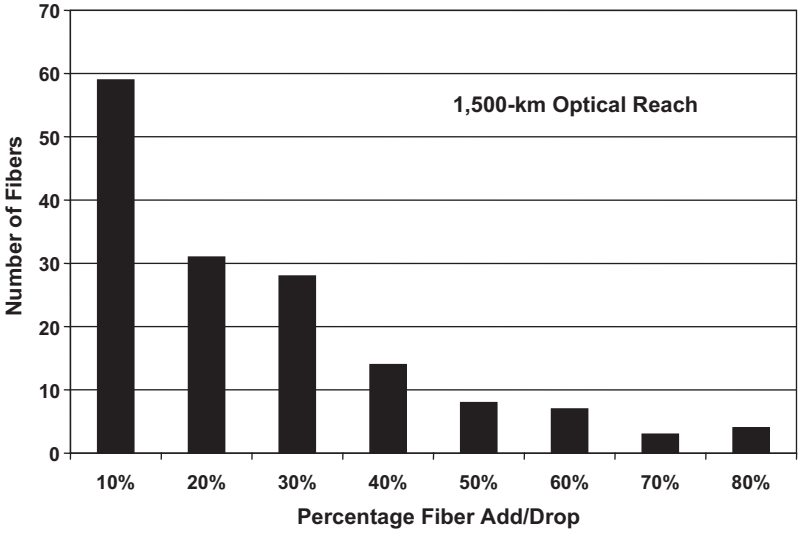


Fig. 10.8 A histogram of the required fiber add/drop percentage (relative to 80 wavelengths per fiber) for the network design with dedicated protection, 1,500-km optical reach, and non-directionless ROADMs

10.5 Optimal Topology from a Cost Perspective

The study of this section investigates the impact of the network topology on the overall network cost. The focus is on an "overbuild" scenario, where a new system is being deployed using existing fiber routes. The assumption is that various links are available to interconnect the network nodes, but that from a capacity and protection perspective, not all of the links are necessary. The cost impact of removing links from the topology depends on whether the system supports optical bypass or whether it is O-E-O based. (In a greenfield scenario, where there is no existing fiber, an equivalent question is how the links should be laid out to meet the capacity and protection requirements, while minimizing cost.)

To understand how the topology affects the network cost, consider the simple network shown in Fig. 10.9, and assume that Link CF is being considered for removal. Removing this link eliminates any associated in-line optical amplifiers and reduces the amount of nodal equipment required at the endpoints. For example, Node C and Node F would become degree-three nodes as opposed to degree-four nodes. However, removing this link also affects the routing. A connection that was routed over Link CF may now be routed over C-B-F or C-E-F, thereby increasing the number of hops in the path.

In an O-E-O network, because regeneration is required at every intermediate node, any extra path hops translate to more regeneration. Thus, whether removing Link CF reduces the network cost depends on whether the additional regeneration cost is offset by the reduced amplifier and nodal equipment costs. In an optical-bypass-enabled network, removing Link CF may not result in any extra regeneration,

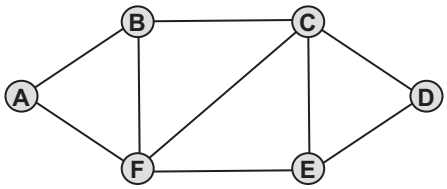


Fig. 10.9 If Link *CF* is removed from this topology, any amplifiers on the link are removed and less equipment is needed at Nodes *C* and *F*. However, traffic that had been routed on this link may now be routed over a path with more hops. In an O-E-O network, this will result in additional regeneration, whereas in an optical-bypass-enabled network it may not

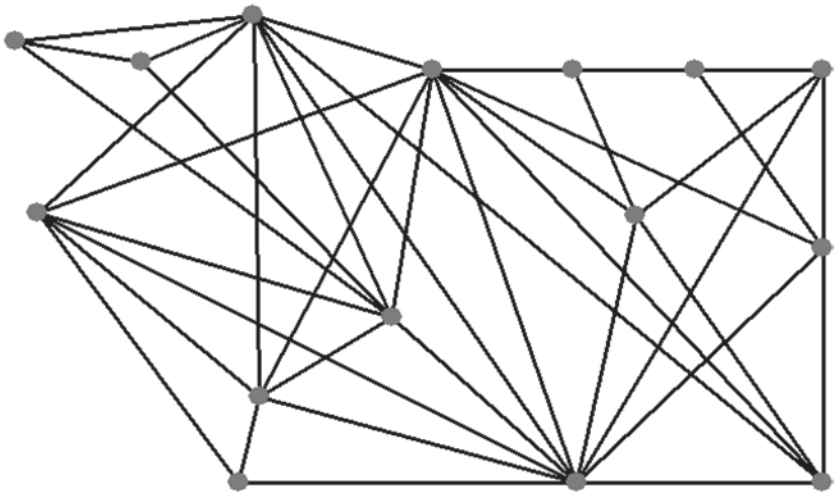


Fig. 10.10 Fifteen-node metro-core network used to study how the network capital costs change as links are removed

depending on the length of the links and the optical reach of the system. Thus, reducing the density of the topology is more likely to produce cost savings in a network with optical-bypass technology.

To quantify this effect, the 15-node metro-core network of Fig. 10.10 was used. This network was modeled after a carrier 13-node metro-core network, which is described in Wilkinson et al. [WBSK03]. With all links shown in Fig. 10.10, the average nodal degree is 5.33. It was assumed that all links were less than 100 km in length, such that no in-line amplifiers were required. A total of 800 Gb/s of subrate traffic was added to the network, where the traffic pattern was similar to that specified in Wilkinson et al. [WBSK03]. The line rate was assumed to be 10 Gb/s. All nodes were assumed to be equipped with OTN grooming switches so that backhauling was not required. All traffic was protected with 1 + 1 dedicated protection at the subrate level.

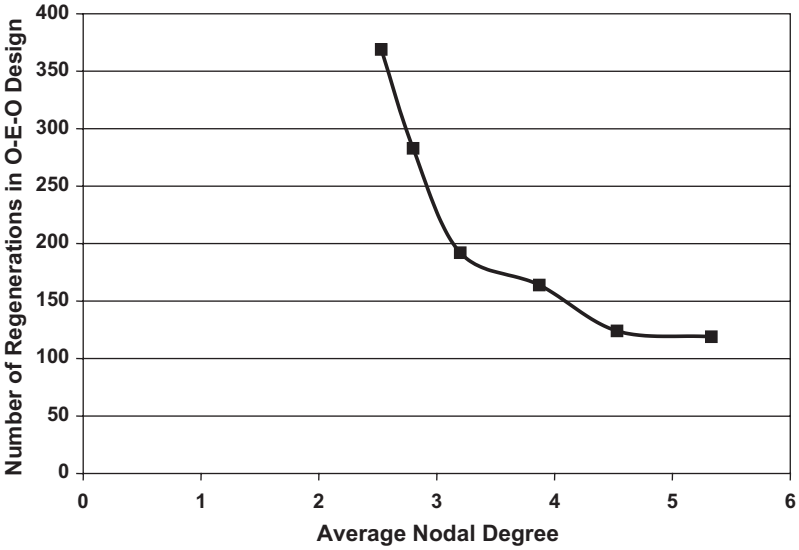


Fig. 10.11 Number of regenerations as a function of the average nodal degree for an O-E-O-based network, based on the original topology shown in Fig. 10.10

Two system designs were performed, one based on O-E-O technology and one based on optical-bypass technology. Both systems were assumed to support a 600-km optical reach. In the O-E-O design, any regenerations were implemented with regenerator cards and patch panels. In the optical-bypass design, the 600-km reach was long enough to eliminate all regenerations (however, some connections still entered the electrical domain at intermediate nodes for purposes of grooming). Links were systematically removed from the topology to determine the effect on required regeneration and network cost. The links that carried the least amount of traffic were selected for removal; however, a link was not removed if it was required to provide a diverse protection path for any connection.

Figure 10.11 plots the number of required regenerations in the O-E-O design as a function of the average nodal degree. As links were removed, the connections were forced to traverse more hops, leading to more regeneration. This was especially true once the average nodal degree was reduced below three. Similar results were presented in Saleh [Sale03] for a 12-node metro-core network. In the optical-bypass design, no regenerations were needed, even as links were removed.

Figure 10.12 plots the normalized capital cost of the network as a function of average nodal degree, for both the O-E-O and the optical-bypass designs. The relative costs shown in Table 10.3 were used. The cost of fiber was not included as it was assumed that the fiber routes already existed. (The required maximum fiber capacity is dependent on the topology, which might have a small effect on cost. This effect was not included in the costs shown in Fig. 10.12, as the effect on cost would be approximately the same for either system.)

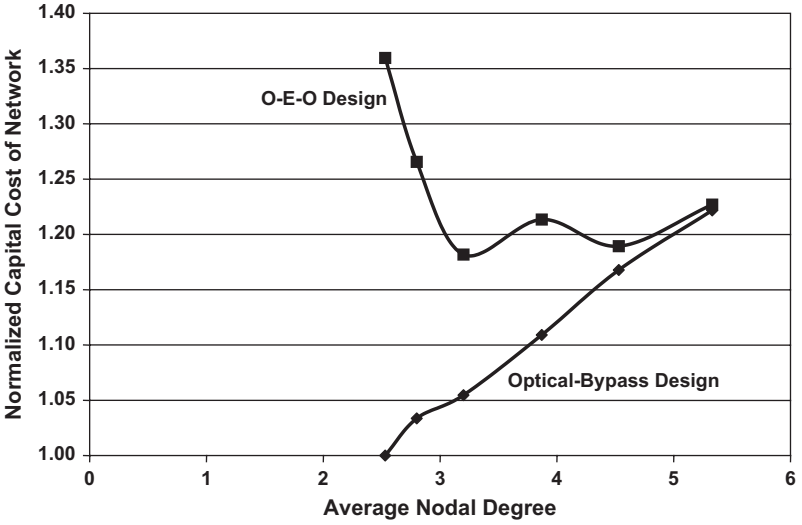


Fig. 10.12 Normalized capital cost of the network as a function of the average nodal degree for both O-E-O and optical-bypass designs, for the original topology shown in Fig. 10.10. With the O-E-O design, the cost wavers somewhat as links are initially removed, and then shoots up. With the optical-bypass design, the cost monotonically decreases

With the O-E-O system, the initial removal of links had only a small effect on the network cost. The savings afforded by the reduction in optical terminals at the nodes was approximately offset by the addition of more regeneration. After the average nodal degree was reduced below three, however, the sharp increase in regeneration more than outweighed the benefits of fewer optical terminals, leading to a spike in network cost. With the optical-bypass system, the network cost monotonically decreased as links were removed due to lower-cost network elements being used. There was no concomitant increase in regeneration to offset this cost reduction.

Note that with the full topology, the difference in cost between the O-E-O system and the optical-bypass-enabled system was very small. In an O-E-O network, each link essentially provides "fiber bypass," where a fiber deployed directly between two nodes can be used to avoid having to traverse intermediate nodes. With the full topology, the amount of fiber bypass achieved by the O-E-O network was significant, so that there was little cost disadvantage relative to a system with optical-bypass elements. However, with the topology pared down to an average nodal degree of 2.53, optical-bypass technology yielded a 35 % cost savings.

This study demonstrates that the topology density has a large impact on network cost. Of course, cost is not the only factor that needs to be considered when laying out the network topology. It is important that enough links be present to meet the capacity requirements. For example, in the full topology, with an average nodal degree of 5.33, the average and maximum link load were 8 and 20 wavelengths, respectively. When the average nodal degree was reduced to 2.53, the average and maximum link load increased to 30 and 51 wavelengths, respectively. If the

maximum supportable load on a link were 40 wavelengths, as is common in metro-core networks, this latter configuration would not be feasible. It is also important that the topology support the availability level required for each demand. Removing links results in longer paths for some connections, which increases their vulnerability to failures. It may also reduce the number of diverse paths for a connection, which may limit the ability to recover from multiple concurrent failures.

Another factor to consider is that of shared risk link groups (SRLGs). As described in Sect. 3.7.4, SRLGs arise when two or more links partially lie in the same fiber conduit. If this shared portion of the conduit is damaged, all of the corresponding links are likely to fail. Thus, the underlying physical diversity of the fibers must be considered when selecting the topology; otherwise, truly diverse paths may not be feasible for some of the demands.

While this study focused on a metro-core network, the same effect occurs in regional and backbone networks. With these geographically larger networks, removing a link not only simplifies the nodal equipment, but typically results in the removal of several in-line optical amplifiers as well. There tends to be less fiber routes from which to choose when designing a regional or backbone network; however, optical-bypass designs can generally benefit from the removal of some of the links, at least from a cost perspective.

10.6 Gridless Versus Conventional Architecture

Chapter 9 discussed a gridless optical-layer architecture, where, *in theory,* the spectrum is partitioned into arbitrarily fine "optical corridors" such that a demand is assigned a capacity that matches its desired service rate. This is in contrast to a conventional architecture, where the spectrum is partitioned uniformly across a fixed number of wavelengths. The line rate of these wavelengths is typically much higher than the service rate of the demands, thereby necessitating grooming. The grooming process packs multiple demands onto a wavelength so that the network bandwidth is used more efficiently. Grooming, which is typically accomplished in the electrical domain, accounts for the bulk of the cost and power consumption in a conventional network. (Transponders account for the majority of the cost and power consumption when considering only the optical layer.) The gridless approach, by better matching allocated capacity to service rate, aims to significantly reduce the amount of required electronic grooming. Another goal of the gridless architecture is to achieve greater bandwidth efficiency, resulting in lower capacity requirements.

This section probes the realism of the anticipated benefits of the gridless architecture. The first challenge is that the granularity of the ROADM filters cannot be made arbitrarily fine. In order to allow a ROADM to add/drop/switch each optical corridor independently of any others, there is a lower bound to the amount of spectrum that must be allocated to each corridor. The second challenge is that guardbands are needed between the optical corridors to reduce crosstalk between the corridors and to minimize spectral clipping due to the ROADM filters. The

Table 10.6 Assumed relative costs of the cards and ports

100 Gb/s TxRx	100 Gb/s muxponder	BVT	IP router port
Y	1.1 Y	1.2 Y	4 Y

combination of these restrictions greatly impacts the achievable benefits, as illustrated in the study below.

The study compares three architectures. First, it considers a conventional architecture, with 100-Gb/s wavelengths at fixed 50-GHz spacing combined with electronic grooming. Second, it considers a gridless architecture, where it is assumed that the optical corridors can be no finer than 12.5 GHz in spectral width. No electronic grooming is utilized in the gridless scenario. The third architecture is a hybrid model, where only a subset of the traffic is carried in optical corridors, with the remainder conventionally groomed and carried in 100-Gb/s wavelengths.

The study was run using Reference Network 2. All 60 nodes in the network were equipped with ROADMs or ROADM-MDs. The optical reach in all scenarios was assumed to be 2,500 km (trading off optical-corridor spectral efficiency for increased reach is also considered below). Regeneration was accomplished using back-to-back transponders. Alternative-path routing was used, combined with First-Fit for both wavelength assignment and spectrum assignment (see Sects. 5.5.1 and 9.5.2). (Routing and spectrum assignment that attempts to minimize fragmentation on a link or maximize the alignment of free spectrum on adjacent links in a gridless architecture provides little improvement [YZZX13].)

Given the continued expected prevalence of demands at 10 Gb/s and below, the following traffic model was utilized in the study: 60 % of the traffic was assumed to be uniformly distributed between 1.25 and 10 Gb/s, at 1.25-Gb/s increments; 30 % of the traffic was uniformly distributed between 11.25 and 40 Gb/s, at 1.25-Gb/s increments; the remaining 10 % of the traffic required 100 Gb/s. (These percentages are with respect to the total traffic bandwidth, not the total number of demands. Note that the nominal bit rate of ODUFlex is $N \times 1.25$ Gb/s, as shown in the OTN hierarchy of Table 1.3.)

A dynamic traffic model was assumed, with Poisson arrivals and exponential holding times. The average load in the network was roughly 16 Tb/s. To focus on the architectures as opposed to protection strategies, all of the traffic was assumed to be unprotected.

Enough capacity was allocated in each architecture to reduce the blocking rate to approximately 10^{-3}. (The demands that were blocked in the gridless and hybrid scenarios tended to be of longer distance and have higher bandwidth requirements than those that were blocked in the conventional architecture.) At this blocking rate, the 16 Tb/s of unprotected traffic required roughly half of the capacity of a conventional 80×100-Gb/s system on Reference Network 2.

The study focused on the number of transponders and network-side grooming ports required in each architecture, along with the total spectrum required. The assumed costs of the various types of cards are shown in Table 10.6. All costs are relative to a conventional 100-Gb/s transponder (TxRx). It is assumed that all electronic

Table 10.7 Card and port counts, total relative cost, and capacity requirements at 0.001 blocking, at an average of 16 Tb/s offered load (unprotected)

Architecture	# of 100 Gb/s TxRx's	# of 100 Gb/s muxponders	# of BVTs	# of IP router ports	Total relative card and port costs	Required spectrum (GHz)
Conventional	180	258	0	826	1.0	2,000
Gridless	0	0	2,501	0	0.8	2,900
Hybrid	176	168	801	489	0.9	2,100

grooming is accomplished in IP routers. The cost of the router chassis is amortized in the cost of the network-side IP router ports. The cost of a router port includes the cost of a wavelength-division multiplexing (WDM) transceiver. The cost assumed for a bandwidth variable transponder (BVT), which is needed to transmit optical corridors, is only a rough estimate as BVTs are still a speculative technology (see Sect. 9.7.3).

The results of the study are shown in Table 10.7. The card and port *counts* are shown in the second through fifth columns for each of the architectures. (Rather than reporting the maximum number of cards and ports ever needed over the course of the simulation, it was assumed that a somewhat smaller number would be deployed, which would produce a tolerable amount of blocking due to unavailable cards or ports.) The total card and port *cost* is shown in the second-to-last column, where the costs are relative to that of the conventional architecture. The final column indicates the amount of spectrum that was required on the most heavily loaded links.

It is emphasized that the relative costs in the table are rough estimates. The more important goals of the study are to show trends and to uncover the strengths and weaknesses of the various schemes in order to provide guidance with respect to future development efforts. (The costs of the ROADMs and amplifiers are not included, as these typically constitute a relatively small proportion of the network cost. The cost of the ROADMs would be somewhat higher for the gridless and hybrid models, to accommodate the variable-sized optical corridors.)

The next three sections provide more details of the various architectures. This is followed by a discussion of the results.

10.6.1 Conventional Grooming-Based Architecture

In the conventional architecture, all wavelengths were at 100 Gb/s, with 50-GHz spacing. Thus, the spectral efficiency was 2 bits/s/Hz. Twenty-five percent of the nodes were equipped with IP routers. The remainder of the nodes used muxponders to backhaul their subrate traffic to a node equipped with an IP router (see Sect. 6.6). (It was assumed that the muxponders were capable of handling the various service rates on the client side. If this is not possible, then a small multiplexing switch would be required at the non-grooming sites.) The 100-Gb/s demands were carried end to end without passing through any routers.

As noted above, 40 wavelengths, or 2,000 GHz of spectrum, were required on the most heavily utilized links to handle the offered traffic with a blocking probability of 10^{-3}. On average, the wavelengths carrying groomed traffic were approximately 80 % filled.

The required card and port counts are shown in the first row of Table 10.7. A relatively small amount of the transponders were used for regeneration. Minimal regeneration was needed due to the majority of the traffic being composed of subrate demands that undergo grooming in the electrical domain (i.e., regeneration/wavelength conversion are essentially obtained "for free" when the traffic is groomed).

10.6.2 Gridless Architecture

The gridless architecture carried all traffic in *flexible* optical corridors. The minimum and maximum spectral widths of a corridor were assumed to be 12.5 and 50 GHz, respectively, with a guardband required between any two corridors. It is not likely that ROADMs can efficiently operate on optical corridors finer than 12.5 GHz; i.e., if the filter passband width were required to be less than 12.5 GHz, then wider guardbands would be required, essentially nullifying any benefit. Given that the assumption regarding minimum corridor size was not overly aggressive, the guardband spectral width was chosen to be 6.25 GHz. (The size of the guardband was fixed, regardless of the size of the optical corridor. Even with this relatively small guardband, the deleterious effect on capacity efficiency was significant.) The system spectral efficiency was assumed to be 2 bits/s/Hz, as in the conventional architecture. Thus, an optical corridor of maximum bandwidth could transport a maximum of 100 Gb/s of traffic.

The optical corridors carried the traffic end to end, with no intermediate grooming. Demands with identical endpoints could be carried in the same corridor; this implicitly assumes that end-to-end multiplexing was performed. Optical corridors were allowed to expand in size, in multiples of 6.25 GHz, assuming there was available spectrum either above or below the existing corridor; i.e., all spectrum allocated to a corridor was required to be contiguous. When a demand was deleted from a corridor carrying other traffic, the corridor was permitted to contract in size, in multiples of 6.25 GHz. Empty corridors were deleted. (Gridless architectures often refer to "slot" size, due to the minimum granularity that is inherently imposed on the architecture by the underlying technology. One could consider the "slot" size used in the study to be 6.25 GHz, with a minimum optical-corridor size of two slots, and a guardband of one slot.) On average, the optical corridors were approximately 65 % full.

The results of the gridless architecture are shown in the second row of Table 10.7. The required spectrum on the most heavily loaded links was 2,900 GHz, which is 45 % *more* than what is required in the conventional architecture. This is largely due to the need for guardbands.

BVTs were the only type of card required in this architecture. Regeneration was accomplished via back-to-back BVTs. Roughly one third of the BVTs were required for regeneration. Due to the lack of any electronic grooming, the amount of required regeneration was significantly higher than that in the conventional architecture. Close to 5% of the regenerations were needed for purposes of "spectrum conversion" (analogous to wavelength conversion). As discussed in Chap. 9, one of the disadvantages of the gridless architecture is that the spectrum is typically partitioned differently on each of the links. As the network fills with traffic, finding spectrum that is free along every link of a subconnection becomes more difficult. This necessitates adding in regeneration to provide more flexibility in assigning spectrum. Periodic defragmentation of the network (see Sect. 9.6), which would alleviate some of spectral contention issues, was not performed.

One option that could be used to reduce the number of required BVTs is to reduce the capacity of an optical corridor in order to increase the optical reach (see the discussion of programmable transponders in Sect. 9.9). For example, if an optical corridor is only half full using a modulation format with a spectral efficiency of 2 bits/s/Hz, the spectral efficiency could be cut in half (e.g., using a modulation format with fewer bits per symbol), which allows a longer optical reach. To gauge how much benefit this potentially could provide, the gridless simulation was rerun using an optical reach of 3,500 km, rather than 2,500 km, for *all* optical corridors. This reduced the number of required BVTs by 20%. In practice, the benefits would be smaller than this, as trading off capacity for reach would be implemented for only a *subset* of the corridors. Thus, such programmability would not reduce the BVT count significantly. (Programmable transponders could be used in the conventional architecture as well; however, the extended optical reach would have even less of a beneficial effect in that scenario, due to the presence of electronic grooming.)

10.6.3 Hybrid Gridless/Grooming Architecture

The hybrid architecture combined aspects of both the conventional and gridless architectures, with both conventional wavelengths and optical corridors supported on a fiber. The spectrum was dynamically apportioned between the two transport mechanisms, depending on the current network state.

As specified for the conventional architecture of Sect. 10.6.1, the wavelengths were assumed to be at a 100-Gb/s line rate, with electronic grooming used to efficiently pack the wavelengths. Again, 25% of the nodes were assumed to be equipped with IP routers, necessitating that any traffic that was both sourced at a non-router site and designated for grooming be backhauled (the design rules for determining whether traffic underwent grooming are discussed below).

The spectral assumptions regarding optical corridors were the same as in Sect. 10.6.2. Guardbands were required between any two optical corridors or between a corridor and a conventional wavelength. Guardbands were not required between conventional wavelengths.

A *soft* partitioning of the spectrum was utilized, where the optical corridors were assigned spectrum starting at the low end and the conventional wavelengths were assigned spectrum starting at the high end. This partitioning minimizes the number of corridors interspersed with conventional wavelengths, which reduces the number of required guardbands. The spectral efficiency was assumed to be 2 bits/s/Hz across the system.

Although the largest-sized optical corridor was capable of carrying a 100-Gb/s demand, the design placed all 100-Gb/s demands in conventional wavelengths, to reduce the need for guardbands. With respect to the subrate traffic, any demand with a service rate of 15 Gb/s or higher was automatically carried in an optical corridor. (Various thresholds were considered; the threshold of 15 Gb/s produced the best results in terms of cost and capacity.) The lower-rate traffic could either be carried in a corridor, using end-to-end multiplexing, or could be groomed and carried on a conventional wavelength. The method that was employed depended upon the state of the network when the demand request arrived; i.e., the design process checked whether there was sufficient unfilled capacity in any existing corridors or wavelengths to carry the new demand.

There was approximately a 3:2 ratio between the number of wavelengths entering IP routers and the number of optical corridors. On average, the optical corridors were 85% filled and the wavelengths carrying groomed traffic were over 90% filled. Both average fill rates are higher than what was achieved in either of the two "pure" architectures. This is partially due to having the option of placing the low-rate demands in either a corridor or a groomed wavelength. For example, rather than provision a new "groomed" wavelength for an arriving low-rate demand, the demand could be placed in an existing corridor with available capacity. Additionally, the minimum bandwidth of an optical corridor (i.e., 25 Gb/s) was a better match for the traffic that was always placed in corridors (i.e., demands requiring 15 Gb/s and higher).

The results of the hybrid architecture are shown in the third row of Table 10.7. Slightly more spectrum was required on the most heavily loaded links as compared to the pure conventional architecture (2,100 vs. 2,000 GHz).

Regeneration again accounted for roughly one third of the required BVTs; about 15% of the regenerations were for purposes of spectral conversion (spectral defragmentation was not performed).

10.6.4 Discussion

First, we examine the results with respect to the need for electronic grooming (i.e., IP routers). As expected, the router ports account for the bulk of the cost in the conventional architecture. The gridless architecture, where all electronic grooming was removed, reduced the cost by 20%. However, it produced an inefficient network that required 45% more spectrum. Ultimately, this would necessitate earlier bandwidth upgrades, likely resulting in a *more* costly network.

The inefficient use of bandwidth in the gridless architecture is largely due to the need for guardbands between optical corridors. Further exacerbating the situation is that the lack of any intermediate grooming resulted in optical corridors that were on average only 65% full, which necessitated the deployment of more corridors, and hence more guardbands.

The hybrid architecture, by employing some amount of electronic grooming, significantly improves the network efficiency. Approximately the same amount of spectrum was needed on the most heavily loaded links as in the conventional architecture. The improved fill rates of the corridors and of the conventional wavelengths balanced out the need for guardbands. Forty percent fewer IP ports were required in the hybrid architecture as compared to the conventional architecture. This compensates for the more costly BVT cards and the need for more regeneration, such that a 10% reduction in total card and port cost was achieved relative to the conventional architecture.

One metric to directly compare costs among the architectures is to calculate *cost/(capacity efficiency),* where lower values are desired. Normalizing to 1.0 for the conventional architecture, this cost metric yields values of 1.15 and 0.95 for the gridless and hybrid architectures, respectively. Thus, only the hybrid architecture provided an advantage over the conventional architecture with respect to this metric.

Another statistic of interest is the large number of BVTs required in the gridless architecture, i.e., approximately 2,500. As discussed earlier, it is possible to reduce the number of required BVTs by using programmable technology that allows trading off capacity for reach (to reduce the number of regenerations), and by periodically defragmenting the network to lessen the need for spectral conversion. Let us assume that these methods combine to reduce the number of required BVTs in the gridless architecture by 15% (based on simulation results, the reduction is unlikely to be more than this). The cost metric for the gridless architecture would be reduced from 1.15 to roughly 1.0. Thus, there is still no cost advantage over the conventional architecture. Furthermore, while programmability and defragmentation reduce the capital costs, they do impose additional operational complexity.

Another means of reducing the BVT count is to make use of virtual transponders, as described in Sect. 9.7.3. This proposed technology would more flexibly allow a single BVT to be used for multiple corridors, where the corridors do not have to be between the same two endpoints. This potentially reduces the BVT count by a significant amount, making it a worthwhile technology to pursue.

10.6.4.1 Effect of Increased Traffic Level

Clearly, the traffic profile and the traffic level that was used in the study had some impact on the results. Note that as the amount of traffic in a network grows, while the number of network nodes remains approximately fixed, the average amount of traffic between node pairs increases. Thus, end-to-end multiplexing should be more effective at packing the optical corridors. To investigate this effect further, the

Table 10.8 Card and port counts, total relative cost, and capacity requirements at 0.001 blocking, at an average of 32 Tb/s offered load (unprotected)

Architecture	# of 100 Gb/s TxRx's	# of 100 Gb/s muxponders	# of BVTs	# of IP router ports	Total relative card and port costs	Required spectrum (GHz)
Conventional	284	420	0	1,540	1.0	3,700
Gridless	0	0	4,055	0	0.7	4,900
Hybrid	274	284	1,370	852	0.8	3,800

simulations were rerun for all three architectures with the offered load doubled; i.e., the average load in the network was 32 Tb/s. (In the gridless architecture, where end-to-end multiplexing was used, doubling the offered load is somewhat similar to, though not equivalent to, doubling the average service rates of the demands.) The results are shown in Table 10.8. The gridless architecture still required more spectrum than the other two architectures; however, the premium was only 35 %, as opposed to 45 % with 16 Tb/s of traffic. This is partly because the average fill rate of a corridor increased from roughly 65 to 75 %. Additionally, the average bandwidth of a corridor increased, such that the guardbands consumed a lower percentage of the capacity. Using the cost metric specified above, and normalizing to 1.0 for the conventional architecture, yields approximately 0.9 and 0.8 for the gridless and hybrid architectures, respectively. While the relative performance of the gridless scheme improves, the hybrid architecture remains more cost effective with respect to this metric.

An alternative architectural model to consider is where all of the traffic is carried in optical corridors but where electronic grooming is utilized to better pack the optical corridors (i.e., the groomed traffic is carried in corridors, not conventional wavelengths). However, given that grooming produces high fill rates of conventional 100-Gb/s wavelengths, there is no real *capacity* advantage to carrying the groomed traffic in an optical corridor instead. Furthermore, carrying the groomed traffic in corridors would require the costlier BVTs and would require guardbands. The one advantage is that somewhat less grooming would be required to efficiently pack the corridors, due to the smallest-sized corridors having a bandwidth of 12.5 GHz as opposed to 50 GHz for a conventional wavelength. Note that there is an inherent trade-off. The smaller the average size of a corridor, the less grooming is required to fill the corridors; however, it also implies that more BVTs and guardbands are required.

Overall, we conclude that the gridless architecture likely does not obviate the need for grooming in the network. Some amount of grooming is warranted, although the level depends on the traffic profile. Second, while it had been anticipated that a gridless architecture would be more capacity efficient, this benefit may not be realized because of the need for guardbands. Third, unless technologies such as virtual transponders are employed, the cost savings as compared to a conventional architecture are not likely to be very significant.

The next section investigates an alternative architecture for reducing the amount of electronic grooming; the scheme is targeted at networks with very high traffic loads.

10.7 Optical Grooming in Edge Networks

As noted in Sect. 10.6.4, as the level of network traffic grows while the number of network nodes remains fixed, the average amount of traffic between node pairs increases. Thus, an increasing amount of traffic can be efficiently packed into wavelengths at the edge of the network without requiring further grooming in the core (i.e., backbone network). This implies that efficient grooming in the edge networks (i.e., regional and metro-core networks) can be used to offload much of the grooming burden from the core network. Furthermore, edge grooming may be able to take advantage of optical-grooming techniques. As discussed in Sect. 6.10, some of the optical-grooming strategies that are being developed are more suitable for edge networks than they are for large core networks, due to these techniques requiring scheduling and/or collision management.

To investigate the efficacy of grooming at the edge, a 100-Tb/s aggregate demand scenario was considered. Conventional wavelengths were assumed. In each design, a certain percentage of the traffic was limited to being groomed solely at the network edge; i.e., intermediate grooming in the backbone core was not permitted for this traffic. This corresponds to end-to-end multiplexing with respect to the core network. The remaining traffic was allowed to undergo intermediate grooming in the core as usual. (Note that this architecture is somewhat similar to the hybrid gridless/grooming architecture of Sect. 10.6.3, except that conventional wavelengths were used for all traffic.)

Given that a portion of the traffic was limited to grooming at the network edge, the wavelength line rate plays an important role in the efficacy of the scheme. A finer line rate generally yields more efficient packing of the wavelengths; however, it generally implies a lower spectral efficiency (SE). (A finer line rate also implies that fewer flows are multiplexed on a given wavelength, such that the aggregate traffic is burstier; see Sect. 6.10.) Two different line-rate scenarios were tested in the study: 40 and 100 Gb/s. As the results below indicate, from the perspective of packing efficiency, 40-Gb/s wavelengths are preferred; however, this line rate likely limits the SE to less than 2 bits/s/Hz. While the packing efficiency decreased somewhat with 100-Gb/s wavelengths, an SE of 4 bit/s/Hz may be achievable. (Current 100-Gb/s systems have an SE of 2 bits/s/Hz. It is conceivable that conventional 100-Gb/s wavelengths could be utilized with 25-GHz channel spacing, without the use of guardbands, yielding a spectral efficiency of 4 bits/s/Hz.)

An explicit cost analysis was not performed for either of the two designs because the cost of optical grooming and the cost of the edge/core interface (to be discussed below) would be too speculative.

In the scenario with 40-Gb/s line rate, roughly 95 % of the traffic was limited to being groomed solely in the edge networks. The remaining 5 % of the traffic was allowed to undergo intermediate grooming in the core network. Despite the limited grooming for 95 % of the traffic, this architecture was still very effective in packing the wavelengths. The overall network capacity requirements increased by only 1 % as compared to a design where all traffic is eligible for intermediate grooming in the core. Furthermore, assuming that all of the traffic was IP and that the traffic that was solely groomed at the edge completely bypassed the IP routers (e.g., via the use

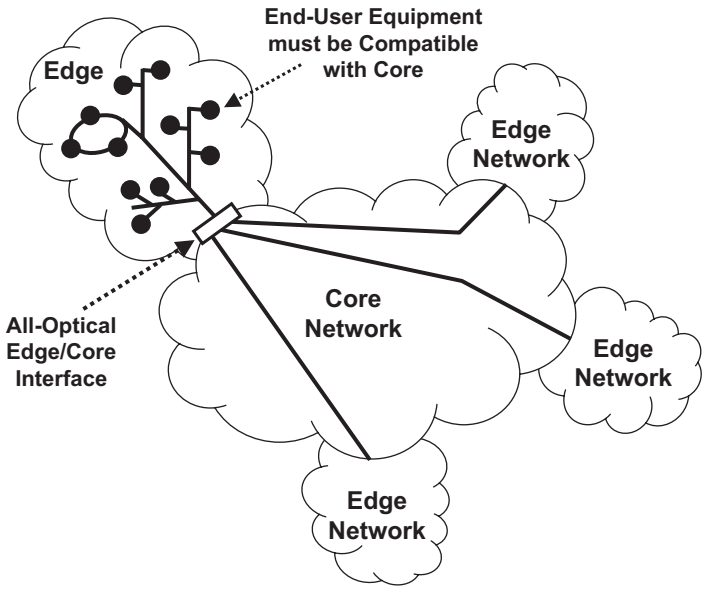


Fig. 10.13 If the edge/core interface is all-optical, then the end users must be equipped with transponders that are compatible with the stringent requirements of the core network. (Adapted from Saleh and Simmons [SaSi06]. © 2006 IEEE)

of optical grooming in the edge networks), then the average required IP router size was reduced by 85 %.

In the second scenario, with 100-Gb/s line rate, only 80 % of the traffic was limited to being groomed solely at the network edge, with the remaining 20 % allowed to undergo intermediate grooming in the core network. More grooming was permitted due to the greater challenge in efficiently packing wavelengths with a higher line rate. With this configuration, the network capacity requirements increased by roughly 5 %; the average required IP router size was reduced by about 65 %, again assuming that the traffic groomed at the edge completely bypassed the IP routers. As expected, the scheme is not as effective with the higher line rate; however, a significant benefit can still be realized.

These results show that grooming at the edge is potentially a viable scheme for the bulk of the traffic in a network with a very high level of traffic. In the remainder of this section, it is assumed that the grooming in the edge networks is performed in the optical domain; e.g., schemes based on fast optical switching or on passive optical broadcasting may be used to aggregate the traffic. (The OBS scheme of Sect. 6.10.4 and the TWIN scheme of Sect. 6.10.5 are possible aggregation architectures that may be suitable for this application.) With optical aggregation at the edge, consideration must be given to the architecture of the interface between the edge and core networks. The following discussion follows that of Saleh and Simmons [SaSi06].

First, assume that traffic is routed in the optical domain between the edge and core networks without any O-E-O conversion, as shown in Fig. 10.13. The advantage

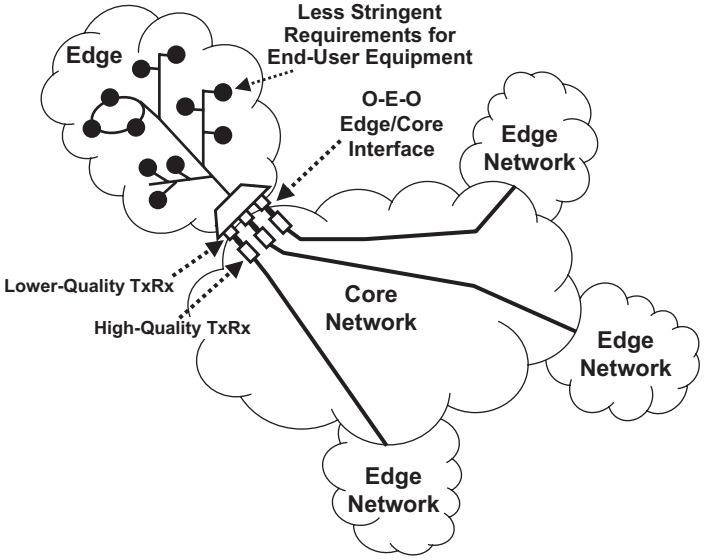


Fig. 10.14 If the edge/core interface is O-E-O based, then lower-cost transponders (*TxRx*'s) can be used in the edge network. (Adapted from Saleh and Simmons [SaSi06]. © 2006 IEEE)

of this approach is that transponders are not needed at the edge/core interface. However, the traffic sources in the edge network would need to be equipped with transponders that are compatible with the transmission system of the core network. Due to the stringent performance requirements in the core, such equipment may be too expensive for customer premises equipment. Another possible disadvantage of an all-optical interface is that any voids between the multiplexed data bursts will remain as signal voids in the core network, which could possibly cause optical amplifier transients.

A second option is to isolate the transmission systems of the edge and core networks via an O-E-O interface, as shown in Fig. 10.14. Edge networks generally have shorter optical reach and lower capacity requirements than the core; thus, lower-cost transponders can be used for transmission solely within the edge network. Transponders at the edge/core interface would be used to convert the aggregated traffic to optical signals that meet the stringent requirements of the core. Furthermore, by converting the aggregated traffic to the electrical domain, bit stuffing can be used to eliminate any voids in the optically multiplexed signal. The trade-off is the cost of the transponders at the edge/core interface. However, these transponders are required for the aggregated traffic, not for the individual traffic streams; thus, this scheme may be ultimately of lower cost than requiring high-quality transponders at all of the traffic sources.

The various optical grooming schemes discussed in Sect. 6.10 need to manage resource contention issues. Due to the complexities of scheduling across large networks, it is assumed that scheduling would be performed independently within each

edge network attached to the core. However, this can lead to conflicts; e.g., two different regions could send traffic on the same wavelength that arrives at the same destination regional network at the same time. The architecture of the edge/core interface affects how such conflicts are managed.

With an all-optical interface, rapidly tunable all-optical wavelength converters could be used to shift one of the conflicting streams to another wavelength. However, this could potentially produce two simultaneous streams that are destined for the same end user. Thus, each customer premise would need to be capable of receiving multiple wavelengths at the same time, e.g., with an array receiver.

With an O-E-O edge/core interface, the wavelengths in the edge networks can be chosen independently from those in the core network such that wavelength contention issues can be minimized (i.e., wavelength conversion can be obtained by tuning the transmitter on the destination edge-network side to a different wavelength). However, an array receiver would still be needed at the customer premises. Alternatively, once the signal is in the electrical domain at the edge/core interface, electronic buffering could be used to resolve conflicts, eliminating the need for array receivers. Clearly, there are still issues that need to be resolved with such optical grooming architectures. However, these schemes provide another example of how operating primarily in the optical domain, with O-E-O conversion at key junctures, can provide an opportunity for reduced electronics, along with the attendant savings in capital and operational costs. It is likely that such applications of optical networking will be the key enablers of continued scalable network growth.

10.8 General Conclusions

As the studies of this chapter have shown, there is no single technology that is optimal in all scenarios. The transmission and switching technologies for a network need to be chosen based on factors such as the topology, the amount of traffic, and the desired amount of network agility.

Electronic-based and optical-based technologies clearly have their associated strengths. Thus far, there has been a trend for optics to scale better than electronics as the line rates increase; however, there are still functions that are best performed in the electrical domain. Thus, it is likely that networks will remain a mix of technologies.

References

[BaGe06] R. Batchellor, O. Gerstel, Cost effective architectures for core transport networks. *Proceedings, Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC'06)*, Anaheim, CA, 5–10 March 2006, Paper PDP42

[RKDG13] F. Rambach, B. Konrad, L. Dembeck, U. Gebhard, M. Gunkel, M. Quagliotti, L. Serra, V. López, A multilayer cost model for metro/core networks. J. Opt. Commun. Netw. **5**(3), 210–225 (March 2013)

[RMTP13] G. Rizzelli, A. Morea, M. Tornatore, A. Pattavina, Reach-related energy consumption in IP-over-WDM 100G translucent networks. J. Lightwave Technol. **31**(11), 1828–1834 (1 June 2013)

[Sale03] A.A.M. Saleh, Defining all-optical networking and assessing its benefits in metro, regional and backbone networks. *Proceedings, Optical Fiber Communication (OFC'03)*, Atlanta, GA, 23–28 March 2003, Paper WQ1

[SaSi06] A.A.M. Saleh, J.M. Simmons, Evolution toward the next-generation core optical network. J. Lightwave Technol. **24**(9), 3303–3321 (September 2006)

[SaSR05] H.P. Sardesai, Y. Shen, R. Ranganathan, Optimal WDM layer partitioning and transmission reach in optical networks. *Proceedings, Optical Fiber Communication (OFC'05)*, Anaheim, CA, 6–11 March 2005, Paper OTuP4

[Simm04] J.M. Simmons, An introduction to optical network design and planning. *Optical Fiber Communication (OFC'04)*, Los Angeles, CA, 22–27 February 2004, Short Course 216

[Simm05] J.M. Simmons, On determining the optimal optical reach for a long-haul network. J. Lightwave Technol. **23**(3), 1039–1048 (March 2005)

[Verb05] S. Verbrugge et al., Modeling operational expenditures for telecom operators. *Proceedings, Conference on Optical Network Design and Modeling (ONDM'05),* Milan, Italy, 7–9 February 2005, pp. 455–466

[WBSK03] S.T. Wilkinson, E.B. Basch, V. Shukla, P. Kubat, S. Raguram, P. Limaye, SONET mesh network architecture. *Proceedings, National Fiber Optic Engineers Conference (NFOEC'03)*, Orlando, FL, 7–11 September 2003, pp. 293–302

[YZZX13] Y. Yin, H. Zhang, M. Zhang, M. Xia, Z. Zhu, S. Dahlfort, S.J.B. Yoo, Spectral and spatial 2D fragmentation-aware routing and spectrum assignment algorithms in elastic optical networks. J. Opt. Commun. Netw. **5**(10), A100–A106 (October 2013)

Chapter 11
C-Code for Routing Routines

11.1 Introduction

The C-code for several useful routing functions is provided in this chapter.[1] The first set of routines uses the breadth-first search method to find a shortest path between a source and a destination. The code for these routines follows the algorithm outlined in Bhandari [Bhan99]. The second set of routines finds the *K*-shortest paths between a source and a destination. This code follows the equivalence method of Hershberger et al. [HeMS03]. The third set of routines finds *N* mutually disjoint paths between a source and a destination. This follows the algorithm outlined in Bhandari [Bhan99]. Various parameters can be set to indicate whether the paths should be link disjoint or link-and-node disjoint. In scenarios where *N* mutually disjoint paths do not exist, the function can be used to find the *N* maximally disjoint paths. The last two sets of routines find a multicast tree among a set of nodes. The first heuristic follows Kou et al. [KoMB81], with the enhancement of Waxman [Waxm88]; the second heuristic follows Takahashi and Matsuyama [TaMa80].

For the most part, memory is pre-allocated for the routines. The maximum size of the network can be adjusted if necessary by redefining the appropriate parameters, which appear at the start of the code. A small *main* function is provided to demonstrate how to specify the network topology.

The routines have been coded for clarity as opposed to run-time speed, although all of the routines demonstrate very good performance on realistic telecommunications networks. A minimal amount of error checking has been added.

Disclaimer *No warranty of any kind is expressed or implied. You use the code at your own risk. In no event shall the author, the agents of the author, or the publisher be liable for data loss, loss of profits, loss of benefits, or other incidental or consequential damages while using this code.*

[1] All code is copyright 2003–2014 Monarch Network Architects LLC.

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4_11,

11.2 Definitions

```
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <limits.h>

/* Adjust these parameters as needed */
#define MaxNodes 120
#define MaxLinks MaxNodes*MaxNodes  // enough links to allow full bi-directional mesh of nodes
#define MaxNodeName 30
#define MaxNodeDegree MaxNodes   // enough to allow full bi-directional mesh of nodes
#define MaxPathHops MaxNodes

#define FALSE 0
#define TRUE 1

/* when performing graph transformations, need to add dummy nodes and links */
#define MaxNodesWithDummies MaxNodes+MaxPathHops
#define MaxLinksWithDummies MaxLinks+2*MaxPathHops

const double CommonNodePenalty = 1500000.0;       /* make greater than sum of link distances */
const double CommonLinkPenalty = 160000000.0;     /* make much greater than
                                                        CommonNodePenalty */
const double INFINITY = 17000000000.0;  /* make much greater than CommonLinkPenalty */
const double SMALL = .0001;  /* small number relative to link distances */

typedef unsigned short USHORT ;
typedef char BOOL ;

typedef struct tagNodeT {
    char Name[MaxNodeName+1];
    USHORT    OutgoingLinks[MaxNodeDegree];
    USHORT    IncomingLinks[MaxNodeDegree];
    USHORT    NumOutgoingLinks;
    USHORT    NumIncomingLinks;
    } NodeT, *NodeTP;

typedef struct tagLinkT {
    USHORT          LinkNode1; /* node at one end of link */
    USHORT          LinkNode2; /* node at other end of link */
    BOOL            Status; /* active link or not (true or false) */
    double          Length; /* any additive metric could be used here */
    } LinkT, *LinkTP;
```

```
typedef struct tagNetworkT {
    NodeT       Nodes[MaxNodesWithDummies];
    USHORT      NumNodes;
    LinkT       Links[MaxLinksWithDummies];
    USHORT      NumLinks;
    } NetworkT, *NetworkTP;

typedef struct tagPathT{
    USHORT      PathHops[MaxPathHops];
    USHORT      NumPathHops;
    double      PathDistance;
    } PathT, *PathTP;

typedef struct tagMST_T{
    USHORT      MSTLinks[MaxNodes];
    USHORT      NumMSTLinks;
    double      MSTDistance;
    } MST_T, *MST_TP;
USHORT PredecessorNode[MaxNodesWithDummies];
USHORT PredecessorLink[MaxNodesWithDummies];
double NodeDistance[MaxNodesWithDummies];
BOOL Marked[MaxNodesWithDummies];
NetworkT TempNetwork,  TempNetwork2;   /* use for graph transformations */
BOOL BreadthFirstSearchShortestPath (NetworkTP NetworkP, USHORT SourceNode,
    USHORT DestinationNode, PathTP ShortestPath);
BOOL BreadthFirstSearchRelax (NetworkTP NetworkP, USHORT NodeA, USHORT Link,
    USHORT FinalDestination);
USHORT KShortestPaths (NetworkTP NetworkP, USHORT Source, USHORT Dest, USHORT K,
    PathTP KPaths);
USHORT NShortestDiversePaths (NetworkTP NetworkP, USHORT Source, USHORT Destination,
    USHORT NumDisjointPaths, BOOL LinkDisjointOnly, BOOL CommonLinksAllowed,
    BOOL CommonNodesAllowed, PathTP NPaths);
BOOL SteinerTreeHeuristic (NetworkTP NetworkP, USHORT * NodeSet,
    USHORT NumNodesInSet, MST_TP Tree);
BOOL SteinerTreeHeuristic2 (NetworkTP NetworkP, USHORT * NodeSet,
    USHORT NumNodesInSet, MST_TP Tree);
USHORT AddLinkToTopology (NetworkTP NetworkP, USHORT Node1, USHORT Node2,
    double Length, BOOL ReverseLinkStatus);
USHORT GetReverseLink (NetworkTP NetworkP, USHORT LinkID);
main ()
{
    USHORT n, NodeSet[10];
    NetworkT network;
```

```
    PathT Paths[10];
    MST_T Tree;

    /* create a simple network to demonstrate the functions */
    strcpy(network.Nodes[0].Name, "A");
    strcpy(network.Nodes[1].Name, "B");
    strcpy(network.Nodes[2].Name, "C");
    strcpy(network.Nodes[3].Name, "D");
    strcpy(network.Nodes[4].Name, "Z");
    network.NumNodes = 5;
    network.NumLinks = 0;
    for (n=0; n<network.NumNodes; n++)
        network.Nodes[n].NumIncomingLinks = network.Nodes[n].NumOutgoingLinks = 0;

    AddLinkToTopology(&network, 0, 1, 10.0, TRUE);
    AddLinkToTopology(&network, 1, 2, 10.0, TRUE);
    AddLinkToTopology(&network, 2, 3, 10.0, TRUE);
    AddLinkToTopology(&network, 3, 4, 10.0, TRUE);
    AddLinkToTopology(&network, 0, 2, 10.0, TRUE);
    AddLinkToTopology(&network, 2, 4, 10.0, TRUE);

    BreadthFirstSearchShortestPath(&network, 0, 4, &Paths[0]);

    KShortestPaths(&network, 0, 4, 10, Paths);

    NShortestDiversePaths(&network, 0, 4, 3, FALSE, TRUE, TRUE, Paths);

    strcpy(network.Nodes[5].Name, "E");
    strcpy(network.Nodes[6].Name, "F");
    network.Nodes[5].NumIncomingLinks = network.Nodes[5].NumOutgoingLinks = 0;
    network.Nodes[6].NumIncomingLinks = network.Nodes[6].NumOutgoingLinks = 0;
    network.NumNodes = 7;

    AddLinkToTopology(&network, 3, 5, 10.0, TRUE);
    AddLinkToTopology(&network, 4, 6, 10.0, TRUE);

    NodeSet[0] = 0;
    NodeSet[1] = 2;
    NodeSet[2] = 5;
    NodeSet[3] = 6;

    SteinerTreeHeuristic(&network, NodeSet, 4, &Tree);

    SteinerTreeHeuristic2(&network, NodeSet, 4, &Tree);
}
```

11.3 Breadth-First Search Shortest Paths

```
/*****************************************************************/
/*      Code for Breadth First Search Shortest Paths             */
/*      Finds the shortest path from source to destination       */
/*      Returns the shortest path in ShortestPathP               */
/*      Returns TRUE/FALSE depending on whether a path exists    */
/*****************************************************************/

BOOL BreadthFirstSearchShortestPath (NetworkTP NetworkP, USHORT SourceNode,
    USHORT DestinationNode, PathTP ShortestPathP)
{
    USHORT n, m, numNodesOnListCurrent, numNodesOnListNew, node, link;
    USHORT nodesOnList1[MaxNodesWithDummies], nodesOnList2[MaxNodesWithDummies];
    USHORT* currentNodeList;
    USHORT* newNodeList;
    USHORT* temp;
    BOOL addedNodeToList1[MaxNodesWithDummies];
    BOOL addedNodeToList2[MaxNodesWithDummies];
    BOOL* currentAdded;
    BOOL* newAdded;
    BOOL* temp2;
    PathT tempPath;

    for (n = 0; n < NetworkP->NumNodes; n++) {
        NodeDistance[n] = INFINITY;
        PredecessorNode[n] = PredecessorLink[n] = USHRT_MAX;
        addedNodeToList1[n] = addedNodeToList2[n] = FALSE;
        }

    NodeDistance[SourceNode] = 0.0;
    numNodesOnListCurrent = 1;
    nodesOnList1[0] = SourceNode;

    currentNodeList = nodesOnList1;
    newNodeList = nodesOnList2;
    currentAdded = addedNodeToList1;
    newAdded = addedNodeToList2;

    while (numNodesOnListCurrent > 0) {
        numNodesOnListNew = 0;
        for (n = 0; n < numNodesOnListCurrent; n++) {
            node = currentNodeList[n];
            currentAdded[node] = FALSE;
            for (m = 0; m < NetworkP->Nodes[node].NumOutgoingLinks; m++) {
```

```
                link = NetworkP->Nodes[node].OutgoingLinks[m];
                if (!NetworkP->Links[link].Status)
                    continue;
                if (!BreadthFirstSearchRelax(NetworkP, node, link, DestinationNode))
                    continue;
                if ((!newAdded[NetworkP->Links[link].LinkNode2]) &&
                    (NetworkP->Links[link].LinkNode2 != DestinationNode)) {
                        newNodeList[numNodesOnListNew] = NetworkP->Links[link].LinkNode2;
                        numNodesOnListNew++;
                        newAdded[NetworkP->Links[link].LinkNode2] = TRUE;
                        }
                }
            }
        numNodesOnListCurrent = numNodesOnListNew;
        temp = currentNodeList; currentNodeList = newNodeList; newNodeList = temp;
        temp2 = currentAdded; currentAdded = newAdded; newAdded = temp2;
        }

    ShortestPathP->NumPathHops = 0;
    ShortestPathP->PathDistance = NodeDistance[DestinationNode];

    if (NodeDistance[DestinationNode] > (INFINITY-SMALL))
        return(FALSE);

    /* the predecessor path traces from the destination back to the root */
    /*    reverse it to have it start at root  */
    node = DestinationNode;
    while (node != SourceNode) {
        if (ShortestPathP->NumPathHops >= MaxPathHops) {
            printf("Need to increase MaxPathHops\n");
            exit(-1);
            }
        link = PredecessorLink[node];
        tempPath.PathHops[ShortestPathP->NumPathHops] = link;
        node = PredecessorNode[node];
        ShortestPathP->NumPathHops++;
        }

    for (m=0; m<ShortestPathP->NumPathHops; m++)  /* reverse the order */
        ShortestPathP->PathHops[m] = tempPath.PathHops[(ShortestPathP->NumPathHops)-m-1];

    return(TRUE);
}
```

```
BOOL BreadthFirstSearchRelax (NetworkTP NetworkP, USHORT NodeA, USHORT Link,
    USHORT FinalDestination)
{
    USHORT nodeB;
    double newDistanceAB;

    nodeB = NetworkP->Links[Link].LinkNode2;
    newDistanceAB = NodeDistance[NodeA] + NetworkP->Links[Link].Length;

    if ((NodeDistance[nodeB] > (newDistanceAB + SMALL)) &&
        (NodeDistance[FinalDestination] > (newDistanceAB + SMALL))) {
            NodeDistance[nodeB] = newDistanceAB;
            PredecessorNode[nodeB] = NodeA;
            PredecessorLink[nodeB] = Link;
            return(TRUE);
            }

    return(FALSE);
}
```

11.4 *K*-Shortest Paths

```
/*********************************************************************/
/*      Code for K Shortest Paths between a source and destination   */
/*      The K paths are returned in KPaths                           */
/*      The function returns the number of paths actually found      */
/*********************************************************************/

#define MaxEquivalences MaxNodes
#define MaxSplit MaxNodes
#define NodeType 0
#define LinkType 1

typedef struct tagEquivalenceClassT{
    char EquivalenceType;
    PathT    PrefixPath;   /* source to most recent split */
    PathT    SuffixPath;   /* split to next split (or to end); Don't track distance for suffix */
    PathT    ShortestPath;
    USHORT     FirstLink[MaxSplit];
    USHORT     NumFirstLinks;
    USHORT     SplitNode;
    } EquivalenceClassT, *EquivalenceClassTP;

EquivalenceClassT Equivalences[MaxEquivalences];
```

```
void CreatePathFromEquivalence (EquivalenceClassTP EquivalenceP, double Distance,
    PathTP PathP);
void FindBestPathInEquivalence (NetworkTP NetworkP, EquivalenceClassTP EquivalenceP,
    USHORT Dest);
void UpdateEquivalences (NetworkTP NetworkP, EquivalenceClassTP Equivalences,
    USHORT* NumEquivalences, USHORT BestPath, USHORT Dest);

USHORT KShortestPaths (NetworkTP NetworkP, USHORT Source, USHORT Dest, USHORT K,
    PathTP KPaths)
{
    USHORT j, n, bestPath, numEquivalences, firstLink;
    double minDist;

    for (j=0; j<K; j++) {
        KPaths[j].NumPathHops = 0;
        KPaths[j].PathDistance = 0.0;
        }

    if (Source == Dest)
        return(K);

    /* first find shortest path */
    if (!BreadthFirstSearchShortestPath(NetworkP, Source, Dest, &KPaths[0]))
        return (0);  /* didn't find any paths from source to dest */

    if (K == 1)
        return(1);

    firstLink = KPaths[0].PathHops[0];

    Equivalences[0].EquivalenceType = NodeType;
    Equivalences[0].FirstLink[0] = firstLink;
    Equivalences[0].NumFirstLinks = 1;
    Equivalences[0].PrefixPath.NumPathHops = 0;
    Equivalences[0].PrefixPath.PathDistance = 0.0;
    Equivalences[0].SuffixPath.NumPathHops = 0;
    Equivalences[0].SuffixPath.PathDistance = 0.0;
    Equivalences[0].SplitNode = Source;

    Equivalences[1].EquivalenceType = LinkType;
    Equivalences[1].FirstLink[0] = firstLink;
    Equivalences[1].NumFirstLinks = 1;
    Equivalences[1].PrefixPath.NumPathHops = 0;
    Equivalences[1].PrefixPath.PathDistance = 0.0;
    Equivalences[1].SuffixPath = KPaths[0];
    Equivalences[1].SplitNode = Source;
```

```
    numEquivalences = 2;
    FindBestPathInEquivalence(NetworkP, &Equivalences[0], Dest);
    FindBestPathInEquivalence(NetworkP, &Equivalences[1], Dest);

    for (j=1; j<K; j++) {
        minDist = INFINITY; bestPath = USHRT_MAX;
        for (n=0; n<numEquivalences; n++) {
            /* could use a heap to make this faster */
            if (Equivalences[n].ShortestPath.PathDistance < (minDist-SMALL)) {
                bestPath = n;
                minDist = Equivalences[n].ShortestPath.PathDistance;
                }
            }
        if (bestPath == USHRT_MAX)
            break;

        CreatePathFromEquivalence(&Equivalences[bestPath], minDist, &KPaths[j]);
        if (j < (K-1))
            UpdateEquivalences(NetworkP, Equivalences, &numEquivalences, bestPath, Dest);
        }
    return(j);
}

void CreatePathFromEquivalence (EquivalenceClassTP EquivalenceP, double Distance,
    PathTP PathP)
{
    USHORT h, numHops;

    *PathP = EquivalenceP->PrefixPath;
    numHops = PathP->NumPathHops;

    if (EquivalenceP->EquivalenceType == LinkType) {
        PathP->PathHops[numHops] = EquivalenceP->FirstLink[0];
        numHops++;
        }

    for (h=0; h<EquivalenceP->ShortestPath.NumPathHops; h++) {
        PathP->PathHops[numHops] = EquivalenceP->ShortestPath.PathHops[h];
        numHops++;
        }

    PathP->NumPathHops = numHops;
    PathP->PathDistance = Distance;
}
```

```
void FindBestPathInEquivalence (NetworkTP NetworkP, EquivalenceClassTP EquivalenceP,
    USHORT Dest)
{
    USHORT h, n, linkID, nodeID, splitNode;
    double minDist, distToAdd;
    PathT tempPath;

    /* make a copy of the network because will perform graph transformations */
    TempNetwork = *NetworkP;

    EquivalenceP->ShortestPath.PathDistance = INFINITY;

    /* eliminate nodes along prefix path so don't get loops */
    for (h=0; h<EquivalenceP->PrefixPath.NumPathHops; h++) {
        linkID = EquivalenceP->PrefixPath.PathHops[h];
        nodeID = TempNetwork.Links[linkID].LinkNode1;
        for (n=0; n<TempNetwork.Nodes[nodeID].NumIncomingLinks; n++) {
            linkID = TempNetwork.Nodes[nodeID].IncomingLinks[n];
            TempNetwork.Links[linkID].Status = FALSE;
            }
        }

    distToAdd = EquivalenceP->PrefixPath.PathDistance;

    if (EquivalenceP->EquivalenceType == NodeType) {
        for (n=0; n<EquivalenceP->NumFirstLinks; n++) {
            linkID = EquivalenceP->FirstLink[n];
            TempNetwork.Links[linkID].Status = FALSE;
            }
        BreadthFirstSearchShortestPath(&TempNetwork, EquivalenceP->SplitNode, Dest,
            &(EquivalenceP->ShortestPath));
        }
    else { /* LinkType */
        /* kick out the vertex node also */
        nodeID = EquivalenceP->SplitNode;
        for (n=0; n<TempNetwork.Nodes[nodeID].NumIncomingLinks; n++) {
            linkID = TempNetwork.Nodes[nodeID].IncomingLinks[n];
            TempNetwork.Links[linkID].Status = FALSE;
            }

        linkID = EquivalenceP->FirstLink[0];
        distToAdd += NetworkP->Links[linkID].Length;

        splitNode = NetworkP->Links[linkID].LinkNode2;
        minDist = INFINITY;
```

```
        for (h=1; h<EquivalenceP->SuffixPath.NumPathHops; h++) {
            /* if there are many hops,  it can be sped up */
            linkID = EquivalenceP->SuffixPath.PathHops[h];
            if (!TempNetwork.Links[linkID].Status)
                continue;
            TempNetwork.Links[linkID].Status = FALSE;
            BreadthFirstSearchShortestPath(&TempNetwork, splitNode, Dest, &tempPath);
            TempNetwork.Links[linkID].Status = TRUE; /* put back the link */
            if (tempPath.PathDistance < (minDist-SMALL)) {
                EquivalenceP->ShortestPath = tempPath;
                minDist = tempPath.PathDistance;
                }
            }
        }

    if (EquivalenceP->ShortestPath.PathDistance < INFINITY-SMALL)
        EquivalenceP->ShortestPath.PathDistance += distToAdd;
}

void UpdateEquivalences (NetworkTP NetworkP, EquivalenceClassTP EquivalencesP,
    USHORT* NumEquivalences, USHORT BestPath, USHORT Dest)
{
    USHORT linkID, splitNode, h, h2, numHops;

    if (EquivalencesP[BestPath].EquivalenceType == NodeType) {
        if (*NumEquivalences >= MaxEquivalences) {
            printf("Need to increase number of equivalence classes\n");
            exit(-2);
            }

        linkID = EquivalencesP[BestPath].ShortestPath.PathHops[0];
        EquivalencesP[BestPath].FirstLink[EquivalencesP[BestPath].NumFirstLinks] = linkID;
        EquivalencesP[BestPath].NumFirstLinks++;

        EquivalencesP[*NumEquivalences].EquivalenceType = LinkType;
        EquivalencesP[*NumEquivalences].FirstLink[0] = linkID;
        EquivalencesP[*NumEquivalences].NumFirstLinks = 1;
        EquivalencesP[*NumEquivalences].PrefixPath = EquivalencesP[BestPath].PrefixPath;
        EquivalencesP[*NumEquivalences].SuffixPath = EquivalencesP[BestPath].ShortestPath;
        EquivalencesP[*NumEquivalences].SuffixPath.PathDistance = 0.0;
        EquivalencesP[*NumEquivalences].SplitNode = EquivalencesP[BestPath].SplitNode;
        FindBestPathInEquivalence(NetworkP, &EquivalencesP[*NumEquivalences], Dest);
        (*NumEquivalences)++;
```

```
        /* can't set new shortest of BestPath until done with previous shortest above */
        FindBestPathInEquivalence(NetworkP, &EquivalencesP[BestPath], Dest);
        }
    else { /* link type */
        if (*NumEquivalences >= (MaxEquivalences-2)) { /* will add three more */
            printf("Need to increase number of equivalence classes\n");
            exit(-2);
            }

        /* find where shortest path diverges from suffix */
        /* suffix starts at branch node; shortest starts at 2nd node */
        for (h=1; h<EquivalencesP[BestPath].SuffixPath.NumPathHops; h++) {
            if (EquivalencesP[BestPath].ShortestPath.PathHops[h-1] !=
                EquivalencesP[BestPath].SuffixPath.PathHops[h])
                    break;
            }

        /* add one for new split node;  position h is where the paths diverge */
        EquivalencesP[*NumEquivalences].EquivalenceType = NodeType;
        EquivalencesP[*NumEquivalences].FirstLink[0] =
            EquivalencesP[BestPath].ShortestPath.PathHops[h-1];
        EquivalencesP[*NumEquivalences].FirstLink[1] =
            EquivalencesP[BestPath].SuffixPath.PathHops[h];
        EquivalencesP[*NumEquivalences].NumFirstLinks = 2;
        EquivalencesP[*NumEquivalences].PrefixPath = EquivalencesP[BestPath].PrefixPath;
        for (h2=0; h2<h; h2++) {  /* add in 1st link plus the part in common */
            linkID = EquivalencesP[BestPath].SuffixPath.PathHops[h2];
            EquivalencesP[*NumEquivalences].PrefixPath.
                PathHops[Equivalences[BestPath].PrefixPath.NumPathHops+h2] = linkID;
            EquivalencesP[*NumEquivalences].PrefixPath.PathDistance +=
                NetworkP->Links[linkID].Length;
            }
        EquivalencesP[*NumEquivalences].PrefixPath.NumPathHops += h;
        EquivalencesP[*NumEquivalences].SuffixPath.NumPathHops = 0;
        EquivalencesP[*NumEquivalences].SuffixPath.PathDistance = 0.0;
        linkID = EquivalencesP[BestPath].SuffixPath.PathHops[h];
        splitNode = NetworkP->Links[linkID].LinkNode1;
        EquivalencesP[*NumEquivalences].SplitNode = splitNode;
        FindBestPathInEquivalence(NetworkP, &EquivalencesP[*NumEquivalences], Dest);
        (*NumEquivalences)++;

        /* add one for one split path */
        EquivalencesP[*NumEquivalences].EquivalenceType = LinkType;
        EquivalencesP[*NumEquivalences].FirstLink[0] =
            Equivalences[BestPath].ShortestPath.PathHops[h-1];
```

```
            EquivalencesP[*NumEquivalences].NumFirstLinks = 1;
            EquivalencesP[*NumEquivalences].PrefixPath =
                Equivalences[(*NumEquivalences)-1].PrefixPath;
            EquivalencesP[*NumEquivalences].SuffixPath.PathDistance = 0.0;
            numHops = 0;
            for (h2=h-1; h2<Equivalences[BestPath].ShortestPath.NumPathHops; h2++) {
                linkID = Equivalences[BestPath].ShortestPath.PathHops[h2];
                Equivalences[*NumEquivalences].SuffixPath.PathHops[numHops] = linkID;
                numHops++;
                }
            EquivalencesP[*NumEquivalences].SuffixPath.NumPathHops = numHops;
            EquivalencesP[*NumEquivalences].SplitNode = splitNode;
            FindBestPathInEquivalence(NetworkP, &EquivalencesP[*NumEquivalences], Dest);
            (*NumEquivalences)++;

            /* add one for other split path */
            EquivalencesP[*NumEquivalences].EquivalenceType = LinkType;
            EquivalencesP[*NumEquivalences].FirstLink[0] =
                Equivalences[BestPath].SuffixPath.PathHops[h];
            EquivalencesP[*NumEquivalences].NumFirstLinks = 1;
            EquivalencesP[*NumEquivalences].PrefixPath =
                Equivalences[(*NumEquivalences)-1].PrefixPath;
            EquivalencesP[*NumEquivalences].SuffixPath.PathDistance = 0.0;
            numHops = 0;
            for (h2=h; h2<EquivalencesP[BestPath].SuffixPath.NumPathHops; h2++) {
                linkID = EquivalencesP[BestPath].SuffixPath.PathHops[h2];
                EquivalencesP[*NumEquivalences].SuffixPath.PathHops[numHops] = linkID;
                numHops++;
                }
            EquivalencesP[*NumEquivalences].SuffixPath.NumPathHops = numHops;
            EquivalencesP[*NumEquivalences].SplitNode = splitNode;
            FindBestPathInEquivalence(NetworkP, &EquivalencesP[*NumEquivalences], Dest);
            (*NumEquivalences)++;

            EquivalencesP[BestPath].SuffixPath.NumPathHops = h;
            FindBestPathInEquivalence(NetworkP, &EquivalencesP[BestPath], Dest);
            }
    }
```

11.5 *N*-Shortest Diverse Paths

```
/*****************************************************************/
/*      Code for N Shortest Diverse Paths                        */
/*      The N paths are returned in NPaths                       */
/*      The function returns number of paths found               */
/*      The paths are not necessarily in order from shortest to longest */
/*****************************************************************/

void KDualPathGraphTranformation (NetworkTP NetworkP, PathTP PathP, USHORT Source,
    USHORT Destination, BOOL LinkDisjointOnly, BOOL CommonLinksAllowed,
    BOOL CommonNodesAllowed, USHORT DummyNodeThreshold);
void GenerateTwoRealPaths (NetworkTP NetworkP, PathTP TempPath1P, PathTP TempPath2P,
    PathTP RealPath1P, PathTP RealPath2P);
void CleanPath (NetworkTP NetworkP, USHORT DummyLinkThreshold, USHORT Source,
    USHORT Destination, PathTP NewPathP);
void AdjustNodeInfoForNewLink (NetworkTP NetworkP, USHORT LinkID);
void ChangeLinkSource (NetworkTP NetworkP, USHORT LinkID, USHORT OldSource,
    USHORT NewSource);
void ChangeLinkDestination (NetworkTP NetworkP, USHORT LinkID, USHORT OldDest,
    USHORT NewDest);

USHORT NShortestDiversePaths (NetworkTP NetworkP, USHORT Source, USHORT Destination,
    USHORT NumDisjointPaths, BOOL LinkDisjointOnly, BOOL CommonLinksAllowed,
    BOOL CommonNodesAllowed, PathTP NPaths)
    /* looks for N paths that are mutually disjoint, or maximally disjoint depending on the settings */
    /* if LinkDisjointOnly is TRUE, then it doesn't try to avoid common nodes among the paths */
    /* if CommonLinksAllowed is TRUE, then if fully disjoint paths can't be found, it will accept */
    /*    paths with common links (although it still minimizes the number of common links) */
    /* if CommonNodesAllowed is TRUE, then if fully disjoint paths can't be found, it will accept */
    /*    paths with common nodes (althouth it still minimizes the number of common nodes) */
{
    USHORT h, n, numFoundPaths;
    int j, k;
    PathTP tempPaths;

    for (n=0; n<NumDisjointPaths; n++) {
        NPaths[n].NumPathHops = 0;
        NPaths[n].PathDistance = 0.0;
        }

    if (Source == Destination)
        return(NumDisjointPaths);

    /* first find shortest path */
```

```
if (!BreadthFirstSearchShortestPath(NetworkP, Source, Destination, &NPaths[0]))
    return (0);  /* didn't find any paths */
if (NumDisjointPaths == 1)
    return(1);

tempPaths = (PathTP)malloc(NumDisjointPaths*sizeof(PathT));
if (tempPaths == NULL) {
    printf("Out of Memory\n");
    exit(-20);
    }

tempPaths[0] = NPaths[0];
for (n=1; n<NumDisjointPaths; n++) {
    /* make a copy of the network because will perform graph transformations */
    TempNetwork = *NetworkP;
    for (j=0; j<n; j++)
        KDualPathGraphTranformation(&TempNetwork, &NPaths[j], Source, Destination,
            LinkDisjointOnly, CommonLinksAllowed, CommonNodesAllowed,
            NetworkP->NumNodes);

    /* run shortest path on the transformed graph */
    if (!BreadthFirstSearchShortestPath(&TempNetwork, Source, Destination, &tempPaths[n]))
        break;

    /* clean path up; may have dummy nodes in it */
    CleanPath(&TempNetwork, NetworkP->NumLinks, Source, Destination, &tempPaths[n]);

    /* paths found above may not be the true paths; need to unravel any interleaving */
    for (j=n; j>0; j--) {
        for (k=j-1; k>=0; k--) {
            GenerateTwoRealPaths(&TempNetwork, &tempPaths[j],
                &tempPaths[k], &NPaths[j], &NPaths[k]);
            tempPaths[j] = NPaths[j];
            tempPaths[k] = NPaths[k];
            }
        }
    }
numFoundPaths = n;
for (n=0; n<numFoundPaths; n++) {
    NPaths[n].PathDistance = 0.0;
    for (h=0; h<NPaths[n].NumPathHops; h++)
        NPaths[n].PathDistance += NetworkP->Links[NPaths[n].PathHops[h]].Length;
    }
free(tempPaths);

return(numFoundPaths);
```

```
}

void KDualPathGraphTranformation (NetworkTP NetworkP, PathTP PathP, USHORT Source,
    USHORT Destination, BOOL LinkDisjointOnly, BOOL CommonLinksAllowed,
    BOOL CommonNodesAllowed, USHORT DummyNodeThreshold)
{
    USHORT n, prevNode, link, newLink, dummyID, forwardLink, reverseLink, reverseLink2;
    int h;
    double tempLength;

    for (h=PathP->NumPathHops-1; h>=0; h--) {
        forwardLink = PathP->PathHops[h];
        if (h == 0) { /* don't split source node, but handle the link */
            if (NetworkP->Links[forwardLink].Length > (CommonLinkPenalty - SMALL)) {
                /* must be a common link in the previous paths that have been found */
                /* increase penalty if use it again */
                NetworkP->Links[forwardLink].Length += CommonLinkPenalty;
                continue;
                }

            reverseLink = GetReverseLink(NetworkP, forwardLink);
            tempLength = NetworkP->Links[forwardLink].Length;

            NetworkP->Links[reverseLink].Length = -1.0*tempLength;
            NetworkP->Links[reverseLink].Status = TRUE;

            NetworkP->Links[forwardLink].Status = FALSE;
            /* add big amount to length to discourage its use */
            NetworkP->Links[forwardLink].Length = tempLength + CommonLinkPenalty;
            if (CommonLinksAllowed) { /* common links OK if can't be avoided */
                NetworkP->Links[forwardLink].Status = TRUE;
                }
            continue;
            }

        prevNode = NetworkP->Links[forwardLink].LinkNode1;

        if (prevNode >= DummyNodeThreshold) {
            /* must be a common node in the previous paths that have been found */
            /* don't add another dummy node */

            if ((!LinkDisjointOnly) && (CommonNodesAllowed)) {
                /* increase penalty if use the node again */
                for (n=0; n<NetworkP->Nodes[prevNode].NumIncomingLinks; n++) {
                    link = NetworkP->Nodes[prevNode].IncomingLinks[n];
```

```
                    if (NetworkP->Links[link].Length > (CommonNodePenalty-SMALL)) {
                        NetworkP->Links[link].Length += CommonNodePenalty;
                        break;
                        }
                    }
                }

            if (NetworkP->Links[forwardLink].Length > (CommonLinkPenalty - SMALL)) {
                /* must also be a common link in the previous paths that have been found */
                /* increase penalty if use it again */
                NetworkP->Links[forwardLink].Length += CommonLinkPenalty;
                continue;
                }

            reverseLink = GetReverseLink(NetworkP, forwardLink);
            tempLength = NetworkP->Links[forwardLink].Length;

            NetworkP->Links[reverseLink].Length = -1.0*tempLength;
            NetworkP->Links[reverseLink].Status = TRUE;

            NetworkP->Links[forwardLink].Status = FALSE;
            /* add big amount to length to discourage its use */
            NetworkP->Links[forwardLink].Length = tempLength + CommonLinkPenalty;
            if (CommonLinksAllowed) { /* common links OK if can't be avoided */
                NetworkP->Links[forwardLink].Status = TRUE;
                }
            continue;
            }

        /* split prevNode */
        if (NetworkP->NumNodes >= MaxNodesWithDummies) {
            printf("Need to increase MaxNodes\n");
            exit(-4);
            }
        if (NetworkP->NumLinks >= (MaxLinksWithDummies-1)) { /* will add 2 links */
            printf("Need to increase MaxLinks\n");
            exit(-5);
            }

        dummyID = NetworkP->NumNodes;
        NetworkP->NumNodes++;
        strcpy(NetworkP->Nodes[dummyID].Name, "DummyAdd");
        NetworkP->Nodes[dummyID].NumOutgoingLinks = 0;
        NetworkP->Nodes[dummyID].NumIncomingLinks = 0;
```

```
        reverseLink = GetReverseLink(NetworkP, forwardLink);
        tempLength = NetworkP->Links[forwardLink].Length;

        NetworkP->Links[reverseLink].Length = -1.0*tempLength;
        NetworkP->Links[reverseLink].Status = TRUE;
        ChangeLinkDestination(NetworkP, reverseLink, prevNode, dummyID);

        NetworkP->Links[forwardLink].Status = FALSE;
        /* add big amount to length to discourage its use */
        NetworkP->Links[forwardLink].Length = tempLength + CommonLinkPenalty;
        if (CommonLinksAllowed) { /* common links OK if can't be avoided */
            NetworkP->Links[forwardLink].Status = TRUE;
            ChangeLinkSource(NetworkP, forwardLink, prevNode, dummyID);
            }

        reverseLink = GetReverseLink(NetworkP, PathP->PathHops[h-1]);
        for (n=0; n<NetworkP->Nodes[prevNode].NumOutgoingLinks; n++) {
            link = NetworkP->Nodes[prevNode].OutgoingLinks[n];
            if (link == reverseLink) {
                continue;
                }
            else {  /* for all other links emanating from prevNode, change the source to dummy */
                ChangeLinkSource(NetworkP, link, prevNode, dummyID);
                n--;
                }
            }

        /* add dummy link to point from dummy to prevNode - give it distance 0 */
        newLink = AddLinkToTopology(NetworkP, dummyID, prevNode, 0.0, FALSE);

        if (LinkDisjointOnly) {
            reverseLink2 = GetReverseLink(NetworkP, newLink); /* prevNode to dummy */
            NetworkP->Links[reverseLink2].Status = TRUE;
            NetworkP->Links[reverseLink2].Length = SMALL;
            }
        else if (CommonNodesAllowed) { /* common nodes OK if can't be avoided */
            reverseLink2 = GetReverseLink(NetworkP, newLink);/* prevNode to dummy */
            NetworkP->Links[reverseLink2].Status = TRUE;
            /* add big amount to length to discourage its use */
            NetworkP->Links[reverseLink2].Length = CommonNodePenalty;
            }
        }
}
```

```
void GenerateTwoRealPaths (NetworkTP NetworkP, PathTP TempPath1P, PathTP TempPath2P,
    PathTP RealPath1P, PathTP RealPath2P)
{
    int i, j, k, lastpos, lastj;
    USHORT link;
    BOOL exchange;

    /* to generate the shortest paths, look for overlapping sections in the two paths */
    RealPath1P->NumPathHops = RealPath2P->NumPathHops = 0;
    exchange = FALSE;
    lastpos = 0;
    for (i=0; i<TempPath1P->NumPathHops; i++) {
        for (j=0; j<TempPath2P->NumPathHops; j++) {
            if (GetReverseLink(NetworkP, TempPath2P->PathHops[j]) ==
             TempPath1P->PathHops[i]) {
                /* found an overlapping section; remove it and then exchange links */
                lastj = j;
                for (i++,j--; i<TempPath1P->NumPathHops, j>=0; i++, j--) {
                    if (GetReverseLink(NetworkP, TempPath2P->PathHops[j]) !=
                        TempPath1P->PathHops[i])
                            break;
                    }
                j++; i--; /* go back to last position of overlap */
                for (k=lastpos; k<j; k++) {
                    link = TempPath2P->PathHops[k];
                    if (!exchange) {
                        RealPath2P->PathHops[RealPath2P->NumPathHops] = link;
                        RealPath2P->NumPathHops++;
                        }
                    else {
                        RealPath1P->PathHops[RealPath1P->NumPathHops] = link;
                        RealPath1P->NumPathHops++;
                        }
                    }
                exchange = !exchange;
                lastpos = lastj + 1;
                break;
                }
            }
        if (j == TempPath2P->NumPathHops) { /* not an overlapping link */
            if (!exchange) {
                RealPath1P->PathHops[RealPath1P->NumPathHops] = TempPath1P->PathHops[i];
                RealPath1P->NumPathHops++;
                }
            else {
```

```
                RealPath2P->PathHops[RealPath2P->NumPathHops] = TempPath1P->PathHops[i];
                RealPath2P->NumPathHops++;
                }
            }
        }
    for (k=lastpos; k<TempPath2P->NumPathHops; k++) {
        link = TempPath2P->PathHops[k];
        if (!exchange) {
            RealPath2P->PathHops[RealPath2P->NumPathHops] = link;
            RealPath2P->NumPathHops++;
            }
        else {
            RealPath1P->PathHops[RealPath1P->NumPathHops] = link;
            RealPath1P->NumPathHops++;
            }
        }
}

void CleanPath (NetworkTP NetworkP, USHORT DummyLinkThreshold, USHORT Source,
    USHORT Destination, PathTP NewPathP)
{
    USHORT h, linkID, numHops;

    numHops = 0;
    for (h=0; h<NewPathP->NumPathHops; h++) {
        linkID = NewPathP->PathHops[h];
        if (linkID >= DummyLinkThreshold)
            continue;
        NewPathP->PathHops[numHops] = linkID;
        numHops++;
        }
    NewPathP->NumPathHops = numHops;
}


USHORT AddLinkToTopology (NetworkTP NetworkP, USHORT Node1, USHORT Node2,
    double Length, BOOL ReverseLinkStatus)
{
    USHORT newLinkID, reverseLinkID;

    /* assumes links always added in pairs */
    /* if reverse link direction not needed, pass in its status as FALSE */

    if (NetworkP->NumLinks >= (MaxLinksWithDummies-1)) {
        printf("Need to increase MaxLinks\n");
        exit(-5);
        }
```

```
    newLinkID = NetworkP->NumLinks;
    NetworkP->NumLinks += 2;

    NetworkP->Links[newLinkID].LinkNode1 = Node1;
    NetworkP->Links[newLinkID].LinkNode2 = Node2;
    NetworkP->Links[newLinkID].Length = Length;
    NetworkP->Links[newLinkID].Status = TRUE;
    AdjustNodeInfoForNewLink(NetworkP, newLinkID);

    reverseLinkID = GetReverseLink(NetworkP, newLinkID);
    NetworkP->Links[reverseLinkID].LinkNode1 = Node2;
    NetworkP->Links[reverseLinkID].LinkNode2 = Node1;
    NetworkP->Links[reverseLinkID].Length = Length;
    NetworkP->Links[reverseLinkID].Status = ReverseLinkStatus;
    AdjustNodeInfoForNewLink(NetworkP, reverseLinkID);

    return(newLinkID);
}

void AdjustNodeInfoForNewLink (NetworkTP NetworkP, USHORT LinkID)
{
    USHORT node;

    node = NetworkP->Links[LinkID].LinkNode1;
    if (NetworkP->Nodes[node].NumOutgoingLinks >= MaxNodeDegree) {
        printf("Need to increase MaxNodeDegree\n");
        exit(-6);
        }

    NetworkP->Nodes[node].OutgoingLinks[NetworkP->Nodes[node].NumOutgoingLinks] =
        LinkID;
    NetworkP->Nodes[node].NumOutgoingLinks++;

    node = NetworkP->Links[LinkID].LinkNode2;
    if (NetworkP->Nodes[node].NumIncomingLinks >= MaxNodeDegree) {
        printf("Need to increase MaxNodeDegree\n");
        exit(-6);
        }
    NetworkP->Nodes[node].IncomingLinks[NetworkP->Nodes[node].NumIncomingLinks] =
        LinkID;
    NetworkP->Nodes[node].NumIncomingLinks++;
}

void ChangeLinkSource (NetworkTP NetworkP, USHORT LinkID, USHORT OldSource,
    USHORT NewSource)
{
    USHORT n;
```

```
    if (NetworkP->Links[LinkID].LinkNode1 != OldSource) {
        printf("Inconsistent\n");
        exit(-7);
        }
    if (NetworkP->Nodes[NewSource].NumOutgoingLinks >= MaxNodeDegree) {
        printf("Need to increase MaxNodeDegree\n");
        exit(-6);
        }
    NetworkP->Links[LinkID].LinkNode1 = NewSource;
    NetworkP->Nodes[NewSource].OutgoingLinks[NetworkP->Nodes[NewSource].
        NumOutgoingLinks] = LinkID;
    NetworkP->Nodes[NewSource].NumOutgoingLinks++;

    /* remove it from OldSource list*/
    for (n=0; n<NetworkP->Nodes[OldSource].NumOutgoingLinks; n++)
        if (NetworkP->Nodes[OldSource].OutgoingLinks[n] == LinkID) break;
    for (; n<NetworkP->Nodes[OldSource].NumOutgoingLinks-1; n++) {
        NetworkP->Nodes[OldSource].OutgoingLinks[n] =
            NetworkP->Nodes[OldSource].OutgoingLinks[n+1];
        }
    NetworkP->Nodes[OldSource].NumOutgoingLinks--;
}

void ChangeLinkDestination (NetworkTP NetworkP, USHORT LinkID, USHORT OldDest,
    USHORT NewDest)
{
    USHORT n;

    if (NetworkP->Links[LinkID].LinkNode2 != OldDest) {
        printf("Inconsistent\n");
        exit(-8);
        }
    if (NetworkP->Nodes[NewDest].NumIncomingLinks >= MaxNodeDegree) {
        printf("Need to increase MaxNodeDegree\n");
        exit(-6);
        }

    NetworkP->Links[LinkID].LinkNode2 = NewDest;
    NetworkP->Nodes[NewDest].IncomingLinks[NetworkP->Nodes[NewDest].
        NumIncomingLinks] = LinkID;
    NetworkP->Nodes[NewDest].NumIncomingLinks++;

    /* remove it from OldDest list*/
    for (n=0; n<NetworkP->Nodes[OldDest].NumIncomingLinks; n++)
        if (NetworkP->Nodes[OldDest].IncomingLinks[n] == LinkID) break;
```

```
        for (; n<NetworkP->Nodes[OldDest].NumIncomingLinks-1; n++) {
            NetworkP->Nodes[OldDest].IncomingLinks[n] =
                NetworkP->Nodes[OldDest].IncomingLinks[n+1];
            }
        NetworkP->Nodes[OldDest].NumIncomingLinks--;
    }

    USHORT GetReverseLink (NetworkTP NetworkP, USHORT LinkID)
    {
        /* Assumes links are always added in pairs */
        if (LinkID % 2 == 0)
            return(LinkID + 1);
        else return(LinkID - 1);
    }
```

11.6 Minimum Steiner Tree

11.6.1 Minimum Spanning Tree with Enhancement

```
/*********************************************************************/
/*      Code for a heuristic to find the Minimum Steiner Tree        */
/*      The nodes to be included in tree are passed in NodeSet       */
/*      The first node in set is treated as root                     */
/*      (It returns same tree regardless of which node is the root)  */
/*      The minimum Steiner tree is returned in Tree (unidirectional)*/
/*      Returns TRUE/FALSE depending on whether a tree is found      */
/*********************************************************************/

BOOL Prim (NetworkTP NetworkP, MST_TP Tree);
void RelaxPrim (USHORT NodeA, USHORT NodeZ, USHORT LinkID, double DistanceAZ);
USHORT MinDistance (NetworkTP NetworkP);

BOOL SteinerTreeHeuristic (NetworkTP NetworkP, USHORT * NodeSet,
    USHORT NumNodesInSet, MST_TP Tree)
{
    USHORT m, n, h, node1, node2, nodeInOrigNet, link, reverseLink, newLink;
    USHORT topologyCNodes[MaxNodes], topologyCPrimeLinks[MaxLinks];
    USHORT CPrimeNodeMap[MaxNodes];
    PathT path;
    MST_T tempTree;
    BOOL addedNode[MaxNodes], addedLink[MaxLinks];
```

```
BOOL tempStatus[MaxLinks];

Tree->NumMSTLinks = 0;
Tree->MSTDistance = 0.0;
if (NumNodesInSet <= 1)
    return(TRUE);

/* create fully connected graph of all nodes in set (Topology B) */
/* store this in TempNetwork */
TempNetwork.NumNodes = TempNetwork.NumLinks = 0;
for (m=0; m<NumNodesInSet; m++) {
    if (NodeSet[m] >= NetworkP->NumNodes) {
        printf( "Invalid node in Steiner Tree set\n");
        return(FALSE);
        }
    TempNetwork.Nodes[m].NumIncomingLinks =
        TempNetwork.Nodes[m].NumOutgoingLinks = 0;
    }
TempNetwork.NumNodes = NumNodesInSet;

for (m=0; m<NumNodesInSet-1; m++) { /* add in links to fully connect network */
    node1 = NodeSet[m];
    for (n=m+1; n<NumNodesInSet; n++) {
        node2 = NodeSet[n];
        if (!BreadthFirstSearchShortestPath(NetworkP, node1, node2, &path))
            return(FALSE);  /* nodes in original graph are not connected */
        /* set distance to distance between nodes in original network */
        AddLinkToTopology(&TempNetwork, m, n, path.PathDistance, TRUE);
        }
    }

if (!Prim(&TempNetwork,  &tempTree) )  /* find Min Spanning Tree on Topology B */
    return(FALSE);

/* create fully connected graph of nodes in expanded Min Spanning Tree (Topology C) */
/* store it in TempNetwork2 */
TempNetwork2.NumNodes = TempNetwork2.NumLinks = 0;
for (n=0; n<NetworkP->NumNodes; n++)
    addedNode[n] = FALSE;
for (n=0; n<tempTree.NumMSTLinks; n++) { /* add in all nodes of expanded tree*/
    link = tempTree.MSTLinks[n];
    node1 = TempNetwork.Links[link]. LinkNode1;
    node2 = TempNetwork.Links[link]. LinkNode2;
    if (!BreadthFirstSearchShortestPath(NetworkP, NodeSet[ node1], NodeSet[node2], &path))
        return(FALSE); /* should not occur */
    link = path.PathHops[0];
```

```
        nodeInOrigNet = NetworkP->Links[link].LinkNode1;
        if (!addedNode[nodeInOrigNet]) {
            TempNetwork2.Nodes[TempNetwork2.NumNodes].NumIncomingLinks =
                TempNetwork2.Nodes[TempNetwork2.NumNodes].NumOutgoingLinks = 0;
            topologyCNodes[TempNetwork2.NumNodes] = nodeInOrigNet;
            TempNetwork2.NumNodes++;
            addedNode[nodeInOrigNet] = TRUE;
            }
        for (h=0; h<path.NumPathHops; h++) {
            link = path.PathHops[h];
            nodeInOrigNet = NetworkP->Links[link].LinkNode2;
            if (!addedNode[nodeInOrigNet]) {
                TempNetwork2.Nodes[TempNetwork2.NumNodes].NumIncomingLinks =
                    TempNetwork2.Nodes[TempNetwork2.NumNodes].NumOutgoingLinks = 0;
                topologyCNodes[TempNetwork2.NumNodes] = nodeInOrigNet;
                TempNetwork2.NumNodes++;
                addedNode[nodeInOrigNet] = TRUE;
                }
            }
        }
    for (m=0; m<TempNetwork2.NumNodes-1; m++) { /* add in links to fully connect nodes */
        for (n=m+1; n<TempNetwork2.NumNodes; n++) {
            node1 = topologyCNodes[m]; /* get node number in original network */
            node2 = topologyCNodes[n];   /* get node number in original network */
            /* get distance from original network */
            if (!BreadthFirstSearchShortestPath(NetworkP, node1, node2, &path) )
                return(FALSE); /* should not occur */
            AddLinkToTopology(&TempNetwork2, m, n, path.PathDistance, TRUE);
            }
        }

    if (!Prim(&TempNetwork2,  &tempTree) )  /* find Min Spanning Tree on Topology C */
        return(FALSE);

    /* form new topology with tempTree expanded into underlying paths (Toplogy C') */
    /* store this in TempNetwork */
    TempNetwork.NumNodes = TempNetwork.NumLinks = 0;
    for (n=0; n<NetworkP->NumNodes; n++)
        addedNode[n] = FALSE;
    for (n=0; n<NetworkP->NumLinks; n++)
        addedLink[n] = FALSE;
    for (n=0; n<tempTree.NumMSTLinks; n++) {
        link = tempTree.MSTLinks[n];
        /* get node numbers in original network */
```

```
        node1 = topologyCNodes[TempNetwork2.Links[link].LinkNode1];
        node2 = topologyCNodes[TempNetwork2.Links[link].LinkNode2];
        if (!BreadthFirstSearchShortestPath(NetworkP, node1, node2,  &path))
            return(FALSE);  /* should not occur */
        if (!addedNode[node1]) {
            TempNetwork.Nodes[TempNetwork.NumNodes].NumIncomingLinks =
            TempNetwork.Nodes[TempNetwork.NumNodes].NumOutgoingLinks = 0;
            CPrimeNodeMap[node1] = TempNetwork.NumNodes;
            TempNetwork.NumNodes++;
            addedNode[node1] = TRUE;
            }
        for (h=0; h<path.NumPathHops; h++) {
            link = path.PathHops[h];
            node1 = NetworkP->Links[link].LinkNode1;
            node2 = NetworkP->Links[link].LinkNode2;
            if (!addedNode[node2]) {
                TempNetwork.Nodes[TempNetwork.NumNodes].NumIncomingLinks =
                    TempNetwork.Nodes[TempNetwork.NumNodes].NumOutgoingLinks = 0;
                CPrimeNodeMap[node2] = TempNetwork.NumNodes;
                TempNetwork.NumNodes++;
                addedNode[node2] = TRUE;
                }
            if (!addedLink[link]) {
                reverseLink = GetReverseLink(NetworkP, link);
                newLink = AddLinkToTopology(&TempNetwork, CPrimeNodeMap[node1],
                    CPrimeNodeMap[node2], NetworkP->Links[link].Length,
                    NetworkP->Links[reverseLink].Status);
                topologyCPrimeLinks[newLink] = link;
                topologyCPrimeLinks[newLink+1] = reverseLink;
                addedLink[link] = TRUE;
                addedLink[reverseLink] = TRUE;
                }
            }
        }

    if (!Prim(&TempNetwork,  &tempTree) ) /* find MST on C' */
        return(FALSE);

    /* temporarily trim down Network to links in final MST */
    for (n=0; n<NetworkP->NumLinks; n++) {
        tempStatus[n] = NetworkP->Links[n].Status;
        NetworkP->Links[n].Status = FALSE;
        }
    for (n=0; n<tempTree.NumMSTLinks; n++) {
        /* Get link in original topology */
```

```
            link = topologyCPrimeLinks[tempTree.MSTLinks[n]];
            reverseLink = GetReverseLink(NetworkP, link);
            NetworkP->Links[link].Status = tempStatus[link];
            NetworkP->Links[reverseLink].Status = tempStatus[reverseLink];
            }

        /* check which links in final MST are needed for Steiner Tree */
        for (n=0; n<NetworkP->NumLinks; n++)
            addedLink[n] = FALSE;
        node1 = NodeSet[0];  /* pick first node in set as root */
        for (n=1; n<NumNodesInSet; n++) {
            node2 = NodeSet[n];
            if (!BreadthFirstSearchShortestPath(NetworkP, node1, node2, &path) )
                return(FALSE);  /* should not occur */
            for (h=0; h<path.NumPathHops; h++) {
                link = path.PathHops[h];
                addedLink[link] = TRUE;
                }
            }

        for (n=0; n<NetworkP->NumLinks; n++) {
            NetworkP->Links[n].Status = tempStatus[n];
            if (!addedLink[n])
                continue;
            Tree->MSTLinks[Tree->NumMSTLinks] = n;
            Tree->NumMSTLinks++;
            Tree->MSTDistance += NetworkP->Links[n].Length;
            }

        return(TRUE);
    }

    BOOL Prim (NetworkTP NetworkP, MST_TP Tree)  /*  finds Minimum Spanning Tree  */
    {
        USHORT  n, d, root, node, nextNode, link;

        Tree->NumMSTLinks = 0;
        if (NetworkP->NumNodes <= 1)
            return(TRUE);

        root = USHRT_MAX;
        for (n = 0; n < NetworkP->NumNodes; n++) { /* initialize all nodes */
            PredecessorNode[n] = PredecessorLink[n] = USHRT_MAX;
            Marked[n] = FALSE;
            if (root == USHRT_MAX) {
```

```
            root = n;  /* take first node as root */
            NodeDistance[n] = 0.0;
            }
        else NodeDistance[n] = INFINITY;
        }
    while (TRUE) {
        node = MinDistance(NetworkP);  /* finds closest node not already in tree */
        if (node == USHRT_MAX)
            return(FALSE);  //* can't find a tree that connects all nodes */
        Marked[node] = TRUE;
        if (node != root) {
            Tree->MSTLinks[Tree->NumMSTLinks] = PredecessorLink[node];
            Tree->NumMSTLinks++;
            if (Tree->NumMSTLinks == (NetworkP->NumNodes-1))
                break;
            }
        for (d=0; d<NetworkP->Nodes[node].NumOutgoingLinks; d++) {
            link = NetworkP->Nodes[node].OutgoingLinks[d];
            if (NetworkP->Links[link].Status != TRUE)
                continue;
            nextNode = NetworkP->Links[link].LinkNode2;
            if (Marked[nextNode])
                continue;
            RelaxPrim(node, nextNode, link, NetworkP->Links[link].Length);
            }
        }
    return(TRUE);
}
void RelaxPrim (USHORT NodeA, USHORT NodeZ, USHORT LinkID, double DistanceAZ)
{
    if (NodeDistance[NodeZ] > DistanceAZ + SMALL) {
        NodeDistance[NodeZ] = DistanceAZ;
        PredecessorNode[NodeZ] = NodeA;
        PredecessorLink[NodeZ] = LinkID;
        }
}
USHORT MinDistance (NetworkTP NetworkP)
{
    USHORT n, minIndex;
    double distance;

    distance = INFINITY;
    minIndex = USHRT_MAX;
```

```
        for (n=0; n<NetworkP->NumNodes; n++) {
            if (Marked[n]) continue;
            if (NodeDistance[n] < distance - SMALL) {
                distance = NodeDistance[n];
                minIndex = n;
                }
            }
        return (minIndex);
    }
```

11.6.2 Minimum Paths

```
/*************************************************************/
/*      Code for a second heuristic to find the Minimum Steiner Tree      */
/*      The nodes to be included in tree are passed in NodeSet            */
/*      The first node in set is treated as root                          */
/*      (The tree returned depends on which node is the root)             */
/*      The minimum Steiner tree is returned in Tree (unidirectional)     */
/*      Returns TRUE/FALSE depending on whether a tree is found           */
/*************************************************************/

BOOL SteinerTreeHeuristic2 (NetworkTP NetworkP, USHORT* NodeSet,
    USHORT NumNodesInSet, MST_TP Tree)
{
    USHORT m, n, h, k, node, treeNode, destNode, numNodesInTree, nodesInTree[MaxNodes];
    double minDistance;
    PathT path, minPath;
    BOOL addedNodeToTree[MaxNodes], addedLinkToTree[MaxLinks];

    Tree->NumMSTLinks = 0;
    Tree->MSTDistance = 0.0;
    if (NumNodesInSet <= 1)
        return(TRUE);

    for (n=0; n<NetworkP->NumNodes; n++)
        addedNodeToTree[n] = FALSE;
    for (n=0; n<NetworkP->NumLinks; n++)
        addedLinkToTree[n] = FALSE;

    // add the first node in the destination set to the tree; this is treated as the source
    nodesInTree[0] = NodeSet[0];
    numNodesInTree = 1;
    addedNodeToTree[NodeSet[0]] = TRUE;
```

```
    for (k=0; k<(NumNodesInSet-1); k++) {
        // find shortest path from a tree node to one of the destinations not in the tree
        minDistance = INFINITY;
        for (m=1; m<NumNodesInSet; m++) {
            destNode = NodeSet[m];
            if (addedNodeToTree[destNode])
                continue;
            for (n=0; n<numNodesInTree; n++) {
                treeNode = nodesInTree[n];
                BreadthFirstSearchShortestPath(NetworkP, treeNode, destNode, &path);
                if (path.PathDistance < minDistance) {
                    minDistance = path.PathDistance;
                    minPath = path;
                    }
                }
            }

        for (h=0; h<minPath.NumPathHops; h++) {
            // the first node in this path is already a tree node
            node = NetworkP->Links[minPath.PathHops[h]].LinkNode2;
            if (!addedNodeToTree[node]) { // the last one added should be a destination node
                nodesInTree[numNodesInTree] = node;
                numNodesInTree++;
                addedNodeToTree[node] = TRUE;
                }
            addedLinkToTree[minPath.PathHops[h]] = TRUE;
            }
        }

    for (n=0; n<NetworkP->NumLinks; n++) {
        if (!addedLinkToTree[n])
            continue;
        Tree->MSTLinks[Tree->NumMSTLinks] = n;
        Tree->NumMSTLinks++;
        Tree->MSTDistance += NetworkP->Links[n].Length;
        }

    return(TRUE);
}
```

References

[Bhan99] R. Bhandari, *Survivable Networks: Algorithms for Diverse Routing*. (Boston, Kluwer Academic Publishers, 1999)

[HeMS03] J. Hershberger, M. Maxel, S. Suri, Finding the k shortest simple paths: A new algorithm and its implementation. *Proceedings, Fifth Workshop on Algorithm Engineering and Experiments*, Baltimore, MD, Jan. 11, 2003, pp. 26–36

[KoMB81] L. Kou, G. Markowsky, L. Berman, A fast algorithm for Steiner trees. *Acta Inform*. **15**(2), 141–145 (1981, Jun)

[TaMa80] H. Takahashi, A. Matsuyama, An approximate solution for the Steiner problem in graphs. *Math. Jpn*. **24**(6), 573–577 (1980)

[Waxm88] B.M. Waxman, Routing of multipoint connections. *IEEE. J. Sel. Areas Commun*. **6**(9), 1617–1622 (1988, Dec)

Appendix

Appendix: Suggestions for RFI/RFP Network Design Exercises

Carriers often issue a network design exercise as part of a Request for Information (RFI) or a Request for Proposal (RFP). System vendors perform the network designs using their respective equipment to provide carriers with information regarding architecture, technology, and pricing. To streamline the design process and to enable carriers to glean relevant information when comparing results, some suggestions to assist carriers in preparing a design exercise are provided here.

1. Provide all data (e.g., nodes, links, demands) in text files or in spreadsheets. Do not require any manual entry of data by the vendors, as this is likely to lead to errors.
2. Provide the latitude/longitudes of the network nodes. Most design tools can use this information to position nodes on the screen, to help visualize the design.
3. If demand sets for multiple time periods are provided, specify whether the demand sets are incremental or cumulative (e.g., are the demands in set #2 added to the demands that already exist from set #1, or does set #2 represent all of the demands). Additionally, when adding demands to the network in subsequent time periods, specify whether the demands already in the network need to be kept fixed, or whether an optimization can be performed over all of the demands.
4. If the design exercise is for extended-reach technology, then fiber span information should be provided, including fiber type, span distances, and span losses. (Information such as PMD can be helpful as well, if available.)
5. Ideally, there would be no more than three design exercises:

    a. A baseline demand set for which the design should be optimized
    b. A modified demand set, where some percentage of the baseline demands have different sources/destinations. The design is run using the equipment configuration that was chosen for the baseline design. For example, if a ROADM-only architecture is being used (i.e., ROADM-MDs are not permitted), the orientation of the ROADMs selected for the baseline design must be used in the modified design. This tests the forecast tolerance of the architecture.
    c. A projected growth demand set, to test the scalability of the design

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4, © Springer International Publishing Switzerland 2014

6. Be specific about what type of protection is desired (e.g., is shared protection suitable, or must dedicated protection be used).
7. If the exercise includes subrate demands, specify whether grooming/routing devices are required at all nodes, or whether traffic backhauling is acceptable. Additionally, be specific about what types of demands can be muxed or groomed together in a wavelength (e.g., what services, what protection types).
8. Provide some guidelines regarding the routing to ensure that comparisons across designs are valid. However, forcing all connections to always use the shortest path or explicitly specifying a path for each connection is too constrictive. It is preferable to specify a guideline such as the routed path for each connection should be no longer than P% longer than the shortest possible path.
9. Request design output in a specified format so that comparisons across system vendors can be readily performed.

Index

J. M. Simmons, *Optical Network Design and Planning,* Optical Networks,
DOI 10.1007/978-3-319-05227-4,