UNIVERSIDADE DO ALGARVE
FACULDADE DE CIÊNCIAS E TECNOLOGIA

Quality of Service in Optical Burst Switching Networks

Tese para a obtenção do grau de doutor no ramo de
Engenharia Electrónica e Computação,
especialidade de Arquitectura dos Sistemas Informáticos

Álvaro de Mascarenhas Pereira do Nascimento de Lima Barradas

Orientadora: Doutora Maria do Carmo Raposo de Medeiros

Constituição do Júri:

Presidente: Reitor da Universidade do Algarve

Vogais: Doutor Vasco Luís Barbosa de Freitas,
Professor Catedrático da Universidade do Minho

Doutor Mário Marques Freire,
Professor Catedrático da Universidade da Beira Interior

Doutor Pedro João Valente Dias Guerreiro,
Professor Catedrático da Universidade do Algarve

Doutor Armando Humberto Moreira Nolasco Pinto,
Professor Auxiliar da Universidade de Aveiro

Doutora Maria do Carmo Raposo de Medeiros,
Professora Auxiliar da Universidade do Algarve

Doutor Amine Berquia,
Professor Auxiliar da Universidade do Algarve

FARO
2009

NOME: Álvaro de Mascarenhas Pereira do Nascimento de Lima Barradas
FACULDADE: Faculdade de Ciências e Tecnologia
ORIENTADOR: Professora Doutora Maria do Carmo Medeiros
DATA: 20 de Novembro de 2009
TÍTULO DA TESE:

Qualidade de serviço em redes com comutação óptica de agregados de pacotes

UNIVERSIDADE DO ALGARVE

Quality of Service in Optical Burst Switching Networks

by

Álvaro Barradas

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

in the

Faculdade de Ciências e Tecnologia

Departamento de Engenharia Electrónica e Informática

2009

"Põe quanto és em tudo o que fazes"

Ricardo Reis

UNIVERSIDADE DO ALGARVE

Resumo

Faculdade de Ciências e Tecnologia
Departamento de Engenharia Electrónica e Informática

Doutoramento

de Álvaro Barradas

A comutação óptica de agregados de pacotes (OBS) promete ser um importante paradigma de suporte às redes da próxima geração. Contudo, o seu desempenho pode ser consideravelmente afectado pela disputa interna por recursos de rede. Existem várias propostas para resolução do problema, algumas exigindo disseminação frequente de tráfego de sinalização mas, na sua maioria, sem considerar um fenómeno intitulado *streamline effect*. A investigação nesta área costuma assumir total capacidade de conversão de comprimento de onda em todos os nós da rede, uma arquitectura presentemente bastante dispendiosa, bem como a adopção de mecanismos de resolução de contenções propensos ao aumento da complexidade dos nós. Com o objectivo de minimizar a contenção, nesta tese propõem-se mecanismos de encaminhamento de tráfego com selecção prévia de caminho, numa abordagem de engenharia de tráfego que utiliza apenas informação da topologia de rede. A ideia subjacente consiste em distribuir o tráfego pela rede de forma a reduzir congestionamentos sem incorrer em acréscimos de sinalização e, nalguns casos, tendo em conta o efeito de *streamline*. Após introdução ao OBS e revisão das principais contribuições no âmbito da sua QoS, são propostas seis estratégias de encaminhamento. A sua avaliação é feita por simulação recorrendo a um modelo de rede OBS especificamente desenvolvido para o efeito segundo o paradigma da Programao Orientada a Objectos (POO), usando quatro topologias diferentes e considerando para cada uma arquitecturas com/sem capacidade de conversão de comprimento de onda. Dependendo da conectividade da rede, os resultados mostram para estas estratégias melhor desempenho que o tradicional encaminhamento via caminho mais curto. A adopção das estratégias propostas não exclui uma utilização ocasional combinada com outros métodos de resolução de contenções por forma a ajudar a rede a recuperar de comportamentos instáveis, favorecendo a sua resiliência.

Palavras-chave: Redes ópticas, Comutação óptica de agregados de pacotes, Algoritmos de encaminhamento, Estratégias de selecção de caminhos, Simulação de redes, Avaliação de desempenho de redes.

UNIVERSIDADE DO ALGARVE

Abstract

Faculdade de Ciências e Tecnologia
Departamento de Engenharia Electrónica e Informática

Doctor of Philosophy

by Álvaro Barradas

Optical burst switching provides a feasible paradigm for the next IP over optical network backbones. However its burst loss performance is highly affected by burst contention. Several methods have been proposed to address this problem, some requiring the dissemination of frequent signaling messages, and most without considering the *streamline effect*. Reported studies also use to assume the existence of total wavelength conversion capacity on all nodes, which is presently a very expensive configuration, and adopt contention resolution schemes liable to increase the complexity on the network nodes. In this thesis we present a traffic engineering approach for path selection with the objective to minimize contention using only topological information. The main idea is to balance the traffic across the network in order to reduce congestion, without incurring into link state dissemination penalties and, in some cases, considering the streamline effect. After introducing OBS and related QoS state-of-the-art, we propose six path selection strategies. This strategies are evaluated by simulations performed on four different network topologies capable of full wavelength convertion on all nodes or forced to strict wavelength continuity constraint. Simulations were done on an OBS model specifically developed for the purpose under an object-oriented approach. Results show that our strategies outperform the traditionally used shortest path routing to an extent that depends on the network connectivity. The proposed strategies can be used alone or combined with other contention resolution schemes, used occasionally to help the network to recover from instability, favoring resilience.

Key-words: Optical networks, Optical burst switching, Routing algorithms, Path selection strategies, Network simulation, Network performance evaluation.

Acknowledgements

I would like to express my sincere gratitude to Doctor Carmo Medeiros, my supervisor, for introducing me to the exciting research area of optical networks. Carmo has been an immensely understanding and supportive mentor, always available for discussions and questions. Her advices and comments significantly enhanced the quality of my work.

I want to thank my research center, CEOT, for the working and computing facilities. I must also thank DEEI, my department at the Faculty of Science and Technology, for having permitted me to conduct research while exempt from teaching responsibilities. I am grateful to FCT, the Portuguese Foundation for Science and Technology, for partially funding my work via research projects at CEOT.

I would like to thank Doctor Marília Pires for having introduced me to the field of Operations Research some years ago when I was an undergraduate student. I owe thanks to Susana and Daniel, friends from "the math side of the world", for some help on complexity issues. I also thank Bastos for the tunneling solution for improved web access through a simple and quite elegant linux command.

Thanks go also to my research colleagues from CEOT, in particular Ricardo, José, Mark and Paulo for their enthusiasm and fruitful collaboration at the Telecommunications Lab. Many thanks to the Ubuntu and OMNeT++ developers for the excellent software products available in the open source domain.

My eternal gratitude to Beti, Edu and Ana for their outstanding support and endless empathy. Sincere thanks to Manu and Manuel for their indirect help in some family matters. I thank Zé for being herself, plenty of hopeful feelings about my major concerns, Sandy for the optimistic messages when they were needed, and all the friends that contributed to keep me sane.

Finally, I am deeply grateful to my wife Noélia who supported and encouraged me with all her love, strength and inspiration.

Contents

List of Figures

List of Tables

Abbreviations

AARA	Adaptive Alternative Routing Algorithm
ACK	Acknowledgment
ADSL	Asymmetric Digital Subscriber Line
ATM	Asynchronous Transfer Mode
BSS	Business Support Systems
BT	British Telecom
CC	Control Channel
CEOT	Center of Electronics Optoelectronics and Telecommunications
CLDR	Contention-based Limited Deflection Routing
CP	Control Packet
CTO	Chief Technical Officer
DC	Data Channel
DP	Drop Policy
DEEI	Department of Electronic Engineering and Informatics
DNA	Deoxyribonucleic Acid
DXC	Digital Crossconnect
EDFA	Erbium Doped Fiber Amplifier
FEC	Forwarding Equivalence Class
FDL	Fiber Delay Line
FDM	Frequency Division Multiplexing
FIFO	First In First Out
FiWi	Fiber-Wireless (networks)

FRR	Forward Resource Reservation
FTTH	Fiber To The Home
GMPLS	Generalized Multiprotocol Label Switching
GPMR	Gradient Projection Multipath Routing
GoOBS	Grid over OBS
GUI	Graphical User Interface
HDTV	High-Definition TeleVision
HHI	Heinrich Hertz Institute
ILP	Integer Linear Programming
IP	Internet Protocol
ISDN	Integrated Services Digital Network
IT	Information Technology
JET	Just-Enough-Time
JIT	Just-In-Time
LAUC	Latest Available Unused Channel
LAUC-VF	Latest Available Unused Channel with Void Filling
LIFO	Last In First Out
MAC	Medium Access Control
MCL	Maximum Congested Link
MEC	Maximum End-to-end Congested path
MPLS	Multiple Protocol Label Switching
NAK	Negative Acknowledgment
NED	Network Description (language)
NGN	Next Generation Network
OADM	Optical Add/Drop Multiplexer
OAMP	Optical Amplifier
OBS	Optical Burst Switching
OCS	Optical Circuit Switching
OEO	Opto-Electronic-Opto
OLS	Optical Label Switching
OLT	Optical Line Terminator

OMG	Object Management Group
OO	Object-Oriented
OOP	Object-Oriented Programming
OOS	Object-Oriented Simulation
OPS	Optical Packet Switching
OR	Operations Research
OSI	Open System Interconnection
OSS	Operations Support System
OT	Offset Time
OXC	Optical Crossconnect
OXS	Optical 'X' Switching, ($X \in \{C, L, B, P\}$)
PON	Passive Optical Network
PXC	Photonic Crossconnect
QoE	Quality-of-Experience
QoS	Quality of Service
QoSDG	QoS Development Group
QoS-MWST	QoS Multiple Wavelength Simultaneous Transmission
RAM	Random Access Memory
REL	RELease (trailing CP)
RNG	Random Number Generator
ROADM	Reconfigurable Optical Add/Drop Multiplexer
RWA	Routing and Wavelength Assignment
SBPR	Streamline Based Pre-planned Routing
SBPR-nPP	Streamline Based Pre-planned Routing without Pre-calculated Paths
SBPR-PP	Streamline Based Pre-planned Routing with Pre-calculated Paths
SDH	Synchronous Digital Hierarchy
SDL	Switched Delay Line
SDP	Service Delivery Platform
SOA	Semiconductor Optical Amplifier
SONET	Synchronous Optical Networking
SP	Shortest Path

STL	Standard Template Library
WDM	Wavelength Division Multiplexing
TAG	Tell-And-Go
TAW	Tell-And-Wait
TCP	Transmission Control Protocol
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
TE	Traffic Engineering
VCR	Virtual Channel Reservation
WADM	Wavelength Add/Drop Multiplexer
WDM	Wavelength Division Multiplexing
WR	Wavelength Routed

In love and gratitude to my wonderful wife
and our two little Algarvians

This thesis is also dedicated

- *to the memory of my father whose words are always with me. The opening quote in the beginning of this thesis was like a code of conduct for him, and became also a dominant principle in my life.*
- *to the memory of my grandmother 'Avó Zita' who have worked as an operator at the cordboard of the local phone company many years ago in Luanda, when the old "party line" was replaced by handmade switching (was it a "disruptive technology" in the 1920s?) leading to increased end-to-end QoS provisioning.*

Preface

The title "Quality of Service in Optical Burst Switching Networks" is so broad in scope that some preliminary considerations are necessary in order to contextualize the research work presented in this thesis. The title of the thesis can be divided into two parts, "quality of service (QoS)..." and "...optical burst switching networks".

About QoS

"QoS is like 'life' - a dictionary may define the word but it does not help us to understand its meaning"(Willis, 2005). This quotation from P.J. Willis[1] shows how a simple and well known expression can be difficult to define completely. In particular when the expression is made by two terms 'quality' and 'service' that are themselves subject to many interpretations. This view corroborates the opinion expressed by L.S. Cardoso[2] in the preface of (Hardy, 2001) where the importance of QoS in telecommunications is emphasized and its multiple interpretation nature is highlighted by the statement "Yet QoS means different things to different people".

In order to help to clarify the expression lets keep in mind that we are dealing with the telecommunication's field. And if 'service' can be generally understood as something that is provided day-to-day for the use of someone (user), under this particular field, a service is a particular capability to communicate with other parties by transmitting and receiving information in a way that is fully specified with

[1]BT's Group CTO for the Next Generation Network architecture.
[2]QoSDG Chairman, Marconi, Portugal

respect to several aspects: how the user initiates a communications transaction, the mode in which the information is exchanged, how the information is formatted for transmission, how are the end-to-end transactions achieved and billed (Hardy, 2001), hence, a specified set of information transfer capabilities provided to a group of users by a telecommunications system under strict technical specifications.

More difficult to describe is the notion of 'quality' that comes associated with the 'service'. Although generally understood as something related to 'excellence', expressed in the singular, and by which the user will form an opinion about how good the service is, 'quality', in this context, is much more plural. In fact, the factors that will determine how good the QoS is are rather multidimensional, both with respect to the attributes of the service and the perspectives on the service (Hardy, 2001). There are many independent attributes able to influence the user's perception of quality. And the inability to meet user expectations with respect to any of them cannot be counterbalanced by exceeding their expectations with respect to the others. This means that an effective measurement of QoS will almost always involve collections of measures rather than "the" measure that would be used to judge how good the QoS is.

What are the qualities users expect to find in a certain telecommunications service? Among other things, it may include value for money, ease of use, friendliness, style, flexibility, productivity, reliability and security, a list that excludes those qualities most network engineers think of when talking about QoS in networks: latency, the time it takes to send a signal across the network; jitter, the variability in the latency; loss, the amount of signal lost in transiting a network; availability, the amount of time the network is available to transfer signals; rate, the rate at which signals can be transferred; errors, even if the traffic unit is delivered on time, it may still be corrupted; delivery failure, which may or may not be associated with loss (Willis, 2005). This differences on how users of the network perceive QoS and network engineers describe QoS show that 'perspective' is also an issue when talking about QoS in the telecommunications field.

The perspective of the engineers can be thought of as Intrinsic QoS and is achieved by the design of the transport network, which determines the characteristics of the connections made through the network, the provision of network accesses, terminations and switch-to-switch links, determining whether the network will have adequate capacity to handle the anticipated demand. Intrinsic QoS is usually gauged by expected values of measures of operational performance and verified by demonstration that those scores compare favorably with analogous scores of competing services. The perspective of quality taken from the users point of view can be called Perceived QoS[3] and results from service usage. It is based on the user's experience of the Intrinsic QoS on their own activities, in their environment, in handling their demand, and reflect how they react to that experience in the light of their personal expectations (Hardy, 2001).

These QoS perspectives are obviously related and are both of great importance. If intrinsic quality is what can make a particular service attractive to a buyer in the first place, the perceived quality is what determines whether that buyer will consider that service acceptable when it is delivered. From these circumstance, it is desirable that engineers design the network with the technical characteristics that will achieve the adequate intrinsic quality in order to guarantee that the resultant perceived quality has a high probability of being considered acceptable by the users. In this thesis, the engineering perspective will be considered.

About the network

One major challenge facing networks in the near future is to provide all kinds of services over single delivery channels in a cost-efficient manner (Raisanen, 2003), for which the internet protocol (IP) has been considered a vital enabler. With its ability to be deployed over a multitude of under-layer protocols (ethernet, integrated services digital network (ISDN), frame relay or asynchronous transfer mode (ATM), for example) and over many network architectures (such as asymmetric

[3] Quality of Experience (QoE) is also used.

digital subscriber line (ADSL) or cable TV), the IP based interconnection is, beyond question, the universal vehicle of choice for accessing information, products, services and recreational activities.

With the increasing importance of the Internet in the global economy, the significance of the IP based interconnection will become even greater in the years to come. Several challenges are posed to IP based interconnection and to the traditional wireline and wireless networks which, notably, are already converging to IP based networks delivering some triple-play combination of data, voice and video (Marcus and Elixmann, 2008). A major challenge that service providers are facing today is how to evolve existing or build new delivery infrastructures capable of rapidly deploying new revenue generating services to grow their business and, at the same time, reduce operational costs. At the present time, however, industry efforts in responding to the service provider's needs are largely fragmented and generally fall into three categories: the first major effort is in the area of the next generation network (NGN) architectures, the second is in the operations and business support systems (OSS/BSS) area, and the third is in the service delivery platforms (SDP) (Hao, 2008).

The high bandwidth requirements generated by the new triple-play (and quad-play) services, in particular for video transmission, are rapidly exhausting the capabilities of legacy copper infrastructures in the access networks. For that reason network operators are installing a new network access infrastructure based on optical fibers (Hehmann and Pfeiffer, 2008). This shift from copper to optical that is presently under way in the access networks[4] is supported by the high bandwidth infrastructure in the backbone network and based in optical technology.

Three major switching techniques have been proposed in the literature for transporting IP traffic over optical networks. Accordingly, IP over optical networks can be further classified as using Wavelength Routed (WR) networks, Optical Packet Switching (OPS) networks, or Optical Burst Switching (OBS) networks.

[4]French cable manufacturer Nexans is predicting 20 million European homes will have a fiber connection by 2015. In Portugal, Sonaecom said it plans to invest about 240 million to build an open-access fiber-to-the-home (FTTH) network that will reach about 25% of the country's population over the next three years.

In WR networks a transparent long lived backbone pipe, usually called *lightpath*, is established between edges for end-to-end transmission. The lightpath concept was introduced in (Chlamtac et al., 1992), and is similar to a circuit in an electronic circuit switching network. This takes the form of optical circuit switching (OCS) without statistical sharing of resources. Both OPS and OBS support subwavelength granularity via statistical multiplexing without opto-electronic-opto (OEO) conversions. In OPS IP traffic is processed and switched on a packet-by-packet basis. Although in this way the bandwidth utilization of the network can increase, many technical challenges remain to be addressed to render this solution viable (Kaheel et al., 2002). The OBS switching technique, that combines the advantages of both WR and OPS, is the one considered in this thesis.

About this thesis

This dissertation is submitted in partial fulfillment of the requirements for the degree of *Philosophiae Doctor* (PhD) at the Department of Electronic Engineering and Informatics (DEEI) of the Faculty of Science and Technology, University of Algarve, Portugal. The presented work has been carried out in the period March 2006 - February 2009 at the Center of Electronics Optoelectronics and Telecommunications (CEOT), and was supervised by Dr. Maria do Carmo Medeiros, auxiliary professor in telecommunications for the Faculty of Science and Technology at the University of Algarve.

The work reported in this thesis is the author's contribution for the global improvement of some operational performance metrics through the use of a traffic engineering (TE) approach based on pre-planned routing strategies applied to OBS networks. Therefore, QoS is to be considered in its intrinsic engineering perspective, and the networks are restricted to the ones using the OBS paradigm. In this context, the service to be taken into account is the correct delivery of bursts to their final destinations and the quality of the service is evaluated in terms of burst loss performance.

During this period I had the pleasure and opportunity to participate in the following research projects as a member of team: IP sobre WDM - POSC/EEA-CPS/59556/2004, and RoFnet - PTDC/EEA-TEL/71678/2006.

I have participated also in the MUSE Summer School "Quality of service in packet-based broadband access networks", organized by the Heinrich Hertz Institute (HHI) for telecommunications and the National Technical University of Athens, co-located with NOC 2006 conference and hosted by HHI, 11-13 July 2006, Berlin.

Álvaro Barradas

1

Introduction

The basic function of a network is to enable communications between the desired end points. Despite the simplicity of this initial assertion, it is well known that the underlying properties of a network can greatly affect its performance. Network capacity, reliability, cost, scalability and operational simplicity are some of the most important benchmarks on which a network is evaluated, and network designers are frequently faced with tradeoffs among these factors in their quest for technological advances (Simmons, 2008).

A major advance was the advent of the fiber-optic cable, a critical milestone in the global telecommunications and information technology revolution. In fact, the ability to communicate worldwide on demand would not have been possible without the development of the silica fiber as a broadband medium for transporting voice, video and data traffic (Glass et al., 2000).

This lightweight cable, through which it is possible to achieve low loss transmission at higher data rates over long distances, has an impressive number of special features for communications: enormous potential bandwidth, small size and weight, electrical isolation, immunity to interference and crosstalk, signal security, ruggedness and flexibility, system reliability and ease of maintenance (Senior, 2008). This is an impressive list of advantages over the more conventional electrical transmission systems. But clearly, one of its most significant (and maybe initially unforeseen) benefits is its tremendous potential networking capacity, giving rise to optical networks and originating the field of optical networking.

An optical network is composed of the fiber-optic cables that carry wavelength channels, combined with the necessary equipment to process the light. Being so, the inherent capacities of optical networks are necessarily tied to the physics of light and the technologies that manipulate lightstreams. As such, the evolution of optical networks has been marked with major shifts as exciting breakthrough technologies are developed. One of the earliest advances was the ability to carry multiple channels of light on a single fiber, with each lightstream, or wavelength, multiplexed onto a single fiber (Simmons, 2008). This gave rise to Wavelength Division Multiplexing (WDM)[1], an approach that can exploit the huge opto-electronic bandwidth mismatch by requiring each end-user equipment to operate at electronic rate, but having multiple WDM channels combined on the same fiber. By allowing multiple channels to coexist on a single fiber, it is possible to exploit the huge fiber bandwidth, a demanding situation claiming for the design and development of the appropriate network architectures, protocols and algorithms (Mukherjee, 2006a).

Another milestone in this field was the development of the optical-bypass technology[2]. Before the introduction of this technology, supported by advances in optical amplification, optical switching, transmission formats and techniques to counteract optical impairments, the optical network served simply as a collection of static pipes. With optical by-pass technology, the network assumes another networking

[1] Key tech: EDFA

[2] Key tech: OXC and OADM

layer with routing and protection supported at the granularity of a wavelength. The optical-bypass technology allowed the elimination of much of the required electronic terminating and switching equipment, permitting a signal to remain in the optical domain for all of its path from source to destination. While providing a scalable trajectory for network growth, it also brought new challenges to optical networking, most notably, the development of new algorithms to assist in operating the network so that the full benefits of the technology could be reached (Simmons, 2008).

Although less obvious, beyond the prodigious bandwidth available in the optical fiber and other physics considerations, one of the most important features of an all-optical path is its complete lack of protocol dependency. Excluding link budget concerns, it does not even matter what is the used bit rate. This valuable digital transparency has the important practical consequence that old protocols may be given “artificial respiration” to extend their lifetimes while new ones are being developed (Green, 2001). Another advantage is functional simplicity, which has important consequences in lowering first costs and offering service lifetime savings by making problem determination and maintenance potentially much simpler than they are with the traditional fiber systems with opto-electronic-opto (OEO) conversion. This way, the large first costs and service lifetime costs of communications software and supporting hardware can be almost confined to the end nodes of the all-optical network. One last advantage is the intrinsic robustness of an all-optical network due to the replacement of the old physical layer. In fact, the new optical physical layer can provide not only basic transport, but also several other networkwide rerouting functions that allow the optical layer to exhibit extremely high service availability. And this can be done without incurring into bit handling operations as required, for example, by the SONET/SDH protection switching or IP packet rerouting deployed today in most telecoms backbones using WDM (Green, 2001; Mukherjee, 2006a).

Emerging all-optical networks are expected to provide optical switched connections, or lightpaths, between edge routers over an optical WDM core network (Alferness et al., 2000). Since the already deployed connections are fairly static, they may not be able to accommodate the bursty nature of the internet traffic in an efficient manner. Ideally, in order to provide the highest possible utilization in the optical core, nodes would need to provide packet switching at the optical level. However such all-optical packet switching is likely to be infeasible in the near future due to technological limitations (Xu et al., 2001). A possible near-term alternative to optical circuit switching (OCS) and to optical packet switching (OPS) is optical burst switching (OBS). In OBS, packets are concatenated into bigger transport units referred to as bursts, which are switched and forwarded through the optical core network without leaving the optical domain. OBS networks allow for a greater degree of statistical multiplexing and are better suited for handling bursty traffic than OCS networks. At the same time, OBS networks do not have as many technological constraints as OPS networks. The main differences between these three all-optical transport technologies will be highlighted in the next chapter where OBS is presented with more detail.

1.1 Motivation

OBS is usually presented as combining the merits of both OCS and OPS while avoiding their shortcomings. However, like the other switching paradigms, OBS does not perform well when the network becomes heavily loaded and contention is likely to increase. Contention occurs when the number of simultaneous reservation attempts exceeds the number of available resources, i.e., when multiple bursts from different input ports are destined for the same output port simultaneously. This is an highly undesirable situation because, due to the lack of sophisticated optical buffers, contention can result in burst loss. Burst loss degrades the global OBS performance since dropping leads to a rescheduling of lost data with significant impact on any end-to-end application running in the upper layers reducing its overall throughput. Therefore, burst loss reduction is considered a key factor for a practical realization of OBS networks. The general solution to burst contention is to move all but one burst “out of the way”. An OBS node has three possible dimensions to move contending bursts: time, space and wavelength. The main corresponding contention resolution schemes are optical buffering, deflection routing and wavelength conversion (Chua et al., 2007a).

Considerable effort has been devoted to the study of different node based QoS improvement approaches to reduce burst loss, with channel scheduling algorithms and burst segmentation schemes added to the contention resolution schemes just mentioned above. These are mainly reactive mechanisms that attempt to resolve contentions rather than avoiding the contentions, and usually requiring extra hardware or software components at each core node, significantly increasing their cost and complexity, leading to scalability impairments. Moreover, some of the research on these schemes revolve around objectives that are somewhat in conflict and for which tradeoffs are becoming difficult to obtain when looking for new improvements (Li and Qiao, 2004). Even without the wished crystal ball that would allow for an insight into the future, it is generally assumed that, concerning the multitude of directions in which technology can be improved, telecoms players would decide in favor of a solution based on a number of very obvious and measurable criteria: trust in a technology and architecture, cost benefits, operational simplification and scalability. “Disruptive technologies” are accepted only in cases where they would drive a fundamental improvement in any dimension of the just mentioned criteria (Berger et al., 2006).

A simple and cost efficient alternative to resolving contention when it occurs is to prevent contention before it happens. With contention avoidance, the goal is to reduce the number of contentions by policing the traffic at the source, or by routing the traffic in a way that congestion in the network is minimized (Thodime et al., 2003). This is a key form of traffic engineering where traffic from congested areas is diverted to lightly loaded areas, and has long been considered an essential feature of any NGN. Load balancing problems can be mathematically formulated as optimization problems, such as integer linear programming (ILP), which is a widely used approach to address both high level and system level synthesis (Mignotte and Peyran, 1997), and can be used in OBS networks for performance improvement through optimized routing path selection. Although research in OBS has taken a multitude of directions, the selection of routing paths is an important area that has received relatively little attention despite the profound impact that routing can have on the overall performance of an OBS network (Teng and Rouskas, 2005b). Driven by this context, the aim of the research presented in this thesis is the development of routing path selection strategies to minimize the global network contention and the overall burst loss of OBS networks using only topological network information.

1.2 Methodology

To accomplish these objectives, the proposed routing path selection strategies are formulated as ILP problems. This approach is generally considered as having very high computation complexity. Therefore, it is mainly intended to be computed offline. However, taking into account the computation times of some of the proposed algorithms, the infrequent update requests expected as a consequence of changes in the OBS backbones whose topologies typically last for long time scales, and the quasi-stationary aggregate traffic demands at optical backbones, which are expected to change relatively slowly (Chen et al., 2008), the online deployment is not completely excluded. This can be carried out by means of an operation process executed during the network initial setup phase, either in a centralized manner, where routes are computed by a central node which has knowledge about the global topology and downloaded to the nodes when the network is booted, or in a distributed manner, if the nodes are equipped with topology searching capabilities.

The routes obtained can be applied as single-path static routes and used alone to provide load-balancing without the need for resource-update signaling messages regarding the congestion status of the network links. Alternatively, they can be combined with some other dynamic contention resolution schemes (like deflection or segmentation, for example) and used occasionally as a default routing option to assume whenever the network needs to recover from instability. This can particularly be done when the activity of multiple dynamic network elements, reacting simultaneously to congestion, may result in oscillation between congestion and de-congestion states on certain links (Thodime et al., 2003). It is widely believed that dynamic operation in optical networks helps to overcome the inefficiencies of static operation. However, a recent study (Li, 2008b) reports several offline routing and wavelength assignment algorithms which perform better than previously studied online algorithms.

Our approach presents also the following *a priori* advantages: no extra hardware or software components are required on the core nodes and no network flooding with signaling messages resulting from (over)active link state update protocols. Moreover, with this approach there is also no place for out-of-order arrivals, a disadvantage of some dynamic contention resolution schemes, also found on multipath routing schemes, typically requiring large memories at the edge nodes for re-ordering operations. These are important features to achieve operational simplification, and to make the architecture of the OBS nodes less complex, contributing to some extent to reduce both their cost and scalability impairments.

Naturally, the adopted approach has also some limitations. For instance, since it uses ILP, some limitations arise for scalability. But considering that in OBS networks, even when taking different approaches, the potential for contention when reserving resources is likely worse as the network increases, this technology may be by itself better suited for smaller regional backbones as opposed to bigger and less connected long-hauls.

The results presented in this thesis show that OBS networks can significantly improve their performance if our proposed routing strategies are adopted and hence increase their quality of service. Thus, considering the aforementioned criteria for technological improvements and the benchmarks for network evaluation previously referred, this methodology can be considered feasible for the development of OBS networks.

1.3 Contribution

The contribution of this thesis for the development of OBS networks includes: (1) a survey on the main OBS technologies and related state-of-the-art, (2) the development of one algorithm for calculation of K shortest paths with less links in common, and the proposal of six algorithms for pre-planned routing optimization entirely based on the network topology information, and (3) the design and implementation of an OBS simulation model specifically developed for testing and evaluation of pre-planned routing algorithms. This contribution can be summarized in the following item list:

1. Survey on main OBS technologies and their "state-of-the-art"
2. Proposal of novel pre-planned routing strategies
   - With full wavelength conversion capability
     - Calculation of link disjoint pre-selected paths
     - Minimize the maximum congested link (MCL)
     - Minimize the maximum end-to-end congested path (MEC)
     - Streamline based pre-planned routing with pre-calculated eligible paths (SBPR-PP)
     - Streamline based pre-planned routing without pre-calculated eligible paths (SBPR-nPP)
   - Without wavelength conversion capability
     - Streamline based pre-planned routing with pre-calculated eligible paths (SBPR-PP)
     - Streamline based pre-planned routing without pre-calculated eligible paths (SBPR-nPP)
3. Simulation model
   - Routing path determination stage

- Primitive routines for route calculation
- Route calculation (optimization)

- Routing path application stage
  - Primitive routines for route selection
  - OBS network (simulation)
  - Routines for data extraction and evaluation

1.4 Published Work

This section lists not only papers published or submited for publication during the period of work for this thesis (marked with a '•' symbol), but also some previously published work (marked with a '○' symbol). Papers more closely related with the subject of this thesis are identified by a '➪' symbol.

JOURNALS:

➪ Alvaro L. Barradas, and Maria do Carmo R. Medeiros, "Pre-planned optical burst switched routing strategies considering the streamline effect," Photonic Network Communications, (under review).

➪ Alvaro L. Barradas, and Maria do Carmo R. Medeiros, "Edge-Node Deployed Routing Strategies for Load Balancing in Optical Burst Switched Networks," ETRI Journal, vol.31, no.1, Feb. 2009, pp.31-41.

• Maria C. R. Medeiros, Ricardo Avó, Paula Laurêncio, Noélia S. Correia, Alvaro Barradas, Henrique J. A. da Silva, Izzat Darwazeh, John E. Mitchell, and Paulo M. N. Monteiro, "RoFnet Reconfigurable Radio over Fiber Network Architecture Overview", Journal JTIT, Jan. 2009, pp.1-6.

○ M.C.R. Medeiros and A.L. Barradas, "Design and Performance Analysis of a Media Access Control Protocol for a Wavelength Division Multiplexing

Circuit-switched Network", WMRC Global Optical Communications 2002, Jun. 2002, pp.89-92.

CONFERENCE PROCEEDINGS:

- ➪ A.L. Barradas and M.C.R. Medeiros, "Analysis of path selection mechanisms in OBS networks", Proc. NOC 2009, Valladolid 10-12 June 2009, Spain. (paper accepted)

- ➪ A.L. Barradas and M.C.R. Medeiros, "Pre-planned Optical Burst Switching Routing Strategies", Proc. ICTON-MW 2008, Sa1.1, Marrakesh 11-13 December 2008, Morocco.

- ➪ A.L. Barradas and M.C.R. Medeiros, "An OMNeT++ Model for the Evaluation of OBS Routing Strategies" Proc. OMNeT++ 2008 Workshop on ACM SIMUTOOLS 2008, Marseille 3 March 2008, France.

- Mark Guerreiro, Claunir Pavan, Alvaro L. Barradas, Armando N. Pinto, and Maria C.R. Medeiros, "Path Selection Strategy for Consumer Grid over OBS Networks" Proc. 10th Anniversary International Conference on Transparent Optical Networks ICTON 2008, 2008, vol.3 pp.138-141, Athens 22-26 June 2008, Greece.

- M. Guerreiro, A.L. Barradas and M.C.R. Medeiros "Path Selection Strategy for OBS Networks Based on a Probabilistic Model for Link Demands" Proc. Cranfield Multistand Conference 2008, Cranfield May 2008, UK.

- ➪ A.L. Barradas and M.C.R. Medeiros, "Path Selection Strategies for OBS Networks Using Topological Network Information", Proc. ICTON-MW 2007, Fr1B.4, Sousse 6-8 December 2007, Tunisia.

- M.C.R. Medeiros, R. Avó, P. Laurêncio, N.S. Correia, A. Barradas, H.J.A. da Silva, I. Darwazeh, J.E. Mitchell and P.M.N. Monteiro, "Radio Over Fiber Access Network Architecture Employing Reflective Semiconductor Optical Amplifiers" Proc. ICTON-MW 2007, pp.1-5, Sousse 6-8 December 2007, Tunisia.

- N.S.C. Correia, A.L. Barradas and M.C.R. Medeiros, "A Novel Restoration Scheme for Optical WDM Networks", LCS 2002 Communications Symposium, London 9-10 September 2002, U.K.

- N.S.C. Correia, A.L. Barradas, M.C.R. Medeiros and J.J. OReilly, "Design of Survivable All-Optical WDM Networks using a Primary-Shared Scheme", CRC2002, 5 Conferncia sobre Redes de Computadores, Sesso de Posters, 26-27 Setembro 2002, Portugal.

- N.S.C. Correia, A.L. Barradas, M.C.R. Medeiros and J.J. OReilly, "Routing Weight Functions for WDM Networks using Wavelength Rerouting", SCI 2001/ISAS 2001, 5th Multiconference on Systemics, Cybernetics and Informatics and 7th International Conference on Information Systems Analysis and Synthesis, (7) 163-168, Florida July 22-25 2001, U.S.A.

- A.L. Barradas, N.S.C. Correia, M.C.R. Medeiros and P. M. Lane, "Performance of a WDM MAC Protocol Under Bursty Traffic", SCI 2001/ISAS 2001, 5th Multiconference on Systemics, Cybernetics and Informatics and 7th International Conference on Information Systems Analysis and Synthesis, (7) 159-162, Florida July 22-25 2001, U.S.A.

- M.C.R. Medeiros, I. Darwazeh, L. Moura, A.L. Barradas, A. Teixeira, P. Andr, M. Lima and J. da Rocha, " Dynamically Allocated Wavelength WDM Network Demonstrator ", ConfTele 2001, 3rd Conference on Telecommunications, 211-215, Figueira da Foz 23-24 April 2001, Portugal.

- N.S.C. Correia, A.L. Barradas, M.C.R. Medeiros and J.J. OReilly, "A Novel Wavelength Rerouting Algorithm for WDM Networks Based on Vertex Coloring Techniques", ConfTele 2001, 3rd Conference on Telecommunications, 211-215, Figueira da Foz 23-24 April 2001, Portugal.

- N.S.C. Correia, A.L. Barradas, M.C.R. Medeiros and J.J. OReilly, "A Lightpath Rerouting Algorithm for WDM Circuit-Switched Networks", 26th European Conference on Optical Communications (ECOC), Workshop on Modelling and Design of Optical Networks and Systems, 19, Munich 3-7 September 2000, Germany.

- N.S.C. Correia, A.L. Barradas, M.C.R. Medeiros and J.J. OReilly, "A Lightpath Rerouting Algorithm for WDM Circuit-Switched Networks", LCS 2000 Communications Symposium, 177-180, London 14-15 September 2000, U.K.

- A.L. Barradas, N.S.C. Correia, M.C.R. Medeiros and J.J. OReilly, "Design and Performance Analysis of a MAC Protocol for WDM Single-Hop Networks", LCS 99 Communications Symposium, 73-76, London 26-27 July 1999, U.K.

1.5 Structure of the Thesis

After this introductory Chapter 1 where the motivation, methodology and contribution of the thesis are presented, Chapter 2 presents an overview on the development of optical networking, including their multiplexing techniques, their deployment trend from the core to the edges of the network, and some challenges that arise from the access side of the network for which OBS is considered a feasible paradigm. In fact, as presented in Chapter 3, where an introduction to the most common switching paradigms for optical transport are introduced, OBS avoids the inefficient resource utilization of OCS and the requirements of buffers, optical logic processing, and the synchronization problems of OPS. OBS is extensively discussed in Chapter 4 starting from its most commonly proposed framework, and following by addressing its burst assembly algorithms, resource reservation schemes, scheduling algorithms and contention resolution approaches. Contention resolution is a quite important subject in OBS networks and several methods have been proposed to reduce burst loss due to contention with their own advantages and disadvantages. Despite the undeniable merits of the research conducted on such methods and the merit of some QoS provisioning proposals presented in Chapter 5, they have also some important drawbacks which pave the way for alternative proposals to reduce contention. One alternative approach to contention resolution is *contention avoidance*, which is the aim of this thesis. Six strategies to avoid contention are proposed in this thesis and evaluated by simulation using an OBS network model specifically developed for the purpose using an object oriented platform for discrete event simulations. This simulation model is presented in Chapter 6. The proposed strategies are presented in pairs. The first two, relying on a link based and on a path based approach, are presented in Chapter 7. Then four strategies follow in Chapter 8 relying on a recently reported phenomenon typical from OBS networks, which is called *streamline effect*. In this chapter, the last two strategies use the streamline effect in a continuity constraint scenario. Finally, some concluding remarks are made in Chapter 9, where an outlook to future work is also pointed out.

2

Optical Networking Overview

Existent and emerging optical networks use different multiplexing techniques to share the bandwidth of the optical fiber. However WDM presents important advantages over the other approaches which make this multiplexing technique the most promising for supporting the next generation Internet. This chapter starts with a succinct description of the enabling technologies that give support to optical networks and presents their main characteristics following the commonly used classification in *generations*. The general tendency of optical networks to be deployed from the core towards the edge is also addressed, and some major concerns regarding unprecedented demands for bandwidth are discussed.

2.1 Multiplexing Techniques

Multiplexing is a technique that allows multiple traffic sources to share a common transmission medium. In the context of optical networks, three main multiplexing approaches have been deployed to share the bandwidth of optical fiber: time division multiplexing (TDM), space division multiplexing (SDM) and wavelength-division multiplexing (WDM).

TDM is a well known technique successfully used in many network electronic architectures throughout the more than 50-year history of digital communications (Green, 1996). However, in the context of high-speed optical networks, TDM is under the pressure of the so-called "electro-optical bottleneck". This happens because the optical TDM signal carries the aggregate traffic of multiple sources and each network node must be able to operate at the aggregate line rate rather than the sub-rate corresponding to the traffic of a specific individual node. Clearly, the aggregate line rate cannot scale to arbitrarily high values but is limited by the fastest electronic transmitting, receiving, and processing technology. Consequently, TDM faces severe problems to fully exploit the huge bandwidth available in the optical fiber.

SDM is a straightforward approach to avoid the electro-optical bottleneck where multiple fibers are used in parallel instead of a single fiber. Each of these parallel fibers may operate at any arbitrary line rate. Although well suited for short-distance transmissions, becomes less practical and more costly for increasing distances due to the fact that multiple fibers need to be installed and operated (Green, 1996).

WDM is widely used to tap into the vast amount of fiber bandwidth. It can be though of as frequency division multiplexing (FDM), where traffic from each source is sent on a different carrier frequency. In optical networks the term wavelength is usually used instead of frequency but the principle remains the same. When using WDM, each transmitter sends on a separate wavelength λ_i where $1 \leqslant i \leqslant T$. At the transmitting side, a wavelength multiplexer collects all wavelengths and feeds

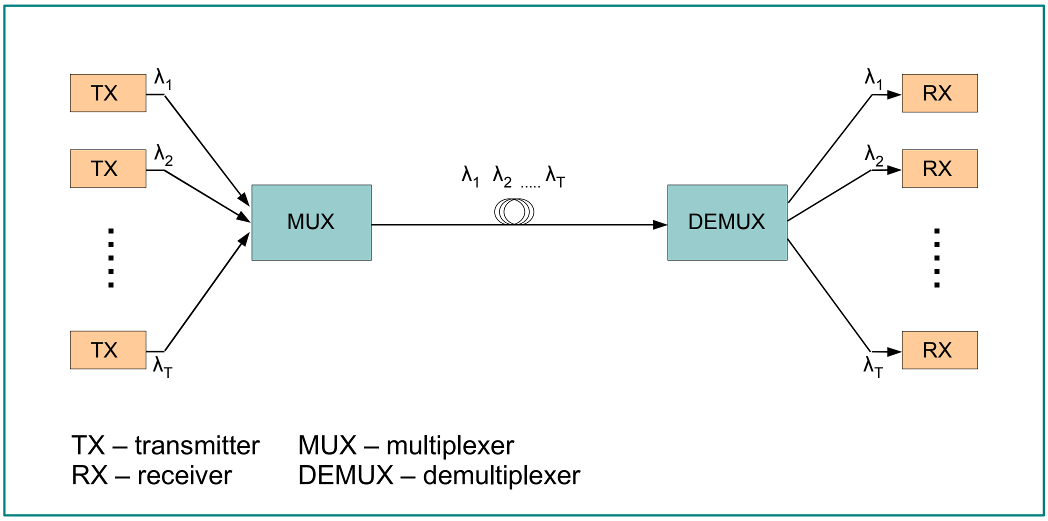

FIGURE 2.1: Wavelength division multiplexing.

them onto a common outgoing fiber. At the receiving side, a wavelength demultiplexer separates the wavelengths and forwards each wavelength λ_i to a different receiver i (see Figure 2.1). Unlike TDM, each wavelength channel may operate at any line rate well bellow the aggregate TDM line rate. Thus, by using multiple wavelengths, the huge potential bandwidth of optical fiber can be exploited. As opposed to SDM, WDM takes full advantage of the potential bandwidth of a single fiber and does not require multiple fibers to be installed and operated in parallel, resulting in significant cost savings (Maier, 2008b).

Driven by the digital revolution and TDM's suitability for the higher data rates (achievable) that would minimize the cost per bit for long distance transmissions, TDM was implemented in several backbone transmission formats. WDM had to wait until TDM rates became limited by transmission impairments and the optical amplifier made WDM a cost effective architecture (Alferness et al., 2000). The existing and emerging optical networks use all three multiplexing techniques, sometimes together, to realize high performance network architectures.

2.2 Development of WDM Networks

The first generation of WDM networks provide only fixed or manually configured point-to-point physical links. In such networks all traffic arriving to a node is dropped, converted from optics to electronics, processed electronically, and converted from electronics to optics before departing from the node. The dropping and adding of traffic at every node in the network incurs significant overhead both in terms of switch complexity and electronic processing cost (Jue and Vokkarane, 2005a). The main technical issues associated with this primitive WDM implementation included the design and development of lasers and optical amplifiers (OAMP) (Glass et al., 2000).

The second generation of WDM is capable of establishing connection-oriented end-to-end lightpaths in the optical layer. Now ring and mesh topologies can be implemented, and lightpaths are operated in a virtual (over physical) topology that can be dynamically reconfigured in response to traffic changes. The key enabling technologies for this achievements were the development of optical add/drop multiplexers and optical crossconnets succinctly described next:

Optical add/drop multiplexer (OADM) also called WADM (Wavelength Add/Drop Multiplexer), is a device that takes in a composite optical signal that consists of multiple wavelengths and selectively drops (and subsequently adds) some of the wavelengths before letting the composite signal out of the output port.

Optical crossconnet (OXC) Is a device with multiple input and multiple output ports. In addition to add/drop capability, it can also switch a wavelength from any input port to any output port. Switching optical signals in an all-optical device is the second approach[1] to realize an OXC. Such a switch is often called a transparent OXC or photonic crossconnect (PXC).

[1]The first is implemented in the electrical domain and called electronic OXC or opaque OXC.

These network elements were developed under the premise that the majority of the traffic that enters a node is being routed through the node to its final destination as opposed to being destined for the node. If that transiting traffic could remain in the optical domain as it traverses the node rather then incurring successive OEO conversions for electronic processing and retransmission, the costs associated with providing high-capacity switching and routing at each node would be substantially reduced (Simmons, 2008). Both OADM and OXC may employ wavelength converters, which nowadays is considered a very complex and expensive device. Besides wavelength conversion, other technical issues of this second generation of WDM networks include the routing and wavelength assignment (RWA) problem, interoperability among WDM networks, network control and management hardware and software (Mukherjee, 2006a).

These first and second WDM generation networks have been deployed in several operational carriers, initially only in the long-haul but after they moved closer to the network edges. The economical reason behind this movement is that as one moves closer to the edge, the cost of a network in a particular hierarchical level[2] is amortized over few end users while increases the price. Because of this difference in price sensitivity there is often a trend to deploy new technologies in the backbone first, to be gradually extended towards the edge of the network as the technology matures and lower prices can be achieved (Simmons, 2008). The deployment of WDM is an example of this trend (see Figure 2.2). An important fact to be aware of, is that, as optics enters the access network enabling the proliferation of high-bandwidth end-user applications, there is potential for huge bandwidth growth throughout the network hierarchy, and telecoms companies must be prepared to handle high volumes of traffic without being overwhelmed.

The third generation of WDM, is also based on the OADM and OXC network elements described above, but this generation is expected to support connectionless optical networking, while providing high levels of statistical multiplexing. The key issues now include the development of an optical access network, and the

[2]Core, metro or access.

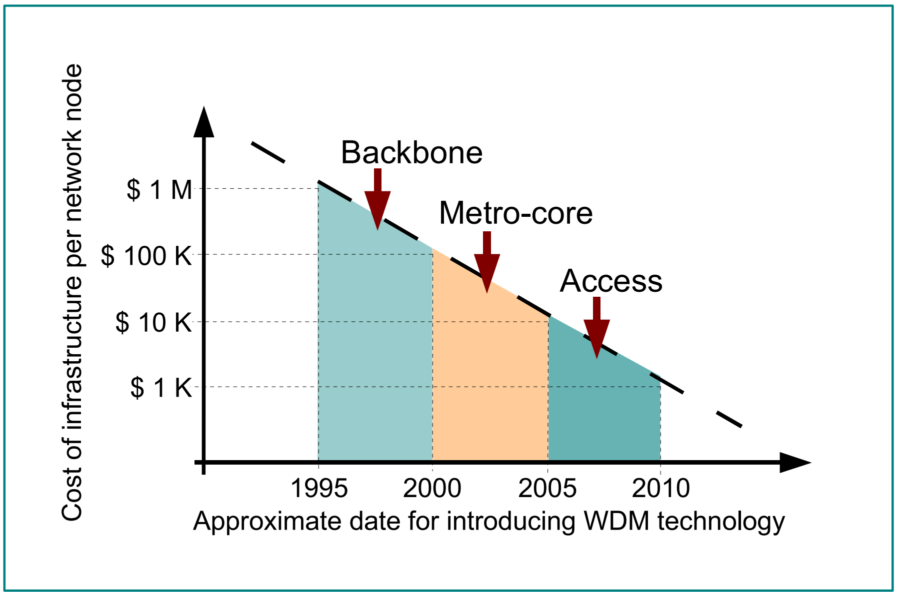

FIGURE 2.2: The effect of cost reduction in WDM infrastructure (Simmons, 2008).

development of optical switching technologies (see Figure 2.3). This next generation network is envisioned to have two main functional parts: an inner core and multiple access networks (Chua et al., 2007a). The access networks today are compatible with the Internet transport architecture and are responsible for collecting, predominantly, IP traffic from end-users. These are build from electronic or low-speed based optical transport technologies such as Gigabit Ethernet, optical rings or passive optical networks (PONs). These access networks are then connected together by the inner core network which consists on a mesh of reconfigurable optical switching elements (OXC and OADM) interconnected by very high capacity long-haul optical links. The connection between the access network and the inner core is established by means of high-speed edge nodes.

2.3 Challenges from the Access Side

The access networks field can be considered relatively marginal to the subject of this thesis. However, access networks are the source of the traffic that the core handles. Therefore, it seams appropriate to highlight in this section one of the

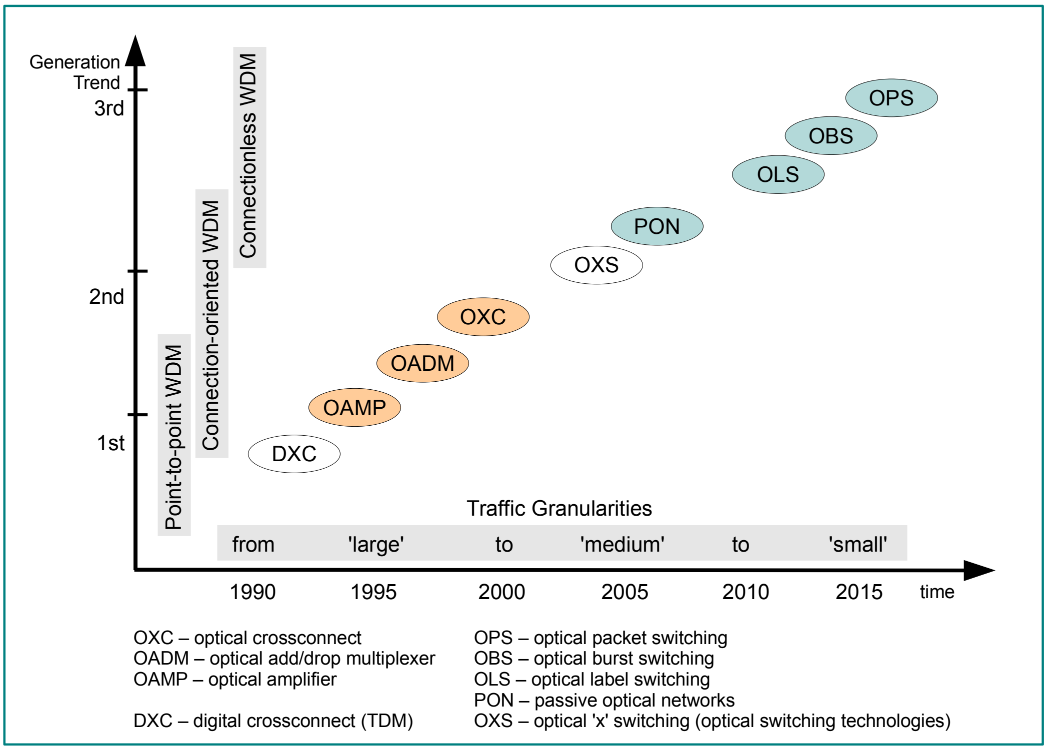

FIGURE 2.3: WDM network evolution (Mukherjee, 2006a).

major concerns of that research area. For more technical content the reader is referred to Effenberger et al. (2007); Cameron et al. (2007); Heron et al. (2008).

Access networks are expected to bring, over the next decade, a demand for bandwidth, mainly driven by video applications, that can seriously challenge the viability of conventional network architectures. The bandwidth available to users in the access networks has grown exponentially over the past 20 years in a self-feeding cycle were the insatiable appetite for new services has driven the need for new innovative higher speed technologies, and the availability of the later has enabled the creation of new imaginative services (Heron et al., 2008). To support this growth the underlying data technologies have evolved from dial-up modems to integrated services digital networks (ISDN), then to successive generations of asymmetric digital subscriber lines (ADSL) and, more recently, to some direct fibers to the home (FTTH). At the same time the traditional telecom operators have undergone a very important move to “all IP”, by which the old amalgam of legacy networks to provide residential and business voice and data have been replaced by an all Internet Protocol (IP) network (Weldon et al., 2008).

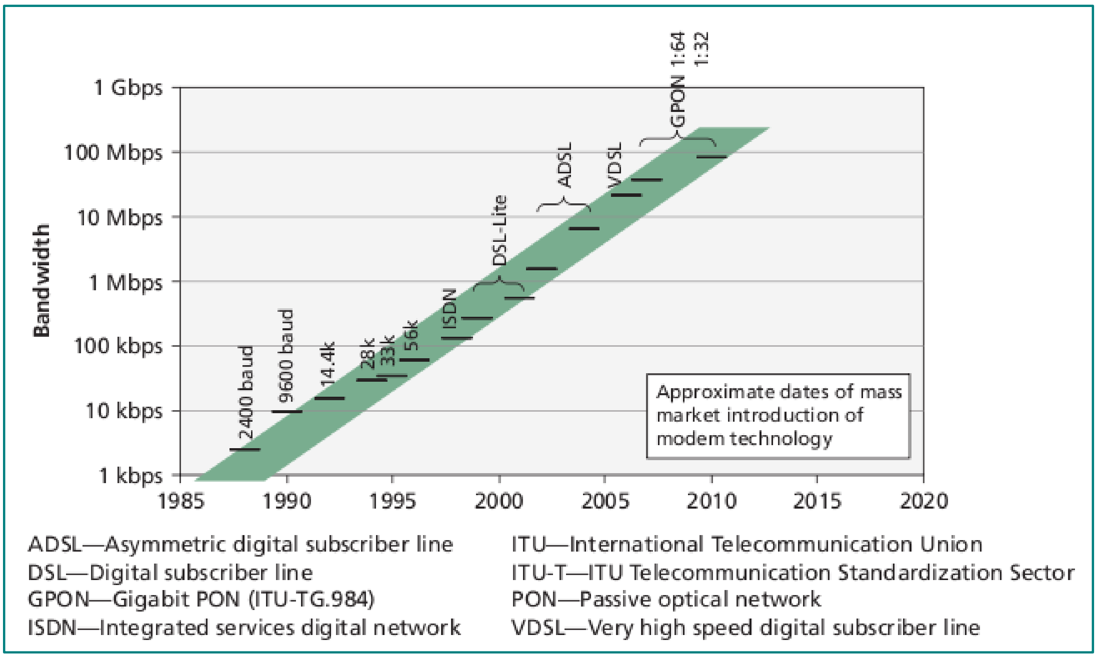

FIGURE 2.4: Bandwidth growth for the access networks (reprinted from Heron et al., 2008, with permission of John Wiley & Sons, Inc. ©2008.).

Historically, there has been a 10-fold increase in bandwidth every 6 years (see Figure 2.4), and there is evidence that this trend will continue in the future (Heron et al., 2008). Today, successful commercial triple-play service packages are offered with a bandwidth in the range of 20 Mbps to 30 Mbps per user[3], for which one could envisage a service set demanding over two or three times that bandwidth for the near future[4]. Although economically challenging, current technologies are quite capable of meeting these near-term bandwidth requirements with the current ADSL driven broadband.

But in future, this situation can be very different. There will be new services and applications which will demand yet more bandwidth, in particular video based. Some could include: transmitting an entire high-definition television (HDTV) video as a file; the delivery of large screen TVs beyond 100 inches by electronic

[3]Todays standard video, voice, and data package consists of 1 HDTV IP video signal (8 Mbps, MPEG-4), 1 SDTV IP video signal (2 Mbps, MPEG-4), 1 phone line (64 kbps), and high speed Internet (5 Mbps). Total bandwidth consumption 15 Mbps (Meis, 2006).

[4]In 5 years, the standard video, voice, and data package is likely to consist of 1 Super HDTV IP video signal (32 Mbps, MPEG-4), 1 HDTV IP video signal (8 Mbps, MPEG-4), 1 phone line (64 kbps), and high speed Internet (20 Mbps). Total bandwidth consumption 60 Mbps (Meis, 2006).

vendors (such as LG, Philips, Panasonic, Samsung, etc.) driving service innovations and increased bandwidth usage; delivery of multiple image video providing multiple angle views of sports events, or a wider angle vision (180 or 360 degree) analogous to the Imax experience; 3D video combined with increased definition; interactive 3D video experiences could be envisaged such as 3D gaming, 3D conferencing, 3D virtual travel, 3D e-learning, 3D acting participative experiences, and much more. Although these service concepts are indeed speculative and their timing uncertain, there is no doubt that competition and technology innovations will produce imaginative services and applications that will exploit additional available bandwidth (Heron et al., 2008). Hence, the potential for growth from a services driven perspective is massive, possibly two or three orders of magnitude greater than the bandwidth from the broadband networks today (Payne, 2008). In this new "e-domesticity" vision[5], simply scaling the technology of today's networks will not produce viable solutions and operators will need to carefully consider new architectural approaches to network build. The role of optical networking to meet this future unprecedented demands will be crucial.

2.4 Summary

This chapter provides an overview of the optical networking field. It starts with a description of the main multiplexing techniques deployed in the context of optical networks, and proceeds with the characterization of the prevailing tendency which is based in WDM. Optical networks are then classified in three generations and their main features are presented. Finally, the driving forces for optical deployment towards the edge of the network are addressed and relevant bandwidth related issues are discussed.

[5] A tag borrowed from WIRED Magazine, issue 10.08 (http://www.wired.com).

3

Switching Technologies for WDM

The development of optical switching technologies is generically referred to as Optical $'X'$ Switching (OXS), where X = {C, P, B, L, F} for circuit, packet, burst, label, and flow, respectively (Mukherjee, 2006a). The three major techniques proposed in the literature for transporting IP traffic over WDM-based optical networks are the OCS, OPS and OBS. They will be described and compared in this chapter.

3.1 OCS Overview

In optical circuit switching (OCS) the principle of operation is connection-oriented and the transmission of data from a source node to a destination node is realized on pre-established WDM channels, also called lightpaths (Chlamtac et al., 1992).

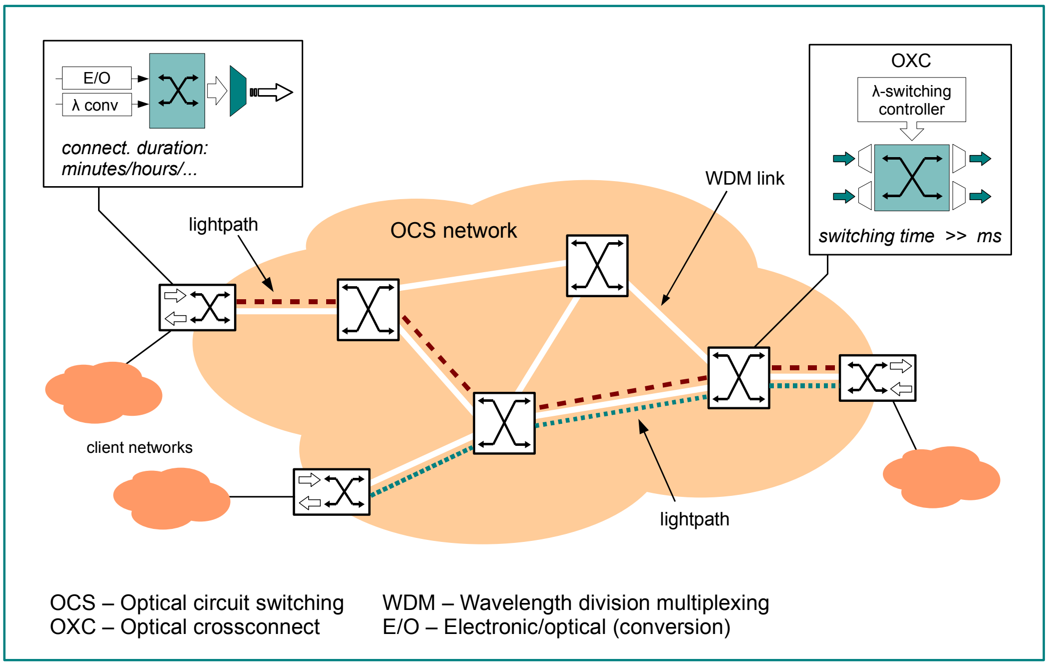

FIGURE 3.1: OCS network (adapted from Klinkowsky, 2007).

These lightpaths are created by reserving dedicated wavelength channels on every link along the physical path (see Figure 3.1), and they can be either static or dynamic. In the former they are setup manually, and often last for several days. In the later, lightpaths are automatically configured according to traffic related requests from the layers above. Lightpaths are switched from one link to another at the intermediate nodes referred to as optical crossconnects (OXCs). These devices are responsible for the all-optical switching of the data being carried on an incoming wavelength (usually denoted by λ) in a given input port, to an outgoing wavelength in an output port. If the OXCs are equipped with wavelength converters, then a lightpath may be converted from one wavelength to another along the route. Otherwise, the wavelength continuity constraint property is applied, i.e., the very same wavelength is used in all the links along the path.

The OCS technique is well-known from the wavelength routed (WR) networks, i.e., networks were the wavelength itself serves as an identifier defining how routing/switching decisions for lightpaths are made. It is a mature technology, already deployed in the core of some major long distance carriers providing good QoS for the admitted traffic. Although this approach can be considered a significant

Table 3.1: Pros and cons shortlist for OCS

√	Mature technology
√	Good QoS for the admitted traffic
×	Large granularity
×	Long reconfiguration delay
×	Inefficient utilization of link resources

improvement over the first generation point-to-point architectures, it has some disadvantages:

1. Large granularity. The smallest switching entity in OCS networks is the wavelength, which can be considered a quite coarse switching granularity. Typical OCS connection durations are expected to be found in a range of some seconds with connection setup and release performed in milliseconds (El-Bawab, 2006).

2. Fairly static lightpath connections. Fixed-bandwidth WR connections may not be able to efficiently accommodate the highly variable and bursty nature of the Internet traffic (Jue and Vokkarane, 2005a). If traffic is varying dynamically, sending this traffic over static lightpaths would necessarily result in inefficient utilization of bandwidth.

3. The number of connections in a network is usually much greater than the number of wavelength, and the required capacity for a connection is much smaller than the capacity of a wavelength. Therefore, despite the spacial reuse of wavelengths, it is neither possible nor efficient to allocate one wavelength to every connection (Chua et al., 2007a).

In Table 3.1 some of the main advantages and disadvantages of OCS networks are summarized.

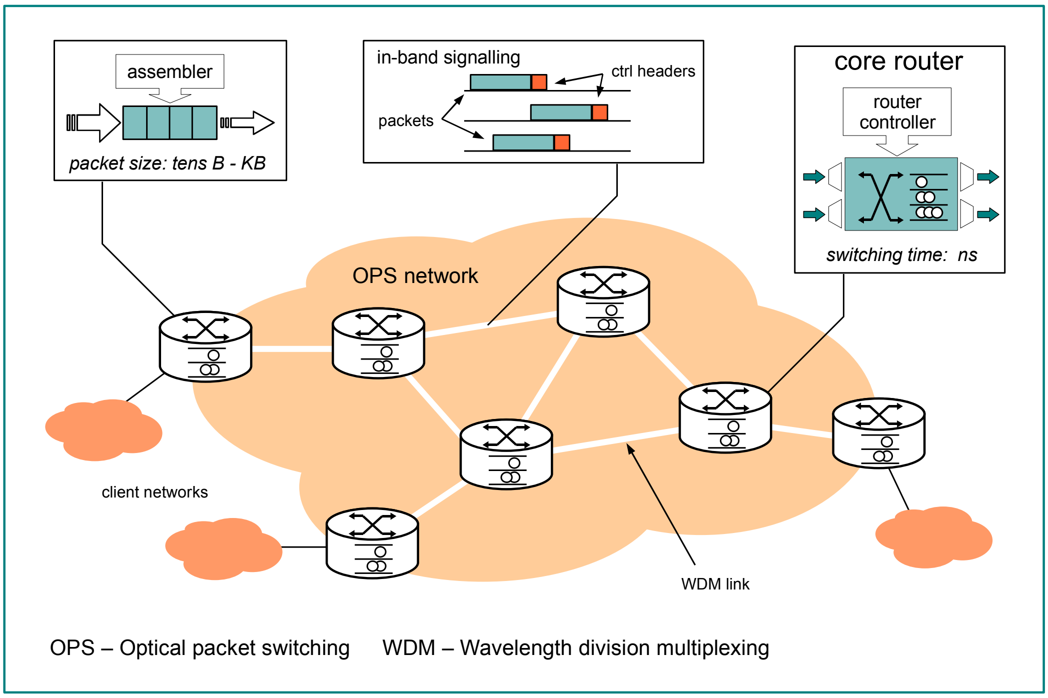

Figure 3.2: OPS network (adapted from Klinkowsky, 2007).

3.2 OPS Overview

The operation principle of the optical packet switching (OPS) networks is based on packet-by-packet processing and switching (Hunter et al., 1999). In OPS, packets containing payload and header are switched and routed independently throughout the network without any prior reservation (see Figure 3.2). In the envisioned photonics paradigm this should be done entirely in the optical domain. However, due to numerous technological limitations, the research community has been investigating different forms of OPS where control, at least in part, is carried out by electronics until optical processing becomes mature. In the meanwhile, the adopted approach whenever a packet is received in an optical switch is to extract and convert the header to its electrical form. The header is then processed in an electronic node controller while the payload is optically delayed using a fiber delay line (FDL), and forwarded after proper switching configuration. Some of the current technological challenges to the development of OPS networks are:

1. Header processing. In this *modus operandi*, each packet has to be optically

extracted, then converted to its electrical form for the electronic processing that will drive the configuration of the switching matrix so that the packet payload can be properly switched in the optical domain. Considering that this operation is to be applied to every single packet, it must be accomplished very fast or a significant amount of overhead is introduced.

2. Synchronization is another challenge in photonic packet switching networks. In OPS networks, synchronization of fixed-length packets at switch input ports is often desired in order to minimize contention. Therefore, it is especially important for OPS that the switching fabric be capable of fast switching of data on a packet-by-packet basis. Only technologies with very high switching speed[1] have been considered as candidates for optical switching in OPS networks. Without neglecting some progress (Mack et al., 2008; Chen, Wolfson, Johansson, Blumenthal and Coldren, 2006), the subject is still in an early stage of laboratory development.

3. Lack of optical random access memory (RAM) for buffering is another limitation. Since network resources are not reserved in advance, packets may experience contention. Contention in packet switching networks is typically handled through buffering. However, presently, there is no optical equivalent of RAM and buffering is realized by limited FDLs, not by fully functional memories. Research continues to report some progress in the area (Wang et al., 2008; Huang et al., 2007) but the use of FDLs continues severely limited. First because even the smallest delays require long lengths of fiber[2]. Secondly, once FDLs are first-in-first-out (FIFO) systems with predetermined delays, they tend to be rather inflexible, unless used in conjunction with switches. But even in that case, if a packet is to be delayed at a certain input port, subsequent packets of the same path may have to go through the same delay as well, degrading the performance of the switching node (Willner et al., 2006).

[1] for example, Semiconductor Optical Amplifiers (SOAs).

[2] 1 μs of delay requires 200m of fiber.

Table 3.2: Pros and cons shortlist for OPS

√	"Familiar" architecture
√	Great flexibility and statistical multiplexing
×	Per packet processing overhead
×	Stringent synchronization and switching requirements
×	Lack of optical RAM

The objective of OPS is to enable packet switching capabilities at rates comparable with those of optical links and thereby replacing wavelength routing in the next generation of optical networks (Chua et al., 2007a). That would bring an impressive increase in network utilization by means of statistical multiplexing for bandwidth sharing. The transmission time of a typical IP packet at 10 Gbps is in the range of tens of ns to 1 μs and further decreases at higher bit-rates. With the current technology this is not possible. The main advantages and disadvantages of OPS are summarized in Table 3.2.

3.3 OBS Overview

Optical burst switching (Qiao and Yoo, 1999; Turner, 1999) is designed to achieve a balance between OCS and OPS. The principle of operation of OBS is tailored to its basic transport unit, the data burst, which is an aggregate message composed of several data packets coming from client networks and assembled at the ingress node in agreement with some relevant criteria. These data bursts, or *bursts* for short, are to be transmitted and switched all-optically throughout the OBS network. The transmission of each burst is preceded by the transmission of a dedicated control packet, on a separate channel, with the purpose of reserving bandwidth along the path for the upcoming bursts. Based on the information carried by the control packet, the intermediate nodes reserve switching resources along a certain path, providing an optical channel through which the bursts can be transmitted, after an adequate delay, from source to final destination without leaving the optical domain (see Figure 3.3).

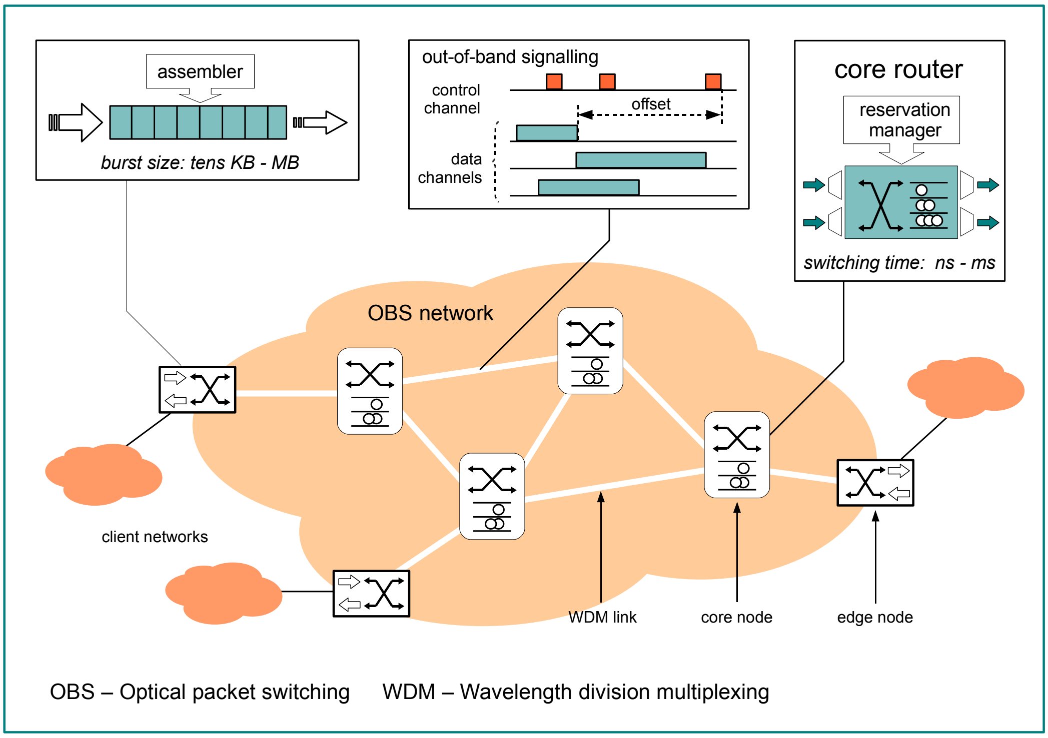

FIGURE 3.3: OBS network (adapted from Klinkowsky, 2007).

With this technology, the network resources can be dynamically reserved for small and very specific periods of time, enabling statistical multiplexing and efficiency in terms of bandwidth allocation. This was unachievable by OCS technology. In addition, since data is transmitted in large bursts, processing overhead can be reduced and the technology requirements for switching are not so hard when compared with the fast switching requirements of OPS. The duration of a typical burst can last from some μs to several hundreds of ms.

While OPS is difficult to implement, OBS presents enormous advantages over OCS (more details of OBS are given in Chapter 4). But besides dispelling the traditional "regard"[3] of some telecom players at the so-called disruptive technologies, several issues need also to be considered before this technology can be deployed in working networks. In particular signaling schemes, contention resolution and QoS, to be addressed later in this thesis. The list of advantages and disadvantages of this technology is presented in Table 3.3.

[3] "... that long fixed look with suspicion (disruptive technologies = money cost)."

TABLE 3.3: Pros and cons shortlist for OBS

√	Moderate processing overhead
√	Asynchronous switching
√	High statistical multiplexing
×	Lack of RAM
×	Moderate/fast switching needed
×	Relatively new, unfamiliar technology

3.4 O$'X'$S Comparison

To make clear the distinguishing features of the three switching paradigms just presented (OCS, OPS and OBS), a comparison will be made considering the following five criteria commonly used in literature (Chua et al., 2007a; Jue and Vokkarane, 2005a): bandwidth utilization, setup latency, switching speed, processing complexity[4], and traffic adaptivity.

Bandwidth utilization: In OCS networks, setting up a lightpath is like setting up a circuit. Either established statically or dynamically, in OCS, the entire bandwidth of each lightpath is dedicated to one pair of source and destination nodes and unused bandwidth cannot be reclaimed by other nodes ready to send data. This means that this transport technology offers no statistical multiplexing gain. In contrast, link bandwidth utilization in OPS and OBS networks can be considerably improved since traffic between different pairs of source and destination nodes are allowed to share link bandwidth, supporting statistical multiplexing.

Setup latency: In OCS networks, a two-way dedicated signaling message need to be sent between each source-destination pair before establishing the end-to-end lightpath. This incurs into long reconfiguration delays. In OPS and OBS networks only one-way signaling is required. Therefore, the connectionless service offered by OPS and OBS helps to reduce network latency, which can be considered low as compared to OCS.

[4]Processing of control headers and related synchronization issues.

Switching speed: In OCS, switching is made at the granularity of an optical wavelength, which means that the smallest switching entity is the entire lightpath. Since the lightpath has a relatively long duration, the required switching speed can be relatively slow. On the contrary, in OPS networks, switching is made at the packet level. Thus, switches need to switch incoming optical packets quickly to different outgoing ports upon their arrival. Therefore, fast switching and reservation capabilities are required for OPS networks. For OBS networks, due to the large granularity of bursts and their associated offset times, the switching requirements can be considered moderate.

Processing complexity: This criteria is closely related with the switching granularity. The processing effort required by long duration lightpaths in OCS networks is relatively low when compared with OPS and OBS networks. In OPS, since the switching entity is an individual packet and each packet has its own header requiring processing, the associated complexity will be quite high. In OBS networks the switching entity is an individual data burst which aggregates multiple individual packets. Therefore the effort put in processing is between that of OCS and OPS.

Traffic adaptivity: OPS and OBS, on the other hand, improve utilization of network resources by statistical multiplexing of traffic streams, in particular when traffic is dominated by data. They are more suited to bursty traffic then OCS, where switching is performed at the granularity of wavelength and latency is quite high.

The differences and similarities just discussed are emphasized in Table 3.4 where the advantages of OBS technology become visible, especially if considered together with the technological improvements foreseeable for the near future. This table also evidences why OBS is usually presented as a technology that combines the merits of both OCS and OPS while avoiding their respective shortcomings.

Table 3.4: Comparative evaluation of optical switching paradigms (adapted from Chua et al., 2007a; Verma et al., 2000)

Criteria	**OCS**	**OPS**	**OBS**
Granularity	coarse	fine	moderate
Bandwidth utilization	low	high	high
Setup latency	high	low	low
Switching speed	slow	fast	moderate
Processing complexity	low	high	medium
Traffic adaptivity	low	high	high

3.5 Summary

In this chapter the three major techniques for IP-over-optical transport, OCS, OPS and OBS, have their concepts described and compared. For each of these techniques, main advantages and disadvantages are weighted up according to specific features and summarized in “pros & cons” tables. The distinguishing features of the three switching paradigms are then discussed with more detail with respect to the following five criteria: bandwidth utilization, setup latency, switching speed, processing complexity and adaptivity to bursty traffic. OBS combines the best characteristics of both OCS and OPS architectures.

4

Optical Burst Switching

Optical burst switching (Qiao and Yoo, 1999; Turner, 1999) is one of the most promising architectures to transmit bursty traffic over an all-optical infrastructure. Based on a burst switching concept[1] initially introduced in Amstutz (1983) and Kulzer and Montgomery (1984), its development in optical communications has arisen as an alternative to a low-flexible OCS operation and to a technologically immature OPS solution. OBS technology relies on advances on several key network elements, including all-optical switches, burst mode receivers and optical wavelength converters (Jue and Vokkarane, 2005b). This switching paradigm has received considerable attention in the past few years, and various solutions have been proposed and analyzed in an attempt to improve its performance, including burst assembly techniques, channel scheduling schemes, contention resolution

[1]Initially for integrated transfer of voice and data over TDM links.

methods and QoS provisioning (Chen et al., 2004). Although still being developed (it has not been standardized yet), OBS is presently being considered the winning network technology to enable the next generation optical Internet (Battestilli and Perros, 2005; Rodrigues, Freire, Monteiro and Lorenz, 2005). Among other reasons, the success of OBS relies on two crucial features: it was built upon one-way signaling schemes to diminish the delay associated with the round trip time, and on its small control overhead for a large amount of payload data. This chapter describes the main features of this technology, its benefits, as well as some of its challenges.

4.1 General Framework

An OBS network consists of OBS capable nodes interconnected by optical fiber links supporting multiple WDM channels. In this meshed networks, nodes can be either edge nodes or core nodes (see Figure 3.3). Edge nodes are primarily responsible for assembling data packets into bursts, and scheduling the bursts for transmission on outgoing wavelength channels. This nodes are commonly referred as edge routers and, just as its name indicates, are to be located at the boundary of the OBS network, providing inter-working facilities between the outside world (client networks) and the optical OBS network. Core nodes are mainly responsible for switching bursts from input ports to output ports and for handling contentions. This nodes, also referred as core routers, compose the inner core of the OBS network, and comprise an optical switching fabric, a switch control unit, and processors for routing and signaling. Users, outside the OBS network, are typically electronic IP routers equipped with legacy networking cards through which packet traffic is forwarded (and received) to (and from) the OBS network.

User data (for example, IP packets) goes through a burst assembly/disassembly process at the edge of the OBS network. The burst, the fundamental transmission and switching unit in OBS networks, is composed of *data payload* and *control*

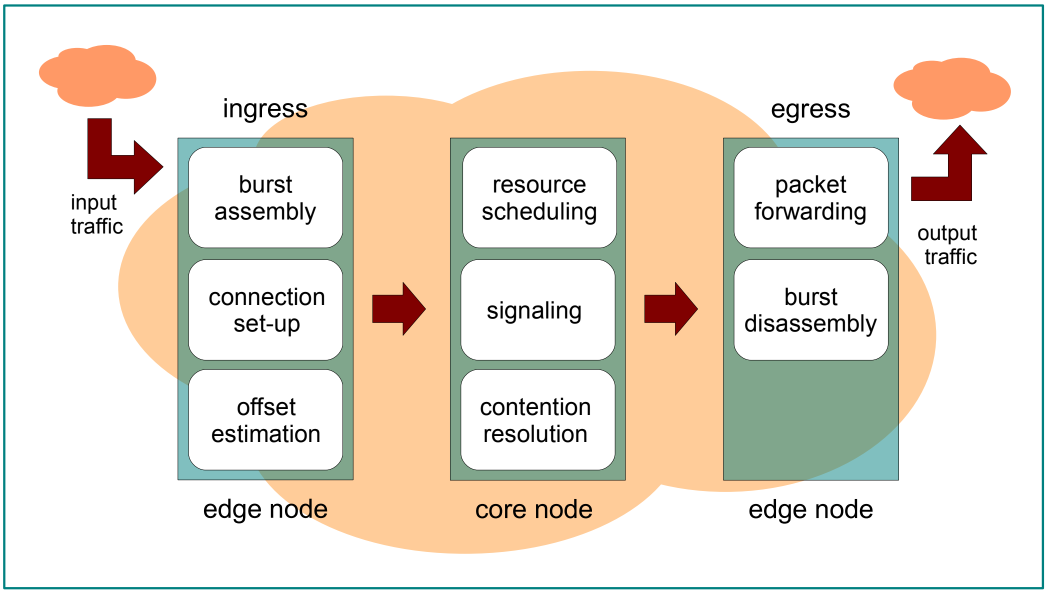

FIGURE 4.1: Main functions of OBS nodes.

packet. The data payload contains the actual data packets after being aggregated and assembled into a bigger transport unit also called data burst. The control packet carries all the relevant information of the corresponding data burst in order to drive several network operations in the core network. The information carried by the control packet can include, for example, the length of the data burst and its class of service. In OBS networks the control packet and the data burst are transmitted separately on different channels (wavelengths). This separation between the control and the data planes allows for great network flexibility and scalability. In OBS networks, while the burst is transmitted all-optically through the core, the control packet undergoes OEO conversion at each intermediate OBS node. These features allow also for two important savings (Chen et al., 2004): first, with the assembly/disassembly process (only) at the edge, data needs to be buffered only at the edge nodes where electronic RAM is available, cheap and abundant; secondly, having data and control signals transmitted separately, costly OEO conversions are only required on a few control channels instead of a large number of data channels. The functions carried out by OBS edge nodes and core nodes (see Figure 4.1) are described bellow.

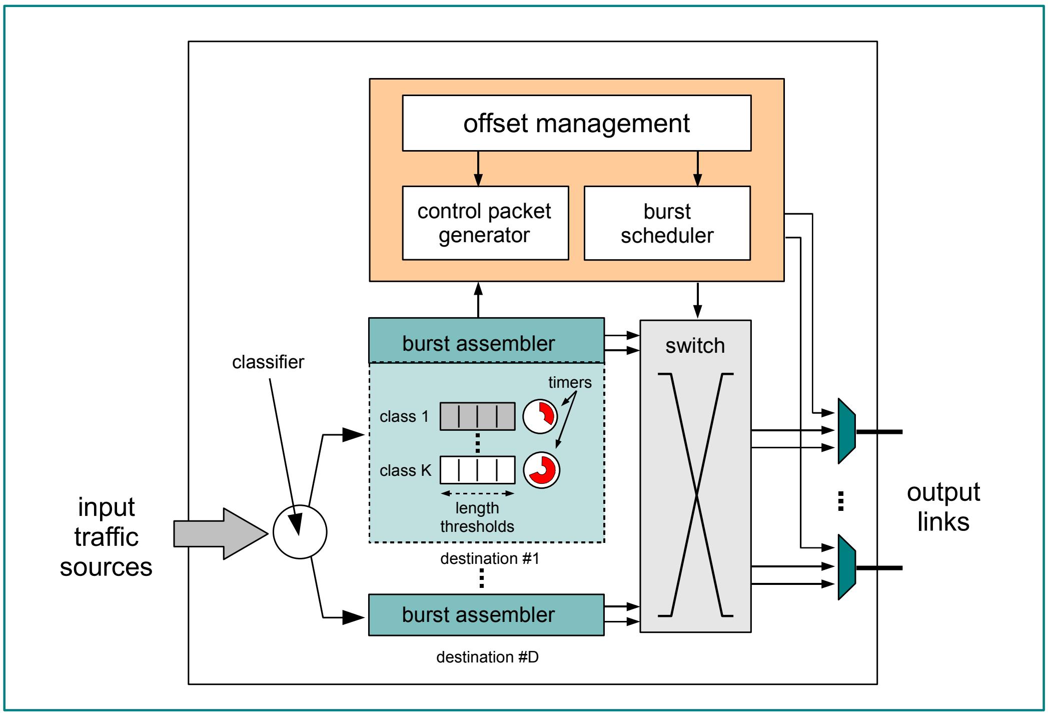

FIGURE 4.2: Ingress OBS edge node (Klinkowsky, 2007).

4.1.1 Edge node

Depending on their sending/receiving status, edge nodes can be viewed either as ingress or as egress nodes. When acting as ingress nodes (see Figure 4.2), edge nodes are responsible for the following functions: burst assembly, connection setup, and offset estimation for the control packet.

4.1.1.1 Burst assembly

This is the process of assembling incoming data packets into bursts (see Figure 4.3). In general, each edge node maintains multiple queues to aggregate the data packets from client networks according to their destinations and QoS requirements. A data burst is then assembled from one of these queues according to a given assembly algorithm whose purpose is to decide when to stop the data aggregation process. Several assembly algorithms have been proposed and investigated up to date, as discussed later in this chapter, but they usually have to take the following parameters into account: a preset timer, minimum burst size and maximum burst

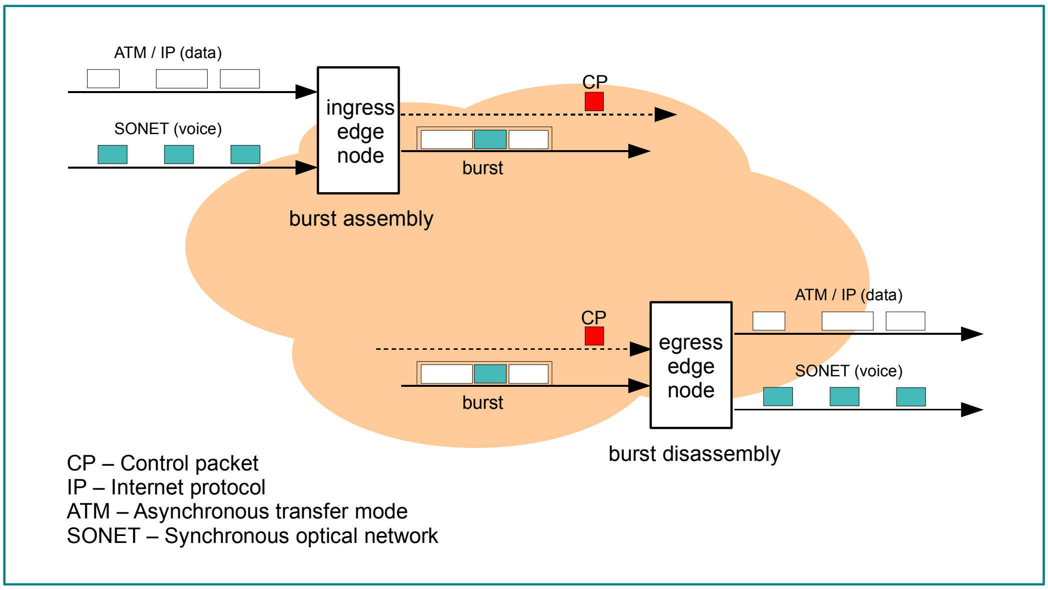

FIGURE 4.3: Burst assembly and disassembly at the edge of an OBS network.

size. The timer is used by the edge node to decide when to assemble a new burst. The minimum and maximum burst size parameters determine the length of the assembled burst. These parameters require careful setting because long bursts can hold network resources for long time periods, prone to higher burst losses, while short bursts result in increased number of control packets.

4.1.1.2 Connection setup mechanisms

For transporting an assembled burst over the optical core a temporary connection must be established between a given pair of edge nodes. This procedure, by which the allocation and configuration of network resources required for burst transmission is accomplished, consists of three main components (Battestilli and Perros, 2003): (1) signaling, used to setup the connection for the burst; (2) routing, used to decide the path of a burst through the network; and (3) wavelength assignment, used to determine on which particular wavelength to transmit the burst. These mechanisms, succinctly introduced next, will be under more detailed attention bellow in dedicated sections of these chapter.

Signaling: is a very important aspect of an OBS architecture. It specifies the protocol by which OBS nodes communicate connection requests to the network and its operation determines whether or not the network resources are utilized efficiently. In OBS networks signaling is implemented out-of-band, either by means of a dedicated wavelength channel or by means of a separate control network, and can rely on a distributed approach using one-way reservation or on a centralized approach using end-to-end reservation.

Routing: is the process of selecting the paths along which to send the network traffic. Thus, it is also considered a crucial functionality for network performance. In OBS networks, the routing process is used to decide the path of a burst through the core network and can be accomplished in three different ways: (1) on a hop-by-hop basis, (2) based on the generalized multiprotocol label switching (GMPLS) approach, and (3) using explicit pre-calculated setup. When routing is done on a hop-by-hop basis, like in the IP networks, a fast table look-up algorithm must be used to determine the next hop. In the second approach routing is done by deploying GMPLS based routing protocols to compute explicit or constraint-based routes at OBS edge nodes as described in Mannie (2004). Labeled OBS (LOBS) suggested in Qiao (2000) is the natural starting framework to design efficient routing strategies of this kind. Although a full GMPLS/OBS integration is still considered difficult to obtain(Pedroso et al., 2007), the cooperation between GMPLS and OBS is attracting attention and is being subject of several recent proposals (Pedroso et al., 2007; Long et al., 2006; Wen et al., 2005). The third approach is to use explicit pre-calculated setup connections, which can be established via constraint-based route label distribution protocol (CR-LDP) or resource reservation protocol with traffic engineering (RSVP-TE). Explicit routing is very useful in a constraint based routed OBS network where the traffic routes have to meet certain QoS metrics such as delay, hop count, bit error rate, or bandwidth (Battestilli and Perros, 2003).

Wavelength assignment: is used to determine on which particular wavelength to transmit the burst. Along the selected path, each link must also be assigned a wavelength on which the bursts are carried. In OBS networks wavelength assignment with and without wavelength conversion at the intermediate nodes is possible, whereby wavelength conversion may be fixed, limited-range, full-range, or sparse (Maier, 2008a). Most studies assume that full wavelength conversion is available throughout the network. In practice, however, economic and technical considerations are likely to dictate a more limited and sparse deployment of wavelength converters in the optical network. Therefore, wavelength assignment policies are expected to be an important component of OBS networks (Teng and Rouskas, 2005a).

4.1.1.3 Offset estimation

As it was already mentioned, the transmission of each data burst is preceded by the transmission of a control packet, whose purpose is to inform each intermediate node of the upcoming data burst so that it can configure its switch fabric in order to switch the burst to the appropriate output port (Teng and Rouskas, 2005a). More precisely, after sending the control packet, the edge node waits for a fixed or variable offset time until it starts transmitting the corresponding burst. The offset is used to allow the control packet to be processed, reserve the required resources, and configure the optical switching fabric at each intermediate node along the adopted path before the corresponding burst arrives (Maier, 2008a). If the control packet succeeds in its resource reservation and switch configuration objectives, the arriving burst can optically bypass each intermediate node without any buffering or processing. Hence, an appropriate estimation of the offset is also crucial for the network performance. Ideally, the offset estimation should be based on the number of traversed OBS nodes and the processing and switch setup times at each of them. In practice, however,the number of intermediate nodes may not be known to the source edge node. Moreover, the current level of congestion in the OBS network can also have an effect on the correct offset estimation. Obviously, with an incorrect estimation, the bursts can arrive at a certain OBS node prior to the proper switch configuration, which would result in burst loss. The offset estimation is, without doubt, a key design feature of all OBS networks to achieve high resource utilization and low burst loss. Despite their differences on how exactly to determine the offset time, estimation methods can be classified into three groups as follows (Battestilli and Perros, 2003):

Fixed offsets: This is the most popular scheme, where the offset time is fixed and equal to the sum of the total processing time at all the intermediate OBS hops plus the switch fabric configuration time of the egress OBS node (Qiao and Yoo, 1999). This offset estimation requires the precise number of hops from source to destination, for which the processing and configuration times are assumed to be the same. In practice, however, these times may vary from node to node due to possible queueing delays in the control channel.

Statistical offsets: This is a variable (statistical) offset generation scheme where each edge node generates transmission tokens based on a Poisson process with a predetermined arrival rate. In this scheme, as soon as a burst is assembled, its corresponding control packet is immediately sent into the OBS network, while the data burst is delayed until it is able to obtain a transmission token. By randomizing the offset generation process it is possible to prevent any undesirable systematic synchronization on the transmission of data bursts that could cause excessive burst loss. This offset estimation method regulates the rate at which data bursts are released into the OBS network (Verma et al., 2000) thus, reducing burst loss probability.

WR-OBS offsets: In this architecture the offset is calculated as the sum of the time it takes an edge node to request resources from the centralized scheduler, the computation time of the routing and wavelength allocation algorithm, and the path signaling time (Duser and Bayvel, 2001).

Most of these methods are based on the assumption that the control packet is to be sent only after the moment at which the entire data burst is assembled. A variation on these techniques is to send the control packet slightly before that moment. This

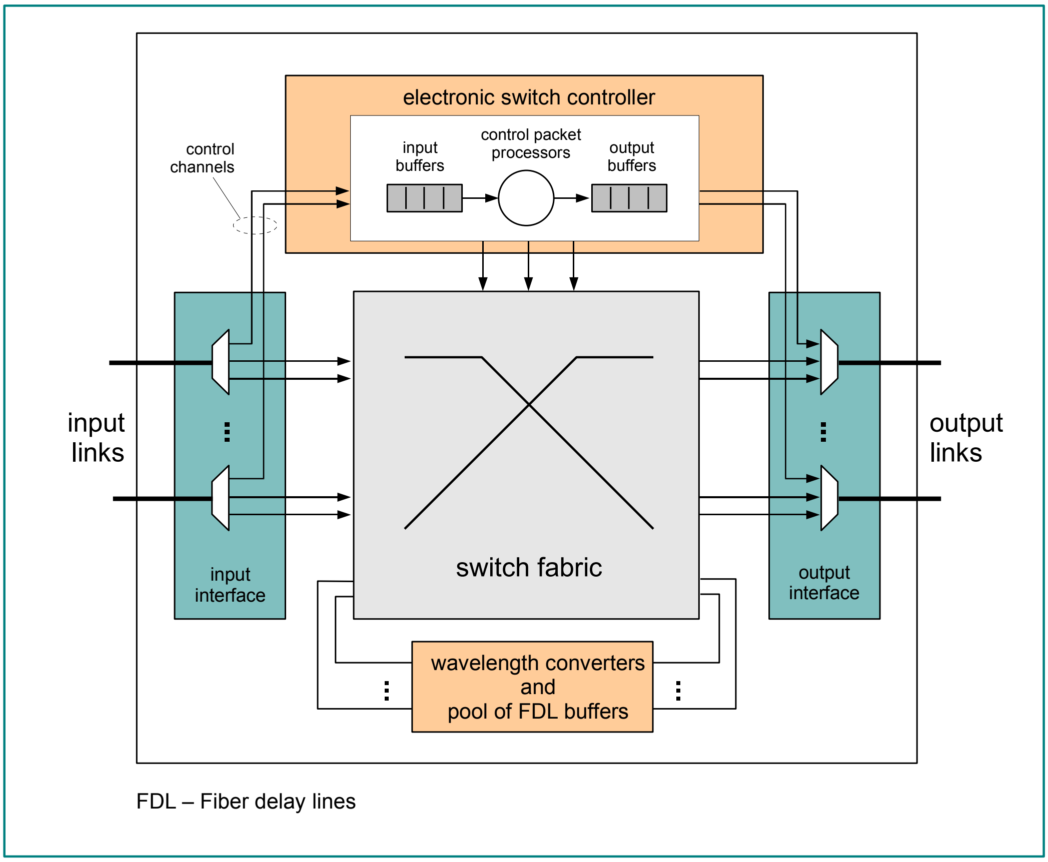

FIGURE 4.4: OBS core node (Klinkowsky, 2007).

can have the advantage of reducing the pre-transmission delay. However, since the exact length of the data burst can not be included in the corresponding control packet, this method can also result in an inefficient utilization of network resources.

4.1.2 Core node

As its name indicates, core nodes are located in the core of OBS networks, where a transparent switching of data bursts from one optical fiber to another takes place (see Figure 4.4). OBS core nodes perform, therefore, the following two main functions: scheduling of resources, and contention resolution. This functions will be explained in the sections bellow. One feature that can be present in OBS core nodes is buffering. Buffering is achieved by means of pools of FDL lines deployed in the architecture of the node to provide contention resolution by time deflection.

In traditional packet switching networks, contention resolution is typically implemented by storing packets in electronic random access memory (RAM) buffers. However, optical RAM-like buffers are still not yet available[2], and light cannot be easily stopped, stored, and forwarded. In all-optical networks buffering is achieved by means of FDLs. An FDL is simply a certain length of fiber into which a burst can be diverted, to emerge later, after a fixed delay period, on the opposite side.

There are important differences between electronic buffering and optical buffering. An electronic buffer can accept a burst if sufficient space is available, and after finding sufficient space the burst can be stored for any arbitrary time period. In contrast, an optical buffer is able to store a burst only for a fixed maximum time period and thus provides a *deterministic* delay to an incoming burst. This means that, once entered the FDL, it is impossible for the passing burst either to be removed earlier or to be held longer[3], two trivial operations for the common electronic RAM. Another reported property of optical buffers is the so-called *balking* property. The balking property of an optical buffer refers to the fact that an incoming burst must be dropped if the maximum delay provided by the FDL is not enough to avoid contention with a burst that is currently being transmitted on a given output line (Lu and Mark, 2004).

This properties bring one fundamental difficulty to optical switch designers, which is, how to implement efficient variable-length buffers from these fixed-length FDLs. One of the most popular approaches for the construction of such optical buffers is to use switched delay lines (SDLs), comprising a combination of concatenated FDLs and photonic cross-bar switches, controlled on a photonic slot basis with the goal of delaying photonic slots until contentions are resolved (Chlamtac et al., 1997).

[2] The goal of replacing electronics with optics for processing data is coming closer through cutting edge research into the mysterious properties of "fast and slow" light. Nevertheless, from "slow light" research testbeds to "frozen light" commercial devices there are serious obstacles to overcome, which are very challenging...

[3] An interesting parallel with ancient computers (in the 1950s) where electrical delay lines based on acoustic waves moving through rods of mercury were used as memory in the first programmable machines (Gevaux, 2007).

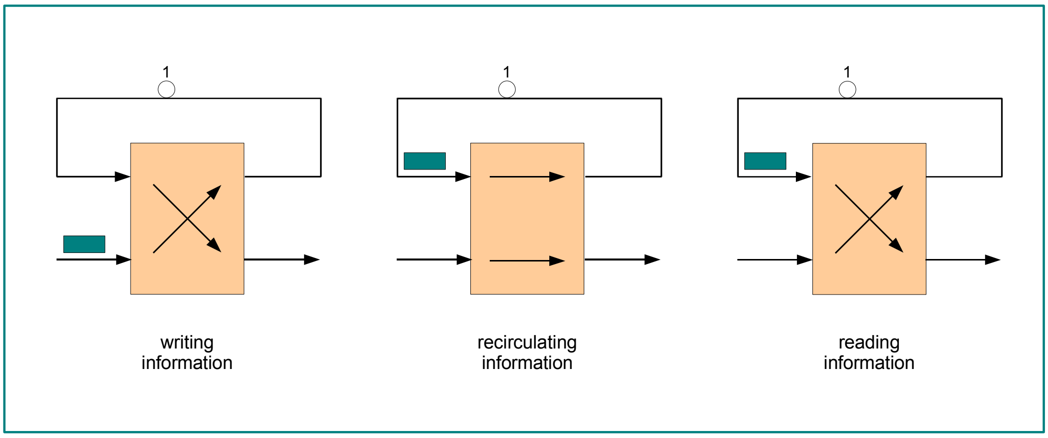

FIGURE 4.5: An optical memory cell.

SDLs are network elements built from crossbar switches and FDLs. They use optical memory cells as basic blocks for the construction of various types of optical queues. As illustrated (see Figure 4.5), we can set a 2 × 2 crossbar switch to the *cross* state to write an arriving burst to the optical memory cell. By doing so, the arriving burst can be directed to the FDL. Once the write operation is completed, the crossbar switch is set to the *bar* state so that the burst directed into the FDL keeps recirculating through the FDL. To read out the information from the memory cell, the crossbar switch is set to the "cross" state so that the burst in the FDL can be directed to the output link (Huang et al., 2007).

4.1.2.1 Scheduling of resources

Reservation of resources in the core concerns the capability to ensure undisturbed switching and transmission of data bursts from input ports to output ports. This process relies on the information carried on each control packet about the corresponding data burst which can include, for instance, the offset and size. Based on this information, core nodes schedule the resources inside the local optical switching fabrics such that bursts can cut through them. The resource scheduling schemes proposed for OBS networks can be classified based on the duration of the reservation, i.e., taking into account the moment when the reservation of a channel started and the moment when that reservation is released, as follows: immediate reservation, delayed reservation, explicit release, and implicit release. These four mechanisms determine the occupancy of resources inside the switching fabrics for a certain data burst and they will be explained in detail later on this chapter.

It should also be noted that the choice of resource scheduling scheme in OBS networks also depends on the burst assembly algorithm used by the edge nodes, since the information that the control packet can provide depends on the assembly algorithm from which it was generated.

4.1.2.2 Contention resolution

OBS networks provide connectionless transport through wavelength channels that support statistical multiplexing. In this working condition it is quite possible that several bursts contend with each other at the intermediate nodes. This can occur because connection requests made by OBS edge nodes generally use one-way reservation protocols in which data bursts are transmitted without confirmation that resources along the path will be successfully reserved. Therefore, if a burst arrives at an OBS node and all local resources are already taken, or if two or more simultaneously arriving bursts compete for the same resource, we have contention and (potentially) burst loss. Burst loss is a critical issue with significant impact on any end-to-end application running in the upper layers reducing its overall performance. Since contention is the primary cause of burst loss, contention resolution is one of the main design objectives in OBS networks.

The general solution to burst contention is to move all but one burst "out of the way". In OBS networks this can be accomplished by actions in the time, space or wavelength domains, or any combination therefrom. In fact, since time, space and wavelength are independent, any technique from any domain can be combined, resulting in a potential high number of different contention resolution schemes. Approaches for contention resolution using each of the above mentioned three domains can include: optical buffering (time domain), deflection routing (space domain), and wavelength conversion (wavelength domain).

4.1.3 IP over OBS

From what has been discussed about edge nodes and core nodes in this general framework, a number of very specific procedures beyond the process of guiding optical carrier waves through fibers of glass are involved in OBS networks. In computer-communication networks the procedures involved in the communication process are usually called protocols. Protocols are sets of rules that must be followed for devices to communicate. They form a communication architecture, sometimes referred as "protocol stack", in which each protocol provides for a function that is needed to make the data communication possible. Usually many protocols are used so that a difficult large problem can be broken into manageable pieces, and so that each protocol can be developed and updated independently, as long as the provided service remains constant. Networks are complex systems and their protocol stacks are normally organized into hierarchical layers of general functionality that can be easily associated to reference models[4].

An OBS network architecture can also be represented in layers as a set of protocols that provide services and exchange data. In fact, a layered architecture with well-defined interfaces between layers is crucial for the deployment of OBS networks, both in terms of their internal functioning and their inter-operability with other networks. As referred in a recent multilayered approach for OBS (Farahmand et al., 2007), in an IP over OBS hierarchy, the IP layer treats the OBS as its link layer, while the OBS operates on top of the optical layer. Therefore, as a data transport system, the OBS network architecture implements the lower three (or two)[5] layers of the reference models, namely physical, data link, and network. In order to understand the way OBS functions, it is important to have some understanding of what happens in this lower network layers.

Figure 4.6 depicts a functional block diagram of the main functions that need to be executed by edge nodes and core nodes at the medium access control (MAC)

[4] For example, the old OSI reference model with 7 layers or, more recently, the TCP/IP model with 5 (or 4) layers.

[5] In the case of reference models with only four layers in which Physical and Link are merged in one unique layer, like it is mentioned in Tanenbaum (Tanenbaum, 2002), Kurose (Kurose and Ross, 2007) or Cisco Academy.

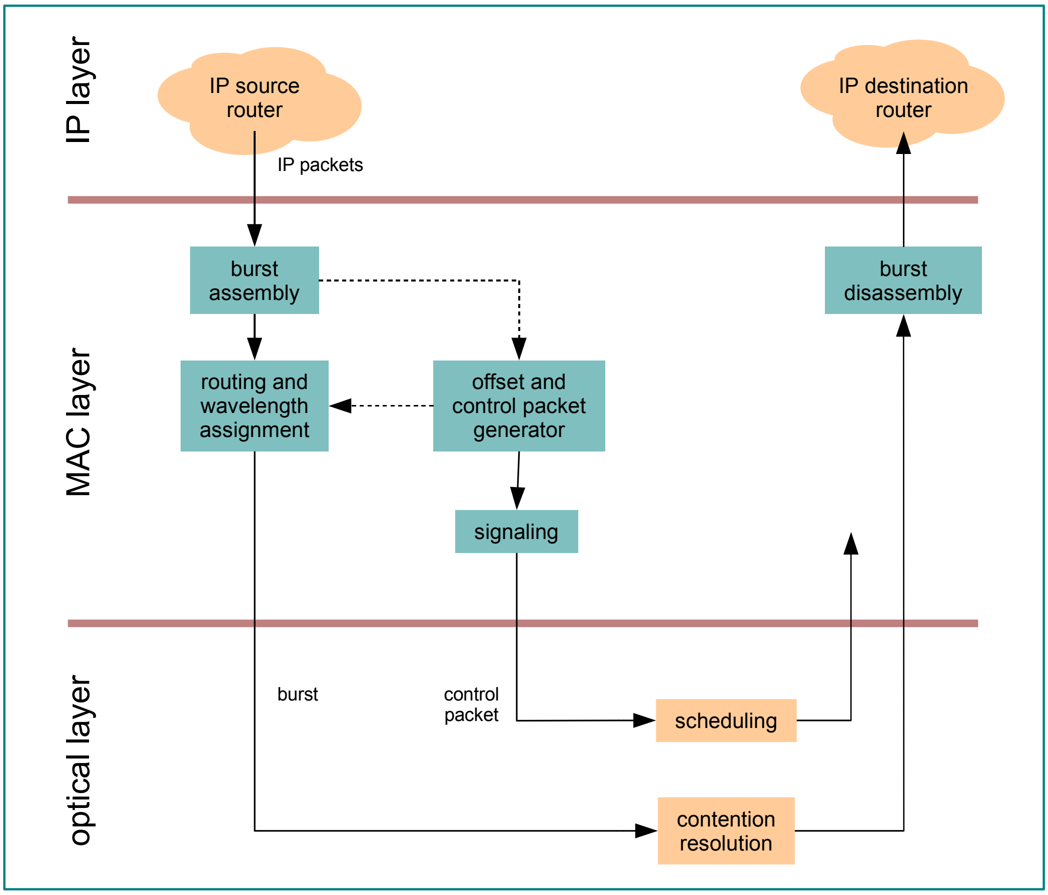

Figure 4.6: Functional block diagram of IP over OBS in three layers (Maier, 2008a).

layer and optical layer (physical). At the MAC layer, source and destination edge nodes perform the functions of burst assembly and disassembly (respectively), offset computation, control packet generation, routing and wavelength assignment and signaling. At the optical layer, intermediate core nodes are responsible for scheduling and contention resolution of bursts in-transit. The MAC layer and the optical layer located bellow offer a set of services that guarantee, for example, certain burst blocking probabilities to the client IP routers located in the layer above (Maier, 2008a). The "quality" of that set of services is, therefore, strongly dependent on the mechanisms implemented in these lower layers. Those mechanisms, represented with functional blocks in Figures 4.1 and 4.6, will be described after a brief discussion on the "quality" perspective that follows.

4.1.4 QoS support

Quality of service (QoS) is a broad term, which as many interpretations. To bring the subject from the "intrinsic engineering perspective" broadly stated in the preface of this thesis to the subject field of QoS support in OBS networks, let us follow the categorization used in Chua et al. (2007a). Usually, offering QoS support to end users, i.e., *end-to-end QoS provisioning*, requires the participation of all network entities along the end-to-end paths. This is so because the network performance perceived by an end user, i.e., its *quality-of-experience* (QoE), is the cumulative result of the service received by the user's packets at network entities along the end-to-end pass. For illustration, consider an application requiring an end-to-end packet loss probability of no more than 1%. If the packet loss probability at a single router on the pass becomes larger than 1%, the QoS requirement cannot be achieved. Entails from this example that OBS networks, which are a provisioning backbone for the next generation Internet, must ensure adequate "quality" across ingress and egress pairs in order to realize edge-to-edge QoS support.

QoS mechanisms can be broadly classified into two major categories (Chua et al., 2007a): QoS improvement and QoS provisioning mechanisms. The first category includes all node based mechanisms that improve the general performance of the network, which have great importance in enabling the network to provide satisfactory QoE. In addition, they allow the network to accommodate more users, therefore, they are also (in part) responsible for reducing the cost of data transmission. The node based mechanisms through which it is possible to improve the overall QoS in OBS networks are the burst assembly process, the signaling schemes for reservation of resources, the scheduling algorithms, and the contention resolution approaches. This mechanisms are discussed in this chapter. After, the QoS provisioning mechanisms will be considered.

4.2 Burst Assembly Algorithms

Burst assembly is the process of aggregating and assembling input packets into bursts at the edge of the OBS network. As packets arrive from client networks, they are aggregated at the edge nodes in local electronic buffers according to a forwarding equivalence class (FEC), which describes client data with similar characteristics, for instance, with respect to their destination, service class, or transmission time window. A data burst is then assembled from the data having the same FEC and according to a given assembly algorithm (Klinkowsky, 2007) whose purpose is to decide when to stop the burst aggregation process.

The key parameter in burst assembly algorithms is the trigger criterion for determining when to create a burst and send the burst into the network (Jue and Vokkarane, 2005c). This trigger criterion is extremely important because it shapes the traffic arrival process into the OBS core network therefore influencing the whole OBS network performance. Various burst assembly algorithms have been investigated in the current OBS literature (Rodrigo and Gotz, 2004; Yu et al., 2004), the most commonly reported using either burst assembly time or burst length, or both, as the trigger criteria. They are usually presented as (1) *timer-based* or (2) *length-based* assembly techniques and can be succinctly described as follows:

1. In timer-based burst assembly approaches, a burst is created and sent out into the optical core at periodic time intervals (Ge et al., 2000). Hence, the network may have variable length input bursts.

2. In length-based (or threshold-based) burst assembly approaches, a limit is placed on the maximum number of packets contained in each burst. Hence, fixed-sized bursts will be generated at the network edge at non-periodic time intervals.

The parameters involved in this assembly algorithms include a threshold time T, a burst length B, and a minimum burst length requirement B_{min}. The threshold time T limits the delay of packets in the assembly queue within a maximum value when the traffic load is low. The burst threshold length B reduces the unnecessary delay of a burst when traffic load is high. By the time the burst is to be sent out, if the assembled burst is shorter than B_{min}, the burst is padded to meet the specified B_{min} requirement (Ge et al., 2000). Both thresholds T and B can be either fixed or adjusted dynamically. Based on these thresholds, burst assembly algorithms can be classified into the following four categories (Yu et al., 2004):

Time-based assembly algorithms: This first category contains time-based assembly algorithms where a fixed time threshold T acts as the primary criterion to send out a burst, i.e., the burst is transmitted after T time units. At the same time the burst is also required to be no smaller than B_{min} and is send out with padding bits up to B_{min} otherwise.

Burst length-based assembly algorithms: The second category contains burst length-based assembly algorithms where a burst length threshold B is used as the criterion, and a burst is send out once the burst length reaches (or exceeds) threshold B.

Both time-based and burst length-based approaches can be considered similar in the sense that at a given constant arrival rate, a burst threshold length value can be mapped to a timeout value and vice versa, resulting in bursts of similar length for each case (Vokkarane, Haridoss and Jue, 2002). Even so they have differences and undergo several shortcomings at low and high traffic loads due to the fact that only one single criterion (either time or burst length) is considered. Under light load, a burst length based algorithm does not provide any constraint on the queueing delay of the packets that wait for being aggregated into a burst of size B. Therefore, the algorithm can not guarantee any bound on the maximum or average queueing delay (see Figure 4.7). In this situation, a time-based algorithm would solve the problem by assembling the burst at point P_2 in the Figure 4.7, after reaching time T, despite the small number of enqueued packets. On the other hand, this time-based algorithm will have longer average queueing delays under heavy traffic. Unlike time-based algorithms, under high traffic load a burst

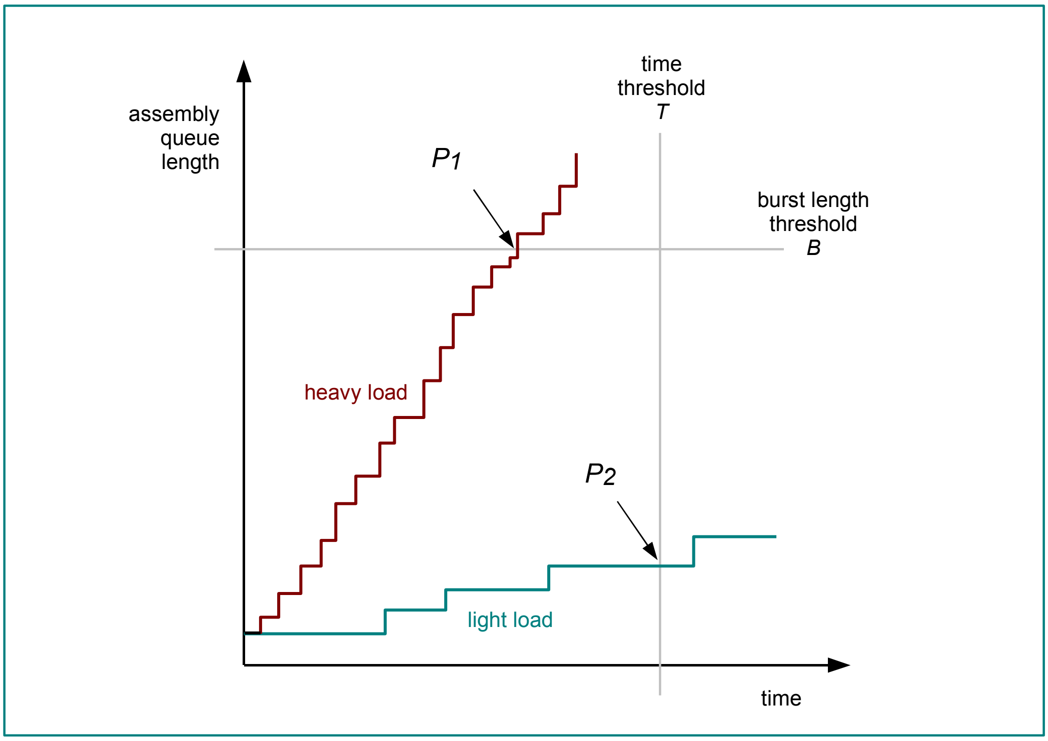

FIGURE 4.7: Burst length and time thresholds for burst assembly algorithms (Yu et al., 2004).

length-based algorithm transmits a burst at point P_1 in the Figure 4.7 as soon as the length threshold B is crossed well before the time threshold T is reached (Maier, 2008a). In addition, the burst length could have a large variance for the time-based algorithm, affecting the scheduling performance at the core nodes. For sure, an algorithm performing well under all load conditions would be very welcome. This thought was the basic idea for the third category of assembling algorithms.

Mixed time/burst length-based assembly algorithms: This third category uses both time threshold T and burst length threshold B to send out a burst. Depending on the traffic load and the values of the parameters T and B, either T or B is crossed first and consequently the burst is transmitted. In general, threshold T will be crossed before B at light traffic loads and the opposite will happen at heavy traffic loads (Yu et al., 2002). It should be noted that, with this mixed approach, if the burst length threshold B is much larger than the average amount of data that arrives in time T, i.e.,

$T \times \lambda_p \ll B$ (where λ_p is the average packet arrival rate to the assembly queue), the length of the assembled burst will never reach B before it is sent at point P_2. In such case, the mixed assembly algorithm will function the same way as the time-based algorithm with a fixed time threshold T. On the other hand, if the burst length threshold B is much smaller than the average data that arrives in time T, i.e., $T \times \lambda_p \gg B$, the time threshold T will never be reached before the length of a burst reaches B at point P_1, as depicted in Figure 4.7. In this case, the mixed assembly algorithm operates like the burst length-based algorithm (Yu et al., 2004).

Dynamic assembly algorithms: The fourth category of assembly algorithms is based on dynamic threshold where either the time threshold T or the burst length threshold B (or both) are adjusted dynamically according to actual network traffic conditions or predictions (Cao et al., 2002). Dynamic burst assembly algorithms are adaptive and therefore achieve an improved performance at the expense of an increased computational complexity compared to the first three categories which used static thresholds (Maier, 2008a).

A significant problem in burst assembly is how to decide on the appropriate timer and burst length threshold values in order to minimize the packet loss probability in the OBS network. The selection of such an optimal threshold value is an open issue (Jue and Vokkarane, 2005c). If the threshold is too high then bursts will be long. Longer burst will reduce the total number of bursts injected into the network core. But then, in the case of contention, the average number of packets lost per contention will increase. On the other hand, if the threshold is too low, bursts will be short. Smaller bursts will increase the number of bursts in the core network leading to a greater number of contentions, but the average packet loss per contention is less. Also there will be increased pressure on the control plane to process quickly the increased number of control packets. In addition, if the switch reconfiguration time is non-negligible then shorter bursts will lead to lower network utilization due to the high switching time overhead for each scheduled burst. Thus, there exists a trade-off between the number of contentions and the average number of packets lost per contention (Jue and Vokkarane, 2005c; Yu et al., 2004).

An interesting reported effect of burst assembly is that it shapes the traffic pattern reducing its degree of self-similarity (Ge et al., 2000). This is generally considered a desirable property since it makes the assembled traffic less bursty with potential benefits for burst loss reduction. However, it is reported in Yu et al. (2002) that long-range dependency in traffic will not change after burst assembly. And the study in Izal and Aracil (2002) shows that the influence of self-similarity on blocking probability is negligible, since the arrival process can be assumed to be Poisson in the time scale of interest for burst blocking.

4.3 Signaling

When a burst is transported over the OBS network, a signaling scheme must be implemented so that the required resources and the optical switches can be properly allocated and configured for the burst at each traversed node. Signaling is a critical aspect on any network. It specifies the protocol by which connection requests are handled, and its operation determines whether or not the network resources are efficiently utilized. For OBS networks it is even more important due to the bufferless nature of the core network where contention for resources can lead to data loss. Several variations of signaling protocols exist for OBS depending on how and when the resources along a route are reserved for a burst. Before discussing the signaling techniques for OBS, the underlying design principles that affect their performance will be explained following the approach in Jue and Vokkarane (2005d); Vokkarane (2007).

4.3.1 Signaling attributes

A particular signaling scheme can be characterized by the following generalized attributes, which will be explained next: (1) direction of the connection setup, (2) the starting point where resource reservation process initiates, (3) Resource, (4) Reservation, (5) Release, and (6) Computation.

Direction: *one-way* or *two-way*

The connection setup phase of a signaling technique can be either one-way or two-way (see Figure 4.8). In the former case, the source sends out a control packet trying to allocate at the intermediate nodes along the path the necessary resources for the data burst. No acknowledgment (ACK) message is send back to the source notifying the success or failure of the resource reservation attempt. In the later, signaling is acknowledgment-based. This means that first, a request for resources is sent from the source to the destination, and second, an ACK message confirming the successful assignment of the requested resources is sent back from the destination to the source. Only if this two-way process of establishing a connection succeeds the data burst can be transmitted. If during the request phase any of the intermediate nodes is busy, then the request is blocked and that particular intermediate node will take the necessary actions to release all the previously reserved links. A negative acknowledgment (NAK) will also be sent to the source indicating the failure of the reservation attempt (Jue and Vokkarane, 2005d).

The primary objective of one-way based signaling is to minimize the end-to-end delay in data transfer operations. However, this objective leads to contention of data bursts inside the OBS core and therefore to high potential data loss. The primary objective of two-way based signaling is to minimize burst loss at the core network. Unfortunately, this objective leads to high data transfer delay due to the round-trip connection setup (Vokkarane, 2007).

Initiation: *source initiation* or *destination initiation*

A signaling scheme can initiate reserving the requested resources at the source or at the destination. In the source initiation technique resources are reserved in the forward path as the control packet travels from the source to the destination. If the allocation of resources in this forwarding direction succeeds, an ACK message may

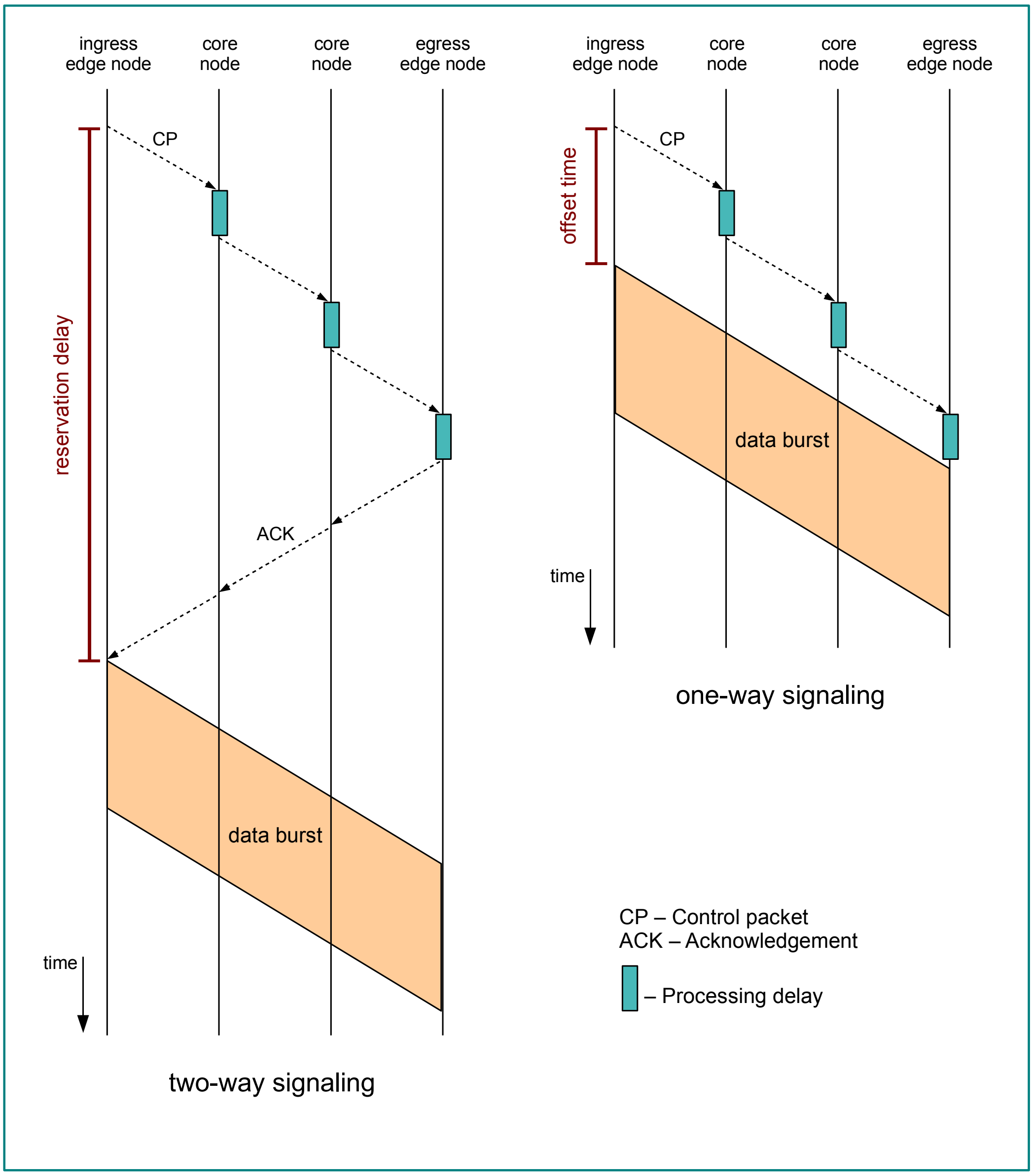

FIGURE 4.8: Centralized and distributed signaling in OBS networks.

be sent back to the source. Upon receiving the ACK confirmation the source transmits the data burst. In the destination initiation technique the source transmits to the destination node a request that collects wavelength availability information on every link along the path. Based on the collected information the destination node will choose an available wavelength, and sends a reservation request back to the source. This reservation request will traverse back the intermediate nodes while reserving the chosen wavelengths(Jue and Vokkarane, 2005d).

Usually source initiated schemes are avid for reserving resources. In order to reduce burst loss, these schemes may reserve more than the necessary wavelengths along the forwarding direction to release the unneeded later in the backward direction. This approach may lead to blocking of other requests due to *lack of resources*, impairing performance. On the other hand, destination initiated schemes first collect the wavelength availability information of all intermediate nodes and then, based on that information, select a wavelength. With this technique, a wavelength selected at the destination may be taken by some other request at any intermediate node along the path during the so called *vulnerable period*[6]. The primary cause of blocking in source initiation is due to lack of free resources, while in destination initiation the loss is due to outdated channel availability information stored at each core node (Lu et al., 2005).

Resource: *persistent* or *non-persistent*

One important decision that each signaling technique needs to make is either to wait on a blocked resource (until it becomes free) or immediately indicate contention and initiate the appropriate connection failure mechanisms (retransmission, buffering, or deflection, for example). In persistent signaling, the control packet waits on a blocked resource and assigns the wavelength when the resource becomes available. This approach leads to minimum loss if the network nodes are provisioned with adequate buffers to store the incoming data bursts. In non-persistent signaling the objective is to have an upper bound on the end-to-end data transfer delay. Thus, each node declares the request to be a failure if the resource is not promptly available (Vokkarane, 2007).

Reservation: *immediate* or *delayed*

Based on when the reservation of a channel is started, signaling schemes can support either immediate reservation[7] or delayed reservation[8]. In immediate reservation, the wavelength channel is reserved and the optical crossconnect is configured

[6]The period between the moment when the resource status for a given intermediate node is collected, and the moment when the reservation message arrives at that node.

[7]also called *explicit setup*.

[8]also called *estimated setup*

immediately from the instant the control packet is received and processed at a node. In delayed reservation, the reservation of a wavelength channel is delayed and, by using information provided by the control packet, the core node estimates the arrival time of the corresponding burst. A wavelength channel is reserved and the optical crossconnet is configured at the core node immediately before the estimated arrival time of the data burst (Maier, 2008a). In order to employ delayed reservation, the control packet must carry the offset time between itself and its corresponding data burst.

In general, immediate reservation is simple and practical to implement but incurs higher blocking due to inefficient bandwidth allocation. On the other hand delayed reservation leads to higher bandwidth utilization but its implementation is more intricate. Delayed reservations can also result in *voids*[9] between the scheduled bursts on the data channels. Scheduling algorithms used during reservation will need to store additional information about the voids, based on which the scheduler must assign a wavelength to a reservation request (Vokkarane, 2007).

Release: *explicit* or *implicit*

An existing reservation can be released in one of two ways: explicitly or implicitly. In the explicit release[10], following the data burst, the source node sends an additional trailing control packet towards the destination to terminate the existing reservation. After receiving the trailing control packet each node releases the reserved wavelength. In the implicit release [11], the core node estimates the end of the burst by using the offset and size information carried in the preceding control packet. Based on this information, a core node is able to estimate the time when to release the reserved wavelength after the corresponding data burst has passed through the switching fabric (Maier, 2008a).

Clearly, the implicit termination approach results in better loss performance due to the absence of any delay between the actual data burst ending time and the arrival

[9] idle periods between data bursts

[10] also referred as *open-ended reservation*.

[11] also referred as *estimated release* or *closed-ended reservation*.

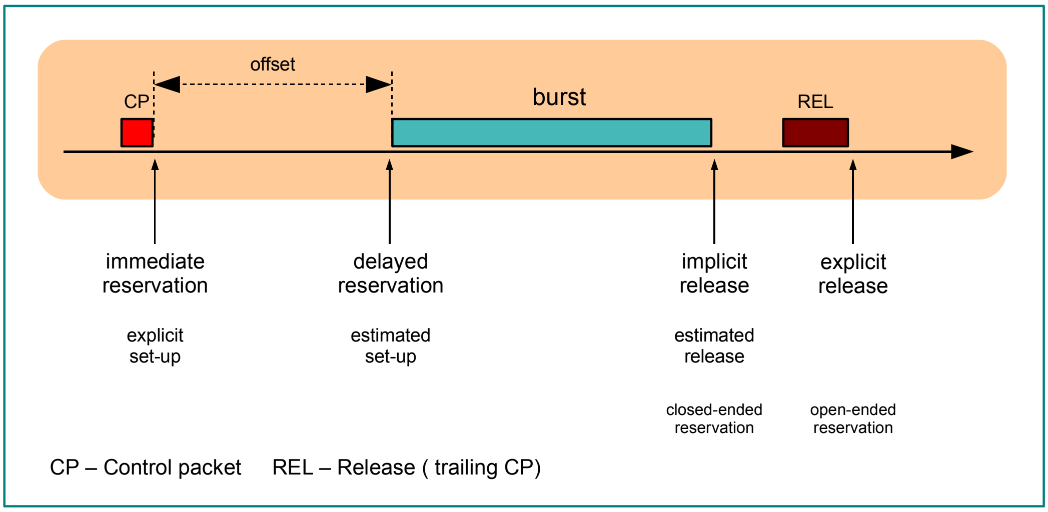

FIGURE 4.9: Reservation and release schemes in OBS (with control packet and data burst shown on the same channel, for simplicity).

time of the trailing control packet at each node. The explicit release technique leads to lower bandwidth utilization and increased control messaging.

These reservation and release schemes (see Figure 4.9) can be combined into four different ways, resulting into the following signaling categories: (1) immediate reservation and explicit release, (2) immediate reservation and implicit release, (3) delayed reservation and explicit release, and (4) delayed reservation and implicit release(Battestilli and Perros, 2003; Xu et al., 2001). Each of this combinations differ in the amount of time the same burst will occupy the switching resources and provide a different performance-complexity trade-off. The tighter the estimate of the start and the end of the burst, the smaller the overhead of keeping the switching elements configured, and the lower the overall burst blocking probability in the network. Clearly, the estimated schemes provide an improved resource utilization over their explicit counterparts. However, the explicit schemes are simpler to implement and can reduce the complexity of the switch scheduler (Baldine et al., 2002; Maier, 2008a).

Computation: *centralized* or *distributed*

In centralized signaling a dedicated centralized request server is responsible for setting up the route and assigning the wavelength on each route for every data burst for all source and destination pairs (Duser and Bayvel, 2002). This centralized scheme may perform more efficiently when the network is small and the traffic is non-bursty. In distributed signaling each node has its own burst scheduler and assigns an outgoing channel for each arrived control packet in a distributed manner. The distributed approach is more suitable for large optical networks and bursty data traffic.

4.3.2 OBS reservation protocols

Although the concept of burst switching have been introduced for centralized time division multiple access (TDMA) systems (Mills et al., 1990) and asynchronous transfer mode (ATM) networks (Widjaja, 1995) in early 1990, protocols suitable for high speed WDM networks were not developed until 1997 (Yoo and Qiao, 1997). In fact, first signaling proposals for OBS adapt the two versions of the ATM block transfer standard proposed for burst switching ATM networks: with delayed transmission, called *tell-an-wait* (TAW), and with immediate transmission, called *tell-and-go* (TAG) (Widjaja, 1995). The TAW protocol relies on a two-way operation mode with end-to-end reservation of resources and ACK. The TAG protocol operates in one-way mode, allocating transmission resources a while before the arrival of the burst payload that needs to be slightly delayed before each intermediate node.

4.3.3 Early overview

The TAG concept forms the basis of the terabit burst switching (Turner, 1999) in which, to compensate for the control packet processing time and prevent a burst from entering the switching fabric before its configuration is finished, a fixed delay is inserted into the data path using an FDL at each input port. On the other hand, the early version of Just-in-Time (JIT) protocol (Mills et al., 1990), initially requiring each burst transmission request to be sent to a central scheduler, can be considered a variant of the TAW approach (Chen et al., 2004). Just-in-time here means that by the time a burst arrives at an intermediate node, the switching fabric has already been configured, a concept that was later extended to the wavelength routed OBS networks (Duser and Bayvel, 2002). Since centralized protocols are generally considered difficult to scale and less robust when compared to the distributed approaches, a JIT version called Reservation with Just-in-Time was proposed (Hudek and Muder, 1995), requiring a copy of the request to be sent to all switches concurrently. With this approach, the schedulers of all switches need not only to be synchronized in time and also to share the same global link status information, which makes the implementation difficult (Chen et al., 2004). Another distributed version of JIT was then proposed in Wei and McFarland (2000) based on a hop-by-hop reservation that adopts some features of the Just Enough Time (JET) protocol (Yoo and Qiao, 1997; Qiao and Yoo, 1999).

4.3.4 Prevailing approach

The prevailing approach for resource reservation in OBS architectures use the distributed model with one-way reservations. With this approach, resources are allocated on-the-fly a while before the data burst arrives to a switching node. Prior to transmitting the data burst, the edge node transmits a control packet out-of-band with relevant information about the burst. The control packet goes through OEO conversion at each intermediate node along the path to the egress node and is electronically processed to configure the underlying switching fabric (Chen et al., 2004). After a pre-specified delay, called *offset*, the corresponding data burst is sent on one of the available data wavelength channels without waiting for an acknowledgment (ACK) that the connection between ingress and egress nodes has been successfully established (Battestilli and Perros, 2003). This approach significantly decreases the connection setup time when compared to the conventional two-way reservation schemes (see Figure 4.8), which becomes equal to the one-way end-to-end propagation delay plus the time required for electronically processing of control packet and configure the optical switch fabric at the intermediate nodes.

In the one-way reservation scheme, however, control packets may not be successful in setting up connections due to congestion on the data wavelength channels. As a result the optical switch cannot be configured as needed and the corresponding burst will be dropped. Thus, with OBS networks using one-way reservation schemes, a certain nonzero burst loss probability must be expected (Maier, 2008a).

A second signaling approach being proposed uses the centralized model with end-to-end reservation procedures. When an ingress edge node has a burst ready to be transmitted, it dispatches a control packet to a central request server which has global knowledge about the current state of all OBS switches and wavelength channels. Based on this global information the central server processes the connection requests and sends positive ACKs to the requesting edge nodes that, upon receiving the ACKs, are allowed to transmit their bursts.

4.3.5 Predominant protocols

The two most popular signaling schemes are JIT and JET. To a certain extent, they can be considered similar protocols, since both adopt the tell-and-go principle, but they differ in the following non-negligible aspect: JIT employs immediate reservation and explicit release while JET employs delayed reservation and implicit release (see Figure 4.9).

4.3.5.1 Just-in-time (JIT)

In JIT signaling (Wei and McFarland, 2000), an OBS node configures its optical switches for the incoming burst as soon as the corresponding control packet is received and processed. This means that resources at the OBS node are made available before the actual arrival time of the burst (Baldine et al., 2002). With

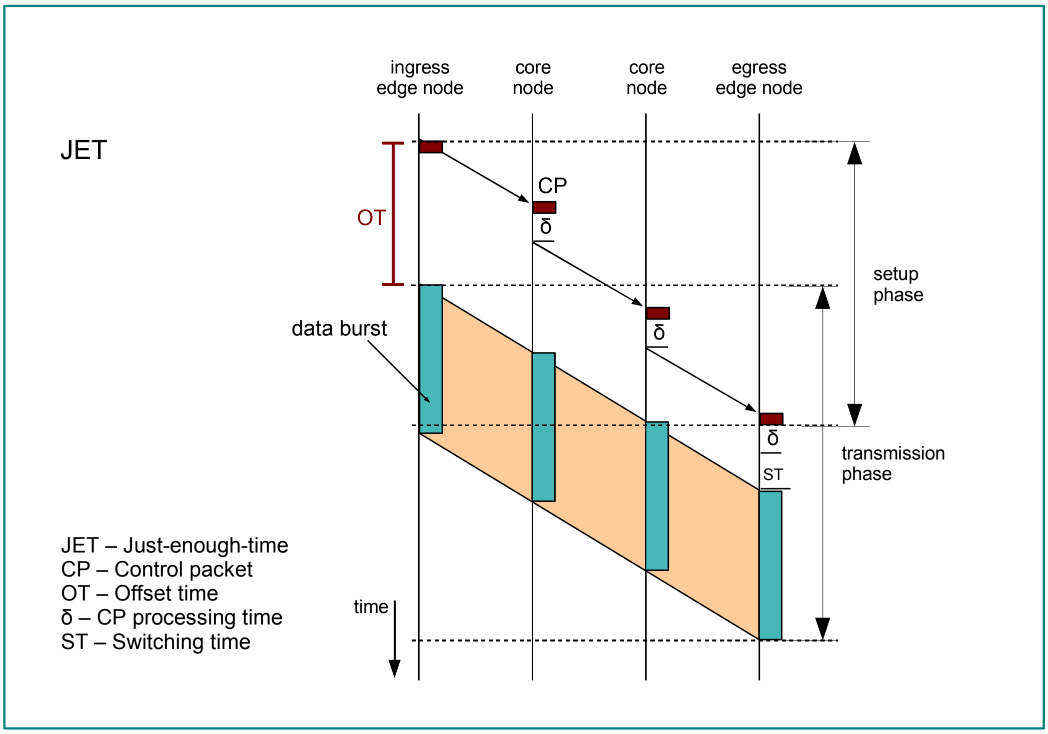

FIGURE 4.10: Just-enough-time (JET) signaling scheme.

this signaling scheme, intermediate nodes do not take into account any offset time information that might be carried in control packets, an aspect that contributes to the simplicity of this technique. Unfortunately, it also contributes to a less efficient use of resources. Since wavelengths are reserved prior to the burst arrival time, channels cannot be utilized during the period between the arrival of the control packet and the arrival of the corresponding data burst (which may be considerable). The overhead of configuring the switching elements for an unnecessary long period of time leads to an increased burst loss probability.

4.3.5.2 Just-enough-time (JET)

JET signaling makes use of the offset time information transmitted in each control packet allowing for higher wavelength utilization due to its delayed reservation feature. With delayed reservation the optical switches at a given OBS node are configured right before the expected arrival time of the data burst (and not immediately after the control packet, like in JIT). When a burst is ready for transmission the JET signaling scheme works as follows (see Figure 4.10): the source node sends a control packet toward the destination node; the control packet is processed at each subsequent node in order to establish an all-optical path (later on) for the corresponding data burst; if this reservation attempt succeeds, the switches along the path will be properly configured prior to the burst arrival; meanwhile, the burst waits at the source, in the electrical domain. Upon expiring a predetermined offset time, the burst is send optically on the chosen wavelength (Qiao and Yoo, 1999). The offset time (OT) is calculated based on the number of hops from source to destination and the switch reconfiguration time of a core node in the following way: $OT = h \cdot \delta + ST$, where h is the number of hops between source and destination, δ is the per-node control packet processing time, and ST is the switching reconfiguration time (Jue and Vokkarane, 2005d). If at any intermediate node the reservation is unsuccessful the burst is dropped.

JET signaling is quite exigent in the information it needs to maintain. For each channel of each output port (of every switch) it is necessary to keep the starting and finishing times of all scheduled bursts, which makes the system rather complex when compared to the JIT approach. However, JET is capable of identifying situations where no transmission conflict occurs, although the start time of a new burst may be earlier than the finishing time of an already accepted burst (see Figure 4.11) (Vokkarane, 2007). This feature makes it possible to schedule a new burst in between two already accepted reservations. Using this technique, the probability of unsuccessful reservations is lower.

Another important feature of JET signaling is that the burst length information carried in the preceding control packet is used to enable close-ended reservation instead of open-ended reservation, which would require explicit release of the configured resources. The close-ended reservation helps OBS nodes make intelligent decisions about whether it is possible to schedule another newly arriving burst. As a result, JET signaling is able to outperform JIT signaling in terms of bandwidth utilization and burst loss probability, at the expense of increased computational complexity (Maier, 2008a). Since implementation simplicity is a relevant feature of any signaling scheme, traditional JIT has been improved in terms of data channel

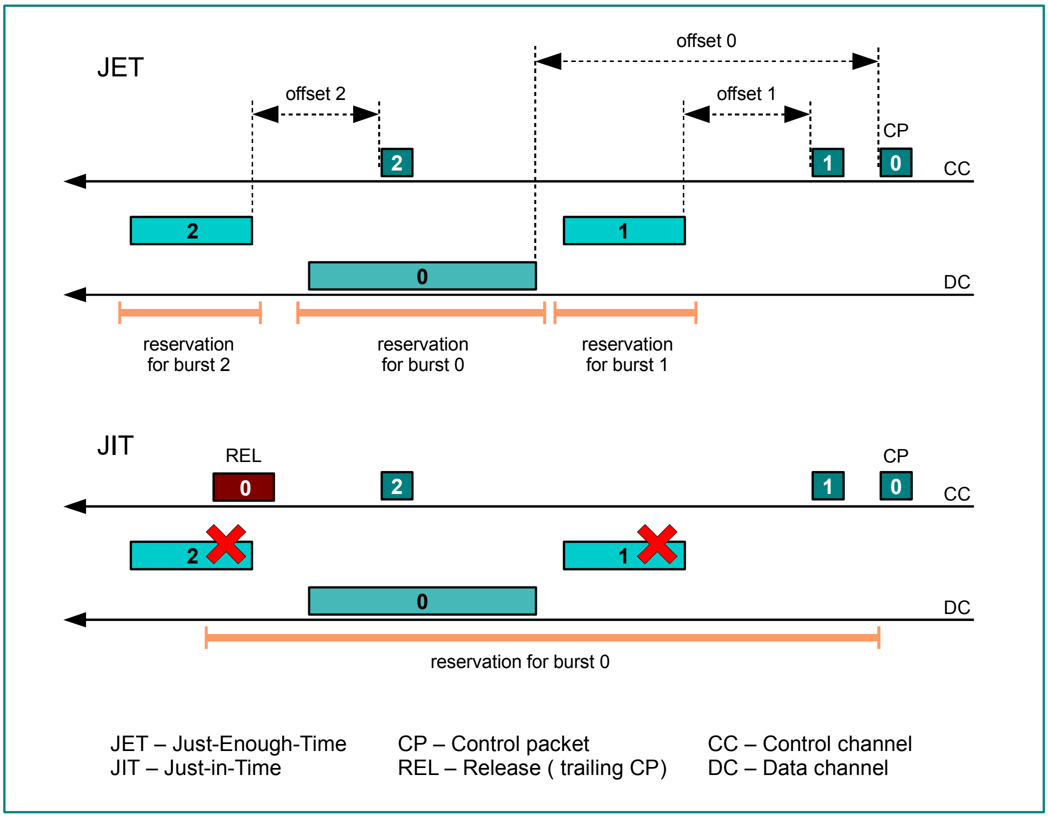

FIGURE 4.11: Comparison of JET and JIT signaling.

scheduling times while keeping simple its implementation simplicity in an enhanced JIT version called E-JIT recently proposed in Rodrigues et al. (2007).

4.4 Scheduling

Whenever a data burst arrives to an OBS node, a wavelength channel must be assigned for the burst on the appropriate outgoing link. The algorithm that undertakes that decision is the scheduling algorithm. Since bursts may have different sizes and offset times, they may not arrive in the same order as their corresponding control packets, and channel occupancy is expected to become fragmented with different voids (idle periods) between data bursts. Combined with fast processing of control packets, an efficient scheduling algorithm should be able to utilize such voids for scheduling newly arriving bursts.

Scheduling algorithms can be roughly classified into two categories (Jue and Vokkarane, 2005e): (1) non-void-filling scheduling algorithms and (2) void-filling scheduling algorithms. In general, a non-void-filling scheduling algorithm is fast but not bandwidth efficient, while on the contrary void-filling scheduling algorithms provide better bandwidth utilization but has much longer scheduling times. To date, a number of scheduling algorithms have been proposed, the simplest ones based on either a round-robin or random selection of resources, but for illustration, two more advanced scheduling policies which have received considerable attention will be briefly described: the *Horizon* (Turner, 1999), also called latest available unused channel (LAUC) in Xiong et al. (2000), and the LAUC with void-filling (LAUC-VF) (Xiong et al., 2000).

4.4.1 Latest available unused channel (LAUC)

In LAUC, which is a Horizon-type algorithm, the scheduler keeps track of the so called horizon of each wavelength channel, defined as the latest time at which the wavelength is currently scheduled to be in use. In Figure 4.12, for instance, time t_1'' is the scheduling horizon for channel C_1. Only the channels whose scheduling horizons precede the arrival time of the new burst are considered available, and the one with the latest scheduling horizon is chosen. In Figure 4.12, channel C_3 would be reserved if Horizon is applied. The horizon is then updated after making the reservation for the next burst. The basic idea for this algorithm is to minimize bandwidth gaps created as a result of new reservations. In doing so, the gap between the current horizon and the starting time of the new reservation period is minimized. Simplicity in both operation and implementation is the main advantage of the Horizon-type algorithms. However, these algorithms suffer from low bandwidth utilization and high burst loss probability because *voids*[12] between two existing reservations (for example, $t_1' - t_1$ in Figure 4.12) are wasted (Chen et al., 2004).

[12]A void is the unscheduled duration (idle period) between two scheduled bursts on a data channel (Jue and Vokkarane, 2005e).

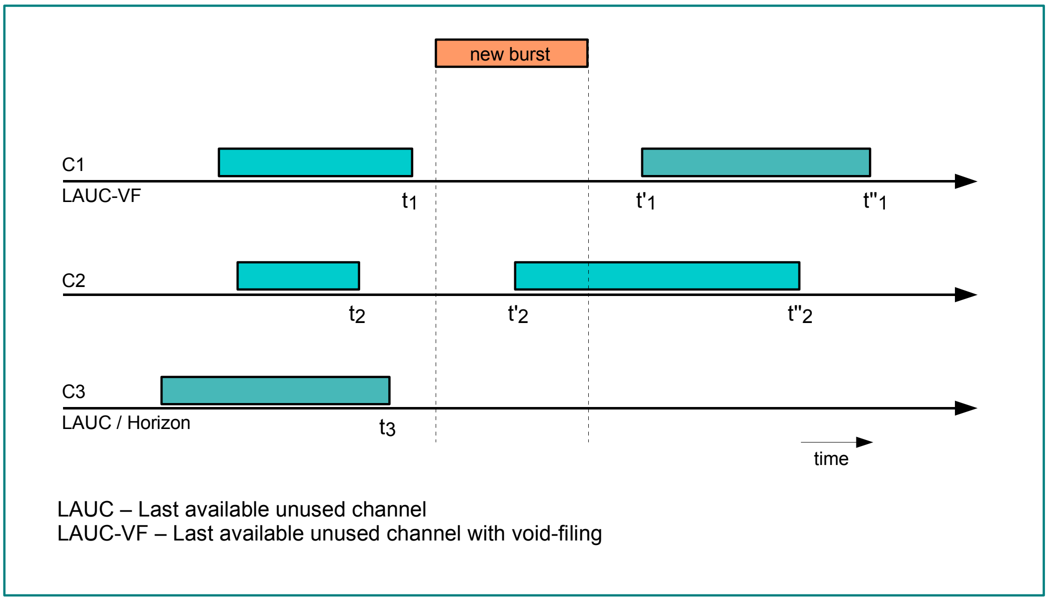

FIGURE 4.12: An illustration of scheduling algorithms.

4.4.2 Latest available unused channel with void-filling (LAUC-VF)

In LAUC-VF, which is a JET-based algorithm, the scheduler keeps track of all existent voids in all wavelength channels. This algorithm assigns an arriving burst a large enough void interval whose starting time is the latest but still earlier than the burst arrival time (Maier, 2008a). Using LAUC-VF, channel C_1 in Figure 4.12 would be chosen. This algorithm provides better bandwidth utilization and burst loss probability than Horizon. Unfortunately, its execution time is also much longer than Horizon.

4.4.3 Other channel scheduling proposals

Several novel scheduling algorithms for OBS networks have been proposed and investigated. Unlike the Horizon and LAUC-VF proposals here described, which primarily aim at optimizing either the running time or the burst loss probability (but not both), some of the new proposals take both performance metrics into account. For a survey and discussion on these and other previously reported algorithms we refer the interested reader to Xu et al. (2004); Jue and Vokkarane (2005e); Xu et al. (2003). For a group of scheduling techniques which are enhanced with the so called *look-ahead processing window* capability, which allows for more collecting information capabilities about the incoming reservations, we refer the reader to Cheyns et al. (2003); Farahmand and Jue (2003, 2006). A review on scheduling techniques can be found in the paper of Li and Qiao (2004).

4.5 Contention Resolution in OBS

Using one-way reservation protocols like JIT or JET, an ingress edge node can send out bursts without having reservation acknowledgments or global coordination. But this requires an intermediate core node to resolve possible contention among bursts. Contention is the primary cause of burst loss in OBS networks, thus contention resolution is a design objective of utmost importance. Contention occurs when more then one burst compete for the same resource at the same time. In bufferless OBS networks, contention among bursts can be resolved in four ways (Chen et al., 2004): deflection, dropping, preemption, or segmentation.

Deflection: Through deflection, a burst is sent to a different output channel instead of the one initially planned. Since contention can only happen when bursts compete for the same wavelength on the same output port simultaneously, deflection can be applied in (1) wavelength, (2) space and/or (3) time domains.

1. When deflection is applied in the wavelength domain, the contending burst is sent on another wavelength through **wavelength conversion**.
2. When utilizing the space domain, the contending burst is diverted to a different output port and will follow an alternate route to its destination (Qiao and Yoo, 1999). This technique is called **deflection routing**.
3. When utilizing the time domain, contending bursts are delayed using FDLs. This is the **optical buffering** domain.

Dropping: If a contending burst cannot be deflected due to the unavailability of any wavelength, output port, or FDL, data loss becomes inevitable and the most common (non-preemptive) approach is to *drop* the incoming data burst.

Preemption: As an alternative to dropping it is possible for the incoming burst to preempt an existing burst based on some priority or service class.

Segmentation: In addition, it is also possible to break either the incoming or the existing burst into segments, and then deflect, drop, or preempt each segment.

4.5.1 Optical buffering

Both FDLs and SDLs can potentially improve network performance and have been explored by several authors to reduce burst loss probability. It should be noted that the presence of this kind of buffers adds another dimension to the reservation and scheduling mechanisms in use and which, in that case, must be revised accordingly. Two different FDL scheduling algorithms are described in Gauger (2003): (1) *PreRes* and (2) *PostRes*.

1. In PreRes, the most commonly adopted, the request to reserve an FDL buffer for an incoming burst is made as soon as the control packet is processed and it is determined that there is no wavelength available on the required output port. Therefore, in the PreRes scheme, the new offset time between the control packet and the burst is increased to the sum of the original offset plus the assigned FDL delay.

2. In the PostRes scheme, the offset time is kept to its original value by delaying both the control packet and its associated burst for a period of time without prior knowledge of whether a resource will be available for the burst upon leaving the FDL buffer.

The addition of the extra dimension resultant from the incorporation of SDLs is also considered in a two dimension reservation scheme proposed in Lu and Mark (2004) and working in two phases: wavelength reservation, and SDL buffer reservation. In the first phase, the scheduler at the OBS node tries to reserve a wavelength for the arriving data burst. If no wavelength is available, the SDL buffer reservation phase begins with the scheduler trying to reserve one SDL buffer having a suitable delay value to hold the arriving burst temporarily, therefore resolving contention at a given output port. With this mechanism the scheduler has to know both the arrival and departure times of the incoming burst. Knowing both, the scheduler is able to compute the delay value required to store the arrival burst in an SDL buffer. If no SDL with appropriate delay is available at the burst arrival time, the contending burst is dropped (Maier, 2008a).

Optical buffers may be also broadly classified either as *single-stage*, having only one block of parallel delay lines, or *multi-stage*, having several blocks of delay lines cascaded together. Single-stage optical buffers are easier to control, but multi-stage implementations may lead to more savings on the amount of hardware used. They can also be classified as *feed-forward* or *feedback* configurations. In the former, delay lines connect the output of a switching stage to the input of the next switching stage. In the later, delay lines connect the output of the switching stage back to the input of the same switching stage. Long holding times and a certain degree of variable delays can be implemented with this configuration by varying the number of loops a burst undergoes (Chua et al., 2007b).

Although on-chip optical buffers based on waveguide delay lines might have significant impact in the development of future optical interconnects, there remain some hurdles that limit their effectiveness. First, by their nature, they can only offer discrete (coarse) delays. The use of recirculating delay lines can give finer delay granularity but it also degrades optical signal quality. Secondly, despite some recent progress reported in Xia et al. (2007), the size of FDL buffers is severely limited, not only by signal quality, but also by physical space limitations[13]. For these

[13]A delay of 1 ms requires over 200 Km of fiber.

reasons, optical buffering (alone) is not usually taken as an effective contention resolution approach to consider under high load or bursty traffic conditions and mixed approaches are receiving attention. For example in Vokkarane, Thodime, Challagulla and Jue (2003), the authors propose a number of data channel scheduling algorithms to reduce burst loss that use FDLs combined with burst segmentation. The proposed scheduling algorithms are classified based on the placement of the FDL buffers in the OBS node and are referred to as *delay-first* or *segment-first* schemes. Recently, in Pedro et al. (2007), a mixed approach with a moderate use of FDL buffers at the core nodes, combined with a priority-based wavelength assignment and a burst scheduling strategy at the ingress nodes is presented with the objective of resolving contentions while matching the performance of a network using full-range wavelength converters.

Among other recent studies, in Huang et al. (2007); Wang et al. (2008), new SDL constructions for optical buffers (FIFO and LIFO) are proposed and evaluated. The study in Rosberg and Vu (2007) compares between the blocking probabilities of several design and buffer scheduler options and, among other results, concludes that feedback FDL design can significantly reduce the number of required buffers when compared with a feed-forward design.

4.5.2 Deflection routing

Deflection routing is considered a contention resolution approach quite appropriated for optical networks whose nodes are usually equipped with very limited buffering capacity, or even with no buffering capacity at all. In fact, in the later case, deflection routing is also known as *hot-potato* routing, an approach in which, if the intended output port for a certain burst is busy, the burst is immediately forwarded through another output port instead of being dropped. In the traditional store-and-forward routing approach, widely used in routers equipped with plenty of electronic memory, this contending burst would be stored and then forwarded through the very same intended port as soon as possible (Maier, 2008a). The performance of these routing strategies has been evaluated for several common regular topologies (Acampora and Shah, 1992; Forghieri et al., 1995; Greenberg and Hajek, 1992) and it is found that deflection generally performs poorly compared to store-and-forward unless the topology in use is very well connected. Even so, the performance of deflection routing can be significantly improved if combined with deflection in the time domain, i.e., with a small amount of buffers.

For an arbitrary topology, the choice of which output links to choose for deflection is critical to the network performance. The existing deflection routing protocols can be classified into the following three categories (Chua et al., 2007b): (1) fixed alternate routing, (2) dynamic traffic aware, and (3) random routing.

1. Fixed alternate routing is the most commonly used. In this method the alternate path is either defined on a hop-by-hop basis (Yao et al., 2003), or by storing at each node both the complete primary path and the complete alternate path from itself to every possible destination on the network (Hsu et al., 2002). Fixed alternate routing can perform well in small topologies.

2. Traffic aware deflection routing takes into account the transient traffic condition in selecting the output links for deflected bursts (Lee et al., 2005; Ogino and Tanaka, 2005; Du et al., 2007). It is an approach that presents similarities with load balancing methods.

3. In Random deflection routing (Cameron et al., 2005) the control packets carry a priority field. Every time a burst is deflected its priority is decreased by one. Normal bursts on their primary paths can preempt those low priority ones. Thus, the worst-case burst loss probability of this method is upper-bounded by that of a standard network. This approach appears to provide a good balance between simplicity and implementation, robustness and performance.

A important aspect to consider is that, typically, deflection routing is triggered by an OBS node using the local status information about its own resources (wavelength availability, or link congestion, for example). Using local status information for selecting a deflection route does not take the global status of network resources into account and may lead to suboptimal network performance (Maier, 2008a). To improve contention resolution by means of deflection routing, the so-called contention-based limited deflection routing (CLDR) protocol was proposed (Lee et al., 2005). In CLDR the nodes periodically exchange local status information about traffic load, burst contention rate, burst blocking probability, etc. The proposed CLDR algorithm runs on all OBS nodes and sequentially performs the following two steps (Lee et al., 2005): (1) based on certain performance criteria, dynamically determines if the burst should be deflected or retransmitted from source; and (2) if the decision is to deflection route the burst, that is done using a path that is based on minimization of a performance measure that combines distance and blocking due to contention. The CLDR routing scheme prevents injudicious deflection routing in OBS networks by using a threshold-based dynamic decision algorithm to decide whether to perform deflection or not. Based on this protocol, a new routing algorithm with on-demand deflection routing for Grid over OBS (GoOBS) is proposed in Guanghong and Zhaohong (2007). Recently, a deflection routing scheme is proposed in Huang et al. (2008) where an analytical model is also developed for its performance evaluation. Simulation results showed that under low or medium traffic loads, deflection routing for multicasting can significantly reduce burst loss while slightly increasing burst delay.

To be applied in OBS networks, deflection routing mechanisms must overcome the problem of insufficient offset time. This is caused by a burst traversing more hops than originally intended as a result of being deflected. Since the offset time between the data burst and its control packet decreases after each hop, the burst may overtake the control packet. Various reported solutions have been proposed (Hsu et al., 2002), such as setting extra offset time, or delaying bursts at some nodes in the path. It has been found that delaying a burst at the next hop after deflection is the most promising option (Chua et al., 2007b). Another issue with deflection results from the fact that deflected bursts may follow a longer path than other bursts on the same flow. This can lead to several problems such as increased delay, degradation of signal quality, increased network resource consumption and out-of-order burst arrivals which are prone to instability.

4.5.3 Burst segmentation

In the contention resolution approaches already presented, when contention between two bursts cannot be resolved through other means, one of the bursts is dropped in its entirety. This is so, even though the overlap between the two bursts may be minimal. For certain applications characterized by stringent delay but relaxed packet loss requirements, it might be desirable to loose a few packets from a given burst rather than losing all the entire burst. To address this issue, in Vokkarane, Jue and Sitaraman (2002) the authors introduced a novel contention resolution technique called *burst segmentation*, which significantly reduces packet loss by partitioning the burst into segments and dropping only those segments which contend with another burst. An important ability of this technique is that it allows bursts to be preempted by other bursts, paving the way for prioritized contention handling.

In burst segmentation a burst is generally divided into basic transport units called *segments*. Each segment may contain a single packet or multiple packets, and the boundaries of each segment represent the possible partitioning points of a burst when the burst is in the optical network. A noteworthy fact is that segment boundaries are transparent in the optical core, i.e., segments are electronic entities not visible in the optical domain (Vokkarane, Jue and Sitaraman, 2002). The optical layer only looks at the burst length stored in the control packet. Thus, the segment size (and respective boundaries), is a key system parameter that must be agreed upon ahead on time.

All segments in a burst are initially transmitted as a single burst unit. However, in a contention between two overlapping bursts, only those segments of a given burst that overlap with segments of another burst will be dropped, instead of the entire burst. The OBS node must know which of the contending burst segments will be dropped. In Vokkarane and Jue (2005) two possible variants are considered (see Figure 4.13): (1) *tail-dropping*, and (2) *head-dropping*. These variants are described next, considering that the burst that arrives at an OBS node first is

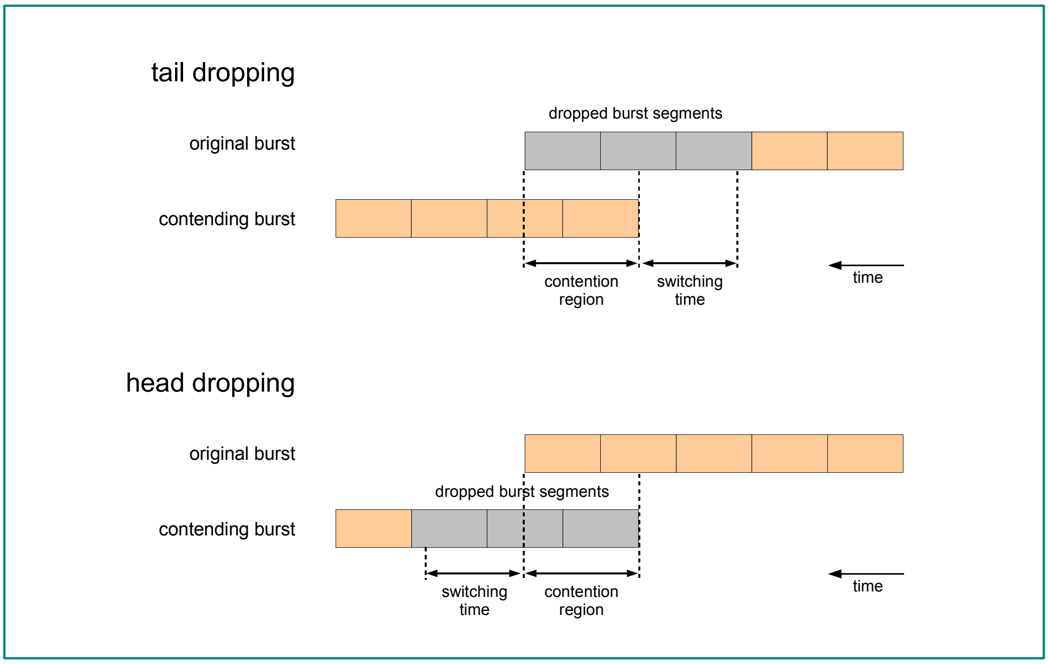

FIGURE 4.13: Burst segment dropping policies.

referred to as the original burst, and the burst that arrives later is referred to as the contending burst.

1. In tail-dropping, the tail of the original burst is dropped, i.e., the last one or more segments of the previously arrived burst are dropped when another contending burst arrives. If switching time at an OBS node is non-negligible, one or more additional segments of the original burst may be lost when the output port is switched from one burst to the other.

2. In head-dropping, the head of the contending burst, which arrives after the original burst, is dropped. The dropped head consists of one or more burst segments, depending on the length of the contention region and switching time.

Tail-dropping provides a better chance of in-sequence packet delivery at the destination, assuming that dropped packets are retransmitted by the edge node at a later time. Although more prone to out-of-order packet delivery, head-dropping policies ensure that, upon arriving at a certain node without contention, the burst is guaranteed to traverse the node without being preempted by a later burst.

While burst segmentation can be used to resolve contention and reduce burst loss, *prioritized* burst segmentation can be used to provide service differentiation by allowing high-priority bursts to preempt low-priority burst. Prioritized burst segmentation was examined in Vokkarane, Vokkarane and Jue (2003) under the assumption of tail dropping, and using a new burst assembly approach for packets of different classes called *composite burst assembly* motivated by the observation that, with burst segmentation, packets toward the tail of the burst are more likely to be dropped than packets at the head of the burst.

A basic extension of burst segmentation is to implement segmentation in conjunction with deflection. In doing so, there are two approaches for ordering the contention resolution policies: (1) *segment-first* or (2) *deflect-first*. They are described in Jue and Vokkarane (2005f) as follows:

1. In the segment-first policy, if the remaining length of the original burst is shorter then the contending burst, then the original burst is segmented and its tail is deflected. If the alternate port is busy, the deflected part of the original burst is dropped. If the contending burst is shorter than the remaining length of the original burst, then the contending burst is deflected or dropped.

2. In the deflect-first policy, the contending burst is deflected if the alternate port is free. If the alternate port is busy and if the remaining length of the original burst is shorter then the contending burst, then the original burst is segmented and its tail is dropped. If the contending burst was found to be shorter, then the contending burst is dropped.

Interesting proposals using segmentation in conjunction with other contention resolution schemes (like optical buffering and deflection, for example) are presented in Vokkarane, Thodime, Challagulla and Jue (2003); Vokkarane and Jue (2005). Service class proposals based on segmentation mechanisms are also under research focus in Vokkarane, Vokkarane and Jue (2003) and more recently in Tan et al. (2006). The effect of segmentation in the upper layers of the protocol stack is being studied in Phung-Duc et al. (2006). Recent studies concerning the impact of segmentation mechanisms on the TCP layer are also presented in Lazzez and Boudriga (2007); Lazzez et al. (2008), where a segment retransmission scheme is proposed in which segments lost in case of contention resolution failure are re-transmitted at the OBS layer.

A possible side effect of segmentation with deflection is that, upon contention, shorter segments will be forwarded as new bursts, which may lead to burst fragmentation and excessive overhead with respect to switching times and per burst packet control. Beyond the accurate maintenance of offsets for the deflected new bursts, another issue when implementing segmentation with deflection is how to handle long bursts that may span multiple nodes simultaneously (Jue and Vokkarane, 2005f).

4.5.4 Wavelength conversion

Wavelength conversion is perhaps the most efficient contention resolution technique. It takes place in the wavelength domain and does not delay signals by FDLs or by deflection routing. Wavelength conversion is the process of converting the wavelength of an incoming signal to another wavelength for transmission on an outgoing channel, thereby increasing wavelength reuse. It is a simple and powerful technique to resolve contentions in OBS networks which can be used either alone[14] or in combination with other contention resolution techniques, such as optical buffering, deflection or segmentation described in the sections above.

With wavelength conversion contention is reduced by utilizing additional capacity in the form of multiple wavelengths per link. In fact, wavelength conversion enables one output wavelength to be used by bursts from several input wavelengths, thereby increasing the degree of statistical multiplexing and burst loss performance. As the number of possible wavelengths per fiber continues to grow, this approach becomes increasingly attractive (Chua et al., 2007b). In practice,

[14]For example, like in the optical composite burst switching (OCBS) networks (Detti et al., 2002).

however, technological feasibility and economic issues do not allow full-range wavelength conversion at all OBS nodes, and the deployment of limited-range and sparse wavelength conversion in OBS networks appears to be more realistic. To be cost-effective, an OBS network will have to be designed with some limitations on its wavelength conversion capabilities. Therefore, contention may occur more frequently in realistic OBS networks, claiming for the deployment of the other contention resolution schemes already mentioned (Maier, 2008a).

The wavelength conversion capability can be classified in the following categories (Maier, 2008b):

Fixed conversion: Static mapping between input wavelength λ_i and output wavelength λ_j.

Limited-range conversion: Input wavelength λ_i can be mapped to a subset of available output wavelengths.

Full-range conversion: Input wavelength λ_i can be mapped to all available output wavelengths.

Sparse conversion: Wavelength conversion is supported only by a subset of network nodes.

With full-range conversion there is no wavelength continuity constraint on the end-to-end connections. It is the most performing and, unfortunately, the most expensive type of wavelength conversion, but it is the assumed by most current OBS (and OPS) proposals.

With limited range conversion, shifting is restricted. This means that not all incoming channels can be connected to all outgoing channels. It is an approach that reduces the cost of the switch at the expense of increased blocking. Nevertheless, in Shen et al. (2001); Eramo et al. (2005); Zhang and Yang (2006) the authors argue that using nodes equipped with limited range conversion it is possible to achieve loss performance values close to those with full conversion capability. The same assertion is supported by a comparative study in Callegati et al. (2006). When using partial conversion capacity two questions need to be answered: how many converters does the network need, and what are the optimal converter locations? This was addressed in Xi et al. (2005) where an heuristic is proposed to find out the necessary number of converters as well as the best placement scheme in order to approximate the best performance offered by full complete conversion.

Fixed range conversion is static. It is a restricted form of limited conversion and does not provide any flexibility with respect to wavelength output. With sparse conversion, only some nodes in the network are equipped with wavelength converters. Although this category is well-studied for WR networks, it has not been widely considered for both OBS and OPS, in part due to the poor loss performance at nodes without wavelength conversion capability. It has been shown in Xiao et al. (2004) that sparse nodal conversion is not able to approximate the best possible performance, especially in case of light load.

A recent study in Maranhao et al. (2008) where the main architectures for wavelength conversion proposed for OCS networks are applied to OBS, concludes that the ideal number of wavelength converters, from a cost benefit perspective, is achieved by using a quite small amount of wavelength converters compared to the full-complete wavelength conversion capability.

4.6 Summary

This chapter is devoted to OBS technology. Starts with the presentation of its generally proposed OBS architecture, which is composed of edge nodes and core nodes interconnected by optical links, and proceeds with a detailed description of the main functionality of its components. A perspective on the IP-over-OBS integration is provided as well. Several node based mechanisms through which it is possible to improve the overall QoS in OBS networks are also discussed, including the burst assembly process, signaling schemes for reservation of resources,

TABLE 4.1: Comparison of contention resolution schemes (Mukherjee, 2006a).

	FDL buffering	**Deflection routing**	**Burst Segment.**	**Wavelength conversion**
Control complexity	lower	lower	high	low
Maturity of technology	yes	yes	no	no
Stability	yes	no	no	yes
Delay	longer	longer	long	shorter
Loss probability	low	high	low	lower
Cost	high	low	high	higher
Efficiency	low	lower	higher	high
Finer granularity	no	no	yes	no
Burst order	yes	no	no	yes

scheduling algorithms, and contention resolution approaches which are summarized in Table 4.1 according to some specific features.

5

Quality of Service

The increased availability of bandwidth has accelerated the growth of a large variety of applications operating over the Internet which demand quality of service (QoS). For instance, applications such as voice and video cannot tolerate long queueing delays and may need to be given higher priority than regular email or browsing traffic[1]. This is a significant issue for next-generation networks (NGNs) since current IP provides only *best effort* service, while some mission-critical and real-time applications might require high QoS support with low values of delay, jitter and loss. Being considered the most promising infrastructure for the NGN, it is desirable for an OBS architecture to support different classes of traffic in the user plane in order to facilitate (or complement) any QoS-enhanced version of IP potentially deployed above.

[1]which are called *elastic* applications

The solution used in WR networks is to treat the optical connection between ingress and egress pairs as virtual links where QoS mechanisms developed for IP networks can be directly applied. However, in OBS networks, wavelengths are statistically shared among many connections between different pairs of nodes, leading to a finite burst loss probability on the transmission path, which renders the WR approach unusable. Although similar to a datagram transport protocol, QoS developments for IP can hardly be adapted to OBS networks due to some of its intrinsic features. Among other differences, OBS lacks electronic buffering, by which service class isolation can be provided in packet switching networks, and optical processing is still in its infancy, making it difficult to implement the typical functions of buffer-based schemes.

Intrinsic OBS features present new challenges for QoS provisioning. Considering the separation between control and data planes, new algorithms were proposed in literature taking advantage of the offset time between control packet and data burst along with reservation, scheduling and contention resolution policies. This chapter focuses on how differentiated QoS is provided in OBS networks.

5.1 Differentiated QoS Provisioning

In general differentiated QoS provisioning can be achieved by introducing *differentiation* at some point in the network. In OBS networks, the various schemes providing QoS can be classified in the following four categories based on the stage at which service differentiation is performed (Praveen et al., 2006):

1. Assembly-time schemes
2. Reservation schemes
3. Scheduling schemes
4. Contention resolution schemes

Some research proposals in each category will be described next. It is noteworthy that some of them involve the joint activity of two or more mechanisms, resulting in combined schemes which can be classified in more than one of these categories.

5.1.1 Assembly-time schemes

In this category, differentiation is performed during the assembly of packets into bursts. In Klinkowski et al. (2005), for example, a burst length differentiation (BLD) technique is proposed to improve the performance when applied together with other QoS mechanisms. The authors considered the application of BLD to assign different burst sizes for different classes of traffic, thus resolving contention and the desired QoS guarantees. Two different approaches will be considered next: prioritized assembly and intentional dropping.

5.1.1.1 Prioritized assembly

Composite burst assembly techniques to be used with burst segmentation are proposed in Vokkarane, Vokkarane, Zhang, Jue and Chen (2002); Vokkarane, Vokkarane and Jue (2003). The basic idea is to assemble a burst using traffic from multiple traffic classes with different priorities in such a way that the data with high priority is put towards the head of each burst and the data with low priority is put towards the end of the burst. Combined with segmentation (Vokkarane, Jue and Sitaraman, 2002) and using a preemptive channel under the assumption of tail dropping, in case of contention the data at the end is more likely to be dropped than the data at the head. A similar approach is followed in Arakawa et al. (2003), but using segmentation under the assumption of head dropping, and a composite burst assembled in the opposite way. In a recent study (Sarwar et al., 2008), the authors propose that high-priority packets should be placed in the middle of a burst and low-priority packets at the head and tail ends of a burst. Working under the assumption of a non-preemptive minimum overlap channel with void-filling, this approach allows for dropping of low-priority packets from the head and the tail of a burst when it contends with two other bursts (more precisely, when the head of the burst contends with one burst, and the tail contends with another).

5.1.1.2 Intentional dropping

In intentional dropping, lower priority bursts are intentionally dropped at assembly time if the percentage of bursts transmitted is not in proportion to that of other classes as predetermined by the proportionality factors of differentiation (Praveen et al., 2006). In Chen et al. (2001) two intentional dropping schemes are proposed to give proportionally differentiated burst loss probability and average packet delay. The first proposal is based on a proportional QoS differentiation model which makes possible to quantitatively adjust the service differentiation of a particular QoS metric to be proportional to the factors that a network service provider sets. In this case, the QoS metric is burst loss. In this scheme a burst is intentionally dropped when a certain equated threshold value is violated, giving more or longer free time periods on the output link capacity, i.e., more opportunity, for high priority bursts to be admitted. The authors also propose a burst assembly scheme that extends the waiting time priority (WTP) scheduling algorithm from Dovrolis et al. (1999) and used in IP networks to provide proportional packet average delay in OBS networks. In this case the QoS metric is average delay experienced by a packet during burst assembly. This proportional differentiation model aims to provide the network operator with the "tuning knobs" for adjusting the quality spacing between classes independently of the class loads, something that cannot be achieved with other relative differentiation models, such as strict prioritization or capacity differentiation (Dovrolis et al., 1999). Similar schemes are proposed in Zhang et al. (2003b,a); Zhang, Zhang, Vokkarane, Jue and Chen (2004) for providing absolute QoS differentiation in OBS networks.

5.1.2 Reservation schemes

This category of QoS schemes provide service differentiation using different reservation policies for different classes. Offset-time based mechanisms, the forward resource reservation (FRR) scheme, a preemptive reservation protocol, and a wavelength grouping scheme will be discussed.

5.1.2.1 Offset-time based

Offset-based schemes provide service differentiation in terms of burst loss probability by setting different offset times for bursts with different priorities. The idea behind this scheme is that different offset times significantly affect the QoS experienced by the data bursts since the larger the offset time for a burst, the higher its chance of successfully reserving a resource by virtue of it being able to reserve earlier (Chua et al., 2007c). An offset-based scheme is proposed in Yoo and Qiao (1998); Yoo et al. (2000) to provide relative QoS differentiation for multiple classes of traffic using an extra offset time for a traffic class of higher priority, thus offering to such bursts a much greater probability of being successfully transmitted. Therefore a class with higher priority will be isolated from the effects of burst arrival from a traffic class with lower priority. Moreover, if the offset time of class i is sufficiently greater than the *max(burst length) + offset* of the immediately lower class, then complete isolation of class i from lower classes is achieved, and its blocking probability is independent of traffic in lower classes (Vu and Zukerman, 2002; Yoo and Qiao, 1999). The offset-based scheme is also used with FDLs in Yoo et al. (2000) with reported improved performance.

Offset-based schemes are easy to implement and highly scalable. They provide a measurable degree of differentiation and complete class isolation can be achieved. However, the additional offset time added may lead to unacceptable end-to-end delays for high priority packets. The method also tends to discriminate against large-sized low priority bursts (Chen et al., 2001). High priority bursts with large offsets tend to break the free capacity periods into small pieces with many voids.

Therefore, among low priority bursts, those with smaller size will have higher probability of being successfully transmitted. Also in the down side, it is reported in Kim et al. (2006) that under certain circumstances the mechanism does not perform well when TCP is used over OBS networks. In order to reduce the delay, a novel algorithm proposed in Mikoshi and Takenaka (2008) transmits a control packet before the burst assembly period expires. This mechanism needs to estimate the burst length, for which the Jacobson/Karels algorithm used in TCP is adopted. To reduce the estimation error, the proposed method uses a head-dropping algorithm for the burst collision procedure.

5.1.2.2 Forward resource reservation (FRR)

This scheme was proposed in Liu et al. (2003) where the authors present a transmission mechanism embedded at ingress nodes to efficiently reduce the end-to-end delay, and a resource reservation algorithm is derived to achieve controllable FRR performance enhancement, including latency reduction capability. On a typical OBS system, the control packet waits for the completion of the burst assembly process before it is transmitted for signaling and resource reservation. After that, the data burst should be further delayed at the ingress node for an offset time before being launched into the core node. Upon this observation, the intuitive idea behind this proposal is that, rather then performing the above two processes in sequence, they should be processed in parallel and thereby minimize their impact on the total end-to-end burst delay. The FRR scheme involves the following three-step procedure (Liu et al., 2003): (1) Prediction, (2) Pretransmission, and (3) Examination.

1. In the prediction phase, as soon as the previous *burstification* is done, a new one starts and its burst length is immediately predicted by a linear estimation method.

2. In the pretransmission phase, instead of waiting for the burst assembly to complete, the control packet is constructed instantly upon the completion of prediction and is launched into the network.

3. In the examination phase, after the burst assembly completion the actual burst length is compared with the reservation that was sent in the pretransmission control packet and there are two cases to consider: if the actual burst length is less then or equal to the prereserved duration, the control packet has reserved enough bandwidth for the data payload and the data burst is transmitted; if the actual burst length exceeds the prereserved duration the control packet is deemed a failure and a new one has to be retransmitted with the actual size.

The FRR scheme does not introduce any extra delay. In this scheme, even a failed forward reservation causes the same latency as a normal transmission not using FRR. However, service differentiation is obtained at the cost of decreased network utilization. Due to an aggressive reservation procedure, under a circumstance of incorrect prediction bandwidth may be wasted. Hence, the success of this scheme largely depends on the accuracy of the prediction, which may be inadequate for highly bursty traffic conditions.

5.1.2.3 Preemptive reservation

In this method, when a node receives the control packet of an high priority burst and fails to schedule it, the node may drop, or preempt, the low priority burst already scheduled to make room for the high priority one. Thus preemption employs the same concept as intentional dropping, but in a more efficient (and elegant) way (Chua et al., 2007d): it drops only those necessary to schedule the high priority burst.

A preemptive multiclass wavelength reservation protocol is presented in Loi et al. (2002) to provide service differentiation in terms of burst loss probability. In this scheme, a usage profile is maintained for each class at each router, and a higher priority burst may preempt a lower one whose usage is above a predefined limit. The overall performance of this scheme may be affected if the limits for the classes are not accurately set. In fact, if the limits of most classes are not met, the service may degrade to that of a classless system (Praveen et al., 2006).

An absolute QoS differentiation scheme called Virtual Channel Reservation (VCR) is proposed in Guan et al. (2005) providing worst case guarantee on the drop probability of higher priority classes. The service differentiation among each priority class in VCR is achieved by applying the concepts of virtual channel reservation and preemption. To address the problem of bandwidth under utilization, which is a potential side effect of preemption, the authors contribute with a new informing header (i-header) mechanism.

Recently, in Phung et al. (2007) the authors propose an absolute QoS framework to overcome a common difficulty from relative differentiation QoS models: the inherent difficulty in communicating information about internal network states to the edge in a timely manner for making admission control decisions. The key idea is to offer quantitative loss guarantees at each hop using a differentiation mechanism and an admission control mechanism. The edge-to-edge loss requirement is then translated into a series of small per-node loss probabilities that are allocated to the intermediate core nodes. The framework includes a preemptive differentiation scheme, a node-based admission control scheme and an edge-to-edge reservation scheme.

More recently, in Marchetto (2008), the authors propose a preemption based service differentiation solution that is suitable for the just-in-time optical burst switching paradigm thanks to the fact that it can minimize the preemption probability in case of contention. The proposed technique combines a conventional preemption scheme at core nodes and an improvement of the recently proposed burst cluster transmission scheme (Tachibana and Kasahara, 2006) at edge nodes. In particular, bursts are created at their ingress node and combined into chains, arranging them in order of decreasing priority.

A variant of preemptive schemes, called probabilistic preemptive burst segmentation (PPBS), preempts and segments low priority bursts in a probabilistic fashion (Tan et al., 2004). This technique is explored in Yang et al. (2003); Tan et al. (2006) to provide relative burst loss rate differentiation by setting different preemption probabilities for multiple classes of traffic. To avoid starvation among the bursts with lower priority and achieve an adequate burst loss ratio among the different traffic classes, preemption is conditioned by a certain probability value p. A flexible burst loss ratio among the different traffic classes can be obtained by changing the value of p. A combined use of probabilistic preemption with segmentation is also proposed in Tan et al. (2006).

5.1.2.4 Wavelength grouping

A wavelength grouping scheme has been proposed in Zhang et al. (2003a) to provide absolute QoS differentiation. In this scheme lower priority bursts are restricted to use only a certain set of wavelengths, while those with higher priority can use a larger, or complete, set of wavelengths. By limiting the wavelength search space for each control packet from lower service classes, differentiated QoS can be provided, and a service class with larger wavelength search space will have lower burst loss probability. The success of this scheme may depend on the relationship between the required degree of differentiation between classes and the cardinal number of the correspondent wavelength sets, which may be difficult to find.

The wavelength grouping approach has the drawback of inefficient wavelength utilization. The reason is that bursts of a given class are restricted to a limited number of wavelengths. Thus, if a class experiences a short period of high burst arrival rate, those burst cannot be scheduled to more wavelengths even though they may be free (Chua et al., 2007d).

Although most proposals assume full wavelength-conversion, wavelength converters are still very complex and expensive. Having that important issue in mind, in Gonzalez-Ortega et al. (2008) the authors propose the QoS Multiple Wavelength

Simultaneous Transmission technique (QoS-MWST), a novel method for providing loss differentiation in OBS networks without wavelength-conversion capability (or even limited capability). In QoS-MWST, multiple copies of each burst are simultaneously sent on different wavelengths, and loss differentiation is achieved by properly sending more copies of the higher-priority bursts.

5.1.3 Scheduling schemes

Differentiated QoS can also be provided by allocating different amount of bandwidth at the OBS nodes to different service classes. A differentiated scheduling scheme is proposed in Liu and Liu (2002) where the processing of a low priority control packet is delayed at the core node by a certain period of time. This allows higher priority control packets which arrive later to reserve bandwidth ahead of the low priority bursts. By controlling this waiting time at a node for low priority control packets, each node can provide proportional QoS to different service classes (Praveen et al., 2006).

A slot-based prioritized scheduling scheme is presented in Yang et al. (2001). In this approach a slotted transmission mode is assumed where data bursts are transmitted in units of slots and control packets in groups, with each group carried in one slot. Among all control packets which arrive in a slot, the ones having higher priority are scheduled first, allowing them a higher probability of being scheduled. This approach does not need a large initial offset for low priority bursts as in Liu and Liu (2002), but the choice of the slot size matters: a small slot size may result in the service degrading to that of a classless system, and a large slot size may bring unfair discrimination against low priority bursts.

Another approach that provides differentiated service using prioritized scheduling is proposed in Kozlovsky and Bayvel (2003). It relies on a centralized request server that performs the scheduling for the entire network and tries to minimize the delay experienced by a request in its queue. This scheme uses a weighted round robin policy where the weights are assigned accordingly to priorities of traffic classes. The main QoS metric of this scheme is blocking probability but it also tries to maintain the delay requirements of the various classes. A drawback of this approach is that the request server may become a bottleneck for the network resulting in poor performance.

5.1.4 Contention resolution schemes

Contention is the major cause of burst loss in OBS networks. Considering that most of the QoS mechanisms of OBS networks have burst loss or blocking probability as their main QoS metrics of interest, then contention resolution is also considered in most of them. When two bursts contend with one another, the burst that was scheduled first is referred to as the original burst and the burst that arrived later is referred to as the contending burst. The method proposed in Vokkarane, Vokkarane, Zhang, Jue and Chen (2002); Vokkarane, Vokkarane and Jue (2003), for example, segments the tail of the original burst during contention, an approach that has the advantage of better in-sequence delivery of packets at the destination. Motivated by this process, during burst assembly, the packets in the burst are arranged in decreasing order of priority, i.e., with lower priority packets at the tail of the burst. In Vokkarane, Vokkarane, Zhang, Jue and Chen (2002); Vokkarane, Vokkarane and Jue (2003), various policies can be applied to resolve contention:

Drop policy (DP): The original burst wins the contention. The entire contending burst is dropped.

Segment and drop policy (SDP): The contending burst wins the contention. The original burst is segmented and the tail segments of the original burst are dropped.

Deflect drop policy (DDP): The contending burst is deflected to an alternate port if an alternate port is available. If no alternate port is available, then the contending burst is dropped.

Segment first and deflect policy (SFDP): The original burst is segmented, and the tail segments of the original burst may be deflected if an alternate port is available, otherwise the tail segments of the original burst are dropped.

Deflect first, segment, and drop policy (DFSDP): The contending burst is deflected to an alternate port if an alternate port is available. If no alternate port is available, then the original burst is segmented and the tail segments of the original burst are dropped, while the contending burst is routed to the original output port.

A similar scheme is proposed in Arakawa et al. (2003), wherein the head of the contending burst is segmented and dropped.

As already stated, in OBS networks, data loss may occur when bursts contend for network resources. There have been several proposed solutions to resolve contentions in order to minimize loss. These localized contention resolution techniques react to contention, but do not address the more fundamental problem of congestion. Hence, there is a need for network level contention avoidance using load balanced routing techniques in order to minimize the loss.

5.2 Traffic Engineering

The ultimate purpose of QoS mechanisms in any network is to provide end-to-end QoS to end users. To achieve this purpose, a wide range of mechanisms must be deployed in the network. They include node based mechanisms, implemented both in edge nodes and core nodes, but also edge-to-edge core network mechanisms. Node based QoS mechanisms such as burst scheduling and service differentiation have been discussed above [2]. In this section, some *traffic engineering* mechanisms

[2] Maybe the reader can realize why, by its very nature, some consider QoS as "a system of managed unfairness", while others argue that "QoS technologies are simply the enablers to organizational objectives"...

within the core network to facilitate end-to-end QoS provisioning, namely *load balancing* and *offline routing optimization* will be discussed.

Traffic engineering (TE) has long been considered an essential mechanism for the next generation network, used to improve its overall network performance. A network without TE is, in most cases, operated under the simplest way to route packets or traffic flows based on the fixed shortest path. Despite the benefits of such simplicity, it can cause some traffic load unbalance in the network, i.e., a large amount of traffic may traverse a certain set of links, while there can be little traffic load on other links. The disadvantage of such an unbalance is poor network capacity utilization. To improve such an unbalanced situation, the concept of TE is therefore proposed. To improve the network capacity utilization, TE tries to best balance link utilization in the network to avoid congestion on some links.

5.2.1 Load balancing

The key form of TE is load balancing, in which traffic from congested areas is diverted to lightly loaded areas. In doing so, load balancing frees up network resources at bottleneck links and helps the network to provide better QoS to end users (Chua et al., 2007e). Several load balancing algorithms have been proposed for OBS networks, some of them (Wen et al., 2003; Zhang, Wang, Zhu, Datta, Kim and Mukherjee, 2004) for example, adopting an offline approach in which the traffic demand among the various source and destination pairs is known. This load balancing problem is usually formulated as an optimization problem and solved by integer linear programming.

In Li et al. (2005); Thodime et al. (2003) the problem of dynamic load balancing for best effort traffic in OBS networks is considered independently by two research groups. The authors propose a load balancing scheme based on adaptive alternate routing aimed at reducing burst loss. The load balancing scheme is based on adaptive alternate routing to reduce network congestion by adaptively distributing the load between two pre-determined link-disjoint alternative paths based on the measurement of the impact of traffic load on each of them.

In Yang and Rouskas (2006) the authors investigate the concept of adaptive path selection in optical burst-switched networks and its potential to reduce the overall burst drop probability. Specifically, the authors assume that each source maintains a list of alternate paths to each destination and uses information regarding the recent congestion status of the network links to rank the paths. Bursts are then transmitted along the least congested path. In this proposal, the authors present a suite of path selection strategies, each utilizing a different type of information regarding the link congestion status, and evaluate them using simulation.

Two recent load balancing proposals (Gonzalez-Ortega et al., 2007; Kiran et al., 2007) address the issue without considering the commonly used network scenario with full wavelength conversion capability. In Kiran et al. (2007), path selection and wavelength selection for OBS networks are formulated as a multi-armed bandit problem, and the difficulties to solve them optimally are discussed. The authors propose algorithms based on Q-learning to solve these problems near-optimally. At an ingress node, the path selection algorithm evaluates the Q values for a set of precomputed paths and chooses a path that corresponds to a minimum burst loss probability. Similarly, the algorithm for wavelength selection selects a wavelength in a pre-routed path such that the burst loss probability is minimized. This proposal does not consider wavelength conversion and buffering at the core nodes. Thus, selection of path and wavelength is done only at the edge of the network. In Gonzalez-Ortega et al. (2007) two approaches for burst loss reduction are addressed considering two adaptive multipath mechanisms (AARA (Li et al., 2005), and GPMR (Lu et al., 2006)) that have been proposed in the literature assuming that all the interfaces have total wavelength conversion capacity. However, it is expected this feature to be gradually introduced, giving rise to heterogeneous scenarios where different types of interfaces coexist. In Gonzalez-Ortega et al. (2007) the authors propose the implementation of a modified GPMR that clearly outperforms AARA in heterogeneous scenarios.

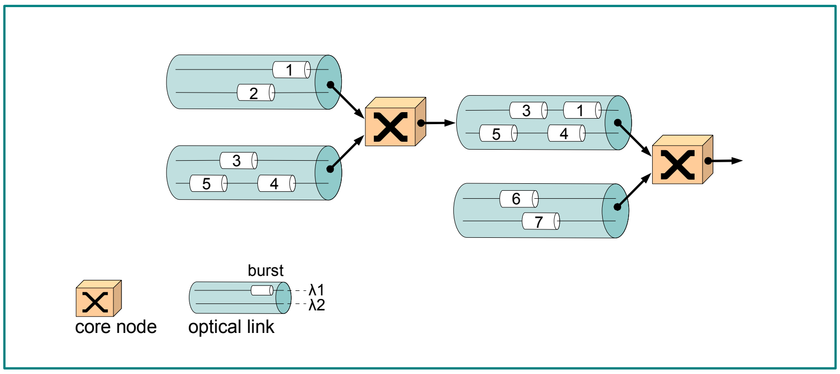

FIGURE 5.1: An illustrative example of the streamline effect.

5.2.2 The streamline effect

An important part of a load balancing algorithm is to collect information about the congestion level at various parts of the network. This is usually indicated by the burst loss probability evaluated at a link level. In fact, if the load balancing algorithm works online and the traffic is best effort in its majority, it is sufficient to measure the burst loss probability at a link to know its congestion level. However, if the algorithm works offline or the traffic is, in its majority, reservation-based QoS traffic, then the approach just mentioned is not possible. In the later case, what is important is the future burst loss probability when the traffic corresponding to a reservation is transmitted. Under this circumstances, the load balancing algorithm is supposed to estimate the burst loss probability at a link.

The traditional estimation approach for performance evaluation in OBS networks uses the M/M/k/k queueing model (Kleinrock, 1975; Yoo et al., 2000). This model assumes that the input traffic to an OBS core node is Poisson, which is equivalent to having an infinite number of independent input streams. However, the number of input streams to a core node is bounded by the small number of its input links, which makes the this model of prediction inaccurate (Chua et al., 2007e). A new and more accurate analytical model to estimate the burst loss probability in OBS links is proposed in Phung, Chua, Mohan, Motani and Wong (2005); Phung et al. (2006) taking into account a newly observed phenomenon called *streamline effect*.

Consider an OBS core node with a number of input links connected to an output link (see Figure 5.1) To facilitate discussion, we define a stream as the aggregate of all burst flows arriving in a common input link and destined for a common output link. The streamline effect is the phenomenon wherein bursts traveling in a common link are streamlined and do not contend with each other until they diverge. The reason is because there is no buffer inside an OBS network. Therefore, once the contentions among them are resolved at the first link where they merge, no intra-stream contention will occur thereafter. This effect is illustrated in Figure 5.1 where burst streams 1 and 3 merge at the first node. After any burst loss that might happen at that node, the remaining bursts are streamlined and no further contention will happen among them. However, they may still experience contention with other burst streams that merge at downstream nodes (Phung, Chua, Mohan, Motani and Wong, 2005).

The significance of this streamline effect is two folds. Firstly, since bursts within an input stream only contend with those from other input streams but not among themselves, their loss probability is lower that the one estimated by the M/M/k/k model. This can have major implications for TE algorithms that need to predict burst loss probability at a link. Secondly, the burst loss probability is not uniform among the input streams. The higher the burst rate of the input stream, the lower its loss probability. Therefore, if traffic within an OBS network is encouraged to form major flows with fewer merging point, the overall loss rate will be reduced (Phung et al., 2006).

The streamline effect was used in Phung, Chua, Mohan, Motani and Wong (2005) in a load balancing algorithm for reservation-based traffic specifically developed for the end-to-end QoS framework in Phung et al. (2004).

5.2.3 Offline route optimization

In Teng and Rouskas (2005b) the authors take a TE approach to path selection with the objective of balancing the traffic across the network links in order to reduce congestion and improve the overall performance. An approximate integer linear optimization problem to solve the problem is presented, as well as a simple heuristic that can be used for large networks.

In Chen, Mohan and Chua (2006b) the problem of route optimization to determine a route layout for a given traffic demand and minimize the overall burst loss is considered with an offline approach. It is argued that route selection based on the traditional Erlang B formula is not efficient because of unique features of OBS networks such as streamline effect. The authors analyze the streamline effect and propose a more accurate loss estimation formula which takes that effect into consideration. Under similar assumptions, the problem of offline route optimization for failure recovery is considered in Chen, Mohan and Chua (2006a) for OBS networks using Multiple Protocol Label Switching (MPLS).

In Chen et al. (2008), the authors assume that the network has MPLS control, and that each node has full wavelength conversion. Offline route optimization is used to determine the routes for the traffic flows in such a way that the overall network burst loss is minimized. In this work the estimated traffic demand is known and assumes a quasi-stationary traffic demand, as the measurements in Internet traffic indicate for aggregated links, where load changes relatively slowly. Two route optimization problems are proposed. The first considering the usual case of normal state where all the links are working properly, and one route is determined for each flow to minimize the overall burst loss. The second, considering failures, where primary and backup paths for each flow are determined in such a way to minimize the expected burst loss over the normal and the failure states.

In OBS networks, signaling protocols are more mature than routing protocols. Usually, routing algorithms for WR networks are directly adopted in OBS networks. However, it may be necessary to dedicate more research effort to the development of specific routing algorithms for the OBS environment (Mukherjee, 2006b). Some of that research effort was presented here, and some is currently being conducted to overcome the issues involved in the subject. Some of those issues will be discussed in the next chapter of this thesis.

5.3 Summary

In this chapter various methods for providing QoS differentiation are discussed based on the stage at which service differentiation is implemented. This includes differentiation performed during assembly time, at the moment of resource reservation, at scheduling time, or when contention resolution is invoked. Several TE mechanisms for end-to-end QoS provisioning are then presented, including load balancing and offline route optimization approaches in which the streamline effect can be incorporated.

6

Simulation Model

The need for performance evaluation and prediction exists from the initial conception of a system to its daily operation, but in the early planning phase of a new network architecture, the manufacturer usually must make two types of predictions: one is to forecast the workload, i.e., the amount of service demands imposed on the system, and which characterization includes not only the traffic arrival model but also the work demands on the multiple resource components of the network; the other is concerned with the choice between architectural design alternatives, based on hardware and software technologies that will be available in the development period of the planned system, and where the selection criteria is often the so-called *cost-performance tradeoff*. The accuracy of such predictions rests to a considerable extent on the capability of mapping the performance characteristics of each network component into the overall network-level performance, a translation procedure that is not straightforward or well established. The techniques used for performance evaluation and prediction during the design phase range from simple hand calculation to quite elaborate simulation. No matter which one of these techniques is followed, it is now widely accepted that the performance prediction and evaluation process should be an integral part of the development effort. In this chapter the simulation approach fallowed in this thesis will be described.

6.1 The Simulation Method

Simulation makes possible the systematic study of problems when analytic solutions are not available and experiment on the actual system is impossible or impractical. The simulation method has been an important means of determining the performance differences between alternative configurations (both in hardware and software) of computing systems and networks, whether it be in the design stage, its installation stage, or its tuning stage (Kobayashi, 1978).

Simulation is the imitation of the operation of a real-world process or system over time. It involves the generation of an artificial history of the system and the observation of that artificial history to draw inferences concerning the operating characteristics of the real system that is represented (Banks, 1998a). The simulation model describes the operation of the system in terms of individual events of the individual elements in the system. The interrelationships among the elements are also built into the model. Then the model allows the computer in which it runs to capture the effect of the element's actions on each other as a dynamic process.

After being constructed, the simulation model is driven either in a probabilistic (self-driven) way, by generating its own input data, or in a deterministic (trace-driven) way, by feeding some representative input data, and simulates the actual dynamic behavior of the system. By repeating this process for various alternative system configurations and parameters, one may identify an optimal system structure. Therefore, simulation can also be considered as a technique of conducting sampling experiments on the model of the system.

The term simulation has a number of connotations. In our discussion, simulation is a technique for conducting an experiment (by a computer) of a system evolving in time. Therefore, in a simulation the concept of time is explicit. A simulation model describes the dynamic behavior of a system, even when the system analyst may be ultimately interested in only the mean value of some measure in the steady state (Kobayashi, 1978).

There are several concepts underlying simulation which will be summarized next following the approach in Banks (1998a). After, a definition of discrete-event simulation is provided.

Model: A model is a representation of the real system. Immediately, there is a concern about the limits or boundaries of the model that supposedly represent the system. Thus, what should be its level of detail? The right level of detail depends on the purpose of the performance evaluation task, the degree of understanding of the system to be modeled as well as its environment, and the output statistics required. The model should be complex enough to answer the questions raised, but not too complex.

Event: An event is an occurrence that changes the state of the system, and there are both internal and external events, also called *endogenous events* and *exogenous events* respectively.

State variables: The system state variables are the collection of all information needed to define what is happening within a system to a sufficient level at a given point in time.

Entity: An entity represents an object that requires explicit definition. An entity can be dynamic if it moves through the system, or can be static in that it serves other entities.

Attributes: An entity may have attributes that pertain to that entity alone and should be considered as local values.

Resource: A resource is an entity that provides service to dynamic entities.

List processing: Entities are managed by allocating them to resources that provide service. This is done by attaching them to event notices (thereby suspending their activity into the future), or by placing them into an ordered list, which is used to represent a queue (often processed according to FIFO, but other possibilities exist).

Activity: An activity is a period of time whose duration is known prior to commencement of the activity. Thus, when the duration begins, its end can be scheduled. The duration can be a constant, a random value from a statistical distribution, the result of an equation, input from a file, or computed based on the event state.

Delay: A delay is an indefinite duration that is caused by some combination of system conditions. When an entity joins a queue for a resource, the time that it will remain in the queue may be initially unknown since that time may depend on another events that may occur.

Based on this concepts is now possible to define a *discrete-event simulation model* as the one in which the state variables change only at those discrete points in time at which events occur. Events occur as a consequence of activity times or delays. Entities may compete for system resources, possibly joining queues while waiting for an available resource. Activity and delay times may "hold" entities for durations of time. A discrete-event simulation model is conducted over time by a mechanism that moves simulated time forward. The system state is updated at each event, along with capturing and freeing of resources that may occur at that time (Banks, 1998b).

The number of areas in which simulation is used is increasing rapidly with benefits that go beyond the one-time remodeling of a facility or the simply providing of a look into the future. These benefits, mentioned in Banks et al. (1996); Law and Kelton (1991), are summarized in Table 6.1 together with some of the disadvantages usually tagged to simulation.

Table 6.1: Pros and cons of using simulation

√	Choose correctly
√	Compress and expand time
√	Understand why
√	Explore possibilities
√	Diagnose problems
√	Identify constraints
√	Develop understanding
√	Visualize the plan
√	Build consensus
√	Prepare for change
√	Invest wisely
√	Train the team
√	Specify requirements
×	Model building requires special training
×	Simulation results may be difficult to interpret
×	Simulation modeling and analysis can be time consuming and expensive
×	Simulation may be used inappropriately

It is worth noting that, most of these disadvantages can be counterbalanced by the appropriate use of packages of simulation software already developed, which are generically called *simulators*. In truth, unless you are an expert, writing software completely "from scratch" is no longer advisable since software systems tend to be complex and several libraries exist for many of the common functions. But even if the aforementioned difficulties persist, simulation is often the only practical solution to a real problem. In that case, according to Pritsker (1998), the next four basic modeling principles should be properly taken into consideration:

1. Conceptualizing a model requires system knowledge, engineering judgment, and model building tools. A modeler must understand the structure and operating rules of a system and be able to extract the essence of the system without including unnecessary detail. The amount of detail to include should depend on the modeling objectives established. A modeling project is normally an interdisciplinary activity and close interaction among project personal is required.

2. The secret to being a good modeler is the ability to remodel. Model building should be interactive and graphical because a model is not only defined and developed, but is continually refined, modified, updated, and extended.

3. The modeling process is evolutionary because the act of modeling reveals important information in stages. Information obtained during the modeling process supports actions that make the model more relevant and accurate. The modeling process continues until additional detail or information is no longer needed for problem resolution (or a deadline is reached).

4. The problem (or problem statement) is the primary controlling element in model-based problem solving. A problem or objective drives the development of the model. Problem statements are defined from system needs and requirements. Data from the system provide the input to the model. The availability and form of the data help to specify the model boundaries and details. The modeler is the resource used to build the model in accordance with the problem statement and the available system data. The outputs from the model support decisions to be made to solve the problem.

Figure 6.1 depicts a set of steps to guide a model builder in a simulation study. For a detailed description of these steps we refer the reader to the book of Banks (1998a).

6.2 Object-Oriented Thinking

The technology of simulation is now a mature and well developed methodology. Although with plenty of room for additional research in some fundamental areas[1], computer simulation techniques have been widely adopted to represent the more diverse types of systems. However, systems tend to become more complex, and the real limits on the adoption of simulation from now on may rest on the ability to

[1] Random number and variate generation, or reliable and appropriate statistics, for example.

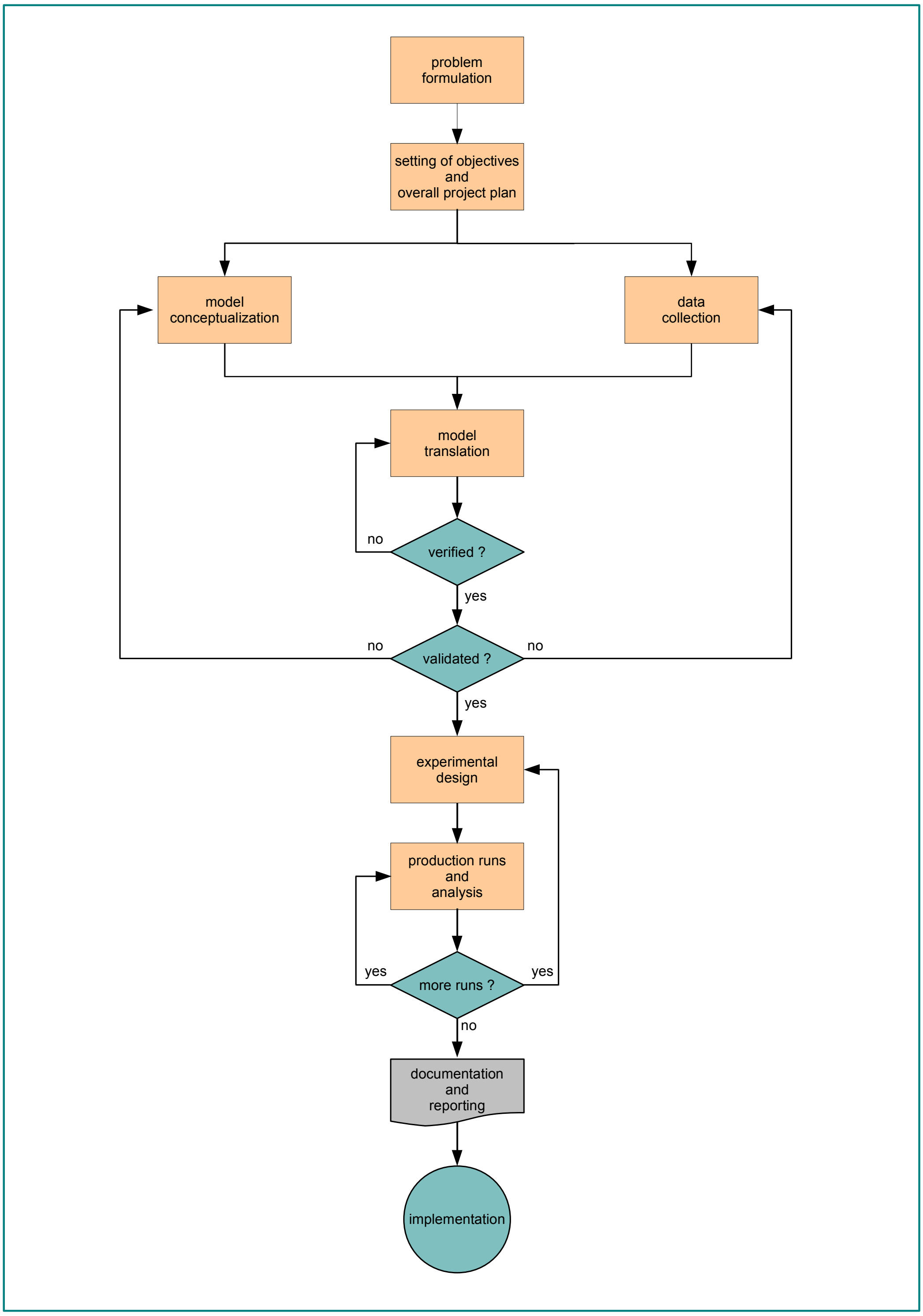

FIGURE 6.1: Steps in a simulation study (Banks, 1998a).

represent complex systems in an easy way (Joines and Roberts, 1999). This concern is also a matter of modeling style, which is closely related with programming style.

Concerning modeling style, an important contribution to improve the modeling of complex systems was the object-oriented (OO) paradigm. This paradigm emerged in the 1960s with the Simula programming language and its underlying concepts of *objects* and *classes* were later adopted by Smalltalk, the first successful programming language to be called "object-oriented" (Kay, 1996). Object-oriented programming (OOP) has been one of the most influential developments in computer programming, gaining momentum in the mid 1980s due to its facility for managing complexity in the ever-growing software systems. Despite the reserved attitude of some OOP purists[2], it was the standardization[3] of C++ and related generic programming facilities that really established, in 1998, the modern era of OOP. The success of this programming style is largely due to the power and popularity of C++ (Jana, 2004), associated to the genius of the Standard Template Library (STL), which demonstrated the utility of the new methodology to a very large audience of developers.

Why has OOP had such a sweeping impact on the software development community? OOP appeals at multiple levels. For managers, it promises faster and cheaper development and maintenance. For analysts and designers, the modeling process becomes simpler and produces a clear, manageable design. For programmers, the clear elegance of the object model and the power of OO libraries makes programming a much more pleasant task, and programmers experience an increase in productivity (Eckel, 1998). If there is a downside, it is the expense of the learning curve, because, thinking in objects is a dramatic departure from thinking procedurally. Nevertheless, the process of designing objects is much more challenging than procedural design, especially if you are trying to create reusable objects.

[2]Often accompanied by a feeling of intense dislike...

[3]ISO/IEC 14882:2003 is the current official standard for the C++ programming language and library.

6.2.1 The abstraction level

All programming languages provide a certain level of abstraction, and all computer programs engage in representations of reality which are based on those levels of abstraction. One can argue that the complexity of the problems we are able to solve is directly related to the kind (and *quality*) of the attained level of abstraction. Assembly language, for example, is a small abstraction of the underlying machine. And many of the so-called imperative languages that followed (such as Fortran, BASIC, or C) are abstractions of assembly language. These languages are big improvements over the assembly language, but their primary abstraction level still requires thinking in terms of the structure of the computer rather than the structure of the problem to solved. In fact, the programmer must establish an association between the machine model (in the "solution space")[4] and the model of the problem being solved (in the problem space)[5]. The effort required to perform this mapping, and the fact that it is extrinsic to the programming language, produces programs that are difficult to write and expensive to maintain Eckel (2000).

The alternative to modeling the machine is to model the problem to be solved. The OO approach goes further with respect to this by providing tools for the programmer to represent elements in the problem space. This representation is general enough so that the programmer is not constrained to any particular type of problem. The elements in the problem space and their representations in the solution space are referred to as objects[6]. The idea is that the program is allowed to adapt itself to the lingo of the problem by adding new types of objects. This way, by reading the code describing the solution, we are reading words that also express the problem, which is more flexible and powerful as language abstraction than the approach used before. Thus, OOP allows the description of the problem in terms of the problem, rather than in terms of the computer where the solution will run. There is still a connection back to the computer, since each object looks

[4]Which is the place where we are modeling that problem, such as a computer
[5]Which is the place where the problem exists
[6]Of course, other objects without problem-space analogs will also be needed.

quite a bit like a small computer: it has a state, and the operations we can ask it to perform (Eckel, 2000).

A pure approach to OOP was expressed by Alan Kay[7] in the following five basic characteristics of Smalltalk (Ingalls, 1981):

1. Everything is an *object*. An object stores data, but we can make requests to that object, asking it to perform operations on itself.

2. A program is a bunch of objects telling each other what to do by sending messages. To make a request of an object, we send a message to that object, i.e., a request to call a function that belongs to that object.

3. Each object has its own memory made up of other objects. Existing objects can be composed together to create new objects (bottom-up approach). Thus, complexity of objects can be built step-by-step by proper abstractions and compositions. Small objects clubbed together constitute bigger objects.

4. Every object has a type. An object is a runtime instance of a conceptual pattern or type called class. In fact every object has an associated type.

5. All objects of a particular type can receive the same messages. In true sense, this means objects of a particular class which may be specialized from another class respond to the same messages to behave similarly. This will lead to *polymorphism* which will send the same interface to a couple of objects belonging to a family of classes, so that the proper method is called in the appropriate class. This is a very powerful concept in OOP.

6.2.2 Fundamental concepts

Object-oriented development is a design technique rather than a coding convention. In developing an OO model, we must focus much more on the design than

[7]Smalltalk was the product of research by a group led by Alan Kay at Xerox Palo Alto Research Center.

the code (this is true, to a certain degree, in structured development as well). To create a solid design, we must first understand the concepts involved in designing the object model (Weisfeld, 2000). In a recent study (Armstrong, 2006) in which extensive efforts were made to gain complete coverage of the subject, diverse material related to OO development was reviewed. From that material, published from 1966 to 2005, a set of more than thirty potential concepts for characterizing the OO approach were collected. The study reveals that eight of those concepts (see Figure 6.2) are cited by more than 50% of the sources and terms them as the "OO *quarks*". These quarks can be considered as fundamental concepts in OO development and will be described next using definitions from (Semaphore, 1995) where the official OO terminology of the Object Management Group (OMG)[8] is adopted:

Inheritance: The language mechanism which allows the definition of a class to include the attributes and methods defined for another more general class. Inheritance[9] is an implementation construct for specialization relations. The general class is the superclass and the specific class is the subclass in the inheritance relation. Inheritance is a relation between classes that enables the reuse of code and the definition of a generalized interface to one or more subclasses.

Object: Anything to which a type applies, i.e., an instance of a class (or type). An instance of a class is comprised of the values linked to the object (object state) and can respond to the requests specified for the class.

Class: An implementation of an abstract data type. A definition of the data structures, methods, and interface of software objects. A template for the instantiation (creation) of software objects.

Encapsulation: The characteristic that an object is a discrete unit of data with a set of permissible operations that control access or modification of the

[8]OMG is an international, open membership, not-for-profit computer industry consortium. OMG Task Forces develop enterprise integration standards for a wide range of technologies.

[9]Also referred to as *class inheritance* or *derivation*.

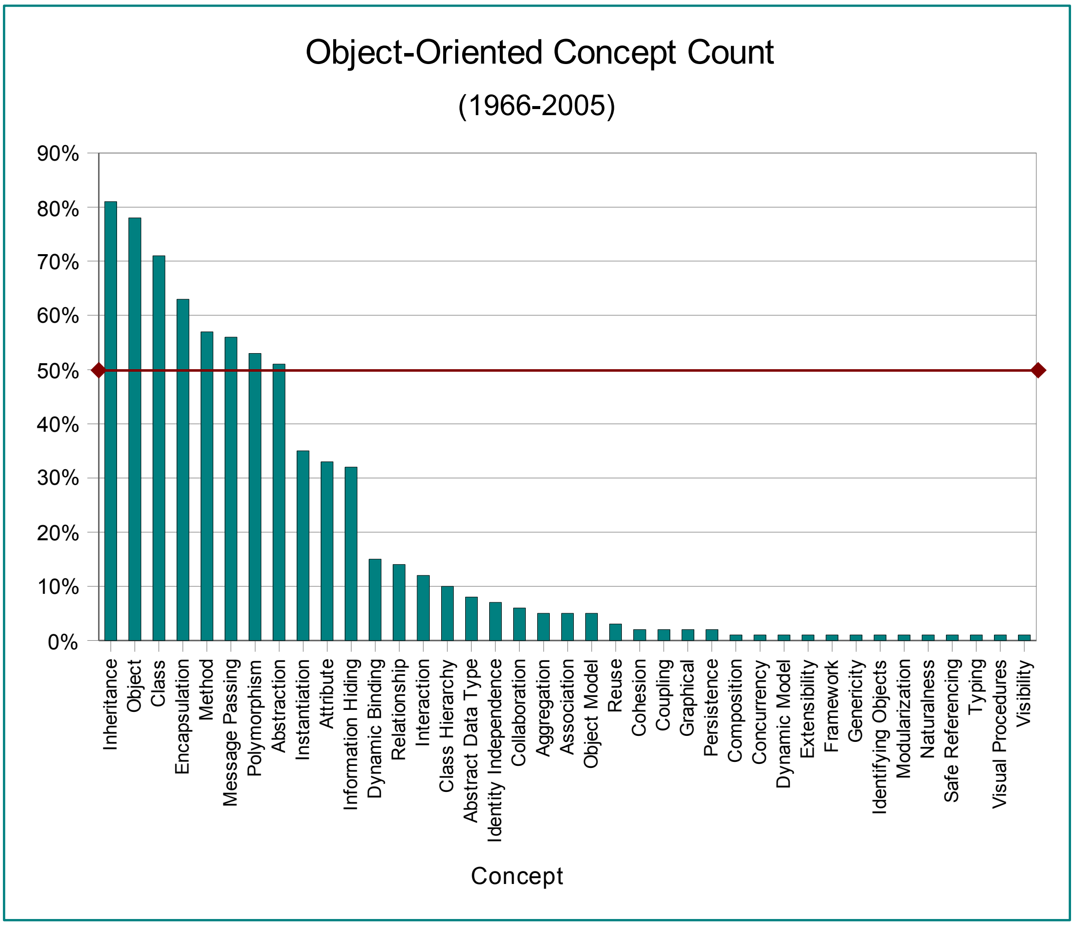

FIGURE 6.2: Object-oriented (potential) fundamental concepts.

values. Encapsulation[10] is a language mechanism by which external aspects are separated from the implementation of an object.

Method: The specific implementation of an operation for a class, i.e., code that can be executed in response to a request. A method can extend or override the behavior defined in the operation. A method is an implementation of a behavior of a type.

Message: A piece of information that identifies an operation to be invoked on an object and any required argument values. The form of the message is defined by a particular request. The response to a message is the invocation of a particular method, or some other action determined by the application.

[10] Also referred to as *information hiding*, or *data hiding*.

Polymorphism A request-handling mechanism that selects a method based on the type of the target object. This allows the specification of one request that can result in the invocation of different methods depending on the type of target object. Most OO languages support the selection of the appropriate method based on the class of the object (classical polymorphism).

Abstraction: The logical form which analogous things have in common. Abstraction[11] is also the process of identifying the characteristics that distinguish a collection of similar objects. The result of the process of abstraction is a type.

It is perceptible from the large number of items present in Figure 6.2 that a certain lack of consensus exist around the fundamental OO concepts. This lack of consensus can also be extended to the adequate way of classify these concepts in order to characterize the OO approach to software development. For the sake of simplicity the proposal in Armstrong (2006) will be presented, and the interested reader is referred to Henderson-Sellers (1992); Rosson and Alpert (1990) for alternative approaches.

The taxonomy proposed in Armstrong (2006) includes the eight most cited concepts above mentioned into two constructs labeled *structure* and *behavior*. The structure construct includes: abstraction, class, encapsulation, inheritance, and object. These concepts are focused on the relationship between classes and objects, and the mechanisms that support the class/object structure. In essence a class is an abstraction of an object. The class/object encapsulates data and behavior, and inheritance allows the encapsulated data and behavior of one class to be based on an existing class. The behavior construct includes: message passing, method, and polymorphism. This construct is focused on object actions (behavior) within the system. Message passing is the process in which an object sends information to another object or asks the other object to invoke a method. Polymorphism enacts behavior by allowing different objects to respond to the same message and implement the appropriate method for that object.

[11] Also referred to as *concept* or *type*.

How do these two constructs fit together? Within the OO approach, behavior (the actions and reactions of the system) is a way of manipulating structure (objects and/or processes within a system and their relations). However, behavior must also support the actions of the system. So this taxonomy emphasizes the interconnected nature of OO development, while providing a simpler approach than the previous ones in Henderson-Sellers (1992); Rosson and Alpert (1990). In Armstrong (2006), it is argued that the two proposed constructs, *structure* and *behavior*, simplify the process of organizing the "OO quarks" into a coherent conceptual scheme, and therefore, rather than complicating the learning process with a complex conceptual scheme, this taxonomy will both simplify the learning process and reinforce the OO approach.

The growth of general OO software and any attempt to enumerating them is beyond the scope of this chapter. However, it should be emphasized that the OO approach has permeated almost every area of computer software development, including in the computer simulation area, which is a fairly narrow interest in the broad spectrum of software development.

6.3 Object-Oriented Simulation

The idea of object-oriented simulation (OOS) has great intuitive appeal because it is very easy to view the real world as being composed of objects. In the networks field, for example, the physical objects can include computers, hubs, routers, cables, and other specific devices. However, some other not so obvious "things" like traffic streams, routing tables, scheduling plans or other information items, can also be viewed as objects. All these objects interact to produce the system behavior. This perspective of simulation, which is quite similar to a real world experience, has made possible by the advent of the OOP paradigm.

Several simulation languages or simulation packages (simulators) exist relying in the OO approach, usually providing the user with a set of predefined classes from

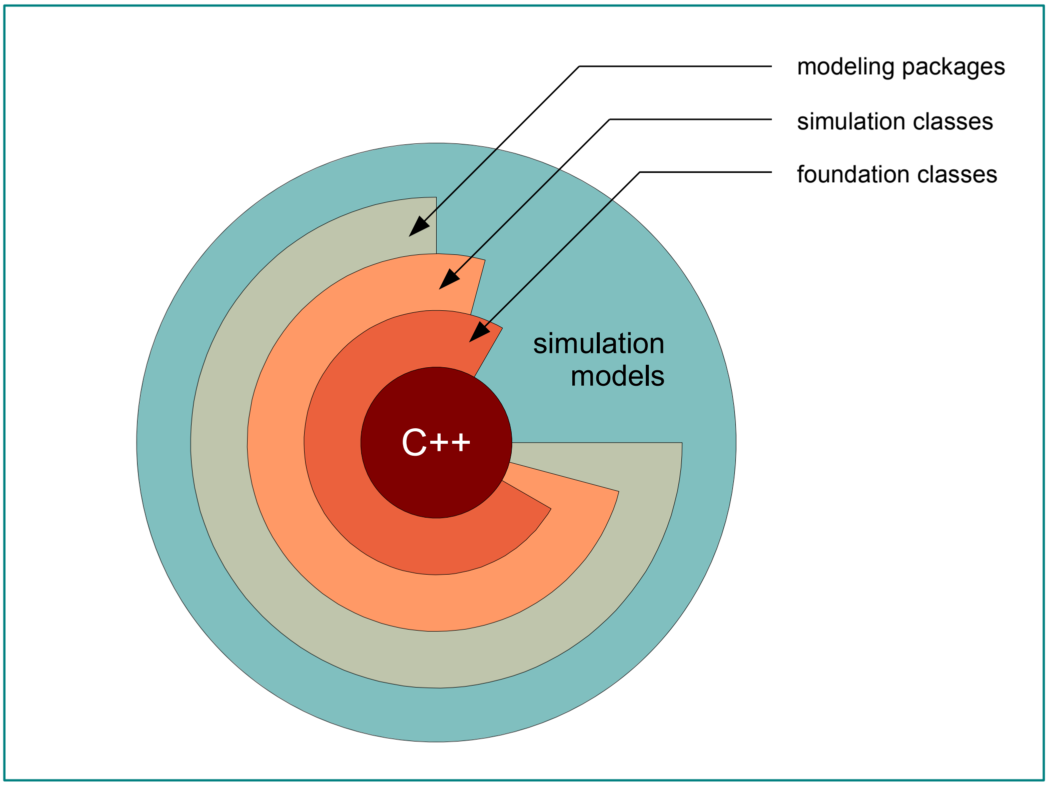

Figure 6.3: General conceptual OOS design (adapted from Joines and Roberts, 1998).

which the simulation modeler can create the needed objects. In general, the conceptual design of an OOS can employ the hierarchical approach illustrated in Figure 6.3. At the outer level, specific simulation models can be directly parametrized by model. At some point during the simulation study the specific model may be insufficient for the application and the modeler will need to resort to more fundamental modeling and simulation concepts and features. At the inner level in the hierarchy the user can employ the general programming language in which the model is based on (C++ in the figure) to implement or modify any required concept or feature. Thus, in an OOS environment the user can relate to the design in different ways, while in the old paradigm a simulation package falls somewhere in the middle of this design with limited opportunity to travel in or out (Joines and Roberts, 1998).

Despite being of fairly narrow interest, the network simulation arena is also becoming very competitive with several old simulation tools being replaced by more recent ones, some of them open-source. NS-2 (www, 2009e), OMNeT++ (www, 2009f), J-Sim (www, 2009c), SSFNet (www, 2009m) and JiST/SWANS (www, 2009d) are, for example, some of the open-source simulators, while OPNET (www, 2009g) and QualNet (www, 2009l) are among the commercial ones. In the open-source domain the most popular is NS-2, which is the most widely used in academic research circles for the analysis of telecommunication networks. From the commercial side, OPNET is usually the most considered both among researchers and companies. Using a significant different approach, OMNeT++ is also being considered one of the most relevant proposals in the simulation field[12] and it was the chosen simulator for the research work presented here. Some of the reasons why preference was given to OMNeT++ over other simulators can be found in the comparison reported in Varga and Hornig (2008), which is partially reproduced in the Appendix B.

6.4 The Adopted Simulator

OMNeT++ (Varga and Hornig, 2008) is an open-source discrete event platform for simulations freely available for the academic community and non-profit use[13]. Its primary application area is the simulation of communication networks, but due to its generic and flexible architecture, OMNeT++ has been successfully used in other areas like the simulation of complex information technology (IT) systems, queueing networks, or hardware architectures, and is rapidly becoming a popular simulation platform in the scientific community as well as in industrial settings (Sekercioglu et al., 2003). Among the distinguishing features of OMNeT++ it is worth to emphasize both its strong OO approach, which promotes well structured and reusable models, and its extensive graphical user interface (GUI) support, through which all internals of the simulation are visible, providing detailed feedback on what is going on. In fact, by using its advanced user interface over conveniently instrumented classes, OMNeT++ makes it possible to traverse objects of

[12] Like OPNET or NS-2, independent peer reviewed OMNeT++ workshops are taking place together with relevant international simulation conferences.

[13] For commercial purposes an OMNEST license from Omnest Global, Inc. is required.

a running simulation and display their internal information (such as name, class name, state variables or other contents). Moreover, it allows control over simulation execution and to intervene by changing variables/objects inside the model. This is a valuable feature not only for the development or debugging phases of the simulation project, but also for classroom presentation or demonstration purposes.

An OMNeT++ model consists of hierarchically nested modules. The depth of module nesting is not limited, which allows the user to reflect the logical structure of the actual system in the model structure. Modules communicate through message passing. Messages can contain arbitrarily complex data structures. Modules can send messages either directly to their destination or along a predefined path, through gates and connections. Modules can have their own parameters which can be used to customize module behavior and to parametrize the model's topology. Modules at the lowest level of the module hierarchy encapsulate the behavior. These modules are termed *simple modules* and are programmed in C++ using the simulation library, like figure 6.3 depicts. OMNeT++ simulations can be performed on several different user interfaces for different purposes (debugging, demonstration or batch execution) and are easily portable between Unix/Linux and Windows environments, using various C++ compilers (Varga, 2005).

How are the two aforementioned basic constructs *structure* and *behavior* embodied in a simulation framework developed with OMNeT++? A brief overview is presented in the following sections, while extensive description can be found in Varga (2005).

6.4.1 Model structure

OMNeT++ provides efficient tools for describing the structure of the actual system. Some of the main features are:

- Hierarchically nested modules
- Module instantiation from module types

- Intermodule communication via messages through channels
- Flexible module parametrization
- Topology description language

OMNeT++ models are often referred to as networks. The top level module is the *system module*. The system module contains submodules, which can also contain submodules themselves in an unlimited depth nesting approach. Modules that contain submodules are termed *compound modules*, as opposed to *simple modules* which are at the lowest level of the module hierarchy. Both simple and compound modules are instances of module types. While describing the model, the user defines module types. Then, in a bottom-up approach, instances of these module types serve as components for more complex module types. Finally, the user creates the system module as an instance of a previously defined module type. All modules of the network are instantiated as submodules (and sub-submodules) of the system module (Varga, 2005).

Modules communicate by exchanging messages. In a network simulation, messages can represent frames, packets, or bursts in a computer network, and they can contain arbitrarily complex data structures. Simple modules can send messages either directly to their destination or along a predefined path, through gates and connections. Gates are the input and output interfaces of modules, thus, messages are sent out through output gates and arrive through input gates. Each connection (also called link) is created within a single level of the module hierarchy, i.e., within a compound module, one can connect the corresponding gates of two submodules, or a gate of one submodule and a gate of the compound module. Due to the hierarchical structure of the model, messages typically travel through a series of connections, to start and arrive in simple modules. Such series of connections that go from simple module to simple module are called *routes*. Compound modules act as "cardboard boxes" in the model, transparently relaying messages between their inside and the outside world (Varga, 1998).

An important feature related with message passing is that the "local simulation time" of a module advances when the module receives a message, either when the message arrives from another module or from the same module. In the later case, messages are termed *self-messages* and can be used to implement *timers*.

Modules can have parameters that can be assigned through different ways, including configuration files. Parameters may be used to customize simple module behavior, and for parametrizing the model topology. Connections can be assigned three parameters, which facilitate the modeling of communication networks: propagation delay, bit error rate and data rate. This can be defined individually for each connection, or through predefined link types to be used throughout the whole model (Varga, 2005).

The structure and topology of an OMNeT++ model is defined by a textual network description (NED) language (Varga, 1998), which contains declarations of simple module types, describes compound module types and contains a network definition that instantiates a compound module. A NED description can contain the following components, in arbitrary number or order: import directives, channel definitions, simple and compound module definitions, and network definitions. The NED language grammar can be found in Varga (2005).

Simulation objects (messages, modules, queues) are represented by C++ classes which have been designed to work together efficiently, creating a powerful simulation programming framework. The following classes are part of the simulation class library: modules, gates, connections, parameters, messages, container classes (for example, queue, array), data collection classes, statistic and distribution estimation classes (histograms, algorithm for calculating quantiles), transient detection and result accuracy detection classes, among others.

6.4.2 Model behavior

The global functioning of the simulation model results from the behavior of their simple modules. In OMNeT++ this behavior is defined by algorithms implemented in C++, taking full advantage in power and flexibility of a general purpose programming language like C++, reinforced by the OMNeT++ simulation class library. The simulation programmer can choose between event-driven and process-style description, and can freely use object-oriented concepts (inheritance and polymorphism, for example) and design patterns to extend the functionality of the simulator.

6.5 Simulation Methodology

The modeling and design of large communication and computer networks has always been an important area to both researchers and practitioners. The high deployment and maintenance costs of networks has stimulated the interest in developing efficient design models and optimization methods, making good network design potentially capable of securing considerable savings (Poíro and Medhi, 2004). In such network design problems the Operations Research (OR) plays a major role as an interdisciplinary branch of applied mathematics that uses methods such as mathematical modeling, statistics, and algorithms to arrive at optimal or near optimal solutions to complex problems.

In the last years, networks have undergone substantial changes caused by the rapid development of new technologies and services, huge growth of data traffic, demand for service availability and continuity, and attempts to integrate different types of technologies and services. This fact increases the challenges that network designers must face. Another relevant aspect is that many network designers have excellent knowledge of networking technologies, including networking techniques, protocols, service implementation and so on, but may not have background in optimization. On the other hand, specialists in optimization and OR may not possess technological knowledge to develop and validate models. How to bridge the gap between

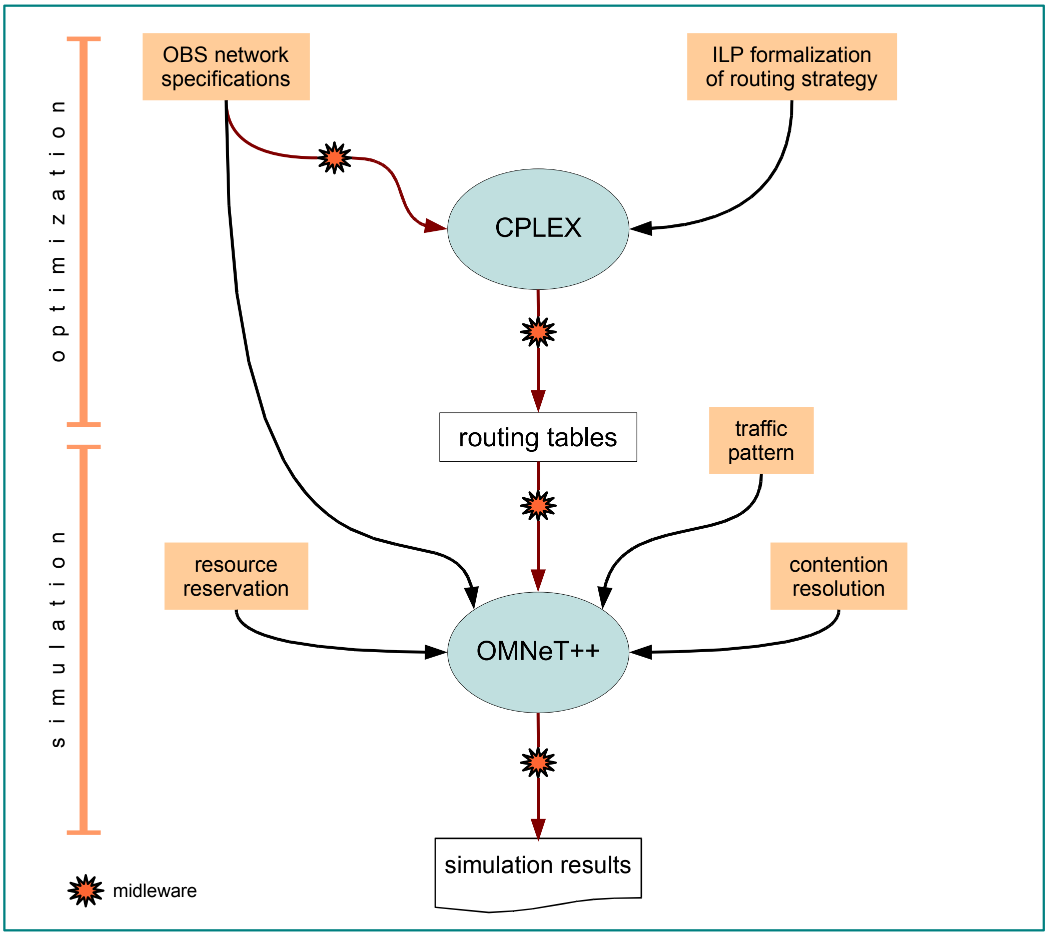

FIGURE 6.4: Conceptual view of the simulation methodology.

these two "parallel" worlds in current and future communication network scenarios becomes, therefore, a challenge for which contributions are welcome.

The aim of the research conducted here is the development of proposals to minimize the global network contention and overall burst loss through the adoption of optimized path selection strategies. Therefore, the work reported in the next two chapters and whose methodology is now described, incorporates both *optimization* and *simulation*, contributing to some extent to bridge the gap referred above.

Based on the combined operation of this two branches of knowledge, the adopted methodology is to be deployed in two stages as depicted in Figure 6.4. The first stage, which comprises the determination of routing paths, is the optimization stage. In this stage the problem is formulated using an ILP approach and solved by the CPLEX optimizer. The second stage, which comprises the application of the routing paths, is the simulation stage. Now the optimized routing paths just produced are incorporated into a realistic OBS network simulation model developed with the OMNeT++ simulator for performance evaluation.

6.5.1 Routing path determination

The determination of the optimized routing paths is to be obtained by optimization, an approach typical from OR. Some principles underlying OR modeling are introduced in Appendix C.

6.5.1.1 Optimization framework

The optimization framework was developed around the CPLEX optimizer. CPLEX is an optimization engine from ILOG[14], based on applied mathematics and computer science, using the most advanced optimization technologies for solving tough business and research problems (www, 2009b). We adopted ILOG Parallel CPLEX (version 10.0) which takes advantage of multiple processors to solve difficult optimization problems, and is pointed out as an optimization option that can substantially reduce the time for solving large linear and difficult mixed integer programs (www, 2009b). An overview on some of its characteristics can be found in Appendix D.

This framework includes also the *midleware* developed for easy integration of CPLEX results into the OMNeT++ based OBS network simulator. This midleware comprises several scripts (bash/sed) and programs (perl/C++) to be executed either before or after the optimization stage. Before the optimization, this software is used to translate a network topology specification and an optimization problem into an ILP formalization tailored for the CPLEX optimizer. After the optimization the software is used to extract the global routing solution from the CPLEX results.

[14]ILOG was recently acquired by IBM for approximately $340 million USD and is now an IBM Company.

The optimization strategies proposed will be explained in detail in the next chapters of this thesis. For the size of the network topologies considered, optimal or near-optimal[15] solutions can be found in a range between few tens of seconds and some minutes, using ILOG Parallel CPLEX 10.0 in a Sun Fire V20z platform with two Opteron 850 CPUs and 2G RAM.

6.5.2 Routing path application

In this stage, the routing paths calculated in the previous stage are applied to several different OBS networks and their performance is evaluated with simulations. For this work, an OBS network simulator was specifically developed using the OMNeT++[16] platform for simulations and some programming effort in C++. Middleware (software) was also developed to read the optimized global routing solution and generate individual node-based files with the routing paths for each source node. This files are used during simulation to populate the routing tables of the OBS nodes.

6.5.2.1 The OBS network simulator

The functional architecture of our OBS model has similar characteristics to the one presented in Xiong et al. (2000); Yu et al. (2004), assuming that each node can support both the new input traffic generated by the client networks and the *in transitu* traffic passing all-optically from source to final destination. The simulation studies were done on different networks, including two from North America (ARPANET and NSFNET) one European (COST239) and one with a random configuration of nodes, whose topologies are depicted in Figure 6.5. Some physical parameters of these networks are presented in Table 6.2, in which the average degree is considered the average number of physical connections per node, and the physical connectivity is defined as the normalized number of bidirectional links

[15] In some of the proposed strategies the optimization stage was terminated after a solution has been found within 1% of the optimal value (1% cutoff, while some proposals in Teng and Rouskas (2005b) use 5% cutoff).

[16] Version 3.3 was used but, as of this writing, an improved version 4.0 is being released.

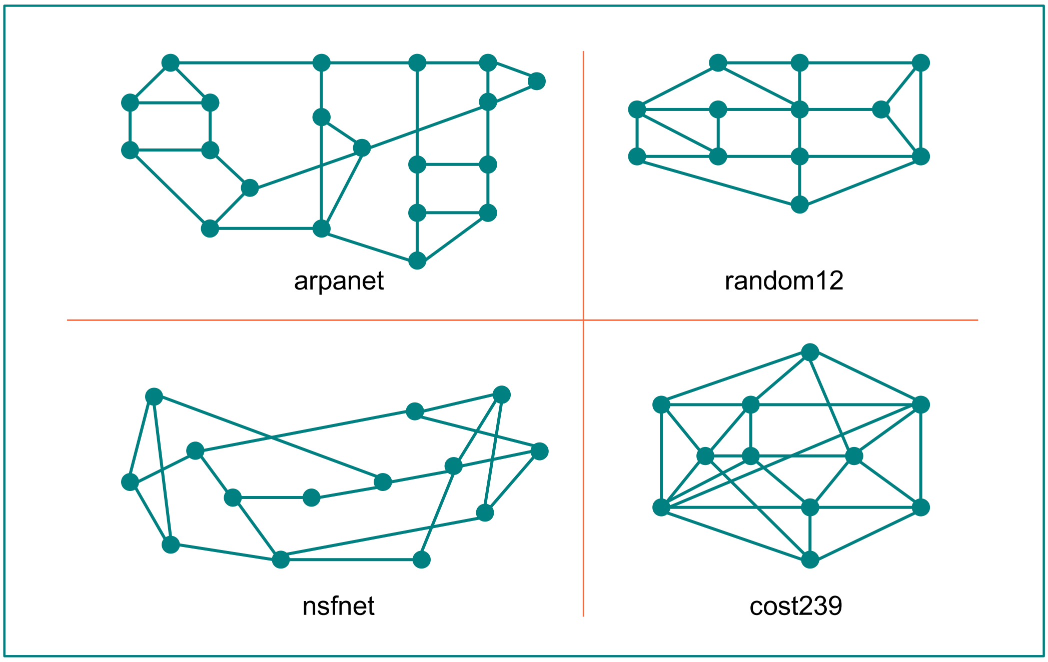

Figure 6.5: Network topologies under study.

Table 6.2: Network physical parameters

	num of	num of	degree		connec-
network	nodes	links	(av)	(stdev)	tivity
ARPANET	20	62	3.1	0.45	0.16
NSFNET	14	42	3.0	0.55	0.23
RANDOM12	12	42	3.5	0.67	0.32
COST239	11	52	4.7	0.65	0.47

with respect to a physically fully connected network of the same size (Baroni and Bayvel, 1997). The nodes are connected by links representing optical fibers with the same relevant characteristics, i.e., having 16 wavelengths per link with a transmission capacity of 10 gigabits per second (per wavelength).

This simulation model was developed to be driven in a stochastic (self-driven) way. Thus, it generates its own input data traffic, representing the load (arriving from legacy client networks) offered to the OBS network to be forwarded (again to legacy client networks). Since in this model we are mainly interested in a comparison evaluation of alternative routing strategies, the issue of packet aggregation and burst assembly is not considered. This means that the traffic used in the network is already burst based traffic after assembly.

Traffic pattern: The traffic characteristics of the OBS traffic are highly dependent on the burst assembly algorithm, an effect studied in Ge et al. (2000); Izal and Aracil (2002); Yu et al. (2004). Both theoretical and simulation results show a *smoothing effect* resultant from a reduction in the degree of self-similarity of the input packetized traffic. In Izal and Aracil (2002) it is argued that, despite of long range dependence, the burst arrival process can be assumed to be Poisson in low timescales. As a consequence, there is nearly no influence of self-similarity on blocking probability, while the influence is significant in optical buffers dimensioning. Therefore, in our approach, the adopted traffic pattern is based on a Poisson arrival pattern assuming a previous burst length threshold assembly method, generating messages that are 100×10^3 bytes.

Signaling scheme: The bursts are forwarded through the core backbone reproducing the relevant actions of the JET signaling scheme, i.e., using delayed reservation and implicit release. The CP processing time is assumed to be 10 μs and average switching time in a core node is also 10 μs, although other values from 12.5 μs to 1 μs could be adopted depending on the technology in place (current state-of-art or foreseeable in the near future). The timings assumed by this configuration parameters are in agreement to the ones widely used in OBS research, like in Vokkarane, Haridoss and Jue (2002); Gjessing and Maus (2005), for example.

Routing: The model employs source routing in which a complete routing decision is taken at the ingress edge node. Like the approach adopted in Teng and Rouskas (2005b); Yang and Rouskas (2006), the path over which the burst must travel is carried by the CP that precedes the transmission of each data burst and is not modified by downstream nodes. The adopted path is fetched from the edge nodes routing tables, previously populated by the results of the path selection strategies discussed in the next chapters.

Contention resolution: The core nodes do not employ any buffering in the data path and they do not use deflection routing. It is generally assumed that the nodes are capable of performing a full wavelength conversion. However, in the last studies presented in this thesis this feature was removed leading to an OBS network operated under strict continuity constraint requirements. Blocking occurs only if there are no free wavelengths available to accomplish the next hop on a predetermined path to a certain destination. If scheduling fails, the burst is simply dropped and no further contention resolution method is adopted.

Together with the network topology description, the OBS simulation model, which is essentially composed of OBS capable nodes interconnected by optical fibers, is based on two main compound modules (derived from OMNeT++ parent classes), namely, *edge node* and *core node*. These modules will be presented from the functional point of view in the next sections.

6.5.2.2 The edge node

Edge nodes connect multiple subnetworks running on top of legacy link layer protocols to the OBS network. They can be considered either ingress or egress nodes. When acting as ingress node, edge nodes are responsible for aggregating the incoming packets into bursts, for taking the initial (and here also permanent) routing decision, and for scheduling the bursts for transmission on outgoing channels. When acting as egress nodes, they perform an inverse operation; that is, they are responsible for disassembling bursts back into packets and for sending them to upper layers for processing. In our model, we assume the burst as the basic transport unit of interest. Hence, the issues of packet aggregation and burst assembly are considered out of scope. It is worth noting that the traffic generator developed here is a burst generator, generating messages based on a Poisson process with a symmetric all-to-all traffic matrix. Thus, whenever a Poisson process timer expires, a new burst is generated, a destination address is chosen at random between all other nodes in the network, a route to the destination node is taken from the source nodes routing table, and an initial wavelength is selected among the free ones. The model uses source-based routing, which is our first way of addressing contentions with an *a priori* action on the network space domain.

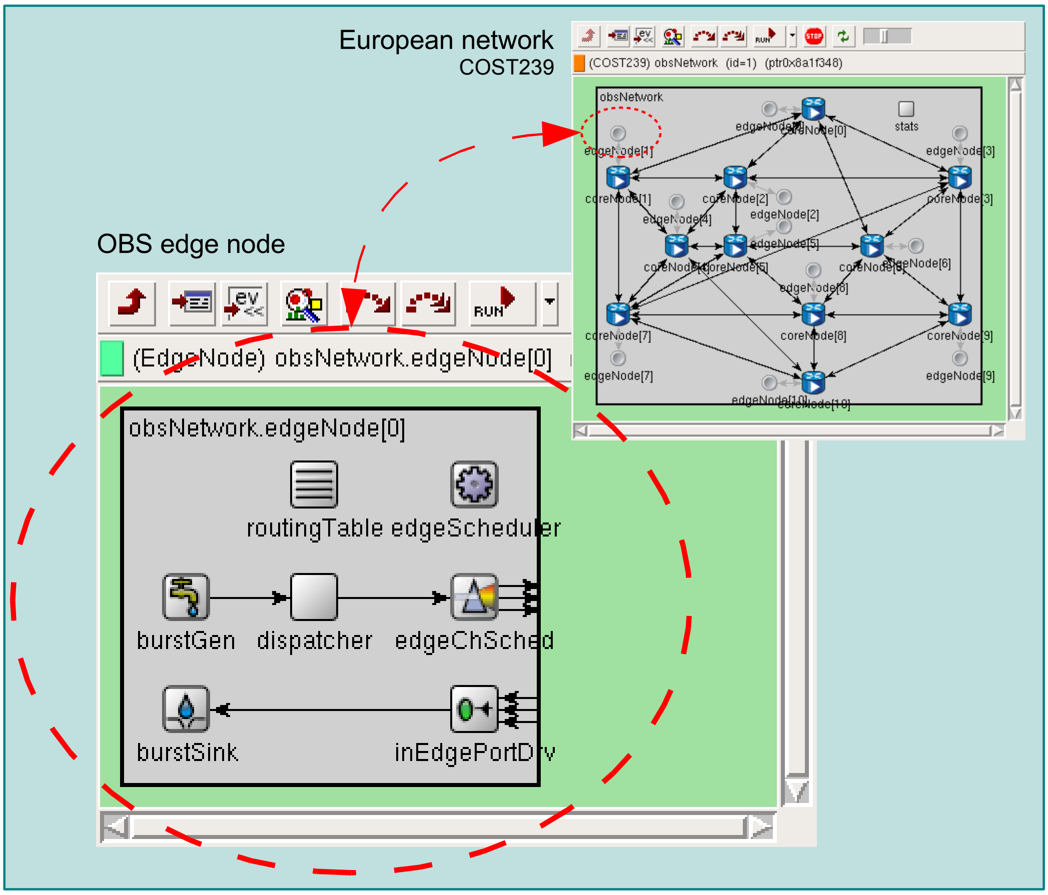

FIGURE 6.6: Screenshot of the OBS edge node (simplified version).

The burst, together with all its relevant information, is then retained in a system queue organized by destination address, and the connection-setup process starts with the sending of a CP on the appropriate dedicated channel. The CP is always transmitted before the corresponding burst and separated from it by the adequate offset time. The model calculates this offset time in order to allow the CP to be processed at each subsequent node, before the burst arrival and in such a way that an optical path can be properly reserved for burst delivery.

A simplified version of the internal structure of the edge node is shown in Figure 6.6, and its functionality is implemented by the following submodules.

RoutingTable holds the routing information of the node and all the related protected and public methods. The routing information comprises complete paths from source to destination extracted from the solution of the optimization problem, and stored into STL vectors loaded during the initialization phase of the simulation. The module takes two parameters: the path selection strategy (e.g., SP for shortest path) and the number of routes to be considered to each source destination pair.

BurstGen is responsible for the traffic that each node generates. This traffic represents already aggregated IP packets assembled into burst units on the access network. The traffic pattern and the loading factor are among the relevant parameters of this submodule.

Dispatcher initiates the signaling process and manages a system of queues where bursts are retained for a certain offset time. The offset time is calculated based on the delays introduced by the core nodes on the downstream and the number of hops along the path.

EdgeChSched is the output port driver of the edge node. This submodule transmits CPs and bursts to the core backbone after finding a free wavelength. The number of available wavelengths and the presence/absence of devices with wavelength conversion capability are among the relevant parameters of this submodule of the access network.

InEdgePortDrv receives CPs and bursts from the OBS backbone. It is the signaling end point, where CP related statistics are obtained and from where data bursts are forwarded to be disassembled.

BurstSink is the submodule that receives the data bursts and collects some data burst related statistics. This submodule represents the place where bursts are disassembled back into IP packets.

Just as an example, the routing table of each edge node, which is based on double structure of vectors from the generic STL library, is defined like this:

```
// routingTable.h

```

```
class RoutingTable : public cSimpleModule
{
  private:
    struct RouteEntry {
        std::vector<int> route;
        int numOfHops;
        bool accessed;
  };

  // container for Routing Table
  std::vector<RouteEntry> routingTable;

  public:
    // ...

  protected:
    // ...
}
```

6.5.2.3 The core node

Core nodes are responsible for processing CP reservations, for switching the bursts from an input to an output port without OEO conversion, and for handling contentions. Signaling in OBS is typically implemented using one out-of-band channel, meaning that CPs are transmitted on a different wavelength from the group of wavelengths used to transmit data bursts. This model uses λ_0 for the transmission of CPs. Several signaling schemes have been proposed for burst scheduling but JIT and JET are two of the most popular protocols using distributed signaling on OBS. These are both one-way and source-initiated signaling schemes, which means that the bursts are sent to the core network without waiting for acknowledgments regarding the success or failure of the reservation attempts. Although they are closely related, they differ in the duration of the reservations. The JIT protocol uses immediate reservation with the data channel being reserved as soon as the

CP reaches the node, while JET delays the channel reservation until the burst arrival. This technique, together with the implicit release, makes JET more efficient than JIT regarding bandwidth utilization, resulting in lower blocking rates and low end-to-end delay (Jue and Vokkarane, 2005a). For these reasons our model runs under a JET-type behavior scheme, (which can easily be converted into a JIT-type behavior).

Together with burst forwarding without leaving the optical domain, core nodes are also responsible for taking contention resolution actions. Contention occurs when multiple bursts from different sources are destined for the same output port at the same time (Chua et al., 2007a). In addition to the initial path selection strategies adopted on the edge nodes, in handling burst contentions, the core nodes are assumed by default to be equipped with devices having full wavelength conversion capability. This means that, by default, no end-to-end wavelength continuity constraint exists, and that any incoming wavelength can be shifted to any outgoing wavelength. As a result, only if there is no wavelength available on the output port the burst will be dropped without any further contention resolution action.

The internal structure of the core node is illustrated in Figure 6.7 with a simplified configuration. Its functions are implemented by the following modules.

InCorePortDrv is the entry point of the core backbone. It is an input port driver that receives both CPs and bursts from the access network and forwards them, after an increment in the number of hops counter, to the proper switch unit. This submodule is also the place from which information related to the node demands is obtained.

SwitchUnit is the submodule where switching takes place and the incoming CPs and data bursts are directed to the proper output ports towards their next hops. Although for data bursts this is done in the optical domain, for CPs OEO conversion is involved. This submodule holds an internal switching table, which is loaded during the initialization phase of the simulation and is

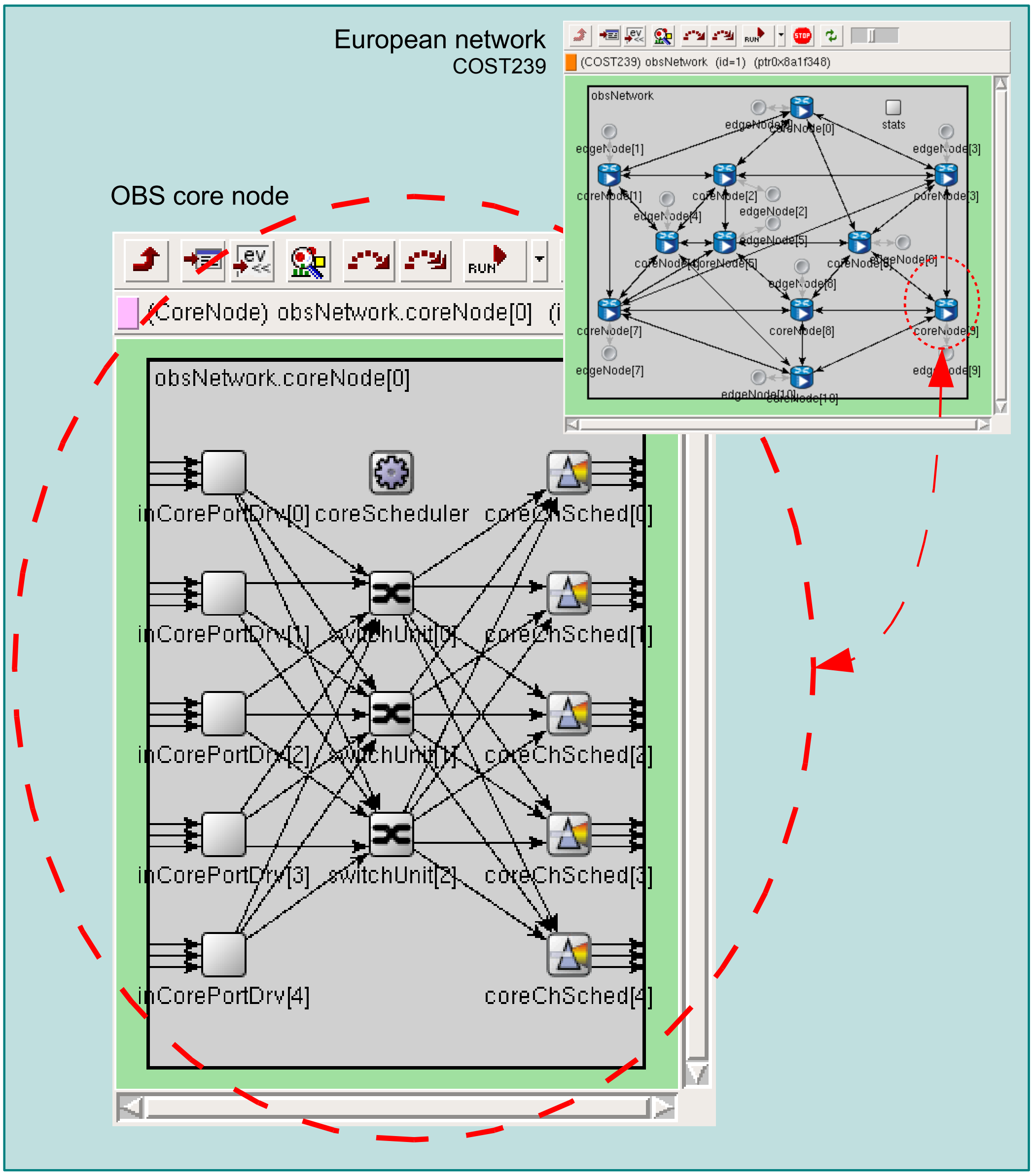

FIGURE 6.7: Screenshot of the OBS core node (simplified version).

stored in an STL map, which relates the target address with their correspondent gate identifications. The CP processing delay, which defaults to $10\mu s$, and the average switching time, which also defaults to $10\mu s$, are two important parameters of this module. Another one is the number of input/output ports of the OBS node.

CoreChSched is the output port driver of the core node. This submodule forwards CPs and bursts to the next core node of the OBS backbone or to the local edge (egress) node of the current core node. In doing this, it defaults to a scheduling scheme that follows the order of arrival (Phung, Chua, Mohan, Motani, Wong and Kong, 2005) and tries to find a free wavelength by checking an availability word of flags implemented through the use of an STL bit set. This is also the place where contention for output resources occurs and the number of drops is obtained. The total number of wavelengths and the presence/absence of wavelength converters, which are independently reconfigurable for each single port in the network, are among the relevant parameters of this submodule.

Just as an example, the global functioning of the core node can be presented in pseudocode as fallows:

```
// coreNode
if burst-W is free
    mark burst-W as busy
    calculates connection holding time
    send burst out through W-channel
    schedule holdingTime selfmessage
    hold the W-channel to be released later
else if any other W is free and W conversion is permitted
    mark new W as busy
    increment num of W changes
    update the header information
    calculates connection holding time
    send burst out through new W-channel
    schedule holdingTime selfmessage
    hold the W-channel to be released later
else, theres NO way to forward, burst will be dropped
    delete burst
    increment drop counter
```

6.6 Summary

This chapter is devoted to the simulation model definition. The chapter begins with a definition of simulation and with a presentation of some of their underlying concepts, such as model, event, entity, attribute, resource, state variable, list processing, and the definition of discrete event simulation. The advantages and disadvantages of using simulation are discussed, and some basic modeling principles presented. The important contribution of the OO paradigm for the modeling of complex systems is also addressed and some fundamental OO concepts are introduced. The chapter proceeds by presenting the simulation platform adopted in this work, and by describing how structure and behavior of a system can be modeled. Finally, our simulation methodology comprising two stages, routing determination and routing application, is presented. In the first stage, some relevant optimization issues are discussed. In the second stage, the OBS network model is described in detail.

7

MCL and MEC Pre-planned Routing Strategies

Due to its bufferless nature, OBS can be highly affected by burst contention. Several methods have been proposed to reduce burst loss due to contention, including burst scheduling, optical buffering, burst segmentation, wavelength conversion, and deflection routing. Despite the undeniable merits of the research conducted on such methods, there are also several important issues that need to be carefully considered. In fact, these contention resolution mechanism are mainly reactive techniques driven by burst contention and requiring extra hardware and/or software components at each core node, significantly increasing their cost and complexity, leading to scalability impairments. A simple and cost efficient solution to resolving contention when it occurs is to deploy the adequate mechanisms at the edge nodes to prevent contention before it happens. This approach has been followed by burst assembly mechanisms (Jin and Yang, 2007; Vokkarane, Vokkarane, Zhang, Jue and Chen, 2002; Vokkarane, Vokkarane and Jue, 2003), by path selection and wavelength assignment (Ganguly et al., 2004; Teng and Rouskas, 2005a; Zhang, Zhang, Vokkarane, Jue and Chen, 2004), and by methods which balance the traffic load between alternate paths (Gonzalez-Ortega et al., 2007; Li et al., 2005; Teng and Rouskas, 2005b; Yang and Rouskas, 2006; Zhang, Wang, Zhu, Datta, Kim and Mukherjee, 2004). Such preventive approach will also be followed by the contention avoidance proposals presented in this chapter and in Chapter 8.

7.1 Contention Avoidance Approach

Path selection mechanisms at the ingress nodes can alleviate contention as compared to shortest path (SP) routing. The comparison of new routing strategies against the SP is an evaluation method commonly adopted in OBS studies for two main reasons: first, because SP is widely used and an easy to implement strategy, and secondly, because it is usually very difficult to make accurate comparisons against other proposals of the research community. The later would result, of course, in a very useful work. However, the reported performance of path selection strategies is, in most cases, evaluated using custom developed applications, and some network operating conditions are not reported. Therefore it is difficult, sometimes impossible, to produce comparative research results (Lenkiewicz et al., 2006).

Although SP routing is successfully used in both circuit switching and packet switching networks, it often causes certain links to become congested while other links remain underutilized (Teng and Rouskas, 2005b). This is highly undesirable in bufferless OBS networks since a few highly congested links can lead to unacceptably high burst loss for the entire network. In Teng and Rouskas (2005b), for instance, a path selection for source routing was obtained using a TE approach aimed at balancing the traffic load across the network. More recently, alternate path selection mechanisms at the ingress nodes have been explored in some studies (Yang and Rouskas, 2006). In this approach, each ingress node maintains a list of alternate paths to each destination ranked according to their congestion status. The authors present a suite of path selection strategies, each utilizing a different type of information regarding the link congestion status. These adaptive and dynamic path selection schemes require a link state signaling protocol. The efficiency of a solution of this type, characterized in terms of burst blocking probability, depends on both the ability of the scheme to provide good performance for a given traffic scenario characterized by a well known statistics and the convergence time of the link state advertisement protocol. The convergence of the link state advertisement protocol is of key importance for a bursty traffic scenario.

The strategies proposed here address similar objectives but utilize only the mapping of the OBS nodes and their interconnections to make the routing decisions from all sources to all destinations in a way that congestion in the network is minimized. The problem is formulated as an ILP problem, which is a technique that is widely used to address both high-level and system-level synthesis (Mignotte and Peyran, 1997). This technique is known to be NP-hard thus, characterized as having very high computational complexity (Garey and Johnson, 1979) and mainly intended to be used offline. However, for the proposals under study, optimal or near-optimal solutions can be reached in a range between a few tens of seconds and few minutes, using regular computation power. Taking into account these computation times and the infrequent update requests expected as a consequence of changes in the OBS backbones whose topologies typically last for long time scales, this approach can be considered feasible for the real production of OBS networks. This can be carried out by means of an operation process executed during its initial setup phase, either in a centralized manner, where routes are computed offline by a central node which has knowledge about the global topology and downloaded to the nodes when the network is booted, or in a distributed manner, if the nodes are equipped with topology searching capabilities.

The routes obtained can be applied as single-path static routes and used alone to provide load-balancing without the need for resource-update signaling messages regarding the congestion status of the network links. Alternatively, they can be combined with some other dynamic contention resolution schemes (like deflection or segmentation, for example) and used occasionally as a default routing option to assume whenever the network needs to recover from instability. This can particularly be done when the activity of multiple dynamic network elements, reacting simultaneously to congestion, may result in oscillation between congestion and decongestion states on certain links (Thodime et al., 2003).

The aim of this proposal is to investigate edge-router-based routing strategies able to minimize burst contention, for a given traffic scenario, using only topological network information. We assume that the network operates with source based routing, that is, the ingress edge node selects a path for a burst that enters the network from a set of K previously calculated paths. Therefore, the strategy comprises two stages: first, the calculation of K eligible paths for each pair of nodes and second, the selection of one path from the set of K eligible paths for each pair of nodes so that the chosen paths minimize the global network contention. For the second stage, two path selection strategies to prevent congestion are proposed and evaluated, the first addressing contention on a link basis, and the second considering the entire path between source-destination node pairs.

7.2 Path Selection Strategies

In the following discussion let $\mathcal{G}(\mathcal{N}, \mathcal{L})$ be the graph representing an OBS network, where $\mathcal{N}$ is the set of optical switches, taken as nodes, and $\mathcal{L}$ represent the set of unidirectional fiber links. We define a path over which a burst must travel, v, as a connected series of directed links, written as $v : s(v) \rightarrow d(v)$, from source node $s(v)$ to destination node $d(v)$. The set of paths that can be used by a burst from s to d is defined as $\mathcal{V}_{s,d} = \{v : s(v) \rightarrow d(v) \mid s = s(v), d = d(v)\}$ and the set including all $\mathcal{V}_{s,d}$ is defined as $\mathcal{V}$. We also define $p_l^v = 1$ if link $l \in \mathcal{L}$ is included in v, $p_l^v = 0$ otherwise, and $q^{v,v'} = 1$ if the two paths v and v' share at least one link. A demand matrix $\mathbf{T}$ is considered, where $t_{s,d}$ represents a relative load from source node s to destination node d. We note that the following formulations are independent of the details of the demand model, which may include the total or average number of demands, over a period of time, or some integer value that reflects the local demand weight over the total network demand.

We propose two path selection strategies: minimizing the maximum congested link (MCL) and minimizing the maximum end-to-end congested (MEC) path. These strategies are formulated in two ILP problems for which a single path for each pair of nodes must be selected. That is, for the overall network, $N(N-1)$ paths must be chosen and resources allocated for burst delivery. For both strategies, input information includes a set of K eligible paths which are computed in advance. The optimization problems are solved in our simulation model by ILOG CPLEX optimizers, and the results obtained are used to populate the routing tables in order to achieve a global contention reduction.

7.2.1 Pre-calculation of eligible paths

The K eligible paths correspond to the K shortest paths with less links in common. That is, if several paths exist with an equal number of hops, then the more non-overlapping ones are chosen. The K shortest paths for a specific pair of nodes (s, d) are computed using the ILP optimization discussed next.

The ILP objective function is written as

$$\text{Minimize} \sum_{k \in \{0,1,\ldots,K\}} \beta_k^{s,d} - \sum_{l \in \mathcal{L}} \frac{\gamma_l^{s,d}}{|L|+1} \tag{7.1}$$

Subject to

$$\sum_{l\in\mathcal{L}:s(l)=m}\epsilon_{k,l}^{s,d}-\sum_{l\in\mathcal{L}:d(l)=m}\epsilon_{k,l}^{s,d}=\begin{cases}1, & if\ s=m\\ -1, & if\ d=m\\ 0, & otherwise\end{cases}, \tag{7.2}$$

$$\forall k\in\{0,1,...,K\},\forall m\in\mathcal{N}$$

$$\beta_k^{s,d}=\sum_{l\in\mathcal{L}}\epsilon_{k,l}^{s,d},\ \forall k\in\{0,1,...,K\} \tag{7.3}$$

$$\gamma_l^{s,d}\leq\sum_{k\in\{0,1,...,K\}}\epsilon_{k,l}^{s,d},\ \forall l\in\mathcal{L} \tag{7.4}$$

$$\psi_{k,k',l}^{s,d}\leq\epsilon_{k,l}^{s,d}+\epsilon_{k',l}^{s,d},\ \forall l\in\mathcal{L},\forall k,k'\in\{0,1,...,K\}:k\neq k' \tag{7.5}$$

$$\psi_{k,k',l}^{s,d}\leq 2-(\epsilon_{k,l}^{s,d}+\epsilon_{k',l}^{s,d}),\ \forall l\in\mathcal{L},\forall k,k'\in\{0,1,...,K\}:k\neq k' \tag{7.6}$$

$$\sum_{l\in\mathcal{L}}\psi_{k,k',l}^{s,d}\geq 1,\ \forall k,k'\in\{0,1,...,K\}:k\neq k' \tag{7.7}$$

$$\sum_{l\in\mathcal{L}:s(l)=i}\delta_{k,l}^{s,d}-\sum_{l\in\mathcal{L}:d(l)=i}\delta_{k,l}^{s,d}=\begin{cases}\beta_k^{s,d}, & if\ i=s\\ -\sum_{l:d(l)=i}\epsilon_{k,l}^{s,d}, & otherwise\end{cases}, \tag{7.8}$$

$$\forall k\in\{0,1,...,K\},\forall i\in\mathcal{N}$$

$$\delta_{k,l}^{s,d}\leq\epsilon_{k,l}^{s,d}\times L,\ \forall k\in\{0,1,...,K\},\forall l\in\mathcal{L} \tag{7.9}$$

$$\sum_{l\in\mathcal{L}:d(l)=m}\epsilon_{k,l}^{s,d}\leq 1,\ \forall k\in\{0,1,...,K\},\forall m\in\mathcal{N} \tag{7.10}$$

$$\epsilon_{k,l}^{s,d},\gamma_l^{s,d},\psi_{k,k',l}^{s,d}\in\{0,1\};\text{non-negative integer: }\beta_k^{s,d},\delta_{k,l}^{s,d} \tag{7.11}$$

The objective function has two components. The first component accounts for the total number of hops of all K paths. The second component accounts for the total number of links used by the K paths. The flow conservation constraint expressed in Eq. (7.2) builds the paths using the binary variables $\epsilon_{k,l}^{s,d}$ and Eq. (7.3) counts the number of hops, $\beta_k^{s,d}$, for each path. Eq. (7.4) determines the value of $\gamma_l^{s,d}$, a binary variable that indicates if link l is used by any path from node s to node d. Eq. (7.5), Eq. (7.6), and Eq. (7.7) prevent the K paths from being equal, that is, two paths must differ at least in one link in the network. Finally, Eq. (7.8), Eq. (7.9), and Eq. (7.10) are loop avoidance constraints.

This stage is independent of the next two path selection strategies. Also note that, instead of the ILP algorithm proposed here to calculate the K most link-disjoint shortest paths, heuristic approaches can be used. However, the most

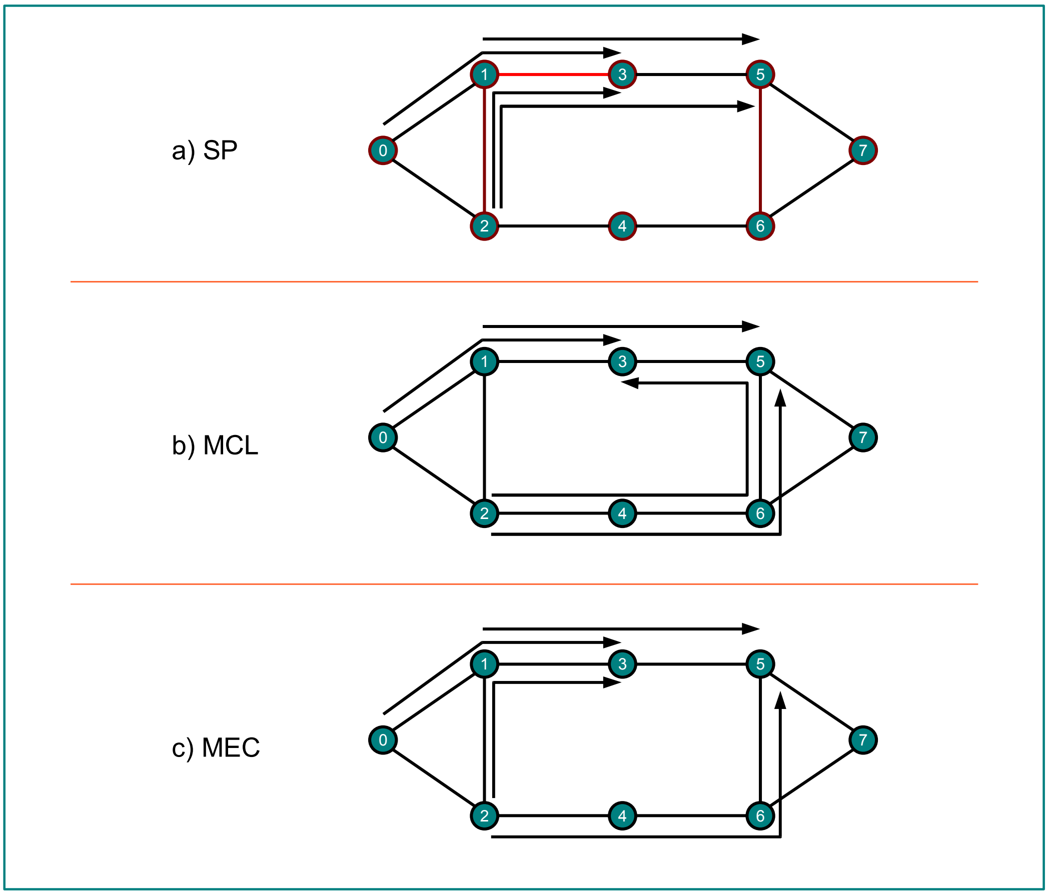

FIGURE 7.1: A small illustrative example of SP, MCL, and MEC.

common heuristics used to compute the shortest path do not fulfill the previously mentioned requirements.

7.2.2 MCL path selection strategy

This strategy is based on the idea that the more a certain link is included in the chosen paths for source-destination pairs, the highest the blocking probability will be. This situation is represented on the small network of Figure 7.1(a) where, considering the paths between nodes 0→3, 1→5, 2→3 and 2→5, the SP algorithm can bring excessive load to the link 1–3 leaving other links underutilized. Therefore, paths for source-destination pairs should be selected with the objective of minimizing the blocking probability of the link with highest expected contention value, denoted by ζ_{MAX}. This is achieved by the following ILP optimization problem:

$$\text{Minimize } \zeta_{MAX} \tag{7.12}$$

Subject to

$$\textstyle\sum_{v\in\mathcal{V}_{s,d}} \sigma^v = 1,\ \forall s,d \in \mathcal{N} \tag{7.13}$$

$$\textstyle\sum_{s,d}\sum_{v\in\mathcal{V}_{s,d}} \sigma^v \times p_l^v \times t_{s,d} \leq \zeta_{MAX},\ \forall l \in \ \mathcal{L} \tag{7.14}$$

$$\sigma^v \in \{0,1\};\text{non-negative integer: } \zeta_{MAX} \tag{7.15}$$

where σ^v is a binary variable that indicates if v is used to carry bursts from node $s(v)$ to node $d(v)$. Eq. (7.13) is a constraint stating that one path must be found for each pair of nodes. Each path is selected from the corresponding set $\mathcal{V}_{s,d}$ of available paths. Eq. (7.14) states that the expected congestion at a link must not exceed ζ_{MAX}.

With this algorithm, and for the same scenario presented, we can see in Figure 7.1(b) that the traffic is now more evenly distributed, increasing the network utilization to avoid the highly congested situation presented in Figure 7.1(a). It should be noticed, however, that on this attempt to spread the traffic throughout the network some paths can became longer (like the case of path 2→3). This observation was the driving force to the next strategy.

7.2.3 MEC path selection strategy

This strategy is based on the idea that blocking may occur at any link traversed by a burst along the path. Therefore, paths for source-destination pairs should be selected so that demands have the smallest probability of contending with other demands at every link from source to destination, minimizing the end-to-end blocking. This is achieved by the following ILP optimization problem, where φ_{MAX} denotes the value of the path having the highest number of contends.

$$\text{Minimize } \varphi_{MAX} \tag{7.16}$$

Subject to

$$\textstyle\sum_{v\in\mathcal{V}_{s,d}} \sigma^v = 1,\ \forall s,d \in \mathcal{N} \tag{7.17}$$

$$\eta^{v,v'} \geq (\sigma^v + \sigma^{v'} - 1) \times q^{v,v'},\ \forall v \in \mathcal{V}, \forall v' \in \mathcal{V} \setminus \mathcal{V}_{s(v),d(v)} \tag{7.18}$$

$$t_{s,d} + \textstyle\sum_{v\in\mathcal{V}_{s,d}} \sum_{v'\in\mathcal{V}\setminus\mathcal{V}_{s,d}} \eta^{v,v'} \times t_{s(v'),d(v')} \leq \varphi_{MAX}, \forall s,d \in \mathcal{N} \tag{7.19}$$

$$\sigma^v, \eta^{v,v'} \in \{0,1\}; \text{non-negative integer: } \varphi_{MAX} \tag{7.20}$$

where σ^v is a binary variable that indicates if v is used to carry bursts from node $s(v)$ to node $d(v)$, and $\eta^{v,v'}$ is a binary variable that indicates if v and v' have both been selected to carry bursts and share at least one link. Similarly to the previous strategy, the constraint expressed on Eq. (7.17) states that one path must be found for each pair of nodes. Eq. (7.18) forces $\eta^{v,v'}$ to be 1 if v and v' share a link and have both been selected to carry bursts. Otherwise, and due to the minimizing nature of the objective function, $\eta^{v,v'}$ will be 0. Eq. (7.19) states that the contending value of a source-destination pair must not exceed φ_{MAX}.

Still considering the same network scenario, we can observe in Figure 7.1(c) that the longest path 2→3 in Figure 7.1(b) no longer exists and that this algorithm also tries to keep the traffic in a distributed manner.

7.3 Evaluation

In this section, we provide simulation results and compare the performance of our proposed path selection schemes with the SP approach. For the four networks represented in Figure 6.5, simulations were done under similar conditions with regard to the total number of bursts generated per source node (1×10^6), arrival pattern,

Table 7.1: Average number of hops

	SP		MCL		MEC	
network	av	stdev	av	stdev	av	stdev
ARPANET	2.71	1,14	2,76	1.18	2.80	1.21
NSFNET	2.13	0.75	2.32	0.93	2.17	0.79
RANDOM12	1.92	0.73	2.02	0.78	1.92	0.73
COST239	1.56	0.55	1.65	0.57	1.59	0.56

traffic load variation, and the number of shortest paths (per pair of nodes) being provided to the path selection strategies (assuming K=2,K=3, and K=4). Figure 7.2 presents the average number of burst drops for all the loads considered and for the different path selection schemes. The values are normalized to the number of bursts that enter the network. As the figure shows, in the majority of the cases, the proposed algorithms behave better than the SP. From the 24 simulations done using MCL or MEC routing paths, 20 present lower than average dropping values, the best of which are highlighted with a star. From these results, it is also possible to conclude that, generally, when K increases, there is a certain degradation of performance, indicating that the algorithms do not benefit from the alternative paths given as input. This means that when longer paths are adopted, this can become a disadvantage resulting in less gain. This was expected since burst scheduling is required at each intermediate node, and longer paths determined by higher values of K also correspond to a greater possibility of contention. The statistical values of Table 7.1 support this conclusion where, for instance, network topologies with lower connectivity present higher numbers of average hops and higher standard deviations when compared with the more connected ones. The exception to this is the case of the COST239 network under the MCL with K=4. This is probably because its high connectivity makes it possible to obtain alternative paths with a similar number of hops.

The best results were obtained with each algorithm for each network as follows: for ARPANET, MCL with K=2 and MEC with K=2; for NSFNET, MCL with K=3 and MEC with K=2; for RANDOM12, MCL with K=2 and MEC with K=2; and for COST239, MCL with K=4 and MEC with K=2. The following evaluation will compare these best cases with the SP approach. In each of the 28 simulations

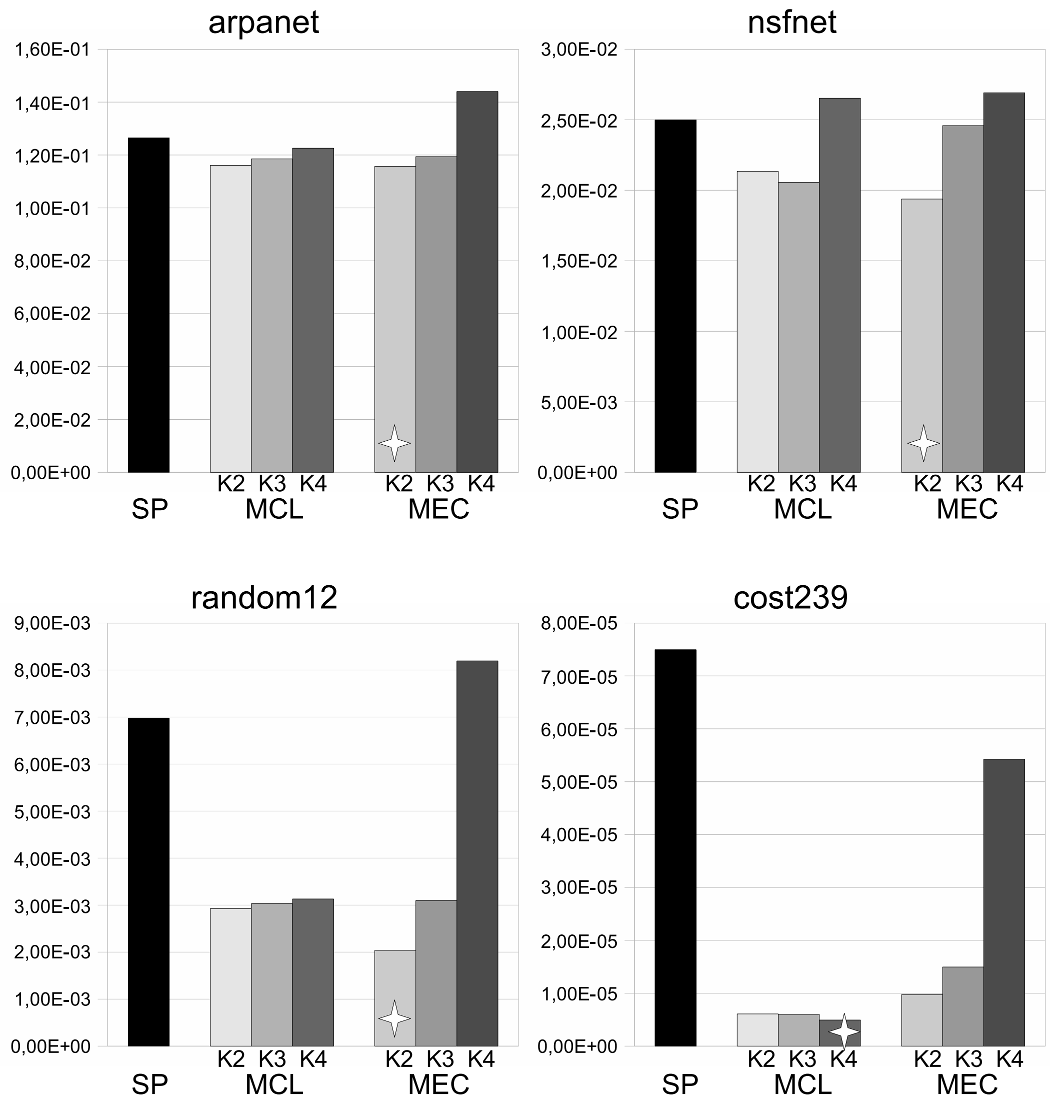

FIGURE 7.2: Normalized burst loss on the OBS backbone (K2, K3 and K4 denote the number of eligible SPs provided to MCL and MEC. The star indicates the best result for the network).

performed, 15 runs were executed with gradually increasing traffic loads. Figures 7.3, 7.4, 7.5, and 7.6 present the results of these runs, plotting the normalized burst loss against the average load. From these graphics it is possible to see that the SP curves are in agreement with the ones generally presented in literature (Teng and Rouskas, 2005b) and that the values of the burst loss for MCL and MEC are always lower than those for SP routing. Moreover, the burst loss improvement increases for both MCL and MEC from Figure 7.3 to Figure 7.6. Considering that these graphics are presented in ascending order of physical connectivity and the same amount of traffic is being generated at each node with the same arrival pattern,

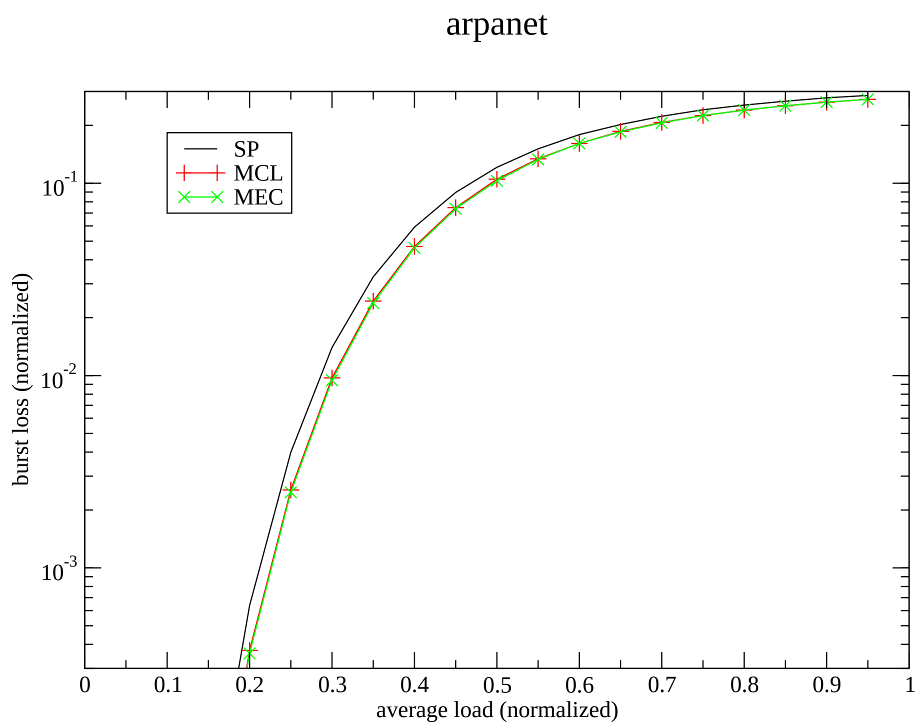

FIGURE 7.3: Proposed schemes versus SP for ARPANET.

we assume that the gain of the proposed algorithms is also related to the way links are connected in OBS networks. This effect corroborates the relevant work in Rodrigues, Freire and Lorenz (2005) where the influence of nodal degree on the performance of OBS networks is analyzed. This means that MCL and MEC are more efficient in highly connected networks, and this explains why they provide better results for the COST239 topology than for the ARPANET topology. The gain of these algorithms is graphically presented in Figure 7.7, where the reduction in burst dropping is easily seen, showing that the MCL and MEC routing clearly outperform the results achieved with the SP.

7.4 Summary

In this chapter, we considered the problem of routing path selection in OBS networks to minimize the overall burst loss. We took a traffic engineering approach which substantially reduces contention using only topological information. We

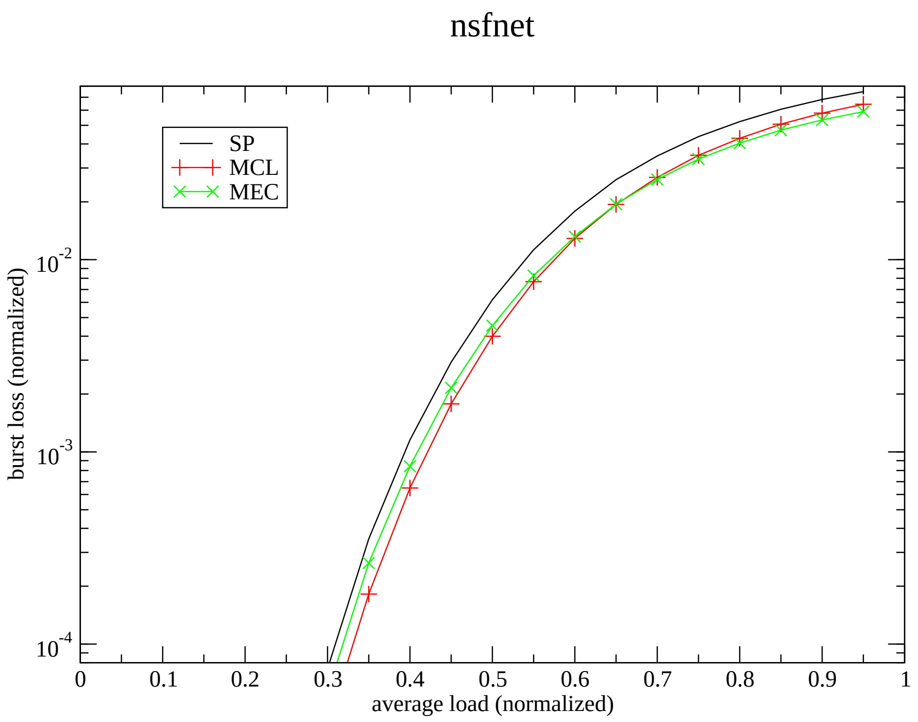

FIGURE 7.4: Proposed schemes versus SP for NSFNET.

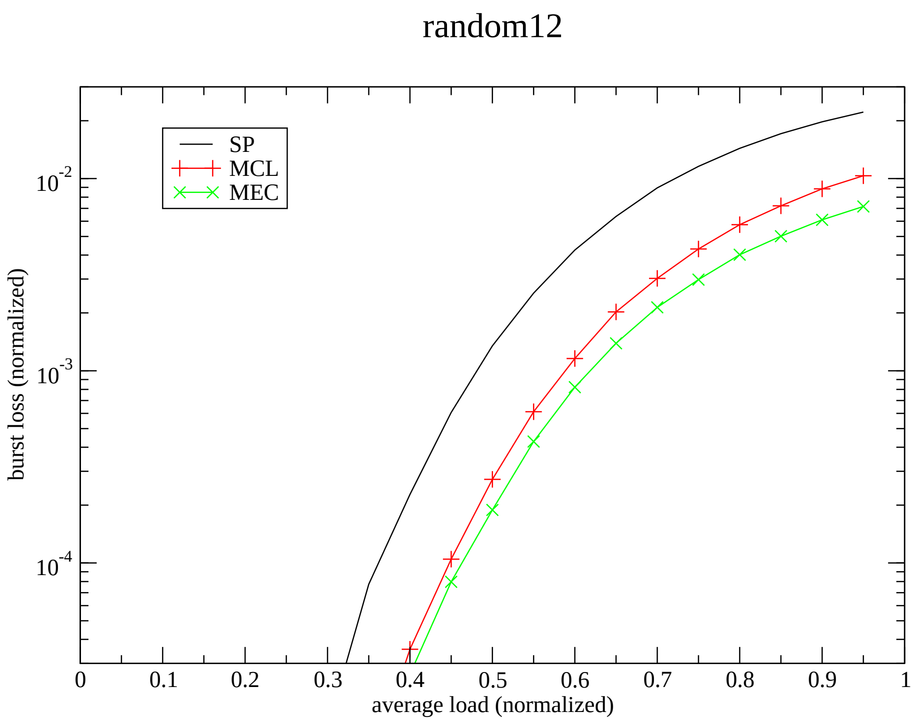

FIGURE 7.5: Proposed schemes versus SP for RANDOM12.

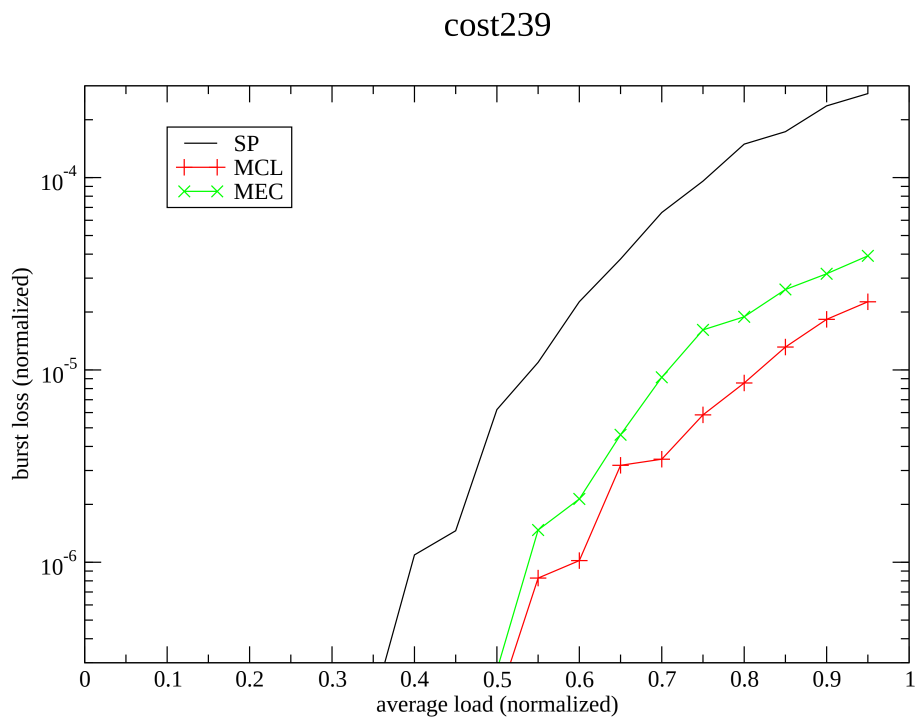

FIGURE 7.6: Proposed schemes versus SP for COST239.

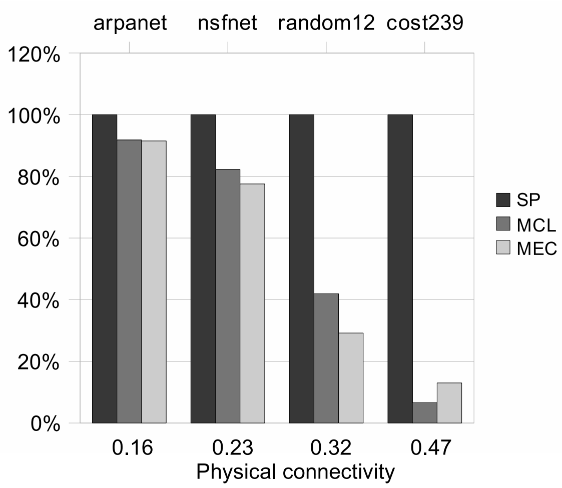

FIGURE 7.7: Average gain in burst loss reduction.

demonstrated that data burst loss can be minimized by appropriately choosing the paths that bursts must follow. That is, an effective choice of paths can lead to an overall network performance improvement. Although the resolution of an ILP problem is involved in the process, since eligible paths are provided as input for the path selection strategies, the solutions can be promptly reached, meaning that they can admit updates if bounded by the reconfiguration requirements of the optical backbones, whose topologies typically last for long time scales. Our results also show that the achieved performance improvement depends on the physical connectivity of the network. More highly connected networks show better performance. This happens because the proposed algorithms take advantage of more, short, alternative paths. Our approach achieves an initial stage of improved performance, measured in terms of burst loss reduction without incurring state dissemination protocol penalties. However, it should be noted that other contention avoidance strategies, including dynamic resolution schemes, can be combined with this offline load balancing methodology.

8

SBPR-PP and SBPR-nPP Pre-planned Routing Strategies

In this chapter we propose and evaluate four contention avoidance strategies which take into account a recently reported phenomenon unique to OBS networks called *streamline effect* (Phung, Chua, Mohan, Motani and Wong, 2005). Most of the methods proposed in literature to address the problem of contention in OBS networks do not consider this phenomenon. In this work we present a TE approach for path selection with the objective of minimize contention considering the streamline effect and using only topological information. The main idea is to balance the traffic across the network in order to prevent congestion and without incurring into link state dissemination penalties. The first two path selection strategies use the streamline effect in networks with full wavelength conversion capability, while the last two rely on networks with imposed continuity constraint. This approach assumes contention at a certain link to be strongly dependent on the number of input streams to the link, in particular when core nodes without converters preclude contention resolution through wavelength changes. Results show that our strategies outperform the traditionally used SP routing while keeping simple and cost-effective the architecture of the core nodes.

8.1 Problem Definition

Although successfully used in both circuit switching and packet switching networks, SP routing does not take into consideration the traffic load offered to the network, and it often causes certain links to become congested while other links remain underutilized (Teng and Rouskas, 2005b). This is highly undesirable in bufferless OBS networks, since a few highly congested links can lead to unacceptably high burst loss values for the entire network, and corroborates load balancing as an important TE feature in the OBS field. Besides the SP routing, some of the congestion avoidance proposals presented in literature also rely on probabilistic predictions based on the M/M/k/k queueing model, assuming an infinite number of incoming sources without considering network connectivity. Such model does not reflect the operation of real nodes with a limited number of fiber links and wavelength channels, and does not consider the streamline effect, a phenomenon unique to OBS networks, described and analyzed in Phung et al. (2006), already presented in Chapter 5 and briefly described bellow.

The approach presented here considers that phenomenon and assumes the following *a priori* advantages: no extra hardware or software components are required on the core nodes and no network flooding with signaling messages resulting from (over)active link state update protocols. Moreover, with this approach there is also no place for out-of-order arrivals, a disadvantage of multipath routing schemes typically requiring large memories at the edge nodes for re-ordering operations.

These are distinguishing qualities to make the architecture of the OBS nodes less complex, contributing to reduce both their cost and scalability impairments.

These strategies are also formulated as ILP problems, which means that the assumptions and considerations already made in Chapters 6 and 7 are also valid here. For the strategies proposed, and for the size of the network topologies considered, optimal or 'near optimal' solutions can be found in a range between few tens of seconds and some minutes, using our ILOG CPLEX based simulation model. Taking into account the computation times involved and the relatively infrequent update requests expected from changes in the OBS backbones whose topologies typically last for long time scales, this approach can also be considered feasible for the real production of OBS networks by means of an operation process to be executed during the initial setup phase, either in a centralized or in a distributed manner. The routes obtained can be used alone, as single-path static routes to provide load-balancing without the need for additional control messages with regard to link status, or combined with other dynamic contention resolution schemes (deflection or segmentation, for example) and used occasionally as a default routing to assume whenever the network needs to recover from instability, favoring the network resilience.

For each of the two strategies proposed in this chapter (with/without wavelength converters) two approaches are followed: one considering that the final path is chosen from a set of K different eligible paths previously calculated, and another considering that the final path is entirely determined by the algorithm.

8.2 Contention Avoidance Strategies

The streamline effect is the phenomenon unique to OBS networks wherein bursts traveling in a common link are streamlined and do not contend with each other until they diverge (Phung et al., 2006). This happens because OBS networks are bufferless and once contentions are resolved at the first link where they merge, no further contentions will happen among them thereafter. However, like was depicted in Figure 5.1, those bursts may still contend with other bursts that merge at downstream nodes (Phung et al., 2006). This phenomenon has impact in load balancing and, for path selection approaches, must be considered differently depending on the presence/absence of wavelength conversion capability on the network nodes and the type of potential joining points obtained from the presence/absence of that feature. This chapter starts by assuming full wavelength conversion. Therefore, routes must be chosen taking into account that, on a contending situation, the wavelength of the contending burst may change at any hop. Therefore, at any particular node, all the arriving bursts can contend with each other regardless of their incoming links, resulting in a somewhat more flexible streamline effect where a global potential joining point exists on each core node. In this case, our approach considers that routes must be chosen with the objective of minimize "load" differences at the output links. In fact, by dividing the total number of incoming links by the number of output links, a lower value of input links per output link is obtained, naturally favoring the creation of a lower number of inbound paths going to the same output link, while favoring the creation of widely dispersed paths at downstream. This is the principle behind the *Streamline Based Pre-planned Routing* (SBPR) strategy presented next, first using K pre-determined paths for each pair of nodes (SBPR-PP), and secondly with no pre-determined path requirements (SBPR-nPP).

8.2.1 SBPR-PP path selection strategy

In the following discussion let $\mathcal{G}(\mathcal{N}, \mathcal{L})$ be a network graph, where $\mathcal{N}$ is the set of nodes and $\mathcal{L}$ is the set of links, and let us define a path over which a burst must travel, v, as a connected series of directed links, written as $v : s(v) \rightarrow d(v)$, from source node $s(v)$ to destination node $d(v)$. The set of paths that can be used by a burst from s to d is defined as $\mathcal{V}_{s,d} = \{v : s(v) \rightarrow d(v) \mid s = s(v), d = d(v)\}$ and the set including all $\mathcal{V}_{s,d}$ is defined as $\mathcal{V}$. We also define $p_l^v = 1$ if link $l \in \mathcal{L}$ is included in v, $p_l^v = 0$ otherwise. When taking a specific node as a reference point, a link coming out of that node can be denoted as l_{out} and a link coming into that node can be denoted as l_{in}. The total number of elements in $\mathcal{N}$, $\mathcal{L}$ and $\mathcal{V}$ is denoted by N, L and V. In the following formulations the demand traffic matrix is ignored.

The main goal of the next objective function is to minimize ζ_{MAX}, a global bound that stores the highest number of routes competing for an output link. The second component of the objective function is a secondary goal so that routes having a small number of hops are chosen if multiple solutions exist for a given ζ_{MAX} value.

$$\text{Minimize } \zeta_{MAX} + \frac{1}{LV} \sum_{v \in \mathcal{V}} \sum_{l \in \mathcal{L}} \sigma^v \times p_l^v \tag{8.1}$$

Subject to

$$\textstyle\sum_{v \in \mathcal{V}_{s,d}} \sigma^v = 1, \ \forall s, d \in \mathcal{N} \tag{8.2}$$

$$\zeta_i \geq \textstyle\sum_{s,d \in \mathcal{N}} \sum_{v \in \mathcal{V}_{s,d}} \sigma^v \times p_{l_{out}}^v, \ \forall i \in \mathcal{N}, \forall l_{out} \in \mathcal{L} : s(l_{out}) = i \tag{8.3}$$

$$\zeta_{MAX} \geq \zeta_i, \ \forall i \in \mathcal{N} \tag{8.4}$$

$$\sigma^v \in \{0, 1\}; \text{non-negative integer: } \zeta_i, \zeta_{MAX} \tag{8.5}$$

where σ^v is a binary variable that indicates if v is used to carry bursts from node $s(v)$ to node $d(v)$. Constraint 8.2 states that one path must be found for each pair of nodes. Each path is selected from the corresponding set $\mathcal{V}_{s,d}$ of available paths. For a specific node i, constraint 8.3 stores in ζ_i the highest number of selected routes competing for an output link of node i. Note that an attempt will be made to evenly distribute the routes among all output links in order to keep ζ_i low. Constraint 8.4 stores in ζ_{MAX} the highest value from all ζ_i. A global bound is being stored in ζ_{MAX}, the variable that will be minimized.

8.2.2 SBPR-nPP path selection strategy

Let $\mathcal{G}(\mathcal{N},\mathcal{L})$ be a network graph, where $\mathcal{N}$ is the set of nodes and $\mathcal{L}$ is the set of links. Now a route must be found for each s-d pair of nodes for burst delivery, meaning that no pre-determined routes for selection exist. Similar to the previous formulation, the main goal is to minimize ζ_{MAX}, the global bound, while the secondary goal is to find routes having a small number of hops. This secondary goal also avoids loops.

$$\text{Minimize } \zeta_{MAX} + \frac{1}{LN(N-1)} \sum_{s,d\in\mathcal{N}} \sum_{l\in\mathcal{L}} \theta_l^{s,d} \tag{8.6}$$

Subject to

$$\textstyle\sum_{l_{out}\in\mathcal{L}:s(l_{out})=i} \theta_{l_{out}}^{s,d} - \sum_{l_{in}\in\mathcal{L}:d(l_{in})=i} \theta_{l_{in}}^{s,d} = \begin{cases} 1, & if\ i=s \\ -1, & if\ i=d \\ 0, & otherwise \end{cases}, \quad \forall s,d\in\mathcal{N}, \forall i\in\mathcal{N} \tag{8.7}$$

$$\zeta_i \geq \textstyle\sum_{s,d\in\mathcal{N}} \theta_{l_{out}}^{s,d},\ \forall i\in\mathcal{N}, \forall l_{out}\in\mathcal{L}: s(l_{out})=i \tag{8.8}$$

$$\zeta_{MAX} \geq \zeta_i,\ \forall i\in\mathcal{N} \tag{8.9}$$

$$\theta_l^{s,d} \in \{0,1\}; \text{non-negative integer: } \zeta_i, \zeta_{MAX} \tag{8.10}$$

where $\theta_l^{s,d}$ is a binary variable that indicates if the route for the s-d pair of nodes, used to carry bursts, uses link l. Constraint 8.7 guarantees flow conservation of every route. For a specific node i, constraint 8.8 stores in ζ_i the highest number of selected routes competing for an output link, similarly to SBPR-PP, and constraint 8.9 stores in ζ_{MAX} the highest value from all ζ_i. The main goal of the objective function is to minimize ζ_{MAX}, the global bound.

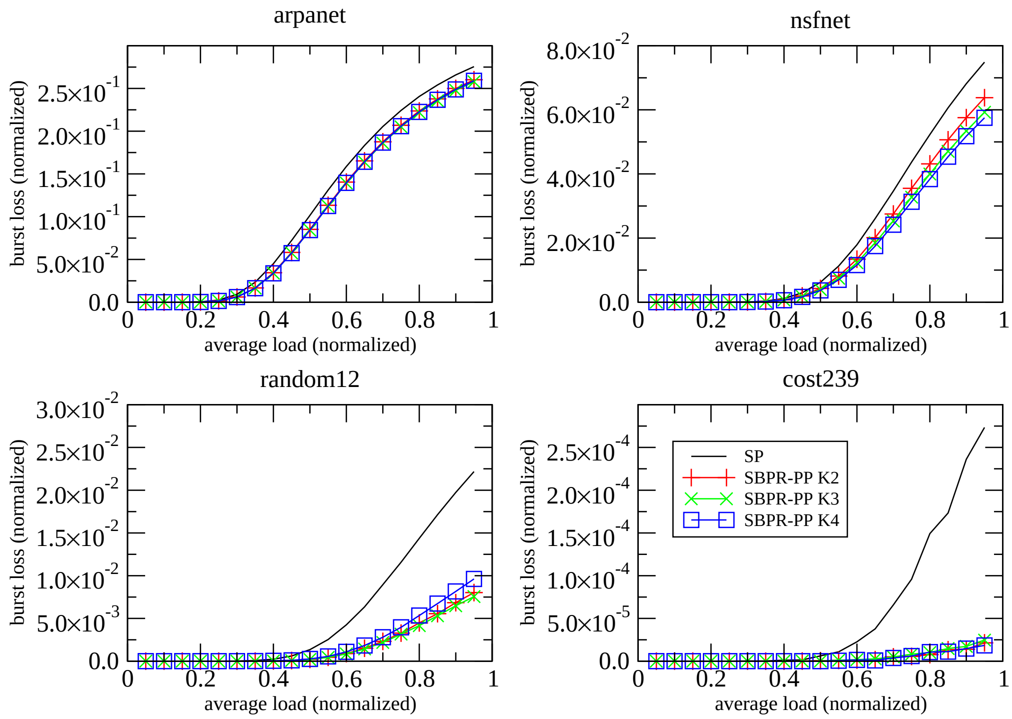

FIGURE 8.1: Burst loss on the OBS backbone (legend of cost239 applies for all networks; K2, K3 and K4 denote the number of eligible SPs provided to SBPR-PP).

8.2.3 Evaluation

Simulations were performed for the same four network topologies represented on Figure 6.5, ARPANET, NSFNET, RANDOM12 and COST239, whose physical parameters were summarized in Table 6.2. For each of these networks, simulations were done with wavelength converters installed. All simulations assume similar conditions with regard to the total amount of bursts generated per source node (10^6), arrival patterns and load variations. When pre-determined paths are required, the strategies adopt the K most link-disjoint shortest paths calculated like in Chapter 7, and assuming the values 2, 3 and 4 for the K alternative eligible paths.

Simulation results to compare the performance of the proposed path selection schemes with the SP routing are plotted first in Figure 8.1 with burst loss values

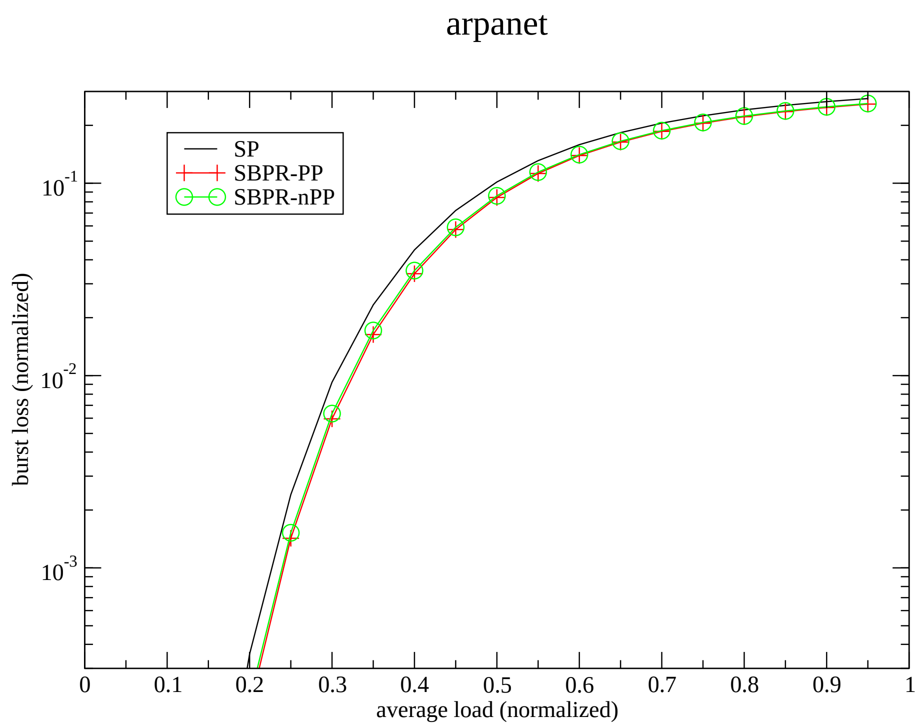

FIGURE 8.2: SBPR-nPP and best SBPR-PP versus SP for ARPANET.

in linear scale, in order to give an idea on the differences and similarities between the SP performance and the three alternatives of the SBPR-PP strategy (with K=2, 3, and 4) for all the networks. These graphics show that the values of burst loss for SBPR-PP are considerable lower than the SP routing, and that the differences among runs with different values of K are generally not relevant.

More detailed burst loss results normalized to the number of bursts that enter the network backbone against the average load are plotted in Figures 8.2 to 8.5. These graphics depict the performance of the best SBPR-PP and the performance of SBPR-nPP compared to the one of SP. The values of burst loss for SBPR-PP and SBPR-nPP are always lower than the SP routing. This means that balancing the potential path-load at the output links when the node is considered like a global joining point can reduce burst loss when compared to the SP. The graphics from ARPANET, NSFNET, RANDOM12 and COST239 (see Figures 8.2 to 8.5) show progressively larger advantage for our algorithms because these last topologies present higher connectivity, which makes it possible to obtain alternative

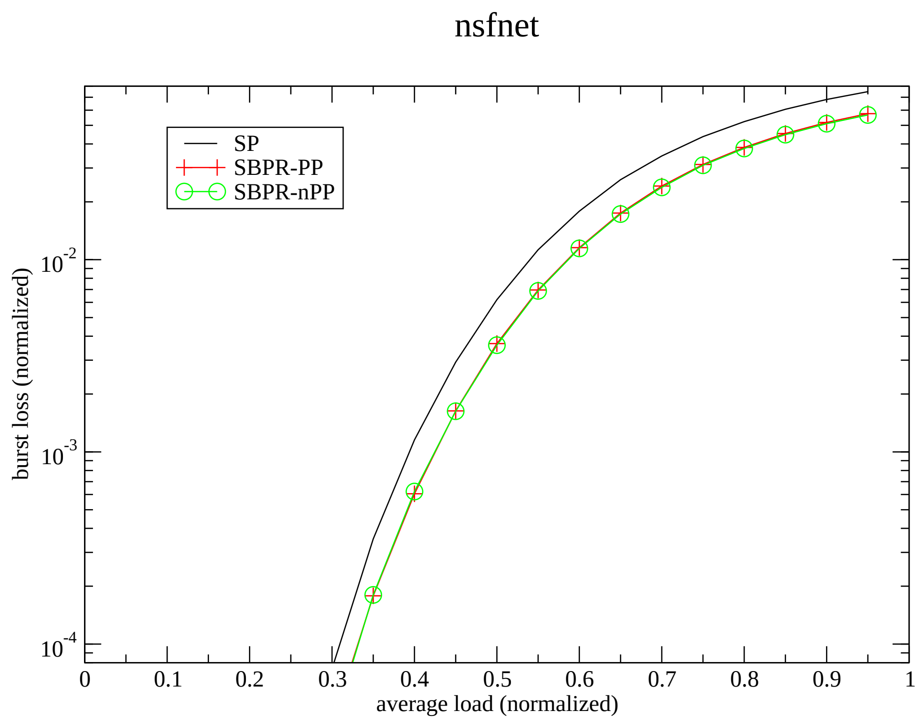

FIGURE 8.3: SBPR-nPP and best SBPR-PP versus SP for NSFNET.

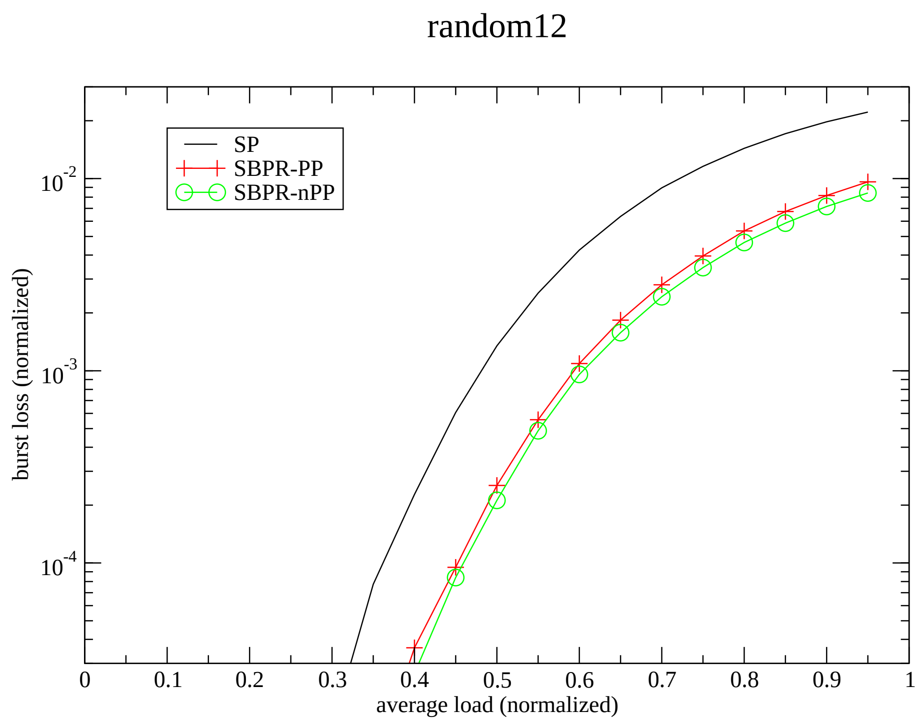

FIGURE 8.4: SBPR-nPP and best SBPR-PP versus SP for RANDOM12.

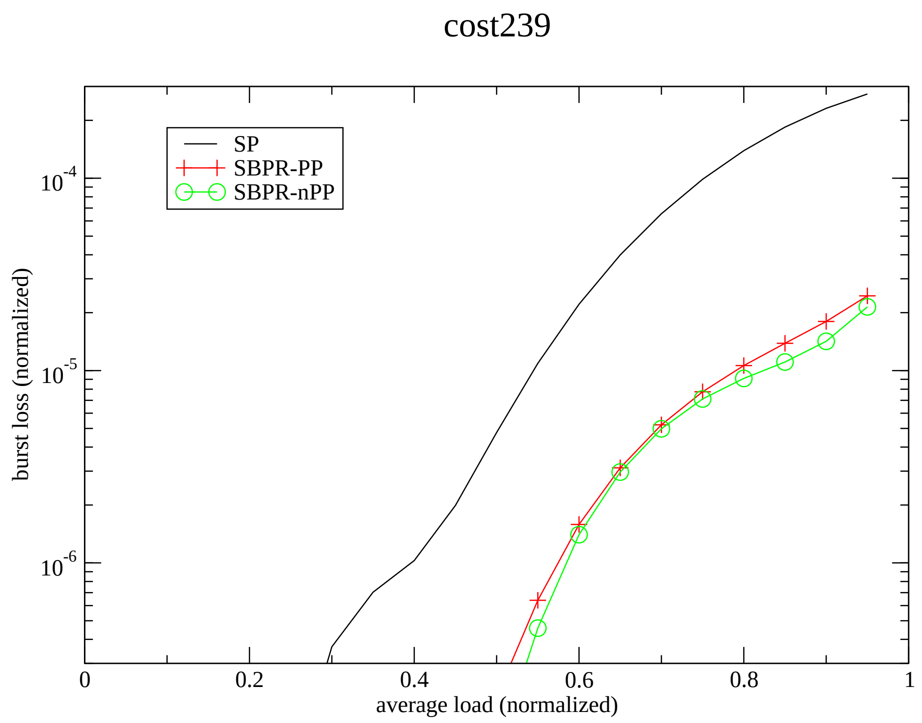

FIGURE 8.5: SBPR-nPP and best SBPR-PP versus SP for COST239.

paths to spread the potential traffic without extending them too much in terms of hop count. But even the ARPANET (see Figure 8.2), despite its low connected topology, can have a small benefit from this approach. Results of this study also show that by using the SBPR-PP algorithm, which is computationally less costly, it is possible to achieve performance results that are quite near the ones obtained by the most costly SBPR-nPP strategy, whose results seem to act like a lower bound.

Figure 8.6 compares the best SBPR result obtained for the COST239 network topology with the best performances obtained with the two previously presented strategies MCL and MEC. The two bottom lines intersect each other near an average load value of 0.65, which means that there is not one unique best strategy for pre-planned routing on this network. In fact, bellow that value MCL is the strategy presenting lower burst loss, while above that value SBPR performs better. This comparison is made only for the European COST239 network because its topology provides the highest (more spread) performance gains.

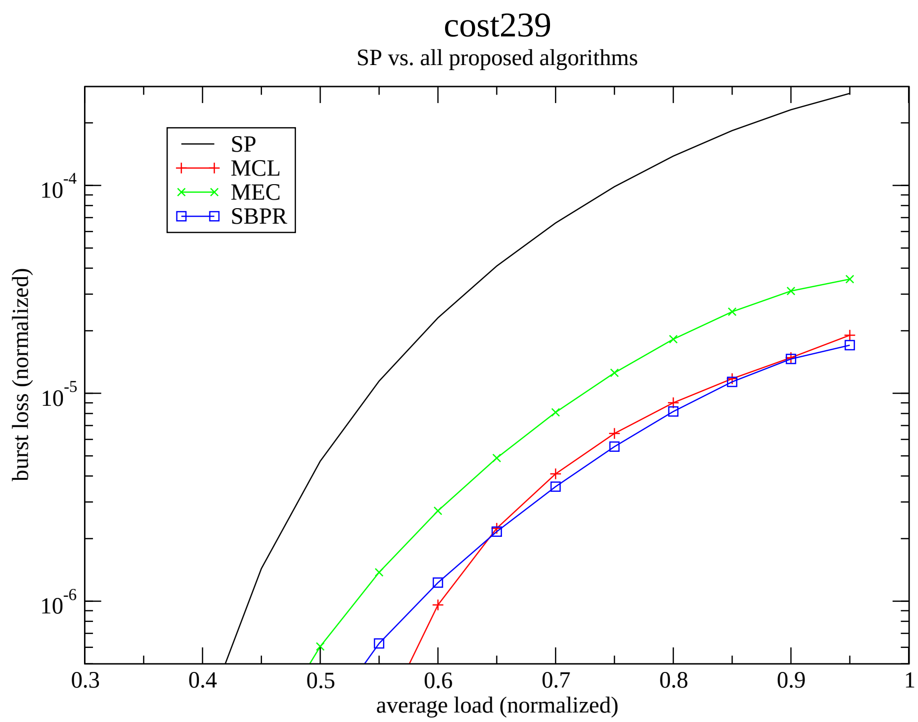

FIGURE 8.6: Best of the three proposed strategies versus SP for European network.

8.3 Contention Avoidance Strategies with Continuity Constraint

This streamline effect has impact in burst contention, in particular when continuity constraint must be respected due to the absence of wavelength converters. Without wavelength converters, successful burst delivery requires not only the opportunity to reserve a wavelength, but to have the same wavelength available in all hops from ingress to egress. In agreement with the above streamline assumptions, there is a potential joining point whenever paths arriving from different links flow to the same output link. Therefore, routes must be chosen in such a way that, for a given node, paths leaving on a certain link have arrived from a number of upstream links as small as possible. This is the basis for the SBPR strategies presented here, whose algorithms are defined next, first with final paths entirely determined by the algorithm (SBPR-nPP), and secondly, considering that the final path is chosen from a set of K eligible paths previously determined (SBPR-PP).

In the following discussion let $\mathcal{G}(\mathcal{N}, \mathcal{L})$ be a network graph, where $\mathcal{N}$ is the set of nodes and $\mathcal{L}$ is the set of links, and let us define a path over which a burst must travel, v, as a connected series of directed links, written as $v : s(v) \rightarrow d(v)$, from source node $s(v)$ to destination node $d(v)$. The set of paths that can be used by a burst from s to d is defined as $\mathcal{V}_{s,d} = \{v : s(v) \rightarrow d(v) \mid s = s(v), d = d(v)\}$ and the set including all $\mathcal{V}_{s,d}$ is defined as $\mathcal{V}$. We also define $p_l^v = 1$ if link $l \in \mathcal{L}$ is included in v, $p_l^v = 0$ otherwise. When taking a specific node as a reference point, a link coming out of that node can be denoted as l_{out} and a link coming into that node can be denoted as l_{in}. We denote by δ the average number of hops when using the shortest paths, and by κ a multiplication factor that indicates how far from δ the average number of hops of the solution can go. The total number of elements in $\mathcal{N}$, $\mathcal{L}$ and $\mathcal{V}$ is denoted by N, L and V. In the following formulations the demand traffic matrix is ignored. The main goal of the next two objective functions is to minimize ζ_{MAX}, a global bound that stores the highest number of paths competing for an output link.

8.3.1 SBPR-nPP path selection strategy

$$\text{Minimize } \zeta_{MAX} \tag{8.11}$$

Subject to

$$\sum_{l_{out} \in \mathcal{L}: s(l_{out}) = i} \theta_{l_{out}}^{s,d} - \sum_{l_{in} \in \mathcal{L}: d(l_{in}) = i} \theta_{l_{in}}^{s,d} =$$

$$= \begin{cases} 1, & if\ i = s \\ -1, & if\ i = d \\ 0, & otherwise \end{cases} \quad , \forall s, d \in \mathcal{N}, \forall i \in \mathcal{N} \tag{8.12}$$

$$\zeta_{l_{in},l_{out}} \geq \theta^{s,d}_{l_{in}} + \theta^{s,d}_{l_{out}} - 1, \;\; \forall s,d \in \mathcal{N}, \forall l_{out} \in \mathcal{L}, \\ , \forall l_{in} \in \mathcal{L} : d(l_{in}) = s(l_{out}) \tag{8.13}$$

$$\zeta_{MAX} \geq \sum_{l_{in} \in \mathcal{L}:d(l_{in})=s(l_{out})} \zeta_{l_{in},l_{out}}, \;\; \forall l_{out} \in \mathcal{L} \tag{8.14}$$

$$\frac{1}{N(N-1)} \sum_{s,d \in \mathcal{N}} \sum_{l \in \mathcal{L}} \theta^{s,d}_{l} \leq \delta \times \kappa \tag{8.15}$$

$$\theta^{s,d}_{l}, \zeta_{l_{in},l_{out}} \in \{0,1\}; \text{non-negative integer: } \zeta_{MAX} \tag{8.16}$$

where $\theta^{s,d}_{l}$ is a binary variable that indicates if the route for the s-d pair of nodes, used to carry bursts, uses link l. Constraint 8.12 guarantees flow conservation of every route. In constraint 8.13 $\zeta_{l_{in},l_{out}}$ indicates if there is at least one selected route v that flows from l_{in} to l_{out}, and constraint 8.14 stores in ζ_{MAX} the highest number of incoming links that have routes flowing into a specific output link. That is, a global bound is being stored in ζ_{MAX}, the variable that will be minimized. Constraint 8.15 imposes a limit on the average number of hops that cannot exceed $\delta \times \kappa$.

8.3.2 SBPR-PP path selection strategy

$$\text{Minimize } \zeta_{MAX} \tag{8.17}$$

Subject to

$$\sum_{v \in \mathcal{V}_{s,d}} \sigma^{v} = 1, \;\; \forall s,d \in \mathcal{N} \tag{8.18}$$

$$\zeta_{l_{in},l_{out}} \geq \sigma^v \times p^v_{l_{in}} \times p^v_{l_{out}}, \ \forall v \in \mathcal{V}, \forall l_{out} \in \mathcal{L},$$
$$, \forall l_{in} \in \mathcal{L} : d(l_{in}) = s(l_{out}) \tag{8.19}$$

$$\zeta_{MAX} \geq \sum_{l_{in} \in \mathcal{L}: d(l_{in}) = s(l_{out})} \zeta_{l_{in},l_{out}}, \ \forall l_{out} \in \mathcal{L} \tag{8.20}$$

$$\frac{1}{N(N-1)} \sum_{v \in \mathcal{V}} \sum_{l \in \mathcal{L}} \sigma^v \times p^v_l \leq \delta \times \kappa \tag{8.21}$$

$$\sigma^v, \zeta_{l_{in},l_{out}} \in \{0,1\}; \text{non-negative integer: } \zeta_{MAX} \tag{8.22}$$

where σ^v is a binary variable that indicates if v is used to carry bursts from node $s(v)$ to node $d(v)$. Constraint 8.18 states that one path must be found for each pair of nodes. Each path is selected from the corresponding set $\mathcal{V}_{s,d}$ of available paths. In constraint 8.19 $\zeta_{l_{in},l_{out}}$ indicates if there is at least one selected route v that flows from l_{in} to l_{out}. Constraint 8.20 stores in ζ_{MAX} the highest number of incoming links having routes flowing into a specific output link. Constraint 8.21 imposes a limit on the average number of hops that cannot exceed $\delta \times \kappa$. The objective function is the same as in the previous case.

8.3.3 Evaluation

Simulations were done under similar conditions for the networks represented on Figure 6.5 whose physical parameters were already summarized and on Table 6.2.

The evaluation is also made against the SP, which is a commonly adopted practice in OBS studies. On Figure 8.7 we plot burst loss values normalized to the number of bursts that enter the network backbone against the average load. Results show that the strategy can reduce burst loss if the network topology presents high

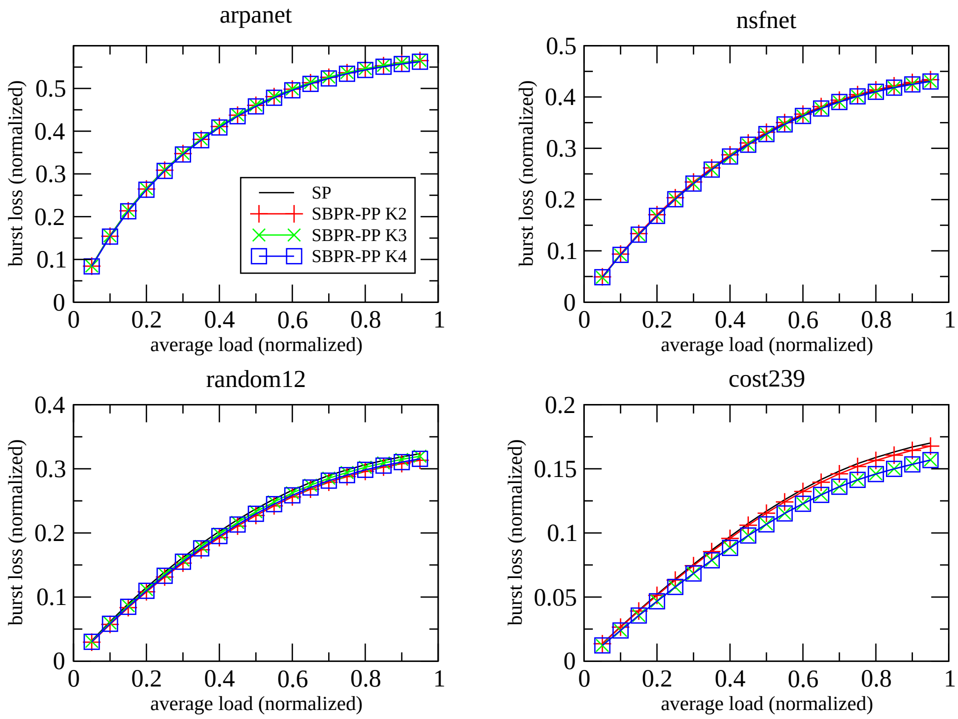

FIGURE 8.7: Burst loss on the OBS backbone (legend of arpanet applies for all networks; K2, K3 and K4 denote the number of eligible SPs provided to SBPR-PP).

connectivity. This is the case of the COST239 topology where all the plotted lines for different values of K are bellow the SP line. However, for networks presenting low connectivity, like the ARPANET or NSFNET, plotted values match point for point, meaning that the SP is already acting like a lower bound and any extra effort to find alternative paths with this strategy is not being productive. Even for RANDOM12, which can be considered moderately connected the plotted values almost coincide. More detailed results are depicted in Figures 8.8 to 8.11 where the best line of the SBPR-PP for each network is compared also with the results obtained from the algorithm without pre-calculated paths SBPR-nPP, showing that SBPR-nPP is acting like a lower bound to which SBPR-PP can come close. Considering the SBPR-nPP algorithm, the upper dashed lines in grey on Figures 8.8 to 8.11, always above the SP lines for all the networks, show that the attempt to find slightly longer paths (permitting an increase in δ of 0.05) immediately

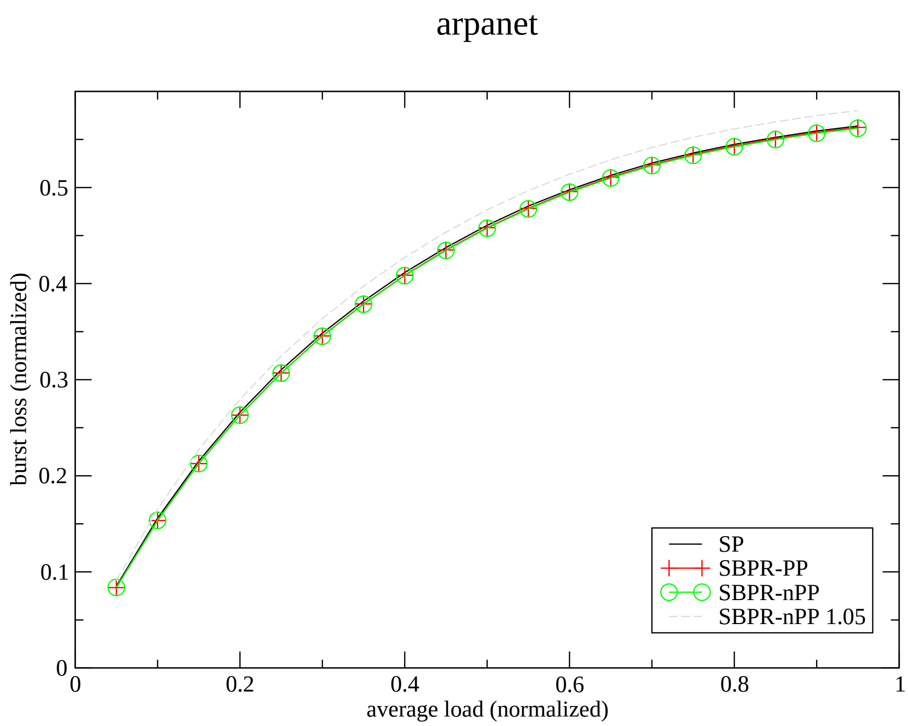

FIGURE 8.8: SBPR-nPP and best SBPR-PP versus SP for ARPANET.

increases burst loss, meaning that any small increase in hop count is a disadvantage. Results in Figure 8.11, where the best SBPR-PP and the SBPR-nPP have coincident lines, show also that it is possible to obtain, with the less expensive in computation cost SBPR-PP algorithm, performance results similar to the ones obtained with the more expensive SPBR-nPP algorithm.

8.4 Summary

In this chapter we consider the problem of routing path selection in OBS networks to minimize the overall burst loss due to resource contention of bursts aimed at the same output links. We take a TE approach aimed at contention reduction in networks equipped with and without wavelength converters, using only topological information, and considering the streamline effect. In the first approach, where each core node is considered like a global joining point, the proposed algorithms behave much better than SP routing for all the networks studied which are

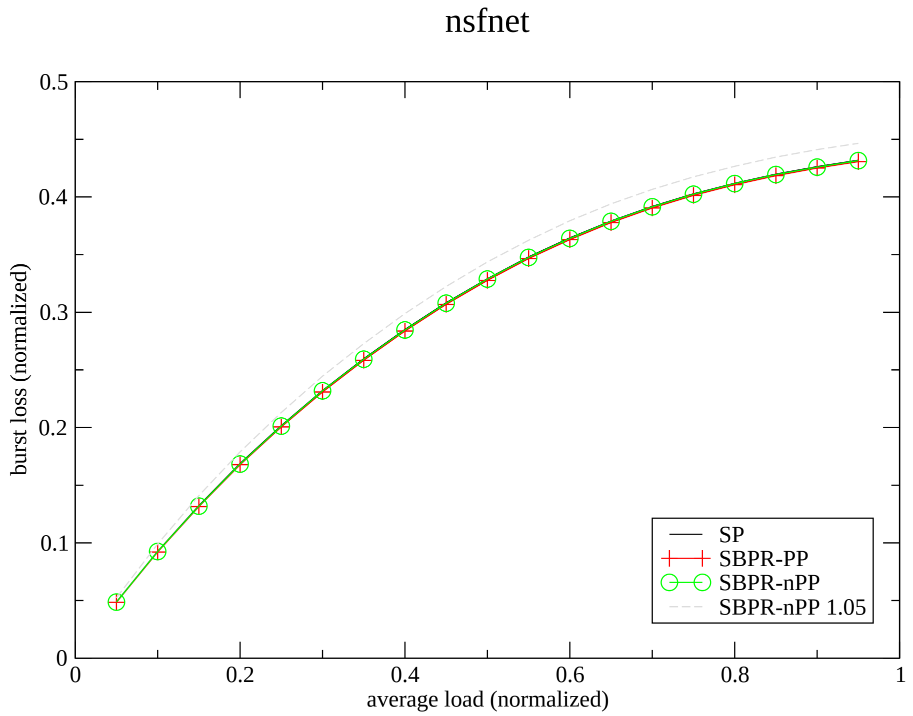

FIGURE 8.9: SBPR-nPP and best SBPR-PP versus SP for NSFNET.

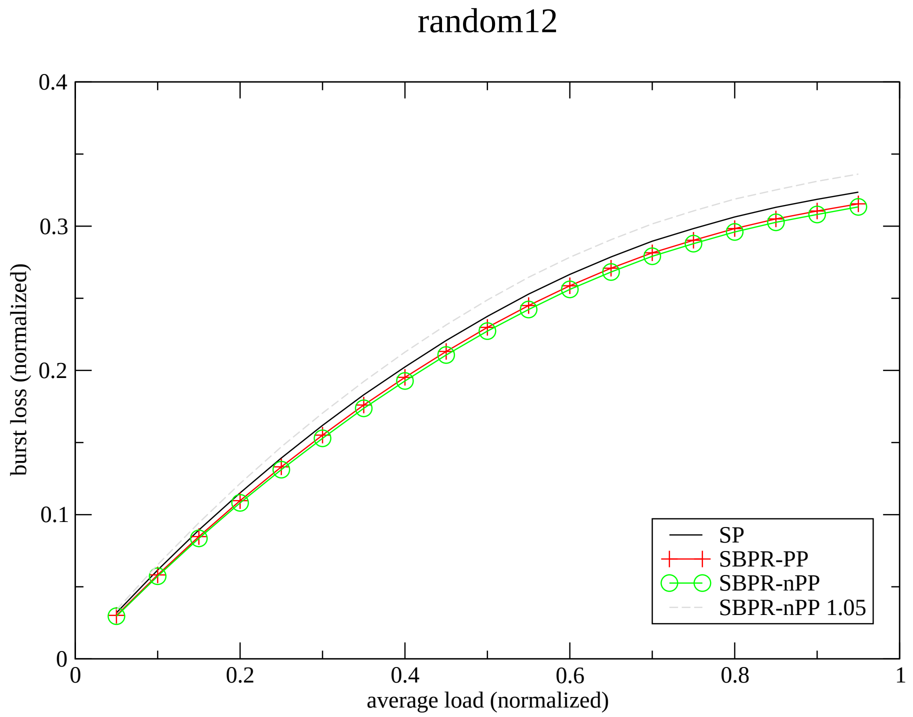

FIGURE 8.10: SBPR-nPP and best SBPR-PP versus SP for RANDOM12.

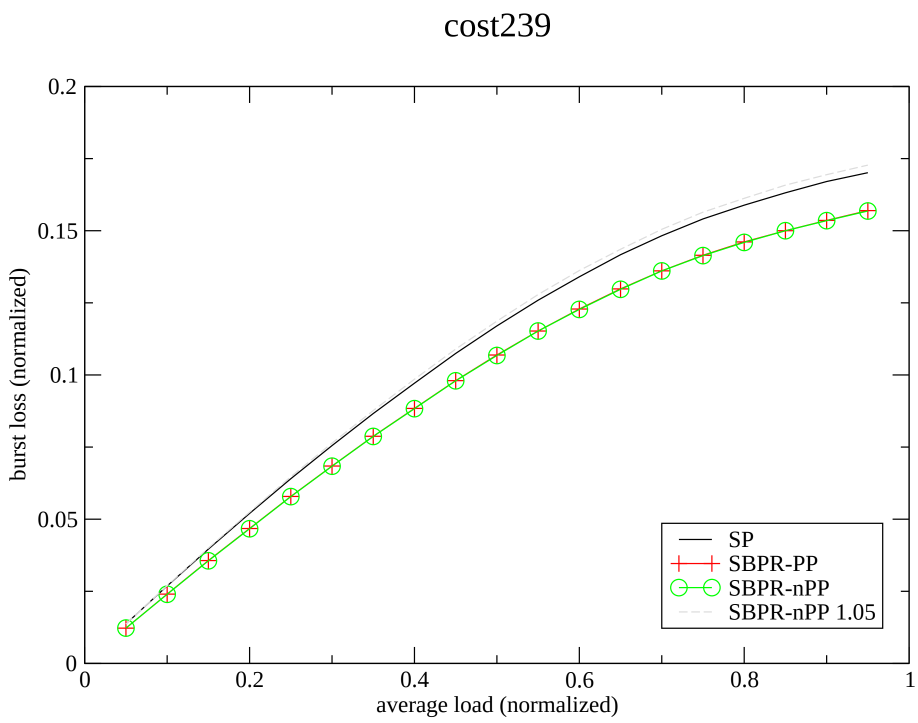

FIGURE 8.11: SBPR-nPP and best SBPR-PP versus SP for COST239.

characterized by having low to high physical connectivity values. In the second approach the performance of the algorithms is highly dependent on the connectivity of the network, but it is demonstrated that by incorporating the streamline effect into pre-planned path selection routing strategies contention can be decreased, in particular on highly connected networks.

9

Conclusion and Future Work

OBS provides a feasible paradigm for the next IP over optical network backbones (see Chapter 2). OBS avoids the inefficient resource utilization of OCS and the requirements of buffers, optical logic processing, and the synchronization problems of OPS (see Chapter 3). The basic transport unit in OBS is a burst containing multiple IP packets going to the same egress node and (if used) grouped by some relevant QoS criteria. Bursts are assembled at the ingress nodes and their transmission is preceded by dedicated CPs transmitted on a dedicated control channel for reserving bandwidth, along the path, for the upcoming data bursts. Based on the information carried on the CPs the intermediate nodes reserve switching resources, providing an optical channel through which bursts can be transmitted from source to final destination, after an adequate offset delay, without any OEO conversion, i.e., making profit from the optical bypass. However, due to its bufferless nature, OBS networks can be highly affected by burst contention. Several methods have been proposed to reduce burst loss due to contention, including burst scheduling, optical buffering, burst segmentation, wavelength conversion, and deflection routing (see Chapter 4). Obviously these mechanisms have their own advantages and disadvantages, which have been summarized in Table 4.1. Despite the undeniable merits of the research conducted on such methods, and besides the merit of some QoS provisioning proposals (see Chapter 5), there are also several important issues that need to be carefully considered, some of them emphasized next:

9.1 Contention Resolution Impairments

Burst scheduling research revolves around two objectives that are somewhat in conflict: one is to address low loss and the other is to achieve fast processing (Jin and Yang, 2007). The opposition between the two objectives is that, to achieve a lower loss rate, the scheduling algorithm need to maintain more information on the scheduled bursts, thus requiring more sophisticated processing.

The presence of optical buffers at the intermediate nodes to reduce contention, presently deployed through the use of FDLs is severely limited. Besides bringing more complexity to the scheduling algorithms mentioned above, FDLs are subjected to signal quality concerns and physical space limitations that can prevent a node from effectively handle high load or bursty traffic conditions (Vokkarane, Vokkarane and Jue, 2003).

By using burst segmentation, when contention happens only the overlapping segments of a burst need to be dropped instead of the entire burst. However, several issues need to be carefully balanced: the choice of the segment length, either fixed or variable, must be a trade-off between the loss per contention and the amount of overhead per burst. The segmentation process must also preclude the excessive burst fragmentation that can occur as a side effect, and that would incur into higher overhead with respect to burst control, switching and synchronization at the receiver (Vokkarane, Vokkarane and Jue, 2003). The decision of which segments to drop (head or tail-dropping) is also an issue, and must consider the problem of in-sequence packet delivery (Jue and Vokkarane, 2005f).

Wavelength conversion is the process of converting the wavelength of an incoming signal to another wavelength for transmission on an outgoing channel. Although most OBS studies assume full wavelength conversion to remove the wavelength continuity constraint, currently, wavelength converters are very expensive and complex devices; and this is expected to remain in the predictable future. Therefore, some limitation is expected on the wavelength conversion capabilities of optical networks with two important consequences: the performance studies relying on the assumption of full wavelength conversion can underestimate the burst drop probability, and the absence of full conversion claims for efficient wavelength assignment policies (Li et al., 2005), and TE approaches.

Deflection routing is a method by which a secondary (typically longer) path may be chosen by a switch whenever the primary intended path is not available. But this method also poses implementation problems, most of them related with the accuracy of the offsets and prevention of out-of-sequence or looping conditions, likely to result in severe degradation of performance. Deflection can be regarded as unplanned multipath routing, but a better method to reduce congestion is probably planned multipath routing or load-balancing (Vokkarane, Vokkarane, Zhang, Jue and Chen, 2002).

9.2 Pre-planned Path Selection Strategies

The above contention resolution techniques are reactive techniques that attempt to resolve contentions rather than avoiding the contentions. Also these contention resolution techniques attempt to reduce the loss based on the local node information. An alternative to this is contention avoidance, in which, instead of react to contention when it occurs one can prevent contention before it happens. With this approach, the goal is to reduce contention by policing the traffic at the source nodes, or by routing traffic in such way that congestion is minimized (Thodime et al., 2003).

In this thesis, we take the preventive approach to contention considering the problem of path selection in OBS networks in order to minimize the overall burst loss. Six pre-planned routing strategies are proposed and evaluated by simulation using an OBS network simulator specifically developed for that purpose using the OOS paradigm (see Chapter 6). The first two strategies provide burst loss reduction based on the assumption that by minimizing the demand for the most congested link (MCL) or for the maximum end-to-end congested path (MEC) during the route calculation stage, a global load balancing with minimized burst loss can be achieved (see Chapter 7). These strategies use pre-determined paths and results show better performance than the SP routing. In addition, it was shown that the gain of the proposed algorithms is also related to the way links are connected in OBS networks, with MCL and MEC providing better results for the COST239 topology than for the ARPANET topology, i.e., being more efficient in highly connected networks.

Four pre-planned path selection strategies are also proposed taking into account the streamline effect, a phenomenon unique to OBS networks wherein bursts traveling in a common link do not contend with each other until they diverge (see Chapter 8). In the approach using wavelength converters in all nodes, the first strategy uses pre-determined paths, while the second entirely determines the final path by itself. Both strategies present similar results, with burst loss values bellow the ones achieved by the SP routing, which means that during the route calculation stage the algorithms are making proffit from balancing the path-load at the output ports of the OBS node. Results also show that two pre-determined paths are enough to reach a near-optimal solution, making unnecessary longer calculations with more alternative paths. In the approach without wavelength converters, i.e, working under strict wavelength continuity constraint, performance improvements with respect to burst loss are much more difficult to obtain. Despite that fact, results show that the strategy can reduce burst loss if the network topology is characterized by having high connectivity values, like the case of COST239.

9.3 Final Remarks

Routing is preferentially done online and dynamically, based on "current" network resource availability. However, to support dynamic routing, the network may need to be flooded by frequent resource update messages, and additional hardware and software may need to be deployed at each OBS node hampering scalability. Pre-planned static routing can avoid both such signaling overhead and more complicated architecture condition. In addition, optimization techniques can be applied in pre-planned routing such that contention can be minimized, or substantially reduced, from a global point of view. Moreover, despite being widely believed that the dynamic operation of RWA helps to overcome the inefficiencies of static allocations (in terms of wavelength utilization), recent simulation studies show static RWA algorithms which perform closely to dynamic ones, in particular in networks without wavelength conversion capability where the benefit of dynamic approaches is not significant (Zapata-Beghelli and Bayvel, 2008). This is in agreement with another extensive performance evaluation recently presented in Li (2008b) where several offline RWA algorithms are referred as performing better than previously studied online algorithms (Li, 2008a).

In this thesis we take a TE approach which substantially reduces contention using only topological information. This approach has benefits in both network architecture (structure), and network QoS (behavior).

Considering the architecture, several advantages deserve to be highlighted: no extra hardware or software components are required on the core nodes for dynamic routing, no network flooding with signaling messages resulting from (over)active link state update protocols, no place for out-of-order arrivals, which are a disadvantage of some dynamic contention resolution schemes, also found on multipath routing schemes, typically requiring large memories at the edge nodes for re-ordering operations. These are important characteristics of the proposed approach that permits to keep simple and cost-effective the architecture of the core nodes, in particular when no wavelength converters are used, which are now very complex and expensive network components.

Considering QoS, the service to be taken into account in this thesis is delivery of bursts to the correct destination and the quality of this service is described by its performance in terms of burst loss, i.e., the lesser the better. We demonstrate that data burst loss can be minimized by appropriately choosing the paths that bursts must follow. That is, an effective choice of paths can lead to an overall network performance improvement. The resolution of an ILP problem is involved in the process, which makes this approach mainly tailored for offline static routing. However, since eligible paths can be provided as input for the path selection strategies, the solutions can be promptly reached. This means that the proposed strategies can admit updates if bounded by the reconfiguration requirements of the optical backbones, whose topologies typically last for long time scales. Our results also show that the achieved performance improvement depends on the physical connectivity of the network, with more highly connected networks showing better performance. This happens because the proposed algorithms take advantage of more, short, alternative paths. Our strategies can provide an initial stage of improved performance, measured in terms of burst loss reduction, without incurring state dissemination protocol penalties. The routes obtained can be used alone, as single-path static routes to provide load-balancing without the need for additional control messages with regard to link status, or combined with other dynamic contention resolution schemes (deflection or segmentation, for example) and used occasionally as a default routing to assume whenever the network needs to recover from instability, favoring the network resilience.

9.4 Future Work

Several research directions can be distinguished for future work. Some of them are summarized next and classified in different time scales. One can argue that, if a common underlying idea exist behind all of them, it is the idea of *cooperative working*.

Near future: Since the best algorithm (among the proposed ones) for some networks is not always the same in all circumstances, one way to increase performance is the adoption of a combined strategy of two (or more) path selection schemes. For the European COST239 network, for instance, while the best performance for low traffic loads is obtained with MCL, for high traffic loads best performance is obtained with SBPR-PP. By making the network assume the correspondent path selection scheme in each proper circumstance a (pseudo)dynamic behavior can be introduced for better performance in the OBS network. This approach is intended to appear shortly.

Another improvement can exploit the cooperative use of the three basic deflection domains (time, space and wavelength). Since they are independent, a combination of any of the three can be considered for contention resolution (some have already been tried and reported in literature). In addition, both static and dynamic approaches can be also explored, resulting in combined mechanisms to drive the network functioning under different conditions. For example, an OBS network can be initially put to run with a dynamic online routing scheme, but be forced to a static offline scheme whenever a certain level of instability is reached.

Mid-term future: Increasing attention is being given to the interaction of higher layer protocols with OBS, in particular the effect of burst assembly and burst loss in TCP performance. TCP is intended to provide a reliable transport layer over an unreliable network layer. It provides acknowledgment mechanisms for data retransmission and flow congestion control mechanisms that reduce the sending rate if congestion is detected in the network. When TCP is implemented over OBS, the loss of one burst may be interpreted by TCP as heavy congestion, although the the loss of a single burst does not necessarily indicate congestion in the OBS network. To study this interactions future work will include the extension of our current OBS simulation model to the packet aggregation and burst assembly processes. This will make it possible to evaluate the impact of different aggregation and assembly policies in the TCP layer.

Another challenging research direction is to extend our OBS model toward the broadband access networks considering its proclaimed bimodal operation where both *optical* and *wireless* technologies are merged. By combining the capacity of optical fiber networks with the ubiquity and mobility of wireless networks, FiWi networks will form a powerful supporting platform for our information society. For this to be possible, flexible, cost-effective, and easy to maintain architectures will have to be deployed, for which effective routing protocols with load balancing will be needed.

Long-term: All around us, world economies are shaky and future is uncertain. However, people's broadband habits are not withdrawable and, as of this writing, global internet traffic continues to grow. Video is playing a major role in traffic growth and this trend is likely to continue for the near and long-term future, in particular with the FTTH becoming a reality in the access networks. Just like MP3 players forever changed the music distribution model, broadband has also changed the video content distribution model and will keep allowing continuing traffic growth, challenging traditional architectures and claiming for new approaches using optical infrastructures. One important difference between video and data traffic is that video traffic is typically transmitted and routed streamlined into flows of packets joined together. While optical processing on a packet by packet basis is not available to handle this packet flows, aggregation paradigms like OBS will have their *momentum* but (disruptive or not) technology development will keep pace towards the "next NGN". The success of such NGN will be based upon a platform sustained by three pillars which are all very important to build a successful optical network: optics, electronics and software. New research directions in any of these areas will be very welcome.

"... à parte isso, tenho em mim todos os sonhos do mundo."

Álvaro de Campos, (Eng.)

The content of this appendix is a partial reproduction of several different sources from which definitions for some technical terms used in this thesis were collected.

A

Glossary

Erbium doped fiber amplifier (EDFA): is the most deployed optical amplifier as its amplification window coincides with the third transmission window of silica-based optical fiber. Two bands have developed in the third transmission window - the Conventional, or C-band, from approximately 1525 nm - 1565 nm, and the Long, or L-band, from approximately 1570 nm to 1610 nm. Both of these bands can be amplified by EDFAs, but it is normal to use two different amplifiers, each optimized for one of the bands. The principal difference between C- and L-band amplifiers is that a longer length of doped fiber is used in L-band amplifiers. The longer length of fiber allows a lower inversion level to be used, thereby giving at longer wavelengths (due to the band-structure of Erbium in silica) while still providing a useful amount of gain (www, 2009a).

Optical add/drop multiplexer (OADM): Is a device that takes in a composite optical signal that consists of multiple wavelengths and selectively drops (and subsequently adds) some of the wavelengths before letting the composite signal out of the output port (Chua et al., 2007a). A traditional OADM consists of three stages: an optical demultiplexer, an optical multiplexer, and between them a method of reconfiguring the paths between the optical demultiplexer, the optical multiplexer and a set of ports for adding and dropping signals. The optical demultiplexer separates wavelengths in an input fiber onto ports. The reconfiguration can be achieved by a fiber patch panel or by optical switches which direct the wavelengths to the optical multiplexer or to drop ports. The optical multiplexer multiplexes the wavelength channels that are to continue on from demultiplexer ports with those from the add ports, onto a single output fiber (www, 2009h). All the light paths that directly pass an OADM are termed cut-through lightpaths, while those that are added or dropped at the OADM node are termed added/dropped lightpaths. An OADM with remotely reconfigurable optical switches in the middle stage is called a reconfigurable optical add-drop multiplexer (ROADM). Ones without this feature are known as fixed OADMs. While the term OADM applies to both types, it is often used interchangeably with ROADM.

Optical amplifier (OAMP): is a device that amplifies an optical signal directly, without the need to first convert it to an electrical signal. An optical amplifier may be thought of as a laser without an optical cavity, or one in which feedback from the cavity is suppressed. Stimulated emission in the amplifier's gain medium causes amplification of incoming light. Optical amplifiers are important in optical communication and laser physics (www, 2009i).

Optical cross-connector (OXC): Is a device with multiple input and output ports. In addition to add/drop capability, it can also switch a wavelength from any input port to any output port (Chua et al., 2007a). Switching optical signals in an all-optical device is the second[1] approach to realize an

[1]The first is implemented in the electrical domain and called electronic OXC or opaque OXC.

OXC. Such a switch is often called a transparent OXC or photonic cross-connector (PXC). Specifically, optical signals are demultiplexed, then the demultiplexed wavelengths are switched by optical switch modules. After switching, the optical signals are multiplexed onto output fibers by optical multiplexers. Such a switch architecture keeps the features of data rate and protocol transparency. However, because the signals are kept in the optical format, the transparent OXC architecture does not allow easy optical signal quality monitoring (www, 2009j). As a compromise between opaque and transparent OXC's, there is a type of OXC called a translucent OXC. In such a switch architecture, there is a switch stage which consists of an optical switch module and an electronic switch module. Optical signals passing through the switch stage can be switched either by the optical switch module or the electronic switch module. In most cases, the optical switch module is preferred for the purpose of transparency. When the optical switch module's switching interfaces are all busy or an optical signal needs signal regeneration through an OEO conversion process, the electronic module is used. Translucent OXC nodes provide a compromise of full optical signal transparency and comprehensive optical signal monitoring. It also provides the possibility of signal regeneration at each node. Thus, such node architecture seems to be the most practical one to deploy the NGN (www, 2009k).

Optical line terminator (OLT): Multiplexes multiple wavelengths into a single fiber and demultiplexes a composite optical signal consisting of multiple wavelengths from a single fiber into separate fibers (Chua et al., 2007a).

Wavelength conversion: In optical transport networks, for flexibility of lightpath establishment and efficiency of wavelength utilization, we need the flexibility of converting an optical signal from one wavelength to another wavelength. Such a conversion process is called wavelength conversion. In general, there are two approaches to realize wavelength conversion, namely (1) OEO-based wavelength conversion and (2) all-optical wavelength conversion.

(1) OEO wavelength conversion employs an optical to electronic and then to optical process to convert an optical signal from one wavelength to another wavelength. Specifically, the optical signal is first converted into electronic format, then the electronic signal is used to modulate a tunable laser to convert to an optical signal. Here tunable laser will be tuned to the wavelength that we want to convert to. OEO wavelength conversion is a quite mature technique. The tunability of tunable laser can be the unique key limitation of the technique. In addition, because the conversion process needs to go though an electronic processing, such a processing could become a bottleneck of optical transmission system as electronic signal processing is generally slower than its optical counterpart.

(2) All-optical wavelength conversion is a more advanced conversion technique. It does not need to convert an signal into electronic format then to optical format. All optical conversion in general utilizes some nonlinear optical effects in optical components to realize wavelength conversion. These effects include Four Wavelength Mixing, SOA saturation, cross-phase modulation (XPM), etc. There is no so-called electronic bottleneck in optical wavelength conversion. However, viewing the immaturity, all-optical wavelength conversion is normally much more expensive than OEO wavelength conversion. Moreover, all-optical wavelength converters are not commercial available yet. Most of them are only experimentally realized in university or institute labs (www, 2009k).

The content of this appendix is a partial reproduction of a section of Varga and Hornig (2008) by Andras Varga and Rudolf Hornig.

B

Comparison of OMNeT++ with other simulation tools

This appendix presents an overview of two of the most widely used simulation tools, one noncommercial and one commercial, and compares them to OMNeT++. The discussion covers only the features and services of the simulation environments themselves, but not the availability or characteristics of specific simulation models like IPv6 or QoS (the reason being that they do not form part of the OMNeT++ simulation package.)

NS-2 (and NS-3) NS-2 is currently the most widely used network simulator in academic and research circles. NS-2 does not follow the same clear separation of simulation kernel and models as OMNeT++: the NS-2 distribution contains the models together with their supporting infrastructure, as one inseparable unit. This is a key difference: the NS-2 project goal is to build a network simulator, while OMNeT++ intends to provide a simulation platform, on which various research groups can build their own simulation frameworks. The latter approach is what called the abundance of OMNeT++-based simulation models and model frameworks into existence, and turned OMNeT++ into a kind of an ecosystem.

NS-2 lacks many tools and infrastructure components that OMNeT++ provides: support for hierarchical models, a graphical editor, GUI-based execution environment, separation of models from experiments, graphical analysis tools, simulation library features such as multiple RNG streams with arbitrary mapping and result collection, seamlessly integrated parallel simulation support, etc. This is because the NS-2 project concentrates on developing the simulation models, and much less on simulation infrastructure.

NS-2 is a dual-language simulator: simulation models are Tcl scripts[1], while the simulation kernel and various components (protocols, channels, agents, etc) are implemented in C++ and are made accessible from the Tcl language. Network topology is expressed as part of the Tcl script, which usually deals with several other things as well, from setting parameters to adding application behavior and recording statistics. This architecture makes it practically impossible to create graphical editors for NS-2 models[2].

NS-3 is an ongoing effort to consolidate all patches and recently developed models into a new version of NS. Although work includes refactoring of the simulation core as well, the concepts are essentially unchanged. The NS-3 project goals [36] include some features (e.g. parallel simulation, use of real-life protocol implementations as simulation models) that have already proved to be useful with OMNeT++.

[1]In fact, OTcl, which is an object-oriented extension to Tcl.

[2]Generating a Tcl script from a graphical representation is of course possible, but not the other way round: no graphical editor will ever be able to understand an arbitrary NS-2 script, and let the user edit it graphically.

OPNET Modeler OPNET Modeler is the flagship product of OPNET Technologies Inc. OPNET Modeler is a commercial product which is freely available worldwide to qualifying universities. OPNET has probably the largest selection of ready-made protocol models (including IPv6, MIPv6, WiMAX, QoS, Ethernet, MPLS, OSPFv3 and many others).

OPNET and OMNeT++ provide rich simulation libraries of roughly comparable functionality. The OPNET simulation library is based on C, while the one in OMNeT++ is a C++ class library. OPNET's architecture is similar to OMNeT++ as it allows hierarchical models with arbitrarily deep nesting (like OMNeT++), but with some restrictions (namely, the "node" level cannot be hierarchical). A significant difference from OMNeT++ is that OPNET models are always of fixed topology, while OMNeT++'s NED and its graphical editor allow parametric topologies. In OPNET, the preferred way of defining network topology is by using the graphical editor. The editor stores models in a proprietary binary file format, which means in practice that OPNET models are usually difficult to generate by program (it requires writing a C program that uses an OPNET API, while OMNeT++ models are simple text files which can be generated e.g. with Perl).

Both OPNET and OMNeT++ provide a graphical debugger and some form of automatic animation which is essential for easy model development. OPNET does not provide source code to the simulation kernel (although it ships with the sources of the protocol models). OMNeT++, like NS-2 and most other non-commercial tools, is fully public-source allowing much easier source level debugging.

OPNET's main advantage over OMNeT++ is definitely its large protocol model library, while its closed nature (proprietary binary file formats and the lack of source code) makes development and problem solving harder.

The content of this appendix is a partial reproduction of the introdutory chapters in Hillier and Lieberman (1990a).

C

The optimization approach

The typical optimization approach seeks for solutions that are optimal for an overall subject matter under study (here, a network) rather then suboptimal solutions that are best for only one of its components. Granted that this approach takes the overall point of view, this does not imply that each problem should be broadened into a study of the entire network. Instead, the objectives used in the study should be as specific as they can be while still encompassing the main goals in demand and maintaining a reasonable degree of consistency with the higher-level objectives of the network (Hillier and Lieberman, 1990b). After formulation, the problem must be expressed in convenient terms for analysis, which is usually done by constructing a mathematical model that represents the essence of the problem.

Models are idealized representations of the world. Typical examples can go from organizational charts to DNA structures, including airplane models, planet globes, or models of the atom. Mathematical models are also idealized representations, but are expressed in terms of mathematical symbols and expressions. The Newton's second law of motion $F = ma$, or the Einstein's mass-energy equivalence $E = mc^2$ are familiar examples of mathematical models. Similarly, the mathematical model of an optimization problem is the system of equations and related mathematical expressions that describe the essence of the problem. Thus, following the definitions in Hillier and Lieberman (1990b), if there are n related quantifiable decisions to be made, they are represented as *decision variables* (say, $x_1, x_2, ..., x_n$) whose respective values are to be determined. The appropriate measure of performance (for example, delay) is then expressed as a mathematical function of these decision variables (for example, $P = 3x_1 + 2x_2 + ... + 5x_n$) which is called the *objective function*. Any restrictions on the values that can be assigned to those decision variables are also expressed mathematically, usually using inequalities or equations (for example, $x_1 + 3x_1x_2 + 2x_2 \geqslant 10$). Such expressions for the restrictions are often called *constraints*. The constants, i.e., the coefficients or right-hand sides in the objective function and the constraints are called *parameters* of the model. The mathematical model might then say that the problem is to choose the values of the decision variables so as to minimize the objective function, subject to the specified constraints.

In a mathematical model if the functions appearing in both the objective function and the constraints are all linear functions we have a *linear programming model*. In many practical problems, however, the decision variables actually make sense if they have *integer* values. For example, it is often necessary to assign people or vehicles to activities and it does not make sense to make the assignment in fractional parts. In networks, for instance, the assignment of connections is also made in integer quantities. If requiring integer values is the only way in which a problem deviates from a linear programming formulation, then it is an *integer linear programming* (ILP) problem (Hillier and Lieberman, 1990c).

Without neglecting its obvious advantage in describing a problem more concisely then its verbal counterpart description, a mathematical model forms an important bridge to the use of mathematical techniques powered by computers to analyze a problem (Hillier and Lieberman, 1990b). In fact, together with optimization, operations researchers can also use probability theory, statistics, queuing theory, game theory, graph theory, decision analysis, and simulation. Because of the computational nature of these fields, OR has also ties to computer science, and operations researchers use either custom-developed and off-the-shelf software tools.

The content of this appendix is a partial reproduction of ILOG CPLEX Product Datasheet.

D

Some ILOG CPLEX features

ILOG CPLEX can solve very large, real-world problems with astonishing speed, and its dependability and stability have been proved through thousands of deployments worldwide. Developers access this power through a library of software components designed to integrate ILOG CPLEX into applications running on many of the leading platforms. Developers can access the ILOG CPLEX algorithms through component libraries or the Interactive Optimizer, a command-line utility that lets users read and write problem files and tune the performance of any ILOG CPLEX algorithm to the needs of a specific problem. All ILOG CPLEX algorithms are tightly integrated with cutting-edge presolve algorithms that reduce problem sizes and solve times without requiring any special user intervention. Each optimizer has numerous options for tuning solving strategies for specific problems.

Fundamental algorithms: ILOG CPLEX comes with the fastest, most reliable implementations of the fundamental algorithms for solving demanding mathematical optimization problems. ILOG CPLEX provides flexible, high-performance optimizers for solving linear programming, mixed integer programming, quadratic programming, quadratically constrained programming and mixed integer quadratically constrained programming problems. It can handle problems with millions of constraints and variables, and consistently sets performance records for mathematical programming.

ILOG CPLEX Mixed Integer Optimizers ILOG CPLEX Mixed Integer Optimizers employ a branch-and-bound technique that takes advantage of innovative cutting-edge strategies to provide high-performance solutions for the hardest mixed integer programs. ILOG CPLEX can solve mixed integer linear, mixed integer quadratic and mixed integer quadratically constrained problems. ILOG CPLEX Mixed Integer Optimizers include the ILOG CPLEX presolve algorithm, sophisticated cutting-plane strategies and feasibility heuristics.

ILOG CPLEX Parallel Optimizers ILOG CPLEX Parallel Optimizers take advantage of multiple CPUs to solve extremely difficult industrial problems. The parallel optimizers include ILOG CPLEX Parallel Barrier Optimizer and ILOG CPLEX Parallel MIP Optimizer. The ILOG CPLEX Parallel Barrier Optimizer achieves significant speedups over its serial counterpart on a wide variety of classes of linear programming problems. The ILOG CPLEX Parallel MIP Optimizer implements a deterministic mode of operation that produces a repeatable, invariant solution path, as well as an opportunistic mode that takes full advantage of parallelism for potentially better performance. ILOG CPLEX Parallel Optimizers can substantially reduce the time for solving large linear and difficult mixed integer programs. In an alternative use of multiple processors, ILOG CPLEX Concurrent Optimizer uses different algorithms on different CPUs in a race to find the best way to solve a problem.

ILOG CPLEX Component Libraries ILOG CPLEX Component Libraries provide the features and flexibility that mathematical programming developers need to create customized applications for solving both simple and complex optimization problems. The libraries include C, C++, .NET and Java programming interfaces that allow developers to use most programming languages to efficiently embed ILOG CPLEX technology directly into their applications. The programming interfaces provide a comprehensive set of routines for defining, solving, analyzing, querying and creating reports for mathematical programming problems and solutions. For example, routines are provided to direct the solution process and completely control ILOG CPLEX messages, and to help developers debug their own ILOG CPLEX applications.

Bibliography

Acampora, A. and Shah, S. (1992). Multihop lightwave networks: a comparison of store-and-forward and hot-potato routing, *IEEE Transactions on Communications* **40**(6): 1082–1090.

Alferness, R. C., Kogelnik, H. and Wood, T. H. (2000). The evolution of optical systems: Optics everywhere, *Bell Labs Technical Journal* **5**(1): 188–202.

Amstutz, S. (1983). Burst switching–an introduction, *Communications Magazine, IEEE* **21**(8): 36–42.

Arakawa, Y., Sakuta, M. and Sasase, I. (2003). QoS scheme with burst dropping in optical burst switching, *Proc. PACRIM Communications, Computers and signal Processing 2003 IEEE Pacific Rim Conference on*, Vol. 1, pp. 397–400 vol.1.

Armstrong, D. J. (2006). The quarks of object-oriented development, *Commun. ACM* **49**(2): 123–128.

Baldine, I., Rouskas, G., Perros, H. and Stevenson, D. (2002). Jumpstart: a just-in-time signaling architecture for WDM burst-switched networks, *Communications Magazine, IEEE* **40**(2): 82–89.

Banks, J. (1998a). *Handbook of Simulation - Principles, Methodology, Advances, Applications and Practice.*, EMP Books and John Wiley & Sons, Inc.

Banks, J. (1998b). *Handbook of Simulation - Principles, Methodology, Advances, Applications and Practice.*, EMP Books and John Wiley & Sons, Inc., chapter Principles of Simulation, pp. 3–30.

Banks, J., Carson II, J. S. and Nelson, B. L. (1996). *Discrete-Event System Simulation*, Prentice Hall.

Baroni, S. and Bayvel, P. (1997). Wavelength requirements in arbitrarily connected wavelength-routed optical networks, *Lightwave Technology, Journal of* **15**(2): 242–251.

Battestilli, T. and Perros, H. (2003). An introduction to optical burst switching, *Communications Magazine, IEEE* **41**(8): S10–S15.

Battestilli, T. and Perros, H. (2005). Optical burst switching for the next generation internet, *IEEE Potentials* **23**(5): 40–43.

Berger, M., Dechet, K., Meyer, G., Rothkegel, W., Sturm, W. and Wilke, B. (2006). Versatile bandwidth management: The design, development, and deployment of lambdaunite(reg), *Bell Labs Technical Journal* **11**(2): 65–81.

Callegati, F., Cerroni, W., Bonani, L., Barbosa, F., Moschim, E. and Pavani, G. (2006). Opn06-6: Congestion resolution in optical burst/packet switching with limited wavelength conversion, *Proc. IEEE Global Telecommunications Conference GLOBECOM '06*, pp. 1–5.

Cameron, A. F., Thorne, D. J., Foster, K. T. and Fisher, S. I. (2007). Fixed access network technologies, *BT Technology Journal* **25**(3-4): 121–131.

Cameron, C., Zalesky, A. and Zukerman, M. (2005). Prioritized deflection routing in optical burst switching networks, *IEICE Trans Commun* **E88-B**(5): 1861–1867.

Cao, X., Li, J., Chen, Y. and Qiao, C. (2002). Assembling TCP/IP packets in optical burst switched networks, *Proceedings of IEEE GLOBECOM*, pp. 2808–2812.

Chen, C.-H., Wolfson, D., Johansson, L., Blumenthal, D. and Coldren, L. (2006). Demonstration of 40 gbit/s optical packet synchronisation using fibre bragg gratings and fast-tunable wavelength converters, *Electronics Letters* **42**(6): 367–369.

Chen, Q., Mohan, G. and Chua, K. (2006a). Route optimization for efficient failure recovery in optical burst switched networks, *Proc. Workshop on High Performance Switching and Routing*, pp. 6 pp.–.

Chen, Q., Mohan, G. and Chua, K. C. (2006b). Offline route optimization considering streamline effect in optical burst switching networks, *Communications, 2006. ICC '06. IEEE International Conference on* **6**: 2562–2567.

Chen, Q., Mohan, G. and Chua, K. C. (2008). Route optimization in optical burst switched networks considering the streamline effect, *Computer Networks* **52**(10): 2033 – 2044. Challenges and Opportunities in Advanced Optical Networking.

Chen, Y., Hamdi, M. and Tsang, D. (2001). Proportional QoS over OBS networks, *Proc. IEEE Global Telecommunications Conference GLOBECOM '01*, Vol. 3, pp. 1510–1514.

Chen, Y., Qiao, C. and Yu, X. (2004). Optical burst switching: a new area in optical networking research, *Network, IEEE* **18**(3): 16–23.

Cheyns, J., Van Breusegem, E., Develder, C., Ackaert, A., Pickavet, M. and Demeester, P. (2003). Performance improvement of an internally blocking optical packet/burst switch, *Proc. IEEE International Conference on Communications ICC '03*, Vol. 2, pp. 1304–1308 vol.2.

Chlamtac, I., Elek, V., Fumagalli, A. and Szabo, C. (1997). Scalable WDM network architecture based on photonic slot routing and switched delay lines, *Proc. IEEE Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM '97*, Vol. 2, pp. 769–776 vol.2.

Chlamtac, I., Ganz, A. and Karmi, G. (1992). Lightpath communications: an approach to high bandwidth optical, *IEEE Transactions on Communications* **40**(7): 1171–1182.

Chua, K. C., Gurusamy, M., Liu, Y. and Phung, M. H. (2007a). *Quality of Service in Optical Burst Switched Networks*, Springer.

Chua, K. C., Gurusamy, M., Liu, Y. and Phung, M. H. (2007b). *Quality of Service in Optical Burst Switched Networks*, Springer, chapter Node-based QoS Improvement Mechanisms, pp. 23–72.

Chua, K. C., Gurusamy, M., Liu, Y. and Phung, M. H. (2007c). *Quality of Service in Optical Burst Switched Networks*, Springer, chapter Relative QoS Differentiation, pp. 73–90.

Chua, K. C., Gurusamy, M., Liu, Y. and Phung, M. H. (2007d). *Quality of Service in Optical Burst Switched Networks*, Springer, chapter Absolute QoS Differentiation, pp. 91–110.

Chua, K. C., Gurusamy, M., Liu, Y. and Phung, M. H. (2007e). *Quality of Service in Optical Burst Switched Networks*, Springer, chapter Edge-to-edge QoS Mechanisms, pp. 111–176.

Detti, A., Eramo, V. and Listanti, M. (2002). Performance evaluation of a new technique for IP support in a WDM optical network: optical composite burst switching (ocbs), *IEEE/OSA Journal of Lightwave Technology* **20**(2): 154–165.

Dovrolis, C., Stiliadis, D. and Ramanathan, P. (1999). Proportional differentiated services: delay differentiation and packet scheduling, *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, ACM, New York, NY, USA, pp. 109–120.

Du, Y., Zhu, C., Zheng, X., Guo, Y. and Zhang, H. (2007). A novel load balancing deflection routing strategy in optical burst switching networks, *Proc. Conference on Optical Fiber Communication and the National Fiber Optic Engineers Conference OFC/NFOEC 2007*, pp. 1–3.

Duser, M. and Bayvel, P. (2001). Performance of a dynamically wavelength-routed, optical burst switched network, *Proc. IEEE Global Telecommunications Conference GLOBECOM '01*, Vol. 4, pp. 2139–2143 vol.4.

Duser, M. and Bayvel, P. (2002). Analysis of a dynamically wavelength-routed optical burst switched network architecture, *IEEE/OSA Journal of Lightwave Technology* **20**(4): 574–585.

Eckel, B. (1998). *Thinking in Java*, Prentice Hall PTR, chapter Introduction to Objects, pp. 25–66.

Eckel, B. (2000). *Thinking in C++ (2nd Edition)*, Prentice Hall.

Effenberger, F., Kramer, G. and Hesse, B. (2007). Passive optical networking update [PON update], *IEEE Communications Magazine* **45**(3): S6–S8.

El-Bawab, T. S. (2006). *Optical Switching*, Springer US, chapter Optical Switching in Communications Networks, pp. 333–379.

Eramo, V., Listanti, M. and Spaziani, M. (2005). Resources sharing in optical packet switches with limited-range wavelength converters, *IEEE/OSA Journal of Lightwave Technology* **23**(2): 671–687.

Farahmand, F. and Jue, J. (2003). Look-ahead window contention resolution in optical burst switched networks, *Proc. HPSR High Performance Switching and Routing Workshop on*, pp. 147–151.

Farahmand, F. and Jue, J. (2006). Analysis and implementation of look-ahead window contention resolution with QoS support in optical burst-switched networks, *IEEE Journal on Selected Areas in Communications* **24**(12): 81–93.

Farahmand, F., Vokkarane, V. M., Jue, J. P., Rodrigues, J. J. P. C. and Freire, M. M. (2007). Optical burst switching network: A multi-layered approach, *Journal of High Speed Networks* **16**(2): 105–122.

Forghieri, F., Bononi, A. and Prucnal, P. (1995). Analysis and comparison of hot-potato and single-buffer deflection routing in very high bit rate optical mesh networks, *IEEE Transactions on Communications* **43**(1): 88–98.

Ganguly, S., Bhatnagar, S., Izmailov, R. and Qiao, C. (2004). Multi-path adaptive optical burst forwarding, *High Performance Switching and Routing, 2004. HPSR. 2004 Workshop on*, pp. 180–185.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H.Freeman and Company.

Gauger, C. M. (2003). *Next Generation Optical Network Design and Modelling*, Kluwer Academic Publishers, chapter Dimensioning of FDL Buffers for Optical Burst Switching Nodes, pp. 117–131.

Ge, A., Callegati, F. and Tamil, L. (2000). On optical burst switching and self-similar traffic, *IEEE Communications Letters* **4**(3): 98–100.

Gevaux, D. (2007). Slow light, fast computers, *Nat Photon* **1**(1): 72.

Gjessing, S. and Maus, A. (2005). Discrete event simulation of a large OBS network, *International Conference on Systems, Man and Cybernetics*, IEEE-SMC.

Glass, A. M., DiGiovanni, D. J., Strasser, T. A., Stentz, A. J., Slusher, R. E., White, A. E., Kortan, A. R. and Eggleton, B. J. (2000). Advances in fiber optics, *Bell Labs Technical Journal* **5**(1): 168–187.

Gonzalez-Ortega, M. A., Lopez-Ardao, J. C., Rodriguez-Rubio, R. F., Lopez-Garcia, C., Fernandez-Veiga, M. and Suarez-Gonzalez, A. (2007). Performance analysis of adaptive multipath load balancing in WDM-LOBS networks, *Computer Communications* **30**(18): 3460–3470.

Gonzalez-Ortega, M., Lopez-Ardao, J., Lopez-Garcia, C., Argibay-Losada, P., Rodriguez-Rubio, R. and Pineiro-Valladares, M. (2008). Loss differentiation in OBS networks without wavelength-conversion capability, *IEEE Communications Letters* **12**(12): 903–905.

Green, P. (2001). Progress in optical networking, *IEEE Communications Magazine* **39**(1): 54–61.

Green, P.E., J. (1996). Optical networking update, *IEEE Journal on Selected Areas in Communications* **14**(5): 764–779.

Greenberg, A. and Hajek, B. (1992). Deflection routing in hypercube networks, *IEEE Transactions on Communications* **40**(6): 1070–1081.

Guan, X., Thng, I. L.-J., Jiang, Y. and Li, H. (2005). Providing absolute QoS through virtual channel reservation in optical burst switching networks, *Computer Communications* **28**(9): 967 – 986.

Guanghong, L. and Zhaohong, L. (2007). Performance analyses of communication paradigm based on anycast in the grid over OBS, *Proc. International Conference on Convergence Information Technology*, pp. 1874–1879.

Hao, Q. M. (2008). Toward a unified service delivery process for next-generation services, *Bell Labs Tech. J.* **12**(4): 5–20.

Hardy, W. C. (2001). *QoS Measurement and Evaluation of Telecommunications Quality of Service*, Wiley.

Hehmann, J. and Pfeiffer, T. (2008). New monitoring concepts for optical access networks, *Bell Labs Tech. J.* **13**(1): 183–198.

Henderson-Sellers, B. (1992). *A Book of Object-Oriented Knowledge*, Prentice Hall.

Heron, R. W., Pfeiffer, T., van Veen, D. T., Smith, J. and Patel, S. S. (2008). Technology innovations and architecture solutions for the next-generation optical access network, *Bell Labs Technical Journal* **13**(1): 163–181.

Hillier, F. S. and Lieberman, G. J. (1990a). *Introduction to Operations Research*, McGrow-Hill International Editions.

Hillier, F. S. and Lieberman, G. J. (1990b). *Introduction to Operations Research*, McGrow-Hill International Editions, chapter Overview of the Operations Research Modeling Approach, pp. 16–25.

Hillier, F. S. and Lieberman, G. J. (1990c). *Introduction to Operations Research*, McGrow-Hill International Editions, chapter Integer Programming, pp. 457–498.

Hsu, C.-F., Liu, T.-L. and Huang, N.-F. (2002). Performance analysis of deflection routing in optical burst-switched networks, *Proc. IEEE Twenty-First*

Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM 2002, Vol. 1, pp. 66–73 vol.1.

Huang, P.-K., Chang, C.-S., Cheng, J. and Lee, D.-S. (2007). Recursive constructions of parallel fifo and lifo queues with switched delay lines, *IEEE Transactions on Information Theory* **53**(5): 1778–1798.

Huang, X., She, Q., Zhang, T., Lu, K. and Jue, J. (2008). Modeling and performance analysis of small group multicast with deflection routing in optical burst switched networks, *IEEE Journal on Selected Areas in Communications* **26**(3): 74–86.

Hudek, G. and Muder, D. (1995). Signaling analysis for a multi-switch all-optical network, *Proc. IEEE International Conference on Communications ICC '95 Seattle, 'Gateway to Globalization'*, Vol. 2, pp. 1206–1210 vol.2.

Hunter, D., Nizam, M., Chia, M., Andonovic, I., Guild, K., Tzanakaki, A., O'Mahony, M., Bainbridge, L., Stephens, M., Penty, R. and White, I. (1999). Waspnet: a wavelength switched packet network, *IEEE Communications Magazine* **37**(3): 120–129.

Ingalls, D. H. H. (1981). Design principles behind smalltalk, *BYTE Magazine* .

Izal, M. and Aracil, J. (2002). On the influence of self-similarity on optical burst switching traffic, *Proc. IEEE Global Telecommunications Conference GLOBECOM '02*, Vol. 3, pp. 2308–2312 vol.3.

Jana, D. (2004). *C++ and Object-Oriented Programming Paradigm*, Prentice Hall.

Jin, M. and Yang, O. W. (2007). Provision of differentiated performance in optical burst switching networks based on burst assembly processes, *Computer Communications* **30**(18): 3449–3459.

Joines, J. A. and Roberts, S. D. (1998). *Handbook of Simulation - Principles, Methodology, Advances, Applications and Practice.*, EMP Books and John Wiley & Sons, Inc., chapter Object-Oriented Simulation, pp. 397–427.

Joines, J. and Roberts, S. (1999). Simulation in an object-oriented world, *Proc. Winter Simulation*, Vol. 1, pp. 132–140 vol.1.

Jue, J. P. and Vokkarane, V. M. (2005a). *Optical Burst Switched Networks*, Springer.

Jue, J. P. and Vokkarane, V. M. (2005b). *Optical Burst Switched Networks*, Springer, chapter Technology and Architecture, pp. 11–22.

Jue, J. P. and Vokkarane, V. M. (2005c). *Optical Burst Switched Networks*, Springer, chapter Burst Assembly, pp. 23–36.

Jue, J. P. and Vokkarane, V. M. (2005d). *Optical Burst Switched Networks*, Springer, chapter Signaling, pp. 37–56.

Jue, J. P. and Vokkarane, V. M. (2005e). *Optical Burst Switched Networks*, Springer, chapter Channel Scheduling, pp. 81–105.

Jue, J. P. and Vokkarane, V. M. (2005f). *Optical Burst Switched Networks*, Springer, chapter Contention Resolution, pp. 57–80.

Kaheel, A., Khattab, T., Mohamed, A. and Alnuweiri, H. (2002). Quality-of-service mechanisms in IP-over-WDM networks, *IEEE Communications Magazine* **40**(12): 38–43.

Kay, A. C. (1996). *History of Programming Languages*, ACM Press, chapter The Early History of Smalltalk.

Kim, M.-G., Jeong, H., Choi, J., Kim, J.-H. and Kang, M. (2006). The impact of the extra offset-time based QoS mechanism in TCP over optical burst switching networks, *Proc. Optical Communication Systems and Networks*.

Kiran, Y., Venkatesh, T. and Murthy, C. (2007). A reinforcement learning framework for path selection and wavelength selection in optical burst switched networks, *Selected Areas in Communications, IEEE Journal on* **25**(9): 18–26.

Kleinrock, L. (1975). *Queueing Systems, Volume 1: Theory*, Wiley Interscience, New York.

Klinkowski, M., Careglio, D., Spadaro, S. and Sole-Pareta, J. (2005). Impact of burst length differentiation on QoS performance in OBS networks, *Proc. 7th International Conference Transparent Optical Networks*, Vol. 1, pp. 91–94 Vol. 1.

Klinkowsky, M. (2007). *Offset Time-Emulated Architecture for Optical Burst Switching - Modeling and Performance Evaluation*, PhD thesis, Universitat Politecnica de Catalunya - Department d'Arquitectura de Computadores.

Kobayashi, H. (1978). *Modeling and Analysis - An Introduction to System Performance Evaluation Methodology*, Addison-Wesley.

Kozlovsky, E. and Bayvel, P. (2003). QoS performance of WROBSS network architecture with request scheduling, *Next Generation Optical Network Design and Modelling*, Springer.

Kulzer, J. and Montgomery, W. (1984). Statistical switching architectures for future services, *Proceedings of ISS'84*, Florence, Italy, pp. 43–A.

Kurose, J. F. and Ross, K. W. (2007). *Computer Networking: A Top-Down Approach*, Addison-Wesley.

Law, A. M. and Kelton, W. D. (1991). *Simulation Modeling and Analysis*, McGraw-Hill.

Lazzez, A. and Boudriga, N. (2007). TCP management over OBS using burst segmentation and segment retransmission, *Proc. ICTON Mediterranean Winter Conference ICTON-MW 2007*, pp. 1–6.

Lazzez, A., Boudriga, N. and Obaidat, M. S. (2008). Improving TCP QoS over OBS networks: A scheme based on optical segment retransmission, *Proc. International Symposium on Performance Evaluation of Computer and Telecommunication Systems SPECTS 2008*, pp. 233–240.

Lee, S., Sriram, K., Kim, H. and Song, J. (2005). Contention-based limited deflection routing protocol in optical burst-switched networks, *IEEE Journal on Selected Areas in Communications* **23**(8): 1596–1611.

Lenkiewicz, P., Hajduczenia, M., Freire, M. M., da Silva, H. J. A. and Monteiro, P. P. (2006). Estimating network offered load for optical burst switching networks, *LNCS, Networking 2006* **3976**: 1062–1073.

Li, J., Mohan, G. and Chua, K. (2005). Dynamic load balancing in IP-over-WDM optical burst switching networks, *Computer Networks* **47**(3): 393–408.

Li, J. and Qiao, C. (2004). Recent progress in the scheduling algorithms in optical-burst-switched networks, *Journal of Optical Networking* **3**(4): 229–241.

Li, K. (2008a). Experimental average-case performance evaluation of online algorithms for routing and wavelength assignment and throughput maximization in WDM optical networks, *J. Exp. Algorithmics* **12**: 1–24.

Li, K. (2008b). Heuristic algorithms for routing and wavelength assignment in WDM optical networks, *Proc. IEEE International Symposium on Parallel and Distributed Processing IPDPS 2008*, pp. 1–8.

Liu, D. and Liu, M. (2002). Differentiated services and scheduling scheme in optical burst-switched WDM networks, *Networks, 2002. ICON 2002. 10th IEEE International Conference on* pp. 23–27.

Liu, J., Ansari, N. and Ott, T. (2003). Frr for latency reduction and QoS provisioning in OBS networks, *IEEE Journal on Selected Areas in Communications* **21**(7): 1210–1219.

Loi, C.-H., Liao, W. and Yang, D.-N. (2002). Service differentiation in optical burst switched networks, *Proc. IEEE Global Telecommunications Conference GLOBECOM '02*, Vol. 3, pp. 2313–2317 vol.3.

Long, K., Yang, X., Huang, S. and Kuang, Y. (2006). A GMPLS-based OBS architecture for IP-over-WDM networks, *Proc. of Network Architectures, Management, and Applications IV*, Vol. 6354, SPIE, p. 63540H.

Lu, J., Liu, Y., Gurusamy, M. and Chua, K. (2006). Gradient projection based multi-path traffic routing in optical burst switching networks, *Proc. Workshop on High Performance Switching and Routing*, pp. 379–384.

Lu, K., Xiao, G. and Chlamtac, I. (2005). Analysis of blocking probability for distributed lightpath establishment in WDM optical networks, *IEEE/ACM Trans. Networking* **13**: 187–197.

Lu, X. and Mark, B. (2004). Performance modeling of optical-burst switching with fiber delay lines, *IEEE Transactions on Communications* **52**(12): 2175–2183.

Mack, J., Burmeister, E., Poulsen, H., Stamenic, B., Bowers, J. and Blumenthal, D. (2008). Photonic integrated circuit switch matrix and waveguide delay lines for optical packet synchronization, *Proc. 34th European Conference on Optical Communication ECOC 2008*, pp. 1–2.

Maier, M. (2008a). *Optical Switching Networks*, Cambridge University Press, chapter Optical burst switching, pp. 103–134.

Maier, M. (2008b). *Optical Switching Networks*, Cambridge University Press.

Mannie, E. e. a. (2004). Generalized multi-protocol label switching (GMPLS) architecture, *RFC 3945*, Network Working Group.

Maranhao, J., Waldman, H., Soares, A. and Giozza, W. (2008). Wavelength conversion architectures in OBS networks, *Proc. IEEE Network Operations and Management Symposium NOMS 2008*, pp. 939–942.

Marchetto, G. (2008). Minimizing preemption probability to efficiently support service differentiation in just-in-time based OBS networks, *Proc. IEEE International Conference on Communications ICC '08*, pp. 5219–5223.

Marcus, J. S. and Elixmann, D. (2008). The future of IP interconnection: Technical, economic, and public policy aspects, *Technical report*, WIK-Consult GmbH, Germany. Study for the European Commission.

Meis, D. (2006). FTTH: The next great household amenity, *Bradband Properties* **2**: 45–49.

Mignotte, A. and Peyran, O. (1997). Reducing the complexity of ilp formulations for synthesis, *ISSS '97: Proceedings of the 10th international symposium on System synthesis*, IEEE Computer Society, Antwerp, Belgium, pp. 58–64.

Mikoshi, T. and Takenaka, T. (2008). Improvement of burst transmission delay using offset time for burst assembly in optical burt switching, *Proc. APSITT Information and Telecommunication Technologies 7th Asia-Pacific Symposium on*, pp. 13–18.

Mills, D. L., Boncelet, G. C., Elias, J. G., Schragger, P. A. and Jackson, A. W. (1990). Highball: a high speed, reserved-access, wide area network, *Technical Report 90-9-3*, University of Delaware, Electrical Engineering Dept., Newark.

Mukherjee, B. (2006a). *Optical WDM Networks*, Springer.

Mukherjee, B. (2006b). *Optical WDM Networks*, Springer, chapter Optical Burst Switching, pp. 845–865.

Ogino, N. and Tanaka, H. (2005). Deflection routing for optical bursts considering possibility of contention at downstream nodes, *IEICE Trans Commun* **E88-B**(9): 3660–3667.

Payne, D. (2008). *ICT Futures - Delivering Pervasive, Real-time and Secure Services*, Vol. ICT Futures: Delivering Pervasive, Real-time and Secure Services, Wiley, chapter The Future All Optical Network - Why we need it and how to get there, pp. 93–114.

Pedro, J., Monteiro, P. and Pires, J. (2007). Efficient optical burst-switched networks using only fiber delay line buffers for contention resolution, *Proc. Fourth International Conference on Broadband Communications, Networks and Systems BROADNETS 2007*, pp. 2–11.

Pedroso, P., Sole-Pareta, J., Careglio, D. and Klinkowski, M. (2007). Integrating GMPLS in the OBS networks control plane, *Proc. 9th International Conference on Transparent Optical Networks ICTON '07*, Vol. 3, pp. 1–7.

Phung-Duc, T., Masuyama, H., Kasahara, S. and Takahashi, Y. (2006). Burst segmentation with upper-layer retransmission and its effect on wavelength utilization for optical burst switched networks, *Proc. First International Conference on Communications and Electronics ICCE '06*, pp. 35–40.

Phung, M., Chua, K., Mohan, G., Motani, M. and Wong, T. (2004). Absolute QoS signalling and reservation in optical burst-switched networks, *Proc. IEEE Global Telecommunications Conference GLOBECOM '04*, Vol. 3, pp. 2009–2013 Vol.3.

Phung, M., Chua, K., Mohan, G., Motani, M. and Wong, T. (2005). The streamline effect in OBS networks and its application in load balancing, *Broadband Networks, 2005 2nd International Conference on* pp. 283–290 Vol. 1.

Phung, M. H., Chua, K. C., Mohan, G., Motani, M. and Wong, T. C. (2007). An absolute QoS framework for loss guarantees in optical burst-switched networks, *IEEE Transactions on Communications* **55**(6): 1191–1201.

Phung, M. H., Chua, K. C., Mohan, G., Motani, M., Wong, T. C. and Kong, P. Y. (2005). On ordered scheduling for optical burst switching, *Comput. Netw.* **48**(6): 891–909.

Phung, M., Shan, D., Chua, K. and Mohan, G. (2006). Performance analysis of a bufferless OBS node considering the streamline effect, *Communications Letters, IEEE* **10**(4): 293–295.

Poíro, M. and Medhi, D. (2004). *Routing, Flow, and Capacity Design in Communication and Computer Networks*, Elsevier.

Praveen, B., Praveen, J. and Murthy, C. S. R. (2006). A survey of differentiated QoS schemes in optical burst switched networks, *Optical Switching and Networking* **3**(2): 134 – 142.

Pritsker, A. A. B. (1998). *Handbook of Simulation - Principles, Methodology, Advances, Applications and Practice.*, EMP Books and John Wiley & Sons, Inc., chapter Principles of Simulation Modeling, pp. 31–51.

Qiao, C. (2000). Labeled optical burst switching for IP-over-WDM integration, *Communications Magazine, IEEE* **38**(9): 104–114.

Qiao, C. and Yoo, M. (1999). Optical burst switching (OBS) - a new paradigm for an optical internet, J. High Speed Networks (JHSN), vol. 8, no. 1, pp. 69–84.

Raisanen, V. (2003). *Implementing Service Quality in IP Networks*, John Wiley & Sons.

Rodrigo, M. d. V. and Gotz, J. (2004). An analytical study of optical burst switching aggregation strategies, *Proceedings of IEEE/SPIE Third International Workshop on Optical Burst Switching (OBSS 2004)*, San Jose, CA (USA).

Rodrigues, J., Freire, M., Garcia, N. and Monteiro, P. (2007). Enhanced just-in-time: A new resource reservation protocol for optical burst switching networks, *Proc. 12th IEEE Symposium on Computers and Communications ISCC 2007*, pp. 121–126.

Rodrigues, J. J. P. C., Freire, M. M. and Lorenz, P. (2005). The role of meshing degree in optical burst switching networks using signaling protocols with one-way reservation schemes, *Lecture Notes in Computer Science* **3420**: 44–51.

Rodrigues, J. J. P. C., Freire, M. M., Monteiro, P. P. and Lorenz, P. (2005). *Encyclopedia of Multimedia Technology and Networking*, Idea Group Reference, chapter Optical burst switching: A new switching paradigm for high-speed Internet.

Rosberg, Z. and Vu, H. L. (2007). Design and performance of fdl buffers in optical switches, *Proc. 9th International Conference on Transparent Optical Networks ICTON '07*, Vol. 3, pp. 213–217.

Rosson, M. B. and Alpert, S. R. (1990). The cognitive consequences of object-oriented design, *Human Computer Interaction Journal* **5**(4): 345–379.

Sarwar, S., Wallentin, L., Franzl, G. and van As, H. R. (2008). Composite burst assembly with high-priority packets in the middle of burst, *Proc. 5th International Conference on Broadband Communications, Networks and Systems BROADNETS 2008*, pp. 140–145.

Sekercioglu, Y. A., Varga, A. and Egan, G. K. (2003). Parallel simulation made easy with omnet++, *Proceedings of the Europeean Simulation Symposium (ESS '03)*, Delft, The Netherlands.

Semaphore (1995). Glossary of object-oriented terminology.

Senior, J. M. (2008). *Optical Fiber Communications: Principles and Practice*, 3/e edn, Prentice Hall.

Shen, G., Bose, S., Cheng, T. H., Lu, C. and Chai, T. Y. (2001). Performance study on a WDM packet switch with limited-range wavelength converters, *IEEE Communications Letters* **5**(10): 432–434.

Simmons, J. M. (2008). *Optical Network Design and Planning*, Springer.

Tachibana, T. and Kasahara, S. (2006). Burst-cluster transmission: service differentiation mechanism for immediate reservation in optical burst switching networks, *IEEE Communications Magazine* **44**(5): 46–55.

Tan, C.-W., Gurusamy, M. and Lui, J.-S. (2004). Achieving proportional loss differentiation using probabilistic preemptive burst segmentation in optical burst switching WDM networks, *Proc. IEEE Global Telecommunications Conference GLOBECOM '04*, Vol. 3, pp. 1754–1758 Vol.3.

Tan, C. W., Mohan, G. and Lui, J.-S. (2006). Achieving multi-class service differentiation in WDM optical burst switching networks: A probabilistic preemptive burst segmentation scheme, *IEEE Journal on Selected Areas in Communications* **24**(12): 106–119.

Tanenbaum, A. S. (2002). *Computer Networks*, Prentice Hall.

Teng, J. and Rouskas, G. (2005a). Wavelength selection in OBS networks using traffic engineering and priority-based concepts, *IEEE Journal on Selected Areas on Communications* **23**(8): 1658–1669.

Teng, J. and Rouskas, G. N. (2005b). Traffic engeneering approach to path selection in optical burst switching networks, *OSA Journal of Optical Networking* **4**(11): 759–777.

Thodime, G., Vokkarane, V. and Jue, J. (2003). Dynamic congestion-based load balanced routing in optical burst-switched networks, *Proc. IEEE Global Telecommunications Conference GLOBECOM '03*, Vol. 5, pp. 2628–2632 vol.5.

Turner, J. S. (1999). Terabit burst switching, *J. High Speed Networks* **8**(1): 3–16.

Varga, A. (1998). Parametrized topologies for simulation programs, *Technical report*, Department of Telecommunications - Technical University of Budapest.

Varga, A. (2005). *OMNeT++ User Manual (Version 3.2).* Accessed on February 2009.
URL: *http://www.omnetpp.org/doc/manual/usman.html*

Varga, A. and Hornig, R. (2008). An overview of the omnet++ simulation environment, *SIMUTools 2008, Proceedings of*, ICST, Marseille, France.

Verma, S., Chaskar, H. and Ravikanth, R. (2000). Optical burst switching: a viable solution for terabit IP backbone, *Network, IEEE* **14**(6): 48–53.

Vokkarane, V. and Jue, J. (2005). Segmentation-based nonpreemptive channel scheduling algorithms for optical burst-switched networks, *IEEE/OSA Journal of Lightwave Technology* **23**(10): 3125–3137.

Vokkarane, V., Jue, J. and Sitaraman, S. (2002). Burst segmentation: an approach for reducing packet loss in optical burst switched networks, *Communications, 2002. ICC 2002. IEEE International Conference on*, Vol. 5, pp. 2673–2677 vol.5.

Vokkarane, V. M. (2007). Intermediate-node-initiation (ini): A generalized signaling framework for optical burst-switched networks, *Optical Switching and Networking* **4**(1): 20 – 32.

Vokkarane, V. M., Haridoss, K. and Jue, J. P. (2002). Threshold-based burst assembly policies for QoS support in optical burst-switched networks, *in Proc. SPIE OptiComm 2002*, pp. 125–136.

Vokkarane, V., Thodime, G., Challagulla, V. and Jue, J. (2003). Channel scheduling algorithms using burst segmentation and fdls for optical burst-switched networks, *Proc. IEEE International Conference on Communications ICC '03*, Vol. 2, pp. 1443–1447 vol.2.

Vokkarane, V., Vokkarane, V. and Jue, J. (2003). Prioritized burst segmentation and composite burst-assembly techniques for QoS support in optical burst-switched networks, *IEEE Journal on Selected Areas on Communications* **21**(7): 1198–1209.

Vokkarane, V., Vokkarane, V., Zhang, Q., Jue, J. and Chen, B. (2002). Generalized burst assembly and scheduling techniques for QoS support in optical burst-switched networks, *Proc. IEEE Global Telecommunications Conference GLOBECOM '02*, Vol. 3, pp. 2747–2751 vol.3.

Vu, H. L. and Zukerman, M. (2002). Blocking probability for priority classes in optical burst switching networks, *IEEE Communications Letters* **6**(5): 214–216.

Wang, X., Jiang, X. and Horiguchi, S. (2008). A construction of shared optical buffer queue with switched delay lines, *Proc. International Conference on High Performance Switching and Routing HSPR 2008*, pp. 86–91.

Wei, J. and McFarland, R.I., J. (2000). Just-in-time signaling for WDM optical burst switching networks, *Lightwave Technology, Journal of* **18**(12): 2019–2037.

Weisfeld, M. (2000). Thinking in objects.
URL: *http://www.developer.com/design/article.php/3347291*

Weldon, M. K., van Landegem, T. and Szurkowski, E. S. (2008). Next-generation access networks: A preview, *Bell Labs Technical Journal* **13**(1): 1–10.

Wen, H., Song, H., Li, L. and Wang, S. (2003). Load-balancing contention resolution in LOBS based on GMPLS, *Parallel and Distributed Computing, Applications and Technologies, 2003. PDCAT'2003. Proceedings of the Fourth International Conference on* pp. 590–594.

Wen, H., Song, H., Li, L. and Wang, S. (2005). Load-balancing contention resolution in OBS networks based on GMPLS, *Int. J. High Perform. Comput. Netw.* **3**(1): 25–32.

Widjaja, I. (1995). Performance analysis of burst admission-control protocols, *IEE Proceedings-Communications* **142**(1): 7–14.

Willis, P. J. (2005). An introduction to quality of service, *BT Technology Journal* **23**(2): 13–27.

Willner, A. E., Khosravani, R. and Kumar, S. (2006). *Optical Switching*, Springer, chapter Optical Packet Switching and Optical Burst Switching, pp. 405–429.

www (2009a). Erbium-doped fiber amplifier (EDFA), http://en.wikipedia.org/wiki/EDFA#Erbium-doped_fibre_amplifiers. Accessed on February 2009.

www (2009b). ILOG CPLEX home page, http://www.ilog.com/products/cplex. Accessed on February 2009.

www (2009c). J-Sim home page, http://www.j-sim.org. Accessed on February 2009.

www (2009d). JiST home page, http://jist.ece.cornell.edu. Accessed on February 2009.

www (2009e). NS-2, home page, http://nsnam.isi.edu/nsnam/index.php/User_Information. Accessed on February 2009.

www (2009f). OMNeT++ home page, http://www.omnetpp.org. Accessed on February 2009.

www (2009g). OPNET home page, http://www.opnet.com. Accessed on February 2009.

www (2009h). Optical add/drop multiplexer (OADM), http://en.wikipedia.org/wiki/OADM. Accessed on February 2009.

www (2009i). Optical amplifier (OAMP), http://en.wikipedia.org/wiki/Optical_amplifier. Accessed on February 2009.

www (2009j). Optical cross-connector (OXC), http://en.wikipedia.org/wiki/OXC. Accessed on February 2009.

www (2009k). Optical network - technical concepts, http://www.optical-network.com/terminology.php. Accessed on February 2009.

www (2009l). QualNet home page, http://www.scalable-networks.com. Accessed on February 2009.

www (2009m). SSFNet home page, http://www.ssfnet.org/homePage.html. Accessed on February 2009.

Xi, K., Arakawa, S. and Murata, M. (2005). How many wavelength converters do we need?, *Proc. Conference on Optical Network Design and Modeling*, pp. 347–358.

Xia, F., Sekaric, L. and Vlasov, Y. (2007). Ultracompact optical buffers on a silicon chip, *Nat Photon* **1**(1): 65.

Xiao, G., Lu, K. and Chlamtac, I. (2004). An evaluation of distributed wavelength provisioning in WDM optical networks with sparse wavelength conversion, *IEEE/OSA Journal of Lightwave Technology* **22**(7): 1668–1678.

Xiong, Y., Vandenhoute, M. and Cankaya, H. (2000). Control architecture in optical burst-switched WDM networks, *Selected Areas in Communications, IEEE Journal on* **18**(10): 1838–1851.

Xu, J., Qiao, C., Li, J. and Xu, G. (2003). Efficient channel scheduling algorithms in optical burst switched networks, *Proc. INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE*, Vol. 3, pp. 2268–2278 vol.3.

Xu, J., Qiao, C., Li, J. and Xu, G. (2004). Efficient burst scheduling algorithms in optical burst-switched networks using geometric techniques, *Selected Areas in Communications, IEEE Journal on* **22**(9): 1796–1811.

Xu, L., Perros, H. and Rouskas, G. (2001). Techniques for optical packet switching and optical burst switching, *Communications Magazine, IEEE* **39**(1): 136–142.

Yang, L., Jiang, Y. and Jiang, S. (2003). A probabilistic preemptive scheme for providing service differentiation in OBS networks, *Proc. IEEE Global Telecommunications Conference GLOBECOM '03*, Vol. 5, pp. 2689–2693 vol.5.

Yang, L. and Rouskas, G. (2006). Adaptive path selection in OBS networks, *Lightwave Technology, Journal of* **24**(8): 3002–3011.

Yang, M., Zheng, S. and Verchere, D. (2001). A QoS supporting scheduling algorithm for optical burst switching WDMM networks, *Proc. IEEE Global Telecommunications Conference GLOBECOM '01*, Vol. 1, pp. 86–91 vol.1.

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

Yoo, M. and Qiao, C. (1997). Just-enough-time (jet): a high speed protocol for bursty traffic in optical networks, *Vertical-Cavity Lasers, Technologies for a Global Information Infrastructure, WDM Components Technology, Advanced Semiconductor Lasers ..., Gallium Nitride Materials, Processing, ..., 1997 Digest of the IEEE/LEOS Summer Topical Meetings*, pp. 26–27.

Yoo, M. and Qiao, C. (1998). New optical burst-switching protocol for supporting quality of service, *in* J. M. Senior and C. Qiao (eds), *All-Optical Networking: Architecture, Control, and Management Issues*, Vol. 3531, SPIE, pp. 396–405.

Yoo, M. and Qiao, C. (1999). Supporting multiple classes of services in IP over WDM networks, *Proc. Global Telecommunications Conference GLOBECOM '99*, Vol. 1B, pp. 1023–1027 vol. 1b.

Yoo, M., Qiao, C. and Dixit, S. (2000). QoS performance of optical burst switching in IP-over-WDM networks, *IEEE Journal on Selected Areas in Communications* **18**(10): 2062–2071.

Yu, X., Chen, Y. and Qiao, C. (2002). A study of traffic statistics of assembled burst traffic in optical burst switched networks, *In Proceedings of Opticomm*, pp. 149–159.

Yu, X., Li, J., Cao, X., Chen, Y. and Qiao, C. (2004). Traffic statistics and performance evaluation in optical burst switched networks, *IEEE/OSA Journal of Lightwave Technology* **22**(12): 2722–2738.

Zapata-Beghelli, A. and Bayvel, P. (2008). Dynamic versus static wavelength-routed optical networks, *IEEE/OSA Journal of Lightwave Technology* **26**(20): 3403–3415.

Zhang, J., Wang, S., Zhu, K., Datta, D., Kim, Y.-C. and Mukherjee, B. (2004). Pre-planned global rerouting for fault management in labeled optical burst-switched WDM networks, *Proceedings IEEE GLOBECOM 2004*, Vol. 3, pp. 2004–2008.

Zhang, Q., Vokkarane, V., Chen, B. and Jue, J. (2003a). Early drop and wavelength grouping schemes for providing absolute QoS differentiation in optical burst-switched networks, *Proc. IEEE Global Telecommunications Conference GLOBECOM '03*, Vol. 5, pp. 2694–2698 vol.5.

Zhang, Q., Vokkarane, V., Chen, B. and Jue, J. (2003b). Early drop scheme for providing absolute QoS differentiation in optical burst-switched networks, *Proc. HPSR High Performance Switching and Routing Workshop on*, pp. 153–157.

Zhang, Q., Zhang, Q., Vokkarane, V., Jue, J. and Chen, B. (2004). Absolute QoS differentiation in optical burst-switched networks, *IEEE Journal on Selected Areas on Communications* **22**(9): 1781–1795.

Zhang, Z. and Yang, Y. (2006). Performance modeling of bufferless WDM packet switching networks with limited-range wavelength conversion, *IEEE Transactions on Communications* **54**(8): 1473–1480.

UAlg
UNIVERSIDADE DO ALGARVE

30 anos
1979 | 2009