Architectures and Protocols for Sub-Wavelength Optical Networks: Contributions to Connectionless and Connection-Oriented Data Transport

by

Joan Triay Marquès

Director:
Cristina Cervelló i Pastor

A Ph.D. Thesis
submitted to the
Department of Telematics Engineering

in partial fulfillment of the requirements for the degree of

Ph.D. in Telematics Engineering

Universitat Politècnica de Catalunya (UPC)

Barcelona, Spain, September 2011

Jury:

Director:	
Dr. Cristina Cervelló i Pastor	Universitat Politècnica de Catalunya (Spain)
External Reviewers:	
Dr. Carla Raffaelli	Università di Bologna (Italy)
Dr. Massimo Tornatore	Politecnico di Milano (Italy)
Defense Committee:	
Prof. Maurice Gagnaire	Telecom ParisTech (E.N.S.T) (France)
Prof. Francisco Javier González Castaño	Universidad de Vigo (Spain)
Prof. Sebastià Sallent Ribes	Universitat Politècnica de Catalunya (Spain)

Abstract

The rapid evolving Internet and the broad range of new data applications (e.g., multimedia, video-conference, online gaming, etc.) is fostering revolutionary changes in the way we communicate. In addition, some of these applications demand for unprecedented amounts of bandwidth resources with diverse quality of service (QoS).

The development of wavelength division multiplexing (WDM) in the 90's made very cost-effective the availability of bandwidth. Nowadays, optical circuit switching technologies are predominant in the core enabling the set up of lightpaths across the network. However, full-wavelength lightpath granularity is too coarse, which results to be inefficient for provisioning sub-wavelength channels. As remarked by the research community, an open issue in optical networking is increasing the protocol transparency as well as provisioning true dynamic bandwidth allocation at the network level. To this end, new architectures are required. Optical burst/packet switching (OBS/OPS) are two such proposed technologies under investigation.

This thesis contributes with three network architectures which aim at improving the sub-wavelength data transport from different perspectives. First, we gain insight into the connectionless nature of OBS. Here, the network dynamics are increased due to the short-lived burst transmissions. Moreover, burst contentions degrade the performance even at very low loads. To cope with them, we propose a proactive resolution scheme by means of a distributed auto load-balancing routing and wavelength assignment (RWA) algorithm for wavelength-continuity constraint networks. In this protocol, the RWA and burst forwarding is based on the exploitation and exploration of switching rule concentration values that incorporate contention and forwarding desirability information. To support such architecture, forward and backward control packets are used in the burst forwarding and updating rules, respectively. In order to analyze the benefits of the new algorithm, four different network topologies are used. Results indicate that the proposed method outperforms the rest of tested RWA algorithms at various margins depending on the topology without penalizing other parameters such as end-to-end delay.

The second contribution proposes a hybrid connectionless and connection-oriented architecture based on a medium access control (MAC) protocol for OBS networks (DAOBS). The MAC provides two main access mechanisms: queue arbitrated (QA) for connectionless bursts and pre-arbitrated (PA) for TDM connection-oriented services. Such an architecture allows for a broad range of delay-sensitive applications or guaranteed services. Results evaluated through simulations show that in the QA access mode highest priority bursts are guaranteed zero losses and very low access latencies. Regarding the PA mode, we report that doubling the offered TDM traffic load increases in more than one order their connection blocking, slightly affecting the blocking of other connectionless bursts. In this chapter, we also tackle two of the issues related with the DAOBS architecture and its operation. Firstly, we model mathematically the lower and upper approximations of the access delay as a consequence of the connectionless queue arbitrated access. Secondly, we formulate the generation of the virtual light-tree overlay topology for the static traffic case.

Finally, the last contribution explores the benefits of a path computation element (PCE)-enabled time-shared optical network (TSON) that uses a centralized sub-wavelength assignment element to schedule the connections avoiding collisions within the network. This architecture enables guaranteed sub-wavelength bandwidth provisioning in a connection-oriented way. We propose three different GMPLS/PCE/TSON architectures and assess them through simulation. Due to the centralized approach, the network performance highly depends on the connection allocation and provisioning. To this end, several time-slot assignment algorithms are proposed and compared against their integer linear programming (ILP) formulation for a static example case. For the dynamic case, simple heuristics on a dual path and sub-wavelength computation engine are also implemented and assessed through simulation. Results show the benefits of providing flexibility both in time and across wavelengths when allocating the sub-wavelength time-slots.

Keywords: Optical WDM network, sub-wavelength provisioning, optical burst switching, MAC, RWA, connectionless, connection-oriented, network architecture.

Resum

La ràpida evolució d'Internet i l'àmplia gamma de noves aplicacions (per exemple, multimèdia, videoconferència, jocs en línia, etc.) ha fomentat canvis revolucionaris en la manera com ens comuniquem. A més, algunes d'aquestes aplicacions demanden grans quantitats de recursos d'ample de banda amb diversos requeriments de qualitat de servei (QoS).

El desenvolupament de la multiplexació per divisió de longitud d'ona (WDM) en els anys noranta va fer molt rendible la disponibilitat d'ample de banda. Avui dia, les tecnologies de commutació òptica de circuits són predominants en el nucli de la xarxa, les quals permeten la configuració de canals (lightpaths) a través de la xarxa. No obstant això, la granularitat d'aquests canals ocupa tota la longitud d'ona, el que fa que siguin ineficients per a proveir canals de menor ample de banda (sub-longitud d'ona). Segons la comunitat científica, és necessari augmentar la transparència dels protocols, així com millorar l'aprovisionament d'ample de banda de forma dinàmica. Per tal de fer això realitat, és necessari desenvolupar noves arquitectures. La commutació òptica de ràfegues i de paquets (OBS/OPS), són dues de les tecnologies proposades.

Aquesta tesi contribueix amb tres arquitectures de xarxa destinades a millorar el transport de dades sub-longitud d'ona. En primer lloc, aprofundim en la naturalesa sense connexió en OBS. En aquest cas, la xarxa incrementa el seu dinamisme a causa de les transmissions a ràfega. A més, les col·lisions entre ràfegues degraden el rendiment de la xarxa fins i tot a càrregues molt baixes. Per fer front a aquestes col·lisions, es proposa un esquema de resolució de col·lisions pro actiu basat en un algorisme d'encaminament i assignació de longitud d'ona (RWA) que balanceja de forma automàtica i distribuïda la càrrega en la xarxa. En aquest protocol, el RWA i la transmissió de ràfegues es basen en l'explotació i exploració de regles de commutació que incorporen informació sobre contencions i encaminament. Per donar suport a aquesta arquitectura, s'utilitzen dos tipus de paquets de control per a l'encaminament de les ràfegues i l'actualització de les regles de commutació, respectivament. Per analitzar els beneficis del nou algorisme, s'utilitzen quatre topologies de xarxa diferents. Els resultats indiquen que el mètode proposat millora en diferents marges la resta d'algorismes RWA en funció de la topologia i sense penalitzar altres paràmetres com el retard extrem a extrem.

La segona contribució proposa una arquitectura híbrida sense i orientada a connexió sobre la base d'un protocol de control d'accés al medi (MAC) per a xarxes OBS (DAOBS). El MAC ofereix dos mètodes d'accés: arbitratge de cua (QA) per a la transmissió de ràfegues sense connexió, i pre-arbitratge (PA) per serveis TDM orientats a connexió. Aquesta arquitectura permet una àmplia gamma d'aplicacions sensibles al retard i al bloqueig. Els resultats avaluats a través de simulacions mostren que en l'accés QA, les ràfegues de més alta prioritat tenen garantides zero pèrdues i latències d'accés molt baixes. Pel que fa a l'accés PA, es reporta que la duplicació de la càrrega TDM augmenta en més d'un ordre la probabilitat de bloqueig, però sense afectar en la mateixa mesura les ràfegues sense connexió. En aquest capítol també es tracten dos dels problemes relacionats amb l'arquitectura DAOBS i el seu funcionament. En primer lloc, es proposa un model matemàtic per aproximar el retard d'accés inferior i superior com a conseqüència de l'accés QA. En segon lloc, es formula matemàticament la generació i optimització de les topologies virtuals que suporten el protocol per a l'escenari amb tràfic estàtic.

Finalment, l'última contribució explora els beneficis d'una arquitectura de xarxa òptica per temps compartit (TSON) basada en elements de càlcul de camins (PCE) centralitzats per tal d'evitar col·lisions en la xarxa. Aquesta arquitectura permet garantir l'aprovisionament orientat a connexió de canals sub-longitud d'ona. En aquest capítol proposem i simulem tres arquitectures GMPLS/PCE/TSON. A causa del enfocament centralitzat, el rendiment de la xarxa depèn en gran mesura de l'assignació i aprovisionament de les connexions. Amb aquesta finalitat, es proposen diferents algorismes d'assignació de ranures temporals i es comparen amb les corresponents formulacions de programació lineal (ILP) per al cas estàtic. Per al cas de tràfic dinàmic, proposem i avaluem mitjançant simulació diferents heurístiques. Els resultats mostren els beneficis de proporcionar flexibilitat en els dominis temporal i freqüencial a l'hora d'assignar les ranures temporals.

Keywords: Xarxa WDM òptica, aprovisionament sub-longitud d'ona, commutaciò de ràfegues òptiques, MAC, RWA, sense connexió, orientat a connexió, arquitectura de xarxa.

Acknowledgments

First and foremost, thanks to my family. Although my dad and mum know little about my work technically speaking, they are the main reason behind this Thesis. Had not been for them, I wouldn't have had the opportunity to get in college and make most of my dreams come true. Thanks for their kindness, support and true love.

Although in the authorship of this Thesis there is just one name -mine- this work would not have been possible without the help from many other people. I will always be thankful to my director, Dr. Cristina Cervelló. Without her aid and guidance, this work would have never materialized. I would also like to thank Prof. Sebastià Sallent and the Fundació i2CAT, particularly Sergi Figuerola, who gave me the opportunity to get involved in my first research and development projects.

I thank Dr. Georgios Zervas and Prof. Dimitra Simeonidou from University of Essex. Throughout my visiting research stay in Essex, I learnt a lot, much more than they would ever think. Georgios provided me many good ideas to further develop which made a huge impact on this Thesis and the publications I had the pleasure to coauthor with them. They also opened me to their circle of friends and made my stay in UK one of the best times in my life. Thanks Eduard, Chinwe, Yixuan and many others for all the great moments in Essex.

I also thank Dr. Vinod Vokkarane and all the team members in the Netlab group at University of Massachusetts, Dartmouth. My visiting research stay in USA helped me sharpen my goals and face the problems from other points of view. The experience I gained in USA will be one my greatest assets. Thanks to Fulbright for sponsoring my research at UMASS and make my stay in USA an unforgettable experience, both professionally and personally.

Last but not least, thanks to Salvador. Despite the ups and downs that this Thesis originated, he has been an endless source of support and understanding. Thanks for encouraging me to overcome all the difficulties I faced over the last four years.

I also acknowledge the Government of Catalonia and the European Social Fund from which I got the financial support through a Ph.D. scholarship.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Contributions 2

1.3 Organization of this Thesis 3

2 Towards Sub-Wavelength Optical Networks: A Background Perspective **5**

2.1 Introduction 5

2.2 Open Issues in Optical Networking 9

2.3 Introduction to Next-Generation Optical Networks 10

2.4 Connectionless Sub-Wavelength Optical Networks 13

2.4.1 Reservation Signaling Mechanisms in OBS 13

2.4.2 Contention Resolution Mechanisms 15

2.4.3 Quality of Service in OBS 16

2.4.4 Routing and Wavelength Assignment 17

2.4.5 Medium Access Control in Sub-Wavelength Optical Networks . . 19

2.5 Towards Guaranteed Sub-Wavelength Transmission 20

2.5.1 Hybrid Connectionless and Connection-Oriented Sub-Wavelength Data Transport 20

2.5.2 Guaranteed Sub-Wavelength Transmission 21

2.5.3 GMPLS/OBS Interworking Multi-Layer Models 22

2.5.4 Connection Sub-Wavelength Time-Slot Assignment 25

2.6 Summary 26

3 RWA for Connectionless Sub-Wavelength Optical Networks **29**

3.1 Introduction 29

3.2 Ant Colony Optimization 30

3.3 Algorithm Implementation 31

3.3.1 Data Structures 31

3.3.2 Algorithm Implementation in OBS 34

3.4 Routing and Wavelength Assignment Algorithm 34

3.4.1 Initialization Algorithm 35

3.4.2 ALBA Routing Algorithm 36

3.5 Wavelength Assignment, Routing and Updating Rule Formulation . . . 38

3.5.1 Initial Wavelength and Routing Selection 38

3.5.2 State Transition Rule 39
3.5.3 Local Updating Rule 40
3.5.4 Global Updating Rule 40
3.5.5 Evaluation of the Updating Functions 41
3.6 Simulation Results 44
3.6.1 Results on the Fish-Like Network 45
3.6.2 Results on the NSFNET 47
3.6.3 Network Topology Performance Comparison 50
3.7 Summary 54

4 Hybrid Connectionless and Connection-Oriented MAC-based Sub-Wavelength Optical Networks 55
4.1 Introduction 55
4.2 Hybrid CL/CO QoS-enabled MAC for OBS 57
4.2.1 Architecture and Basic Operation 58
4.2.2 Queue Arbitrated Access 61
4.2.2.1 Queue Arbitrated Operation Mode 62
4.2.2.2 Queue Arbitrated Module Architecture 62
4.2.2.3 Queue Arbitrated State Machine 64
4.2.2.4 Queue Arbitrated Wavelength Assignment 65
4.2.3 Pre-arbitrated Access 66
4.2.3.1 Pre-arbitrated Operation Mode 66
4.2.3.2 PA/QA Slot Scheduling 68
4.3 Simulation Results 70
4.3.1 Queue Arbitrated Access Mode Without QoS 71
4.3.2 Queue Arbitrated Access Mode With QoS 72
4.3.3 Results With Mixed Connectionless and Connection-Oriented Burst Services 76
4.3.4 PA/QA Access Mode Transmission Decision 79
4.4 Delay Model of a DAOBS Lightpath 81
4.4.1 Mean Delay in the Local Queue 82
4.4.2 Mean Delay in the Distributed Queue 83
4.4.3 Numerical Results 88
4.5 Light-Tree Topology Optimization Without Wavelength Conversion . . . 88
4.5.1 General Problem Statement and Formulation 90
4.5.2 Problem Formulation Without Fractional Capacities 91
4.5.3 Problem Formulation With Sub-wavelength Light-tree Capacity Demands 94
4.5.4 ILP Results 97
4.6 Summary 99

5 Guaranteed Sub-Wavelength Provisioning on Time-Shared Optical Network 101
5.1 Introduction 101
5.2 Related Work on Sub-Wavelength Network Provisioning 104
5.3 PCE/TSON Architectures 106
5.3.1 Time-shared Optical Network Data Plane 107
5.3.2 Daisy-chain Centralized SLAE Time-shared Optical Network . . 108

5.3.3 Daisy-chain Distributed Time-shared Optical Network 110
5.3.4 Tree-based Time-shared Optical Network 110
5.3.5 Simulation Performance GMPLS/PCE/TSON Architectures . . 112
5.4 Sub-Wavelength Assignment General Problem Statement 116
5.4.1 Problem Formulation 118
5.4.2 Multi-Wavelength Non-Continuous Policy 120
5.4.3 Tunable Non-Continuous Policy 122
5.4.4 Single-Wavelength Non-Continuous Policy 123
5.4.5 Single-Wavelength Continuous Policy 123
5.4.6 Tunable Continuous Policy 126
5.4.7 ILP Numerical Results 127
5.5 Heuristics for the Static Scheduling Scenario 130
5.5.1 SWFFC: Heuristic for the SW-C Policy 130
5.5.2 LUSCT: Heuristic for the T-C Policy 133
5.5.3 Heuristic Numerical Results Under the Static Traffic Scenario . . 134
5.5.3.1 ILP and Heuristics Performance Comparison 135
5.5.3.2 Heuristics Performance for Larger Scenario 135
5.6 PCE and SLAE Routing, Wavelength and Time-slot Assignment in TC-TSON 141
5.6.1 SWFFC Algorithm 142
5.6.2 LUSCT Algorithm 145
5.7 Simulation Results 146
5.7.1 Blocking Probability and Throughput Performance 148
5.7.2 Average Lightpath Length 151
5.7.3 Wavelengths Used per Connection 154
5.7.4 Fragmentation per Connection 155
5.8 Summary 157

6 Conclusion 159

7 Future Research and Work 163
7.1 Chapter 3: ALBA 163
7.2 Chapter 4: DAOBS 164
7.3 Chapter 5: TSON 164

8 Publications 167
8.1 Publications Related to this Thesis 167
8.2 Other Publications 169

Bibliography 171

List of Figures

2.1 Cisco VNI global consumer Internet traffic forecast. Source [1]. 6

2.2 Schematic of network. 7

2.3 IP-over-WDM evolution. 8

2.4 OBS's Data and Control planes. 12

2.5 Examples of reservation signaling: (a) explicit setup and explicit release, and (b) estimated setup and estimated release. 14

2.6 Comparison of packet loss probability results of different contention resolution schemes: (a) with wavelength conversion, (b) with optical buffering and deflection routing, and (c) with a combination of schemes. Source [2]. 16

2.7 A light-trail. Based on [3]. 23

2.8 GMPLS/OBS control plane inter-working. 24

3.1 Optical switch with 3 links and 4 wavelengths. 32

3.2 OBS core node architecture with offset time emulation. Source [4]. . . . 33

3.3 High-level RWA algorithm diagram. 35

3.4 Evaluation of (3.2) with $\beta = 1.0$ and $\beta = 2.0$. 39

3.5 Global updating rule analysis with two different SRC values (0.1 and 1.0). 42

3.6 Function evaluation of (3.19). 44

3.7 Fish-like network with 8 nodes. 45

3.8 Simulations on the fish-like network with 2 wavelengths: (a) burst blocking comparison of different RWA protocols, (b) link and wavelength utilization of the ALBA protocol, (c) burst blocking of ALBA with $\alpha_1 = 0.001$ and $\beta = 1.0$, and (d) burst blocking of ALBA with $\alpha_1 = 0.001$ and $\beta = 2.0$. 46

3.9 Convergence time to cancel burst collisions on the fish-like network with 8 nodes and 2 wavelengths using ALBA: (a) as a function of ρ_1 and β with $\alpha_1 = 0$, (b) as a function of ρ_1 and β with $\alpha_1 = 0.001$, and (c) as a function of ρ_1 and β with $\alpha_1 = 0.01$. 48

3.10 ALBA burst blocking probability evaluation on the NSFNET with 16 wavelengths: (a) as a function of ρ_1 with $\alpha_1 = 0.001$ and $\beta = 3.0$; and (b) as a function of β with $\alpha_1 = 0.001$ and $\rho_1 = 0.001$. 49

3.11 Set of network topologies under test: (a) NSFNET, (b) SmallNET, (c) EON, and (d) Waxman Random. 51

3.12 Burst blocking probability on different network topologies: (a) NSFNET, (b) EON, (c) Waxman Random, and (d) SmallNet. 52
3.13 Comparison among topologies: (a) blocking probability vs. network load, and (b) mean route length (MRL). 53
3.14 Evaluation of the average Δl value: (a) in global updating rule with load = 5.0 Er/wl, and (b) in local updating rule with load = 5.0 Er/wl. . . . 53

4.1 DAOBS protocol and network architecture. 58
4.2 Example of a DAOBS light-tree. If node N1 needs to transmit to any ToB, it requests for free slots to HoB, or eventually waits for a free slot coming on the downstream direction. 59
4.3 Data channel PA/QA slot example. 60
4.4 Format of the control packets: (a) burst control packet, and (b) request control packet. 61
4.5 Flow diagram for the HoB slot processing. 61
4.6 DAOBS QA modular architecture. Entity with three burst priorities on wavelength λ_k. 63
4.7 QA DASM flow at priority i. 64
4.8 Example of a multi-layer PA service setup and allocation of slots in the MAC layer. 67
4.9 Pre-arbitrated slot scheduling algorithms. 69
4.10 QA access mode without QoS for different LQ lengths: (a) burst blocking probability, and (b) mean access delay. 72
4.11 QA access mode with QoS and LQ=5 bursts for the two distributions in Table 4.1: (a) burst blocking probability, and (b) access delay. 73
4.12 QA access mode with QoS and LQ=5 bursts for the first distribution in Table 4.1: (a) HoB transmission probability, (b) average path length, and (c) end-to-end delay. 74
4.13 QA access mode burst blocking probability as a function of the network load. 75
4.14 Dual PA/QA access modes: (a) Mean burst/call blocking probability (BP) in traffic distribution 20-80%, (b) BP with traffic distribution 40-60%, (c) Mean access delay for connectionless bursts and mean connection setup delay in traffic distribution 20-80%, and (d) HoB connectionless burst transmission and HoB connection setup rates in traffic distribution 20-80%. 77
4.15 BP for different connection types in traffic distribution 20-80%. 79
4.16 BBP comparison between standalone QA access with QoS and PA/QA dual access modes. 80
4.17 DAOBS queue model. 82
4.18 Waiting delay faced by an arriving burst. 83
4.19 Access delay: (a) lightpath and ring of 8 nodes and 4 wavelengths, (b) lightpath and ring of 16 nodes with 8 wavelengths, and (c) lightpaths setup on NSFNET. 89
4.20 Optimization results with full-wavelength capacity demands. 97
4.21 Optimization results with sub-wavelength demands and switching constraints for *case (a)*. 98

4.22 Optimization results with sub-wavelength demands and switching constraints for *case (b)*: (a) original light-trees generated, and (b) aggregated light-trees. 99

5.1 Time-shared optical network application scope. 102
5.2 Alternative architectures. 105
5.3 Client frame aggregation. 108
5.4 DC-TSON architecture: connection request setup. 109
5.5 DD-TSON architecture: connection request setup. 110
5.6 TC-TSON architecture: connection request setup. 111
5.7 (a) Madrid's network topology, and (b) network node architecture used throughout simulation. 113
5.8 Performance results comparison on a metro network with 16 wavelengths: (a) mean connection blocking probability vs. offered load, and (b) mean connection blocking probability vs. network load. 114
5.9 Performance results comparison on a metro network with 16 wavelengths: mean connection setup (time-to-service) delay. 115
5.10 Performance results comparison on a metro network with 16 wavelengths: (a) mean lightpath length vs. offered load, and (b) mean packet end-to-end delay vs. offered load. 116
5.11 Node's normalized add port utilization. 116
5.12 Time-slot assignment policies: (a) MW-NC, (b) SW-NC, (c) SW-C, (d) T-NC, and (e) T-C. 118
5.13 Continuous scheduling constraint. 124
5.14 6-node topology. 127
5.15 ILP performance results: (a) connection blocking probability, (b) average lightpath cost, and (c) average lightpath length. 129
5.16 Heuristics performance results: (a) connection blocking probability, (b) average lightpath cost, and (c) average lightpath length. 136
5.17 Heuristic results on 6-node network with 8 wavelengths: (a) BP vs. number of connections for k=1, and (b) for k=3. 138
5.18 Heuristic results on 6-node network with 8 wavelengths: (a) average lightpath length (in number of hops) vs. number of connections for k=1, and (b) for k=3. 139
5.19 Heuristic results on 6-node network with 8 wavelengths among the proposed heuristics: (a) average lightpath length (in number of hops) vs. number of connections for k=1, and (b) for k=3. 139
5.20 Fragmentation vs. number of used wavelengths. 140
5.21 Heuristic results on 6-node network with 8 wavelengths: (a) average number of fragments per connection vs. number of connections for k=1, and (b) for k=3; and (c) average number of wavelengths used per connection vs. number of connections for k=1, and (d) for k=3. 141
5.22 PCE and SLAE tandem time-shared bandwidth computation. 142
5.23 Simulation topologies for the dynamic scenario: (a) TID network, and (b) EON. 147
5.24 Connection blocking probability results in TID: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths. 149

5.25 Blocking probability wavelength comparison of (a) SWFFC and (b) LUSCT in TID network. 149
5.26 Connection blocking probability results in EON: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths. 150
5.27 Blocking probability wavelength comparison of (a) SWFFC and (b) LUSCT in EON. 150
5.28 Normalized throughput per add port with 16 fixed wavelength transmitters: (a) TID network and (b) EON. 151
5.29 Average path length in TID: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths. 152
5.30 Average path length in EON: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths. 153
5.31 Average number of wavelengths used per connection in TID: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths. 154
5.32 Average number of wavelengths used per connection in EON: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths. 155
5.33 Average fragmentation per connection in TID: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths. 156
5.34 Average fragmentation per connection in EON: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths. 157

List of Tables

2.1 Comparison of OCS, OBS and OPS networks. 12

3.1 Switching rule concentration (SRC) table. 32
3.2 Notation and parameterizations. 38
3.3 Parameters values used in the simulations. 47
3.4 Parameter performance on NSFNET. 49
3.5 Network topologies. 50

4.1 Traffic distribution configurations. 72
4.2 Applications' QoS requirements [5]. 76
4.3 DAOBS Classes of Service (ConnectionLess (CL), Connection-Oriented (CO)). 81
4.4 Light-tree optimization with full-wavelength capacity demands. 97
4.5 Light-tree optimization with sub-wavelength capacity demands and switching constraints for *case (a)*. 98
4.6 Light-tree optimization with sub-wavelength capacity demands and switching constraints for *case (b)*. 99

5.1 Summary of variables and constraints needed by each scheduling policy. 126
5.2 Sub-wavelength connections and results for the MW-NC, T-NC and SW-NC policies. 128
5.3 Sub-wavelength connections and results for the T-C and SW-C policies. 128
5.4 Scheduling policies and heuristic implementation correspondence. 135

Acronyms

ACK	Acknowledgement
ACO	Ant Colony Optimization
ACS	Ant Colony System
ALBA	Auto Load-Balancing with Acknowledgement
ASON	Automatically Switched Optical Network
BCP	Burst Control Packet
BCP-ACK	Burst Control Packet Acknowledgement
BHP	Burst Header Packet
DAOBS	Distributed Access for OBS
DQDB	Distributed Queue Dual Bus
DSMT	Directed Steiner Minimum Tree
EON	European Optical Network
EoS	Ethernet over SDH/SONET
FDL	Fiber Delay Line
GFP	Generic Framing Procedure
GMPLS	Generalized Multi-Protocol Label Switching
ILP	Integer Linear Programming
IP	Internet Protocol
IPTV	Internet Protocol Television
JET	Just-Enough-Time
JIT	Just-In-Time
LAN	Local Area Network
LCAS	Link Capacity Adjustment Scheme
LRM	Local Resource Manager
LSP	Label Switched Path
MAC	Medium Access Control
MAN	Metropolitan Area Network
MPLS	Multi-Protocol Label Switching
NSFNET	National Science Foundation NETwork

O/E/O	Optical-to-Electrical-to-Optical
OBS	Optical Burst Switching
OCS	Optical Circuit Switching
OPS	Optical Packet Switching
OSPF	Open Shortest Path First
OTN	Optical Transport Network
PA	Pre-Arbitrated
PCE	Path Computation Element
PSTN	Public Switched Telephone Network
PXC	Photonic Cross-Connect
QA	Queue Arbitrated
QoS	Quality of Service
RCP	Request Control Packet
ROADM	Reconfigurable Optical Add-Drop Multiplexer
RSVP	Resource Reservation Protocol
RSVP-TE	RSVP - Traffic Engineering
RWA	Routing and Wavelength Assignment
RWTA	Routing, Wavelength and Time Assignment
SDH	Synchronous Digital Hierarchy
SLA	Service Level Agreement
SLAE	Sub-Lambda Assignment Element
SONET	Synchronous Optical NETwork
SRC	Switching Rule Concentration
TAG	Tell-And-Go
TDM	Time Division Multiplexing
TED	Traffic Engineering Database
TNRC	Transport Network Resource Controller
TSON	Time-Shared Optical Network
VCAT	Virtual Concatenation
VoIP	Voice over IP
WAN	Wide Area Network
WC	Wavelength Conversion
WCC	Wavelength-Continuity Constraint
WDM	Wavelength Division Multiplexing

Chapter 1

Introduction

Optical networking has experienced a tremendous revolution in order to support high-bandwidth demanding applications. In spite of the growing traffic demands which can be aggregated to fully occupy the wavelength capacity, we also need to address the all-optical bandwidth provisioning at the sub-wavelength[1] level. Two are the main drivers that support such motivation, as we will introduce.

1.1 Motivation

The motivation of this thesis is the growing need for sub-wavelength provisioning in optical WDM networks. Two are the main drivers: first, the ever-increasing wavelength capacity which is expected to provide Terabit transmission on a single fiber, and second, most applications, even high-bandwidth demanding, do not generally consume/generate enough traffic load to fill the whole wavelength capacity. This last statement is especially true as we move closer to client networks.

Currently, such sub-wavelength provisioning is realized by grooming in the electrical domain different client connections and/or packets, both in metro ring networks [6] and in mesh networks [7]. That is, electronic switching equipment in the IP/MPLS layer is required for grooming traffic at the core nodes. This stage is necessary in order to guarantee that individual wavelength capacity is fully exploited. Commonly, the problem also comprises network planning and virtual topology design for static traffic demands [8], or dynamic lightpath establishment for dynamic scenarios [9]. However, this has two important issues. On the one hand, the re-encapsulation and frame formatting across layers produces loss of protocol transparency. Also, multiple framing stages acting like “russian dolls” result in poor bandwidth efficiency due to the bytes used by multiple consecutive headers, sometimes including duplicated functions (e.g., checksum, addressing, etc.). On the other hand, multiple optical-to-electrical-to-optical (O/E/O) conversion to groom the client signals turns out to be energy-inefficient.

[1] In this Thesis we will use *sub-wavelength* and *sub-lambda* indistinctively.

Research on sub-wavelength provisioning is not new. Over the last fifteen years, several architectures and protocols enabling the multiplexing of packets or bursts have been proposed. The majority have focused on the connectionless transmission of data by idealizing the packet-based nature of several Internet applications. Although this is true, other applications can also benefit from connection-oriented bandwidth provisioning. As a matter of fact, many optical network architectures introduced over the recent years propose using multi-granular architectures for the channel transmission as a way to cope with the application-driven bandwidth requirement diversity, and the provisioning of hybrid connectionless and connection-oriented guaranteed channels to the clients or end-users.

As far as connectionless is concerned, the main issue to tackle is the packet or burst contention in the core of the network, especially when being constraint to use the same wavelength from origin to destination. Several techniques have been proposed in recent years as we will see in Chapter 2, but there is still a lot of work to do and solutions to explore. One question to answer is whether we can rely on mechanisms able to reduce contentions yet at the same time not increasing the capital expenditure, which unfortunately is required by some of the proposed solutions in the past.

With regards to connection-oriented provisioning, this has basically been monopolized by circuit switching, and in particular, wavelength-routed optical networks. As we have introduced, not all applications demand for such bandwidth capacity, and therefore, unless we accept to lose network resources and capacity on the network, other sub-wavelength solutions need to also be explored.

Another main unexplored scenario is hybrid CL/CO in sub-wavelength optical networking. Some approaches have dealt with the partitioning of network resources to provide some sort of dual connectionless and connection-oriented provisioning. However, these solutions assume that both CL and CO are strictly assigned their own resource partition not allowing one or the other to use resources from the other partition. This can lead to undesired inefficient resources utilization due to over-provisioning resources as to guarantee some level of QoS.

1.2 Contributions

The contributions of this Thesis can be summarized as architectures for sub-wavelength provisioning in optical network with connectionless, connection-oriented and hybrid CL/CO capabilities. We will explore throughout this Thesis different network architectures to enable provisioning new applications the necessary bandwidth with finer granularity than the wavelength capacity. More specifically, our aim is to fill the existing gap in literature regarding hybrid CL/CO provisioning and control plane implementations for guaranteed sub-wavelength transmission.

Following the classification of provisioning and transport mechanisms we have introduced, we divide this Thesis into three main parts. The first contribution deals with the connectionless nature of optical burst switching and how to tackle the data contentions in order to improve the performance of the network. We propose a proactive mechanism based on a more efficient load-balancing RWA.

The second contribution proposes a hybrid connectionless and connection-oriented MAC-based optical network architecture. Based on a light-tree virtual topology (or overlay) and a MAC implementation on top of it, nodes are able to transmit bursts by means of a dual access mode: (i) a queue arbitrated connectionless mode, and (ii) a pre-arbitrated mode which emulates connection-oriented channels sharing the wavelength capacity together with the rest of connectionless bursts.

The last contribution tackles the third provisioning and data transport mode, the connection-oriented paradigm. We see from the first contribution that connectionless transmission cannot guarantee the bandwidth provisioning due to limitations with the lower-layer optical enabling technologies. This is an issue, especially from the network operator's perspective who needs to guarantee an SLA with the customers. To this end, we design a general time-shared optical network framework which is used in three different network architectures. The framework is based on an augmented control plane inter-working between a standardized PCE/GMPLS control plane an a proprietary time-shared optical network. The aim is to guarantee the bandwidth provisioning to clients. The contributions are also backed up by the lack of formal formulations of the RWTA and practical implementations of sub-wavelength channel provisioning mechanisms compatible with up-to-date control plane and PCE implementations.

1.3 Organization of this Thesis

The remaining of this doctoral Thesis is organized in six chapters. In Chapter 2 we review the research background concerning the topic of this Thesis. We develop more in detail some of the open issues in optical networking and introduce some of the existing network architectures that consider sub-wavelength provisioning. Later in the chapter, some more specific topics regarding RWA, hybrid CL/CO and time-slot assignment for guaranteed bandwidth are discussed.

Chapter 3 presents our first contribution. The chapter, divided in 7 different sections, introduces the basis of the proposed load-balancing RWA for OBS and its algorithm implementation. This RWA includes an extensive wavelength assignment, routing and updating rule formulation, which is detailed thoroughly before presenting the simulation results and concluding the chapter with its specific summary.

In Chapter 4 we introduce the hybrid CL/CO MAC for OBS. The design and implementation of such architecture is detailed. Results are presented on an extensive simulation scenario. Before concluding the chapter, we take on two of the issues concerning the DAOBS architecture: the access delay mathematical modeling and the light-tree virtual topology optimization.

Chapter 5 is devoted to our last contribution about guaranteed connection-oriented sub-wavelength provisioning. We start by introducing some of the existing technologies related to TSON. We propose three specific architecture implementations based on different control plane inter-working schemes. We then pick the most flexible architecture among the three and derive the time-slot assignment policies we consider and their mathematical formulations. We also propose some related heuristics. We finally assess the chosen architecture under a dynamic traffic scenario before presenting the summary of the chapter.

Chapter 6 condenses the findings of this Thesis. We discuss the main contributions proposed and report our conclusions. Although in this Thesis we present several protocols and architectures, more work can still be performed. For this reason, in Chapter 7 we present some future work topics related to this Thesis which aim to extend some of the findings or even gain a better insight into the issues addressed throughout.

Last but not least, Chapter 8 references the contributions made throughout the course of this Thesis that have been published (or are currently under review) in journals, international conference and workshops.

Chapter 2

Towards Sub-Wavelength Optical Networks: A Background Perspective

The rapid evolving Internet is fostering the development of new applications demanding more and more bandwidth. As such, the supporting network technologies need to evolve in order to deliver this bandwidth growth. Specifically, optical network technologies provide huge bandwidth thanks to the capabilities of the wavelength division multiplexing (WDM), which provides the ability to transmit on several optical channels through the same fiber. This thesis aims at contributing with new sub-wavelength optical network communication architectures. To do so, we review in this chapter the background in the field to support the contributions introduced in the following chapters.

2.1 Introduction

Internet has experienced –and it is still experiencing– a huge application-driven growth through innovation of new services and improvement of older ones. Internet Protocol (IP) provides a common framework that has allowed a tremendous expansion of the communication channels all over the world with the development and supply of a large number of services over multiple network technologies. Therefore, IP has become a converged layer able to adapt even typical connection-oriented applications, like Voice over IP (VoIP) as a substitute to older public switched telephone network (PSTN). As a result, this increase in number of services, and also of users, demands more and more bandwidth, which becomes a big issue for current lower-layer network technologies.

There is a general agreement in the research community and industry that the future growth in network capacity will be mostly oriented towards the support of data traffic [10]. The norm is expected to be the expansion of real-time and streaming applications with an IP or IP-like traffic form. For instance, Fig. 2.1 shows the increase in bandwidth forecasted by Cisco. The bandwidth demand of IP-based applications is expected to exponentially grow between 2009 and 2014. It is worth noting that video applications will be the main contributor, alongside with other bandwidth-hungry applications like file sharing.

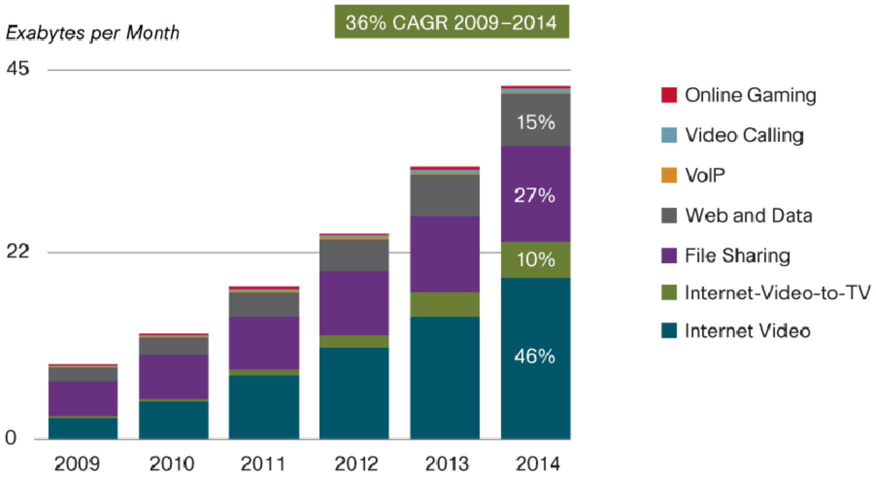


Fig. 2.1: Cisco VNI global consumer Internet traffic forecast. Source [1].

The deployment of these new services will require the provision of a predictable performance able to guarantee the desired quality of service (QoS). That is, even when using a connectionless packet-based transport system, some features like guaranteed transmission from connection-oriented services will be a required capability. Therefore, we can identify future Internet-enabling network technologies should rely on the following premises:

- Reliability and availability.
- Scale, diversity, manageability and interoperability.
- Support for new applications/services and service transparency.
- Economic and cost effectiveness.

The evolution of Internet can be closely related to the *revolution* of its enabling networking technologies. This revolution has pivoted over three main points: the expansion of mobility capabilities, the coming up of the Internet of data and the increasing and cost-availability of bandwidth. The last has to do, principally, with the availability of optical communication technologies [11].

The evolution and development of photonics technology, together with the increased need for an inter-networking broadband communication environment, pushed in the decades of the 90's to the direction of the development of more sophisticated optical networks, in general oriented towards the optical packet switching technology. Those networks are now interconnected and cooperating with other network technologies such as wireless [12], and have also expanded from the long-haul transport networks to the access network environment as fiber to the home deployments [13].

But, what is an optical network? Some misunderstandings and disagreements fall into this topic [14]. An optical network is not necessarily all optical or packet switched though the transmission must be optical. The switching can be optical, electrical or hybrid; and it can be at different granularities: from packets, to bursts or circuits. However, a device is labeled all optical if it performs an operation in the optical domain what traditionally would be expected to be performed in the electrical domain [15]. In practice, the term is often used when dealing with transparent networks. All-optical devices do not necessarily provide *strict transparency*. Traditionally, these devices provide certain degree of full protocol and bit-rate transparency.

Optical networks are present in each network area. On local area network (LAN) environments (see Fig. 2.2), where the primary function is the aggregation of traffic, Ethernet has become a very good and cost-effective transport technology thanks to its broad standardization and the support from both industry and academia. Nowadays, the presence of Ethernet using optical technologies is not strange. As a matter of fact, with the international standardization of passive optical network (PON) technologies and its early deployment in some Asian countries, i.e. Japan and South Korea, Ethernet will still play and interesting role in the evolution of the current networked world. However, to satisfy the demands of new high-end applications in the next years, Ethernet will need to improve its performance, e.g. the provisioning of newer admission control mechanisms to prevent upstream congestion in the network. But this seems not to be a critical problem without solution.

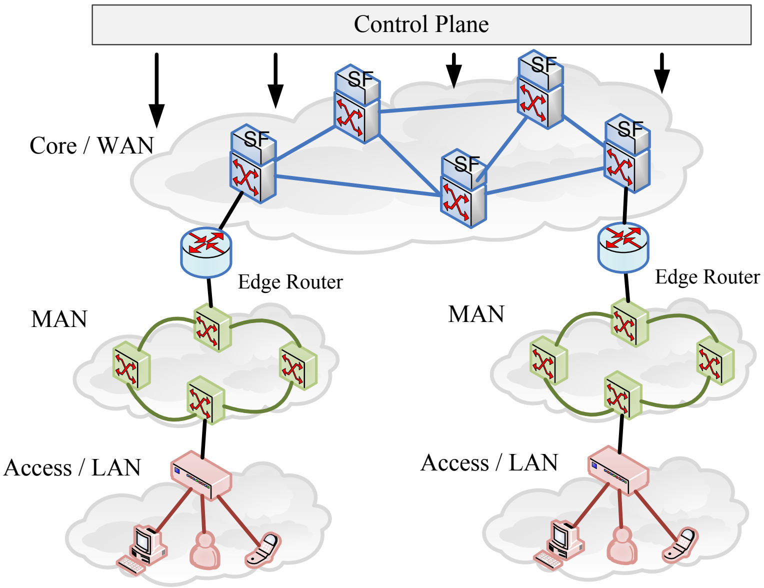


Fig. 2.2: Schematic of network.

To cope with expensive wavelengths in the core, currently we need a prior phase of grooming, which is commonly done in the metropolitan area network (MAN). In this environment, the predominant technologies are Synchronous Digital Hierarchy (SDH) and Synchronous Optical Network (SONET) and Ethernet over SDH/SONET (EoS) using generic framing procedure (GFP). Virtual private network (VPN) and other circuit services can be provided by Generalized Multi-Protocol Label Switching (GMPLS). Finally, the wide area network (WAN) or core of the network is currently dominated by SDH/SONET over WDM point-to-point links.

Fig. 2.3 shows the evolutionary multi-protocol stack for wide-area and core networks [16]. In the past, it was common to have an intermediate stage of cell encapsulation of IP packets using Asynchronous Transfer Mode (ATM) and SDH/SONET, previous to transmit the data through the optical layer.

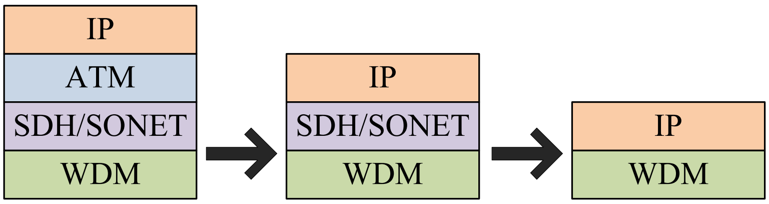


Fig. 2.3: IP-over-WDM evolution.

Despite the nice capabilities of ATM regarding its QoS capabilities and traffic engineering, this extra layer added too much complexity, inefficiency and bandwidth-wastage due to the protocol headers. An initial step was to extend SDH/SONET to map various client layer data services by means of GFP [17], bridging the gap between IP and the optical layer. Also, flexible virtual concatenation (VCAT) [18] and adaptive link capacity adjustment scheme (LCAS) [19] enabled a more efficient transport capacity with the so-called next-generation SDH/SONET. Nevertheless, still this architecture, which is the current one in use by the network operators [20], is not optimized for bursty data traffic and cannot scale to support the rapid growth of Internet traffic. Moreover, its synchronous time-division multiplexing (TDM) nature renders it inefficient for connectionless data. The ultimate step in the evolution is truly realizing the "optical Internet" by transmitting the packets directly on the optical layer or through a "thin" optical framing layer [16].

Optical networking research is closely related to the continuous demand to improve the transmission capacity, configuration capabilities, and flexibility of networks based on fixed optical fibers, and at the same time, sharply reducing operational costs. Several advances in optical transport network (OTN) technologies and protocols [21] have made available a new generation of devices and equipment with a high degree of functional integration. This behavior can be seen in the innovations from the use of single-channel fibers to point-to-point WDM links, to dense WDM (DWDM) transmission systems, and to WDM with multiple add/drop points. Recently, reconfigurable optical add-drop multiplexer (ROADM), photonic cross-connects (PXC) and wavelength tunable lasers have been developed to meet the current networking needs. As a result, the automatic provision of transmission systems provides flexible adjustment of bandwidth in the order of seconds and restoration and self-healing at the network level.

In order to meet the growing bandwidth demands in metropolitan or long-haul environments, transport systems that support fast resource provisioning and that can handle bursty traffic need to be also developed. The rapid increase in data traffic also suggests that all-optical WDM networking technologies, capable of switching at sub-wavelength granularity, are attractive for meeting diverse traffic demands of the next-generation networks. The intelligent optical network seems to be the only solution that can provide not only large capacity sub-wavelength links in a flexible, dynamic and cost-effective way, but also to overcome any bottlenecks and limitations arising from the electronic processing. Optical burst switching (OBS) and optical packet switching (OPS) are two such promising methods for transporting traffic, in particular Ethernet frames, directly over an optical WDM network [22, 23].

2.2 Open Issues in Optical Networking

According to [24] there exist some near-term research issues, which can be resumed as:

- Expanding the media.
- Increasing the protocol transparency.
- Realizing a complete photonic integration.

These three points cover several topics in optical communications research, from the simple proposal of the foundations of the optical networking, through the engineering of new optical devices, to the end-design of new networking systems and architectures. Focusing on this last topic about network protocol and architecture research, the research community will need to address three significant challenges:

1. The provision of true dynamic bandwidth allocation at the network level.
2. Impart intelligence to the network subsystems with automation.
3. Design of all optical transport and transmission network with self-configuring, self-managing and self-organizing capabilities supplying high efficiency as well as robustness and security.

All these topics fall within the borders of the next-generation services and applications that will demand more architecture flexibility, scalability and manageability. Readers are also referred to [25] for a list of trends of optical network architectures.

On solving these new networking demands, the control plane plays one of the key roles. Expanding unregenerated distances, providing optically transparent switching capabilities and multiway optical switching are unresolved challenges from the network perspective in the optical networking evolution [24]. Other important aspects to take into account are the management of the wavelength routing in real-time, the allocation of application-specific bandwidth and the extension of current and new services both closer to the network edge and deeper into the network.

All these aspects are closely related to the variability of the mid-term packet traffic patterns that need to be taken into account in order to maintain the routing stability in the network. For instance, GMPLS needs to overcome the inefficiencies in supporting non-uniform bursty traffic due to its coarse granularity based on circuit switching, long setup delay and slow network reconfiguration [26].

Understanding networking engineering decisions that drive optical transport network reconfiguration is also a challenge to be faced. In sum, the control plane will have to manage the network so that it provides the demanded fail-over/recovery times, yet providing greater capacity and reliability of services deployed across provider boundaries, and intrinsic distributed adaptability using, for instance, the nature of autonomic communication systems.

2.3 Introduction to Next-Generation Optical Networks

Optical communication networks have evolved from a static operation to a more dynamic one thanks to the introduction, principally, of two standards, automatically switched optical network (ASON) [27, 28] and GMPLS [29]. Both automate the signaling of connection-oriented circuits in optical circuit-switched (OCS) networks. Nevertheless, there still exists a gap to achieve the benefits of a true dynamic network able to provide sub-wavelength communication provisioning and support the traffic diversity created by new applications and multimedia services.

OBS and OPS have attracted lot of interest over the last decade being proposed as future optical transport technologies. As such, sub-wavelength technologies like OBS and OPS may represent the next step towards a more transparent bandwidth provisioning from current OCS. However, the poor efficiency of the OBS due to the great number of burst contentions [30] and the lack of enabling optical technologies, e.g., optical buffers for OPS, have attracted some interest to design and architecture more efficient dynamic optical circuit switching (DOCS) [31], especially as a backbone network solution. This would realize the transparent optical network paradigm, hence being blind to the format within each wavelength and letting the MAN do the finer grain grooming. Another possible option is the combination of different switching transport technologies in hybrid and multi-granular network architectures [32].

As introduced, over the last decade three optical switching network architectures have centered the research community attention: optical packet switching, optical burst switching and optical circuit switching. In comparison with OCS, OPS and OBS provide sub-wavelength granularity enabling the multiplexation of packets and bursts, respectively. OPS and OBS mainly differ in their control plane implementation and the length of the data packet, microseconds in the case of OPS, and milliseconds in the latter.

In OCS, all traffic from a connection follows the same path, a previously reserved one, so that client packets arrive in the order they were sent. As the channel needs to be reserved, transfers may suffer from long setup times, in comparison to the connection holding time. Furthermore, the capacity of the circuit or reserved channel must be equal to the peak data rate in order to provide lossless transmission at the ingress edge buffer. This capacity can be orders of magnitudes higher than the average data rate. Therefore, for bursty sources, which are commonly observed in the Internet, OCS results in low utilization [33]. To serve such a packet based network, OPS and OBS aim at improving the network utilization by allowing statistical multiplexing.

OPS uses in-band control information -the data block follows the header of the packet- so there is no reservation possible. Control information processing can be done in the electrical domain [34], although other alternatives propose on an entirely optical approach [35]. Processing and re-insertion of packet headers with strict timing constraints are required, due to the short packet duration -typically around μs. Contention resolution is usually achieved by a combination of wavelength conversion, buffering and, in rare cases, deflection routing.

An interesting technology option which provides flexible and dynamic resources allocation for the future optical Internet is optical burst switching. OBS [36, 37] is a fast circuit switching technology that provides granularity in between wavelengths and packets. OBS can satisfy the future bandwidth requirements avoiding the inefficient resource utilization of OCS and the requirements of OPS in terms of optical buffers, fast processing and implementation complexity.

In OBS, client packets are assembled in edges nodes, transported and switched through the optical network in optical bursts. A key characteristic is the hybrid plane approach: control information is signalled out-of-band using a control packet (BCP), also know as burst header packet (BHP) and processed electronically, while data bursts stay in the optical domain until they reach the egress node. At the egress, bursts are disassembled into their original data packets and forwarded to their destination. Typically, an offset time between the BCP and the optical burst allows the former to be processed at the intermediate nodes and set up the optical cross connects to all-optically switch the corresponding burst. See Fig. 2.4 to fully understand the separation between control and data planes.

Another key concept of OBS is one-way reservation, i.e. burst transmission is initiated shortly after the burst has been assembled and the control packet sent out, without requiring confirmation.

Table 2.1 gives a short comparison between the three types of optical networks introduced in this section. It is interesting to realize how OBS sets a first step to the

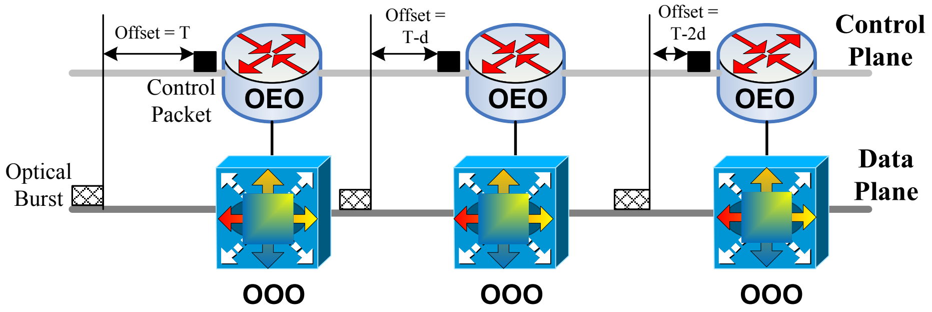


Fig. 2.4: OBS's Data and Control planes.

more desirable, although more complex, OPS. OPS might be the preferred technology that establishes a real future option to present electronic packet-switched networks. Data sizes gradually move from gigabytes (OCS) to hundreds of kilobytes (in OBS) and finally microsecond packet sizes (OPS). The corresponding service or holding times go from seconds or even hours in OCS, to milliseconds in OBS, and finally microseconds for the OPS case.

Table 2.1: Comparison of OCS, OBS and OPS networks.

Property	**OCS (including setup)**	**OBS**	**OPS (with buffers)**
Data sizes of	> GB	tens of kB	100-1500 B
Transfer Guarantee	Possible	More mechanisms required	More mechanisms required
Loss type	Setup request rejection	Loss of burst	Loss of packet
Control	Out-of-band	Out-of-band	In-band
Buffering	No	Possible (FDL)	Typically
Latency given by	Setup + propagation	Propagation + burst assembly + offset	Propagation + packet processing
Disordering	No	Possible	Possible
Control overhead	Connection setup	Maybe	Possible

OCS can possibly guarantee the transfer of data, while the other two technologies need additional mechanisms. This is an important drawback in future optical networks that force them to integrate complex, and sometimes expensive, low loss transport architectures. Note that a loss of burst in OBS implies the loss of multiple data packets contained in it.

As previously introduced in this chapter, OBS offers an out-of-band control channel, that permits a simpler and easier way of processing the control information electronically. This is achieved by using control packets send in advance to the optical burst. Disordering does not occur in OCS as the channel is permanently established for a given traffic flow. On the other hand, as reservations in OBS and OCS are done at a finer granularity, i.e., bursts and packets, disordering can occur in these two.

Performance evaluation amongst OCS, OBS and OPS is an open issue. Several authors have come to compare the OCS and OBS performance. While some have pointed out the benefits of OBS in comparison to OCS with regard to better link utilization [33, 38, 39], other authors remark that bandwidth savings of OBS over OCS are not always achievable and should not be taken for granted [40]. In fact, in backbone networks with predictable and aggregated bandwidth traffic, circuit switching can achieve higher utilization and cost-savings [41]. However, other authors point out the contrary by concluding that the cost advantage of OBS over OCS grows as the capacity of the core network increases [42].

Also, some evaluation studies [30] suggest that strict wavelength-continuity constraints may severely limit the utilization achievable with OBS, and that the true benefit of OBS is on the wavelength-space domain. Since wavelength conversion technology may not be cost-effective, real implementations of OBS, as is, should be properly assessed. As such, a possible solution is towards a hybrid switching approach [43] that may prove more efficient to accommodate diverse IP services by using indistinctly circuit, burst or even packet switching.

2.4 Connectionless Sub-Wavelength Optical Networks

In contrast with OCS, optical burst and packet switching are based on the connectionless transmission of bursts and packets, respectively. Although some authors have proposed the use of two-way reservation mechanisms to cope with the increasing burst or packet blocking probability, most of the proposed architectures make use of one-way reservation, also denoted tell-and-go (TAG). Wavelength-routed OBS [44] is one of the systems that uses two-way reservations. Next, we present some of the most popular one-way mechanisms.

2.4.1 Reservation Signaling Mechanisms in OBS

The transmission of the burst control packet is separated in time to the optical burst transmission. During this time, various mechanisms allow resource reservation for data burst transport, such that bursts can cut through at core nodes entirely at the optical domain. Most of these reservation schemes are based on explicit and/or estimated setup and release. Two are the main signaling protocols: just-in-time (JIT) and just-enough-time (JET). Both adopt the TAG principle.

In TAG, the source first sends a control packet on a separate control channel that immediately reserves bandwidth along a path for the following optical data. In case no resources are available, the burst is blocked. From the edge, data is sent without waiting for an acknowledgment (one-way reservation).

In the original JIT proposed by Wei and McFarland [45], a tear-down control signal or packet is sent out-of-band to release the connection. Fig. 2.5(a) shows an example of explicit setup and release in a network path of four nodes. Two types of signals, or messages (setup and release), must be explicitly used as shown in the figure. Also, a possibility is to send out an in-band termination signal to teardown the reservation. In this case, the burst length still remains unknown until the terminator signal arrives.

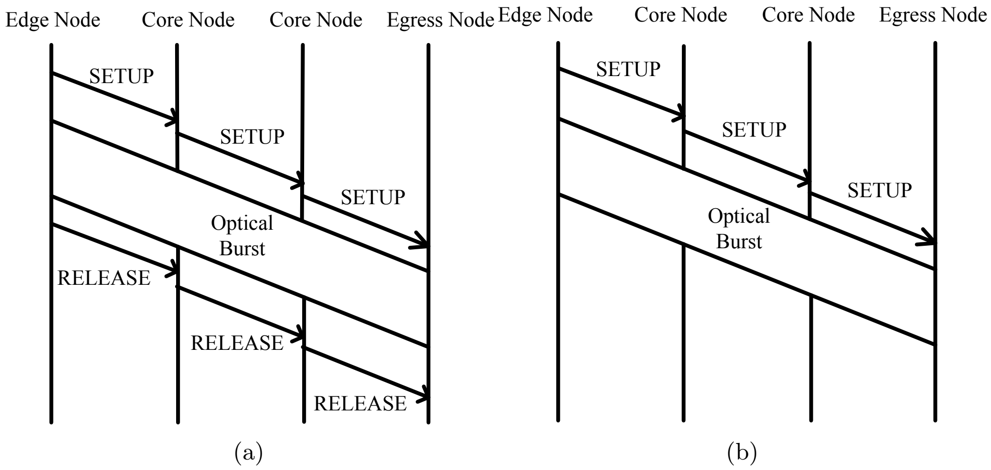


Fig. 2.5: Examples of reservation signaling: (a) explicit setup and explicit release, and (b) estimated setup and estimated release.

One big drawback of using the explicit setup and release is the significant bandwidth waste that can be produced when the release message or signal is lost. However, due to its scheduling simplicity, the complexity of the control units implementing it are relatively low. In general, the time between arrival of the control packet and of the data burst cannot be exploited by core switches to switch other bursts, and therefore, it can occur some waste of bandwidth producing a decreasing on the utilization performance. JumpStart by Baldine et al. [46] is also another specific implementation of JIT wherein the authors consider both, explicit setup/explicit release and explicit setup/estimated release schemes.

The fixed-duration reservation mechanisms, like JET proposed by Qiao and Yoo [47], exploit the control packet's information of the burst length, and only reserve resources for the duration of the burst. Thus, this type of reservation scheme can be classified as estimated setup and estimated release and permits to overlap a preceding burst transmission as long as there is no reservation conflict among other data burst reservation (refer to Fig. 2.5(b)). A newly arriving burst can then be reserved in a gap left by already reserved bursts. This can be done thanks to using the estimated setup and estimated release approach. This type of schemes need that control packets carry information of burst's offset and length to calculate the exact reservation start and end times. Moreover, this scheme reduces blocking in case of varying offset times compared to other schemes [48] as burst reservations can be allocated between reservation voids.

JET-based schemes can be extended for QoS support. Some authors propose [49] to employ larger offset times and delay the reservation process until shortly before the transmission. They use this additional time for an optimized reordering of reservations. However, this can be expected to increase both complexity and delay. If fixed size fiber delay lines (FDL) are not used to prevent the processing related offset decrease at every node, JET suffers from an increased burst loss ratio of the bursts that approach the egress node (and thus have consumed more network resources). This can be inverted by increasing the offset time of long-traveling bursts, thereby increasing their priority [50].

2.4.2 Contention Resolution Mechanisms

Although advanced reservation schemes may be implemented, blocking still occurs due to statistical multiplexing. Focusing in OBS, contentions are produced whenever from two different input ports, two bursts have to leave the optical switch through the same output port, on the same wavelength and at the same time [51].

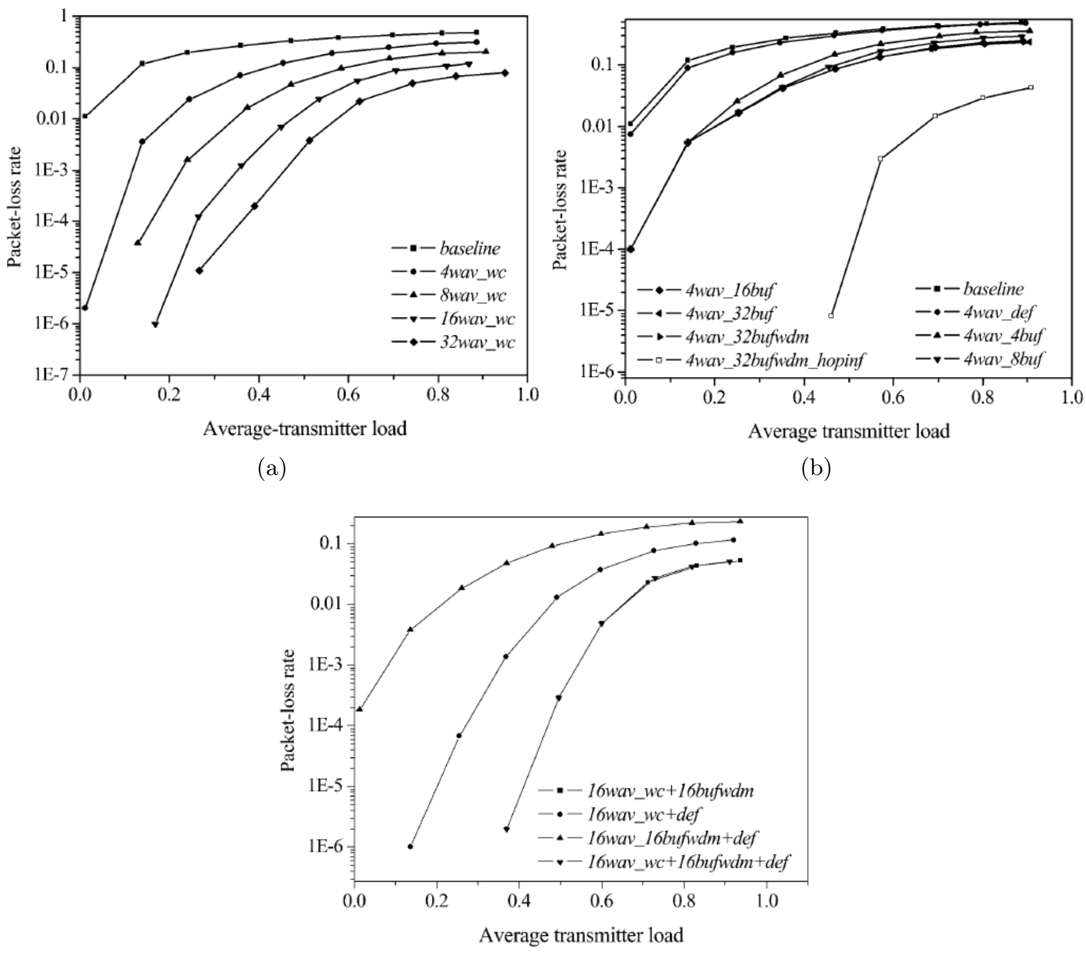
An OBS node that is not able to handle an incoming burst reservation will delete/drop the burst when it arrives, hence, retransmission may need to be initiated by a higher layer protocol, such as Transmission Control Protocol (TCP). Retransmission in the optical layer, would probably be too complex, considering the enormous amount of data that would need to be stored. Note that this contention problem, in the electrical domain is handled by electronic switches with the store-and-forward technique, where data packets can be stored in memory, and sent out at a later time when the desired output port becomes available. This is possible because of the availability of electronic random access memory (RAM).

Effective contention resolution is critical in OBS networks to restrict losses to a reasonably low level. To cope with these contentions, two are the main trends [52, 53]:

- Reactive contention resolution.
- Proactive contention resolution.

The reactive mechanisms try to resolve the contention once this occurs. In this group, three are the main procedures which use the time, wavelength and space domain to minimize loss: (a) fiber-delay lines (FDL) [54, 55], (b) wavelength conversion (WC) [2, 56], and (c) deflection routing [57, 58]. Also, burst segmentation [59] is used to minimize packet loss by only dropping the overlapping packets from one of the contending bursts. When it comes to proactive contention resolving, two are the main mechanisms: (a) call-admission control (or medium access control protocols), and (b) enhanced routing and wavelength assignment mechanisms.

Fig. 2.6 by Yao and Mukherjee [2] shows the results of a packet based optical switching network. Results would not greatly differ from a burst switched network.



(a)

(b)

(c)

Fig. 2.6: Comparison of packet loss probability results of different contention resolution schemes: (a) with wavelength conversion, (b) with optical buffering and deflection routing, and (c) with a combination of schemes. Source [2].

These graphs give results of packet-loss rate when different contention resolution techniques are applied to the network. Baseline applies to the results obtained from a network without any contention resolution scheme applied on. As shown, wavelength conversion is the one that provides the best results, followed by the fiber delay line solution. Deflection routing is the poorest performance solution, although it is the most cost-effective one because the network operator does not have to invest on expensive equipment like wavelength converters. Fig. 2.6(c) shows results when some of the schemes are combined, and in this case, depending on the order of combination, results can be improved in comparison to using an isolate contention scheme.

2.4.3 Quality of Service in OBS

QoS differentiation in OBS networks is also an important issue. Certain types of QoS techniques applied in traditional store-and-forward electronic networks, such as active queue management, are no longer the best way to provide service differentiation in OBS unless we accept to lose the optical data transparency. Thus, other types of techniques need to be realized.

In OBS, QoS differentiation can be achieved by using many different strategies in order to guarantee different quality parameters, although most of them are based on a per-class approach [60]. Additionally, in the per-class method, QoS parameters can be differentiated as [61] absolute (a fixed quality in terms of a certain parameter) or relative guarantees (the quality of each class is qualitatively or proportionally guaranteed between classes). Offset-based, preemption-based and restriction-based schemes are three of the main burst loss differentiation QoS strategies for OBS.

Offset-based schemes [62] rely on the fact that bursts with a greater offset time should have more time to search for free resources on the network and be scheduled for switching with higher probabilities than those bursts with a smaller offset time. Thus, high priority bursts are given a greater offset. Although this scheme increases the delivery rate of high priority bursts, it also increases the latency of them, and for this reason, it is not a feasible scheme when both delay and loss rate need to be guaranteed at a reasonable level.

In the preemption-based QoS scheme [63], high priority bursts are able to take over (preempt) the resources taken by low priority bursts, while these last ones can never preempt high priority bursts. Hence, on average, high priority bursts see more available resources which results in a lower loss rate.

Resource restriction-based schemes exclusively reserve a subset of the available resources for high priority traffic only. An example is *wavelength grouping* [64], which pre-reserves wavelengths for high priority bursts that can only be used by them even if the wavelengths are available for lower priority traffic. Intuitively, the more wavelengths are reserved for high priority traffic, the smaller its blocking probability will be.

2.4.4 Routing and Wavelength Assignment

In optical networking, routing and wavelength assignment (RWA) involves the selection of the best route and wavelength for the transmission of a data unit or the setup of a connection channel (lightpath) between two nodes. The problem is even more challenging in wavelength-continuity constraint optical networks. In this case, the optical switching nodes do not have wavelength converters, hence if at the output, the same wavelength needs to be switched from two different input ports at the same time, there occurs a collision. As introduced before, WC is used, among others, as a reactive-based contention resolution scheme that greatly improves the efficiency and throughput of the network.

Instead on solely relying on the use of WC or FDLs, more efficient RWA can be engineered to improve the throughput of the network by decreasing the burst blocking probability. Shortest path routing has been proved to successfully decrease the amount of resources to enable the transport of packets, bursts or even circuits. However, it does not take into account the offered traffic load to the network which can easily, and often, cause certain links to become congested, increasing the amount of data losses, while other links remain underutilized [65].

In this field, algorithms that enhance the optical control plane with load-balancing properties are specially interesting. One of their benefits is that they permit lowering the cost of the network since less hardware equipment is required and most of the enhancements are an integral part of the control plane software. However, the majority of these load-balancing schemes rely on the hypothesis that the traffic is known and stable, hence they cannot adequately react to variations of traffic.

Load-balancing stands for the ability to balance the traffic across different links with the aim to increase the total amount of bandwidth available and the overall throughput of the network. The concept has been extensively applied on optical networking [66, 67], and is intimately related to the RWA problem. Some of the present works on OBS networks make use of feedback and self-learning searching packets to modify the route utilization probabilities. Others apply weights to several routes as a function of their length and blocking probability.

There exist in the literature several RWA algorithms for OBS networks [68]. Some of them indeed focus on the load-balancing properties of these algorithms and their benefit for decreasing the blocking probability on the network. As established by [68], routing algorithms can be grouped as non-adaptive (static) or adaptive (dynamic). Moreover, adaptive algorithms can be further divided into three groups: centralized (or global), isolated (or local) and distributed. In the centralized scheme, the routing decisions are taken by a single entity with information collected from the entire network. On the contrary, in the local approach, each node runs a local algorithm only with local information available in the node. Finally, the distributed scheme represents a compromise between centralized and local.

Regarding the route management, there also exist two possibilities: explicit routing and hop-by-hop. Both ways have their advantages and drawbacks. While explicit routing uses a set of predefined routes, in the hop-by-hop the routing decision is taken at each intermediate node and essentially, the selection of the output port (i.e. next node) is the only action to take. As a result, the hop-by-hop mechanisms can react promptly to changes on the network such as failures or congestion; however, traffic-engineering techniques cannot be applied by the source node, which is one of the advantages of explicit routing. Arguably, the authors in [68] state that explicit routing with LSP identifier is preferable for highly dynamic OBS networks due to the less processing required. Nonetheless, the results show that multi-path source based load balancing optimization techniques are very time-consuming and most of the time require the traffic conditions to be static or lightly change throughout the evaluation period. The proposed methods in [68] are designed for networks with full wavelength conversion capability.

With the aim at improving the multi-path source routing with adaptation capacity in order to react more adequately to abrupt and frequent traffic changes, [69] proposes a modification of the cost function of a gradient projection multipath routing protocol [70] to be valid in scenarios without or with little wavelength conversion capacity. The adaptive mechanism is implemented by means of an iterative algorithm that splits the traffic among alternative LSPs. Nonetheless, the authors also point out that in order for the algorithm to work properly, the relevant time-scale of the traffic pattern must be greater than the iteration length, restricting the use of the algorithm for scenarios where the traffic changes vary slowly.

Other solutions, like the proposed in [71], use feedback and self-learning searching packets to modify the probabilities to transmit on a certain route. However, due to its formulation, stagnation problems can occur when the number of transmissions is large. Another work [72] also applies weights to several routes between source and destination depending on their length and blocking probability. Nonetheless, the authors suppose the use of link-state protocols able to disseminate back to the source nodes this blocking probability. Therefore, it may happen that routes are outdated when the acknowledgement message is still in transit back to the source node. Finally, other authors [73, 74] also propose the use of self-learning agents to update the routing tables in sub-wavelength optical burst switching networks with full wavelength conversion.

2.4.5 Medium Access Control in Sub-Wavelength Optical Networks

As we introduced, another way to cope with contention in the network, an in OBS in particular, is to actually prevent them to occur by controlling the way the nodes transmit on the network. This is the main purpose of medium access control protocols.

In terms of MAC protocols for OBS networks, these have been given little attention, and almost all existing studies are focused on OBS metro-ring networks [75]. For instance, [76] proposes a simple MAC protocol called beforehand bandwidth reservation for OBS ring networks to reserve the empty slots in the next big-slot cycle.

Likewise, in [77] a loss-free OBS metro ring architecture (CORNet) is designed along with a distributed MAC protocol to integrate the support of differentiated service and fairness access. The proposed QoS provisioning mechanism adopts a bandwidth-reservation approach which combines real-time transmission establishment and termination routines. Furthermore, a credit-based fairness control scheme is defined to guarantee the transmission fairness of best-effort traffic.

And recently, [78] proposed a multiple-token-based MAC protocol for OBS ring networks using a tunable transmitter and one tunable receiver (TT-TR). In order to avoid or reduce the receiver contentions, tokens manipulate the wavelength accessibility and the destination queues decide on the burst scheduling.

2.5 Towards Guaranteed Sub-Wavelength Transmission

The traffic diversity created by modern applications and the rapid advance of optical technologies has driven the development of new optical network architectures able to provide flexible and dynamic resources allocation [79].

The connectionless nature of OBS and OPS are an important asset to provide delay-sensitive services and enable on-demand provisioning with multiplexing capabilities among different traffic sources. However, despite the extensive work to minimize data contentions, these are still a very important issue.

In the provisioning of connection-oriented sub-wavelength data transport, GMPLS may play a very active role. An obvious benefit of GMPLS is the standardization process behind it and its industry support. GMPLS provides protocol transparency on optical networks by supporting traffic engineering and switching at different layers: packets at layer 3, layer 2 frames, TDM circuits and wavelength paths [29]. This allows for multi-granular provisioning of bandwidth. However, despite the strong connection between GMPLS and OBS in terms of being both technologies widely used for optical networking, the research dealing specifically on its interworking is fairly limited.

2.5.1 Hybrid Connectionless and Connection-Oriented Sub-Wavelength Data Transport

Multi-granular networks offer the greatest service diversity by enabling the setup or transmission of data under different forms, connection-oriented or connectionless transmission, and for different granularities, like circuits, bursts or packets. In this sense, the authors of [80] propose an optical hybrid circuit/burst switching architecture that supports OCS and OBS. The origins of this kind of approaches comes from the so-called dynamic OCS [81], wherein connections are set up and taken down frequently, the OBS with acknowledgement [82] or the proposed wavelength routed OBS [44].

Some of the first works to facilitate the implementation of OCS and OBS and other potential technologies on the same network are [26, 83]. The authors investigate into the OBS and GMPLS extensions to seamlessly combine different scheduling tecniques for a variety of applications, yet managing a single framework. This framework, Polymorphous, Agile and Transparent Optical Networks (PATON), based on polymorphous OBS (POBS), allows multi-granularity bandwidth provisioning and integrated control and management. As described by the authors, PATON aims at achieving three main objectives: 1) transparency and low latency through single-hop optical data lightpaths, 2) single network infrastructure and framework to support multiple applications with different bandwidth and delay requirements, and 3) cost-effective utilization of resources.

Following the idea introduced in PATON, [84] performs a simulation evaluation of the common and integrated signaling system using POBS, and compares these results against the two-hop and multi-hop OCS. Also, [85] gives an exhaustive analysis of the best-effort burst-based services performance when some channel capacity is allocated to TDM services in POBS scenarios. Nevertheless, neither of them provide a detailed and specific architectural protocol analysis to support POBS.

In the analyzed papers, contentions cannot be totally avoided; hence contention resolution methods need to be also included. A possibility to consider is to reserve multiple tunnels between source-destination pairs (e.g., equal cost multi-path) in order to provide routing alternatives to the OBS control plane which implements routing deflections in case of contention.

2.5.2 Guaranteed Sub-Wavelength Transmission

Lately, other proposals towards guaranteed sub-wavelength tranmission have been proposed. Some take principles from OBS and other proposed technologies, and others try to redefine the network architecture. Some bring together known technologies like GMPLS, as introduced before.

One of the first works to relate both paradigms (label and burst switching) is labeled optical burst switching (LOBS) [86]. LOBS simply augments each OBS node with an IP/MPLS controller in order to achieve IP-over-WDM integration. Each control placket is sent as an IP packet containing a label, as part of its control information along a pre-established LOBS path, similar to a label-switched path (LSP). In [86], a general description of this architecture is given; however, there are no results to check the performance of this design.

This work is related to a previous introduced one [83], wherein the authors propose an integrated signaling and control framework based on the integration of GMPLS and OBS to support diverse applications and services. One of the benefits of this network architecture is the emulation of transparent TDM circuits with sub-wavelength capacity. Although some GMPLS extensions such as fast connection provisioning, scheduled reservation, and new types of transparent sub-wavelength LSPs (both asynchronous and synchronous) are discussed in the paper, it does not give any specific extension design to the current standard status.

More recently, Qiao et al. [87, 88] introduce the LOBS with Home circuits (LOBS-H) to decrease the number of wavelength in comparison with OCS. However, again, there is not explicit implementation description of the GMPLS/OBS integration even though LOBS is assumed.

Another proposal is [89] which specifies an inter-working architecture based on the use of a hybrid control plane using two different control networks logically and physically separated. The OBS control plane is used for OBS specific tasks while the GMPLS control plane manages the OBS background tasks. The virtual topology management is part of the OBS background tasks. This module is responsible for setting up, maintaining and tearing down the LSPs between the edge nodes. However, no reservation is made. GMPLS establishes the LSPs in a two-way mode to distribute the label among the nodes. Once the label is established, the resource reservation is made on a per burst basis by means of the BCP over the pre-established LSP. This is part of the OBS specific tasks. This approach gives a connection-oriented basis to the burst transmission.

A more recent paper [90] extends the previous work by giving a more detailed description of changes made to the Resource Reservation Protocol - Traffic Engineering (RSVP-TE) messages. It also introduces the concept of shared wavelengths among different tunnels using virtual assignment. All in all, these two last proposals try to cope with the inter-working issues of GMPLS and OBS, however, the solutions provided are not completely lossless and the sub-wavelength extensions given to GMPLS are rather defined.

To improve the burst blocking probability, [91] implements two different methods, one using BCP and the other through link management protocol (LMP), to disseminate the utilization statistics to the neighbour nodes. This scheme also uses a similar hybrid control plane as in [89] that avoids the use of the Open Shortest Path First - Traffic Engineering (OSPF-TE) due to its high complexity and slow convergence.

Another approach is the one presented in [92] and [93] where an overlay GMPLS/OBS model is proposed. In such architecture, the GMPLS domain is used to interconnect OBS client networks by means of group label-switched paths (G-LSP) that successfully enable the BCP and burst to be transmitted over a GMPLS-controlled transparent OCS network. In spite of the valuable experimental results, both schemes do not support sub-wavelength OBS transmission.

Finally, we also have light-trail [94]. This sub-wavelength network architecture can span over ring and mesh topologies using optical buses or light-trails (see Fig. 2.7) enabling the participating nodes to share the capacity of the channel in a spatial reuse time-shared basis. It defines the figure of a controller, the *arbitrator*, which is responsible for assigning network resources using an explicit request-granting process in a per slot basis. To cope with possible contentions, light-trail uses electronic buffers.

2.5.3 GMPLS/OBS Interworking Multi-Layer Models

Based on the background presented in the previous section, we can identify two main interworking models: the overlay approach (i.e. GMPLS on top of OBS control plane) (see Fig. 2.8(a)) and the GMPLS/OBS integrated approach (i.e. G.LOBS) (refer to

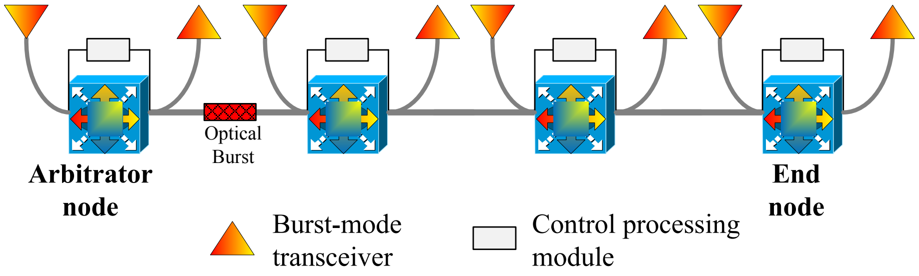


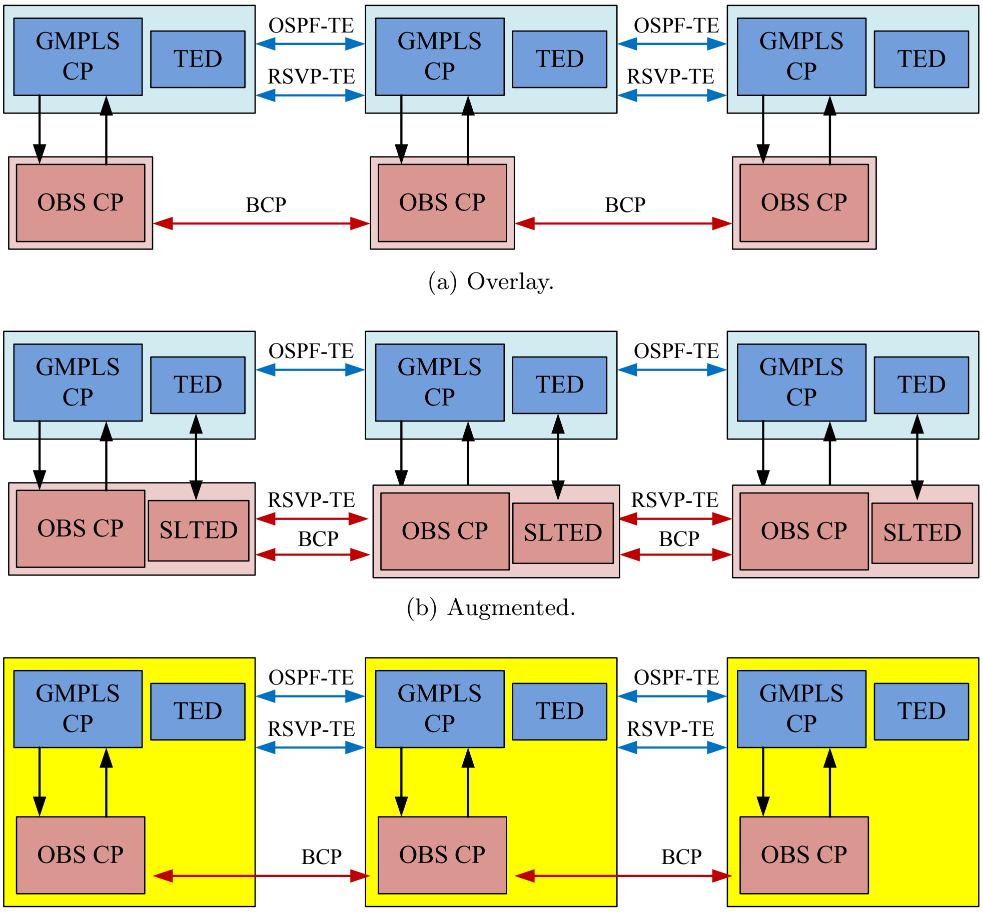
Fig. 2.7: A light-trail. Based on [3].

Fig. 2.8(c)). A third possibility is the augmented approach [95]. Actually, these three inter-working multi-layer approaches are identified by the GMPLS architecture [96]. However, their terminology is strictly applied to the support of multi-layer GMPLS networks and the application of these models to the separation of MPLS and GMPLS control plane islands.

As identified by [96], we can have:

- In the peer model, MPLS and GMPLS nodes run the same routing instance, and routing advertisements within islands of one level of protocol support are distributed to the whole network. Signaling in the peer model may result in contiguous LSPs, stitched LSPs [97].
- The overlay model preserves strict separation of the routing information among network layers. Despite the separation of signaling information between network layers, some interaction in signaling between these may exist. Moreover, this model requires the establishment of control plane connectivity for the higher layer across the lower layer.
- The augmented model allows limited routing exchange from the lower-layer network to the higher-layer network. This assumes that nodes provide some form of mapping or aggregated routing information from the lower-layer. Signaling between layers is required.

If we look more in detail about the GMPLS and other sub-wavelength (e.g., OBS), the overlay inter-working establishes the GMPLS to signal the macro-flow or long-life connections while the OBS signals and reserves the micro-flows on a per burst basis. This approach requires rules to aggregate and flood the sub-lambda resources. Moreover, this design would require implementing several OSPF-TE extensions to model the sub-lambda resources and determine the aggregation mechanisms from the transport network resource controller (TNRC) to the local resource manager (LRM) in order to finally flood this information with the OSPF-TE. Some extensions would also be needed in the link state advertisement (LSA) object, namely, how to announce finer sub-lambda bandwidth availability. The distribution of aggregate available bandwidth per priority or the use of wavelength bitmasks is too coarse. It is not the best solution



(a) Overlay.

(b) Augmented.

(c) Peer or integrated.

Fig. 2.8: GMPLS/OBS control plane inter-working.

due to the duplicated functions between the two control planes and the rather poor sub-wavelength capabilities. Finally, in the overlay approach, RSVP-TE extensions are also necessary. For instance, a dedicated sender TSpec for OBS traffic would need to be defined with information of maximum and average bandwidth, QoS, etc.

The integrated GMPLS/OBS (G.LOBS) is based on the signalling of OBS directly implemented by RSVP-TE with a finer sub-lambda resource modeling. That is, all layers are controlled by a unified contol plane, and decision are taken considering the whole network information. This approach requires, not only to extend, but also modify the GMPLS protocols to cope with the very short-life LSP sessions, which requires, for instance, to provide mechanisms for fast signalling, fast routing information, local deflections, delay lines, etc. In this mode, the OBS nodes are directly controlled by a single control plane, so GMPLS acts directly on the OBS forwarding tables. The main disadvantage of this approach is its increased and critical implementation complexity. Specifically, regarding the OBS LSP sessions, these impose to speed up the signaling mechanisms to shorten the current three tiers reservation model (PATH-RESV-CONFIRM) to two-tiers or even just one.

Another possible control plane integration for the provisioning of guaranteed sub-wavelength bandwidth services is the augmented approach (see Fig. 2.8(b)). In this architecture, different control plane instances run on each layer but some information is exchanged between them aiming at improving the network bandwidth allocation. Some functionality may still be duplicated; however, this approach ensures an easier compatibility with different underlying sub-wavelength transport technologies, yet enabling GMPLS standardization on the top.

The augmented model maintains a separation between optical and routing topologies; unlike integrated model approach, where topology information is shared between sub-wavelength and coarser optical domains. This allows the augmented model to be more efficient in resource utilization than overlay model, such that router and optical domain resource can be optimized. At the same time, it can yield more efficient use of resources, similar to the full peer model.

2.5.4 Connection Sub-Wavelength Time-Slot Assignment

In the provision of sub-wavelength guaranteed services, the optimality of time and slot sub-wavelength allocation in time-shared optical networks is a key factor. In this regard, [98] established the foundations for the provisioning of slot-based all-optical WDM mesh network. The authors investigated the all-optical routing issues and proposed an approach whereby packets from different sources but same destination node, and possibly transmitted on different wavelengths, are routed through the network as an integrated unit. However, it is worth noting that the approach followed by the authors is based on a totally connectionless paradigm which differs from the sub-wavelength connection-oriented proposed later in this thesis (refer to Chapter 5).

Our approach for sub-wavelength time-sharing is more inline with the so-called time-driven switching (TDS) networks [99]. TDS applies both time division multiplexing and frame structures to provide end-to-end sub-wavelength circuits. Time is divided into time frames of equal size grouped in time cycles. On setting up a connection between a source and destination node, a free time frame is searched in the cycles associated to each link along the path. If found, the time frames are reserved, and this sequence form a synchronous virtual pipe. To make this possible, the whole network must be synchronized to recognize the time division, which in the TDS case, the authors assume a global system provides absolute time reference with high accuracy, e.g., form a GPS system.

An example of TDS networks is fractional lambda switching (FλS). In FλS [100], the capacity of an optical carrier is divided into a larger number of sub-channels based on time frames and the data content of each time frame is independently switched using TDS techniques.

Huang et al. [101] introduced the concept of long-lived *time-sharing wavelength* *channels* for all-optical transport networks. The authors propose an optical transport network architecture based on the use of time-wavelength-space routers. On such network, connections are established by constructing a time-slot based lightpath between the source and destination routers. In the article, a heuristic algorithm, named *greedy-rwta*, is presented to cope with the NP-completeness of the RWTA problem. As defined in [102], the RWTA is NP-complete by the fact that when the number of wavelengths per fiber is 1 ($W = 1$), the slot assignment approximates the RWA problem, which is known to be NP-complete.

Regarding the time-slot assignment in all-optical networks, the authors in [103] present some algorithms for the disjoint RWTA problem (i.e., the problem is divided disjointly into routing, wavelength and time-slot selection subproblems [104]). A new least resistance weight function, which incorporates link load and path-length information, is used for computing the cost paths, along with a time-slot-allocation algorithm based on the least loaded (LL) wavelength.

Likewise, [105] introduced the first-fit (FF) approach, in which the first available time-slots along a route and wavelength are reserved. Later on, [106] presented a fine slot assignment based on the least constrained slot (LC). This approach constrains the number of fixed routes a slot can be used, and for a given route, it selects the least constrained slots. The algorithm is compared to the LL yielding a better perform, specially in multi-fiber network scenarios, and a close computation performance in comparison with FF. The same authors presented recently a distributed algorithm [107] based on the same least constrained slot allocation in GMPLS-enabled networks.

Finally, a different approach is proposed in [108]. In a TDM/WDM network with no time-slot interchanging capabilities, the authors design a RWTA scheme wherein logical connections are allocated on the physical network making use of a set of *super-lightpaths*. A super-lightpath is able to carry all connections between a single source node to a number of different destinations, which are able to extract the sub-channels directed to themselves and forward to an output port the remaining until the last destination is reached.

2.6 Summary

The development of WDM marked a revolution for optical communication networks, not only by increasing the bandwidth per fiber several orders of magnitude, but also by decreasing the cost per bit. Wavelength-routed optical networks have been the primary benefited from this development. However, one of the open issues for future optical networking is to increase the transparency of these networks and enable reliable sub-wavelength provisioning to satisfy.

In this chapter, we have come across several sub-wavelength network technologies that would ultimately realize this finer granularity all-optical data transport. We have principally introduced OBS and its associated protocols and seen that its main issue is the number of contentions it can produce. As a consequence of this, lately, new solutions such as hybrid switching architectures and GMPLS/OBS inter-working have been proposed.

Chapter 3

RWA for Connectionless Sub-Wavelength Optical Networks

With new optical transport systems able to provide sub-wavelength granularity, the dynamic characteristics of the network are expected to sharply increase. Furthermore, such properties are highly related to the underlying physical network topology, which in turn require from the control plane important features such as scalability, dynamism and automatism. Following the outline of this thesis, in this chapter we start by analyzing the properties of a distributed routing and wavelength assignment algorithm for connectionless data transport on wavelength-continuity constraint sub-wavelength optical networks. The protocol is evaluated through its implementation in optical burst switching on four different transport network topologies.

3.1 Introduction

Connectionless data transport in electronic networks is not only a fast way of communicating two end-points on the network, but also seen as reliable means of transmitting data due to its self-resiliency capabilities. The electronic processing of packets allows greater flexibility on providing to each data packet a special forwarding treatment. This, in turn, can be used to recover from traffic congestion on a specific path or from link failures. However, connectionless transport on sub-wavelength optical networks is yet an open issue. The lack of efficient optical buffering techniques and the use of one way reservation protocols produce a great amount of packet or burst contentions even at very low network loads.

As introduced in the background chapter, contentions can be resolved or minimized reactively or proactively. Regarding the latter, two are the main mechanisms: (a) call-admission control (or medium access control protocols), and (b) enhanced routing and wavelength assignment mechanisms.

In this chapter, we gain insight into a distributed load-balancing RWA algorithm based on the information supplied from successful/unsuccessful burst transmissions using acknowledgment control packets. In comparison with other source-based routing algorithms, the proposed protocol goes a step further by establishing a hop-by-hop route in a totally distributed way taking advantage of the local information in the core nodes to determine the best possible forwarding port towards destination. The load-balancing capabilities are established among the different input-output port possibilities at every network node. That is, the connectionless nature of the burst transmission is reinforced by avoiding the source node to decide the end-to-end path. Furthermore, in order to provide a complete assessment of the protocol, we investigate its performance on four different network topologies and derive its performance mathematically from the algorithm formulae under certain restricted conditions.

As introduced in the previous chapter, OBS improves the network utilization under unconstrained-wavelength conversion scenarios. However, we forecast that the availability of full wavelength conversion cannot be totally realized mainly due to the expensive equipment required. Therefore, we propose our RWA scheme as an evolutionary solution for wavelength-continuity constraint networks while wavelength conversion is not fully available and cost-effective.

The fundamentals of the proposed auto-load balancing routing protocol derive from ant colony optimization (ACO). ACO has been successfully used in diverse routing problems, both on electronic and optical networks.

The remainder of the chapter is organized as follows. Section 3.2 introduces the ant colony optimization and the main differences of our solution with respect to ACO. Section 3.3 describes the necessary steps to implement the proposed algorithm in OBS networks. Section 3.4 formulates the RWA algorithm whilst Section 3.5 describes the formulation used by the algorithm's forwarding and updating rules. Results through simulations are analyzed in Section 3.6. The chapter concludes Section 3.7 by presenting a short summary.

3.2 Ant Colony Optimization

Ant colony optimization [109] comprises a set of algorithms used for the optimization of several problems. In essence, ACO algorithms try to emulate the biological behavior of ant colonies on their task of foraging for food. In the real world, ants initially wander randomly. Upon finding food, they return to their colony, and on their way back to the colony, they lay down pheromone trails which can be used by other ants to find more food. Eventually, if these ants also find food, they return back to the colony reinforcing the same trail. Over time, the pheromone trails tend to evaporate, thus reducing their attractive strength. However, if the ants use shorter paths, they can come back to the colony more rapidly, thus reinforcing this kind of paths in front of longer ones.

Due to its distributed nature, ACO is an interesting solution to support the control plane mechanisms on very dynamic and connectionless optical networks.

The proposed RWA for connectionless sub-wavelength optical networks is based on the well-known ant colony system (ACS) algorithm by Dorigo et al. [110]. ACS differs from previous ant-based optimization systems on three main aspects [110]: first, the state transition rule balances the exploration of new edges and the exploitation of accumulated knowledge about the problem; second, the global updating rule is applied only to edges which belong to the best ant tour; and finally, while constructing a solution, ants in ACS apply a local pheromone updating rule. Originally, ACS was defined for resolving problems such as the traveling salesman problem (TSP).

Taking as a basis part of the formulation and ideas in [110], we engineer and extend ACS with a two-fold contribution: (1) by enhancing the protocol to support the auto-load balancing RWA in wavelength continuity-constraint sub-wavelength optical networks, and (2) extending and adapting conventional path-scoring methods using both switching congestion information and path length to efficiently choose the optimum route and wavelength for every burst transmission. One of the main differences of our solution and ACS is how the global updating rule is applied. While in ACS, only the globally best ant, i.e., the ant which constructed the shortest tour from the beginning of the iteration, is allowed to deposit pheromone using the global updating rule, in the proposed algorithm, every ant (control packet) is allowed to update the pheromone. The main reasoning behind this is that we couple the ant-based optimization with the burst transmission, which results in the algorithm becoming an integral part of the optical network control plane.

3.3 Algorithm Implementation

The RWA algorithm auto load-balancing with acknowledgment (hereafter simply denoted ALBA) comprises two different but dependent stages: (1) the selection of the next hop (and also wavelength at the source node), and (2) the update of switching rule concentration (SRC) values. Briefly, the former uses a set of rules to select the output port using the SRC values updated by the latter. As we just introduced, the algorithm is an integral part of the sub-wavelength optical network control plane, and therefore, routing and burst transmission are coupled together.

3.3.1 Data Structures

As introduced above, ALBA makes use of a particular value to decide on the forwarding and routing process of the data units (i.e., optical data bursts in this particular case). These values are the switching rule concentration (SRC) and are denoted by τ_{ijk}. An SRC defines the strength of using a certain switching/routing configuration in the optical switch. In order words, to what extend a specific input-to-output port and wavelength configuration on the switch is suitable for switching a given packet/burst. In essence, they provide a measure of the attractiveness to follow a path. For comparison purposes, SRC values are to ALBA what pheromone concentrations are to ACO. Regarding the SRC sub indexes $\{ijk\}$, i denotes the input port, j the output port, and k the wavelength.

For fast look up and update of the SRC values, these are conveniently stored in an SRC table in each network node. The size of the table is directly related to the physical capacity of the optical switch and the number of wavelengths on the network. Let assume $G = (V, E, W)$ represent the graph topology of the network, where V is the set of nodes (vertices) in the network, E represents the edges and W is the set of wavelengths per link. In such a case, the worst-case space complexity of the SRC table is $O(|V|^2|W|)$ if we assume nodes are directly connected to each remaining node on the network. Table 3.1 shows the SRC values of the non-blocking switch in Fig. 3.1. The switch has three input ports and four wavelengths per link.

Table 3.1: Switching rule concentration (SRC) table.

		Output link		
Input Link	**Wave**	**Link1**	**Link2**	**Link3**
Link1	**1**	$\tau_{1,1,1}$	$\tau_{1,2,1}$	$\tau_{1,3,1}$
	2	$\tau_{1,1,2}$	$\tau_{1,2,2}$	$\tau_{1,3,2}$
	3	$\tau_{1,1,3}$	$\tau_{1,2,3}$	$\tau_{1,3,3}$
	4	$\tau_{1,1,4}$	$\tau_{1,2,4}$	$\tau_{1,3,4}$
Link2	**1**	$\tau_{2,1,1}$	$\tau_{2,2,1}$	$\tau_{2,3,1}$
	2	$\tau_{2,1,2}$	$\tau_{2,2,2}$	$\tau_{2,3,2}$
	3	$\tau_{2,1,3}$	$\tau_{2,2,3}$	$\tau_{2,3,3}$
	4	$\tau_{2,1,4}$	$\tau_{2,2,4}$	$\tau_{2,3,4}$
Link3	**1**	$\tau_{3,1,1}$	$\tau_{3,2,1}$	$\tau_{3,3,1}$
	2	$\tau_{3,1,2}$	$\tau_{3,2,2}$	$\tau_{3,3,2}$
	3	$\tau_{3,1,3}$	$\tau_{3,2,3}$	$\tau_{3,3,3}$
	4	$\tau_{3,1,4}$	$\tau_{3,2,4}$	$\tau_{3,3,4}$

The SRC value is computed according to two sources of information: (1) the congestion level as a measure of the number of contentions through that specific output port and wavelength, so the more the contentions, the lower its value, and (2) the path

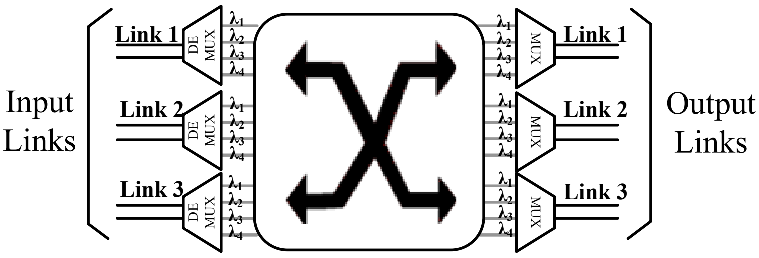


Fig. 3.1: Optical switch with 3 links and 4 wavelengths.

length, in the sense that shorter paths through a given switching configuration shall increase the SRC value, i.e., the attractiveness of a route. Shorter paths imply less allocation of optical resources for a single packet/burst transaction, and as a consequence, more resources can be available for other transmissions. Therefore, shorter paths can decrease the congestion level among different traffic flows, as long as these flows, in turn, are correctly balanced and distributed over a number of paths.

In wavelength-continuity constraint optical networks, the availability of lambdas over a specific output port is no longer a good measure for the goodness of the wavelength assignment. The actual availability of a wavelength on a specific link does not infer enough information about its availability on the remaining links along the path. Alternatively, congestion level or the number of burst contentions through a specific switching configuration on a given output port provide a more useful measure.

Originally, in OBS the transmission of the burst is delayed by an offset time which normally depends on the burst route length. During this time, a BCP is transmitted and processed electronically by the intermediate nodes along the path from the origin to the destination of the burst. It is worth noting that this process involves an optical-to-electrical-to-optical conversion in order to process the BCP as it is shown in Fig. 3.2. When the offset time elapses, the burst is transmitted and switched all-optically without acknowledging that the optical resources have been reserved by the BCP. Due to this one-way signaling system, losses can occur in the core of the network.

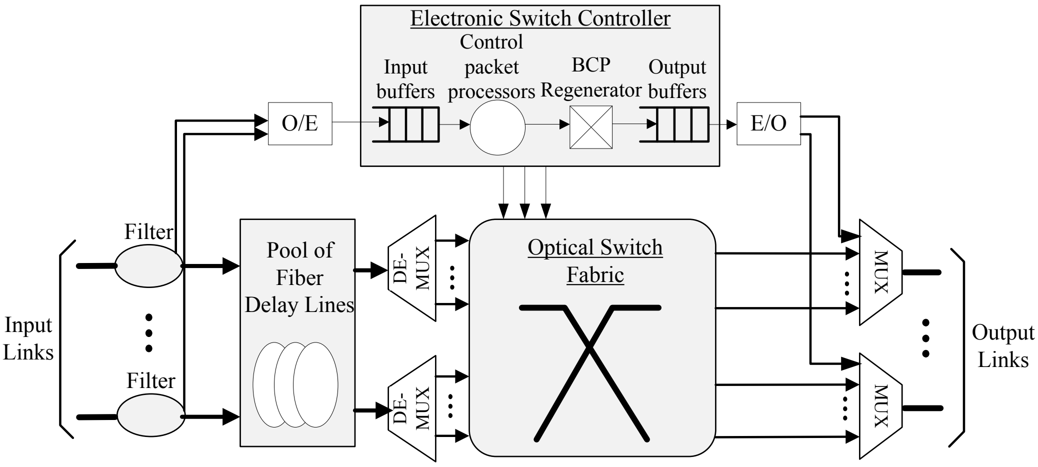


Fig. 3.2: OBS core node architecture with offset time emulation. Source [4].

Two different control packets are used to support the burst forwarding and update of SRC values: the BCP, which provides information about the burst, such as length, QoS level, offset time, wavelength, etc., and the Acknowledgment BCP (BCP-ACK). The BCP-ACK is mainly used on the backward update process and announces the nodes involved on a given burst transmission whether the delivery of the burst has been successful or not. As pointed by the connectionless approach of OBS [111], BCP-

ACKs are not used to confirm the reservation of resources prior to the transmission of the burst. In the present case, this kind of packets is only used to acknowledge the positive or negative delivery of the burst to its destination.

3.3.2 Algorithm Implementation in OBS

In ALBA, and due to the connectionless operation, the end-to-end path from the origin OBS node to the destination node is not known in advance. During the burst forwarding, the next output port is calculated hop by hop. As such, the path may comprise the least number of possible hops using the shortest path (or equivalent) or also take a longer route. Therefore, in some cases, the path may be longer than expected, so making it difficult to calculate the best offset time in order to avoid the burst loss due to the expiration of the offset, and at the same time, reduce the end-to-end delay of the burst. Consequently, to enable the protocol on an OBS network, an offset time-emulated scheme [4] is implemented along with the extra processing and scheduling tasks of ALBA. As a result, with the emulated scheme, the offset value no longer depends on the route length, so that, a specific time value is not needed. However, in order to give enough time to the switching and reservation controllers to reserve the optical resources, the upcoming burst is delayed at each input port of the network node by means of an FDL pool of a length equivalent to the BCP processing time.

Fig. 3.2 shows a diagram of a core OBS network node with offset emulation. As the figure shows, a pool of FDLs are placed at the input data ports before the WDM demultiplexers to delay the bursts and give time to the control packet to be processed. Previously, before the data channels get into the FDLs, the control wavelength is filtered out to the optical-to-electrical (O/E) converter in order to process the control packets electronically. In any case, this node architecture does not greatly differ from the common architecture, and in this work does not only fulfill the purposes of ALBA, but also provides many other benefits. For instance, it also enhances fairness in resource allocation for the bursts as the offset time is constant and not decreased along the path. For more information about the offset time-emulated architecture readers are referred to [4].

3.4 Routing and Wavelength Assignment Algorithm

The fundamental operation of ALBA is supported by two procedures: the data routing, which makes use of the forwarding rules, and the update of the forwarding decision data which is executed by the acknowledgement control packets. The algorithm can be further divided into four main sub-procedures, as shown in Fig. 3.3:

1. Initial wavelength assignment and next hop selection.
2. In network core next hop selection based on state transition rules.

3. In network core local updating rules.
4. Global updating rule, which can take two values: (a) positive global updating if the burst is delivered to its destination, and (b) negative global updating if the burst is blocked due to contention.

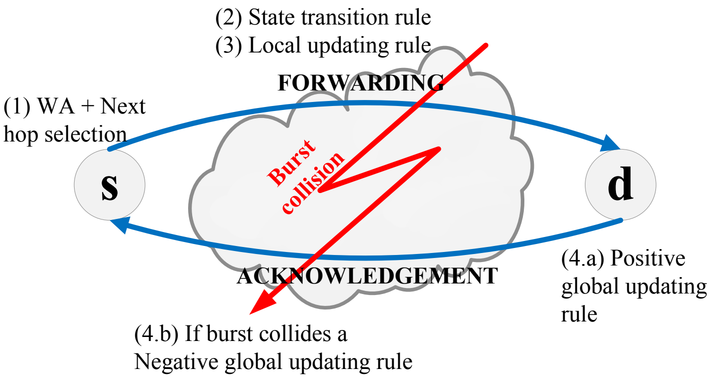

(2) State transition rule
(3) Local updating rule
FORWARDING
(1) WA + Next hop selection
s
Burst collision
d
ACKNOWLEDGEMENT
(4.a) Positive global updating rule
(4.b) If burst collides a Negative global updating rule

Fig. 3.3: High-level RWA algorithm diagram.

Additionally, the algorithm makes use of some pre-defined values in some of the transition rules, which need to be set in advance before any of the network nodes can start transmitting bursts. This process is realized by the initialization algorithm.

3.4.1 Initialization Algorithm

The initialization algorithm initializes the protocol and node parameters to their initial value. Algorithm 1 shows the steps followed in the initialization process. After setting the list of parameters used in the protocol in line 2 (we will describe in detail each one of these parameters in the following sections), the algorithm initializes the routing tables for each node on the network. A shortest path routing algorithm like Dijkstra [112] is used in this step to compute the shortest path from each source node to each remaining destination node using a number of candidate output ports. That is, using the shortest routes, a node n builds a candidate list N_n^m to each destination node m (line 7). Additionally, the initial shortest path routes are also used to calculate the desirability values η_{nj} for the candidates N_n^m. Likewise, the list of wavelength candidates W_n^m is also initialized, which for simplicity is set to W, all the wavelengths available on the network.

In order to reinforce good global solutions that promote the use of less resources on the network per burst transmission, the selection (or transition) rules use also a variable called desirability and represented by η_{nj}. This value gives heuristic information about the attractiveness or desirability of a certain "move". For instance, in the case where the mean path length shall be minimized. Particularly, in the present algorithm version the desirability is defined as $\eta_{nj} = \frac{1}{f(x)}$, where $f(x)$ is equal to the length of the

Algorithm 1 Node's ALBA initialization.

```
variables: $W$ wavelengths, $N$ nodes
Initialize parameters $\alpha_1$,$\rho_1$,$\beta$,$\omega$,$\phi$,$r_0$
for each node $n \in N$ do
    Initialize routing tables
    $m \leftarrow N - \{n\}$
    for each possible destination $m$ do
        Initialize candidate nodes list $\mathcal{N}_n^m$
        Compute initial $\eta_{nj}$ using $\mathcal{N}_n^m$
        Initialize candidate lambdas list $\mathcal{W}_n^m \leftarrow W$
    end for
end for
```

shortest path from node n, through output port j to destination m, as defined by $f(x) = |\, x_{n,j}^{m+}(t)\,|$.

Finally, once all previous steps are over, the network nodes are ready to process data transmissions.

3.4.2 ALBA Routing Algorithm

Algorithm 2 shows the steps involved in the burst forward and backward stages. Initially, the node runs the RWA process to select the next node and wavelength on which the data is going to be transmitted. In this step, the node makes use of (3.1), which is going to be described in detail in the next section.

After this initial stage, the burst is routed to the destination using the forwarding transition rule. In this step, the computation of a random number permits to diversify the routing towards routes either exploiting previous information (SRC values) or exploring for new possible routes. If at the current node, the BCP exploits past knowledge, it checks the list of candidate output ports and calculates the next node to process the burst using (3.2). Alternatively, the lookup for new routes is controlled using an empirical probability distribution that incorporates information about the desirability of each output candidate link towards the destination.

Once the next output port is selected, the node's controller schedules and reserves resources to switch the burst all-optically. Eventually, a collision can happen using this port. In this case, the burst is blocked and the global updating rule is initiated using a negative feedback on the reverse path. On the contrary, if the burst is finally delivered at destination, it means that the data transaction has been successful, which initiates the global updating rule, but now using a positive feedback. In both cases, either having a positive or a negative feedback, the same global updating rule (see (3.6)) is used by the BCP-ACK messages on the reverse path followed by the forward BCP.

The algorithm finishes when the BCP-ACK arrives the data source completing all the return path. This whole process is repeated for every single burst transmission.

Regarding the complexity of the algorithm, let the undirected graph $G = (V, E, W)$

Algorithm 2 Auto load-balancing with acknowledgment (ALBA) algorithm.

```
$x^m(t) \leftarrow \oslash$
$Burst_{blocked} \leftarrow false$
if $x^m(t) = \oslash$ then
    run initial RWA using (3.1)
end if
repeat
    if $\exists j \in \mathcal{N}_n^m(t)$ then
        $r \leftarrow random()$
        if $r \leq r_0$ then
            for all $j \in \mathcal{N}_n^m(t)$ do
                Choose $u \in \mathcal{N}_n^m(t)$ using (3.2)
            end for
        else
            for all $j \in \mathcal{N}_n^m(t)$ do
                Calculate empirical prob. dist. $f_J^m(j)$ using (3.4)
            end for
            Choose $u \leftarrow f_J^m(j)$
        end if
        $x^m(t) \leftarrow x^m(t) \cup \{link(n,u)\}$
        if Burst reservation on $link(n,u)$ is false then
            $Burst_{blocked} \leftarrow true$
        end if
        if $Burst_{blocked}$ is false then
            run positive local updating rule using (3.5)
        end if
    else
        $Burst_{blocked} \leftarrow true$
    end if
until (Burst arrives destination) OR ($Burst_{blocked}$ is true)
repeat
    create BCP-ACK
    $x^m(t) \leftarrow x^m(t) - \{link(n,u)\}$
    run global updating rule using (3.6)
until $x^m(t) = \oslash$ OR BCP-ACK arrives origin node
```

model the network, where V represents the set of vertices (nodes), and E represents the set of edges (links) each one with W wavelengths. The complexity can be calculated as follows: the initial wavelength assignment, which implies the lookup of the best wavelength and output port to transmit the burst, adds a cost of $O(|V||W|)$. Once the lambda is assigned, the computational complexity added on the forward transition and local updating rule is $O(|V|^2)$. These operations are assumed to be run in constant time ($O(1)$), but must be applied at every possible node of the candidates list, which at the most can be in the order $O(|V|)$, and for each node on the route $O(|V|)$. As a result, the forward and backward processing (BCP + BCP-ACK) of a burst transaction adds a complexity of $O(2|V|^2)$. Altogether, the worst case algorithm complexity produced by a burst throughout its traveling period (BCP and BCP-ACK) will be $O(|V||W|+2|V|^2) \sim (|V|^2)$, if $|V| \sim |W|$.

It is worth noting that the per-hop forwarding in ALBA is more computationally intensive than a single packet routing table lookup, which can take $O(|V|)$, in the average case. Nonetheless, in OBS, the processing is per burst control packet instead for every single packet assembled into the burst. Let assume k to be the average number of packets contained in a burst. The complexity of forwarding all those packets one by one would be $O(k|V|)$. Recalling that bursts can be assembled from thousands of packets, which can be one or two orders greater than the number of network nodes, then $|V| \ll k$. As a result, the average forwarding complexity of ALBA is $O(|V|^2) \ll O(k|V|)$, smaller than the packet-by-packet forwarding.

3.5 Wavelength Assignment, Routing and Updating Rule Formulation

The algorithm runs four different rules or functions. Two of them are used in the wavelength selection and forwarding process and the other two for updating the SRC values. Hereafter we use the notation (t) on some of the protocol parameters to remark the possibility that their values can change throughout time.

3.5.1 Initial Wavelength and Routing Selection

The first function is the *initial wavelength and routing selection* which is used to choose the wavelength (λ) and output port (u) in the burst transmission stage at the origin node. Equation (3.1) shows the formula used to this end. The result of the initial RWA is the greatest value of the product between the SRC value (τ_{ijk}) and the desirability of using such output port to the power of β.

$$\{u, \lambda\} = \{\arg\max\{\tau_{ijk}(t)\eta_{nj}^{\beta}(t)\} \mid j \in \mathcal{N}_n^m(t), k \in \mathcal{W}_n^m(t)\} \tag{3.1}$$

The β parameter, as described in Table 3.2, is used to emphasize the value of $\eta_{nj}(t)$, and therefore, whether or not to reinforce the use of shorter paths, e.g., an output port that belongs to the shortest route between the source and destination nodes.

Table 3.2: Notation and parameterizations.

$\mathcal{N}_n^m(t)$	Output candidates list at node n for destination m.
$\mathcal{W}_n^m(t)$	Lambda candidates list at node n for destination m.
β	Potential of choosing an output port.
ρ_1	SRC update rate.
α_1	SRC local update rate.
ω	Power of the exponential decay.
r_0	Exploitation/exploration factor.

3.5.2 State Transition Rule

The second function is the *state transition rule.* This rule is run by the core nodes in the BCP processing and burst forwarding stage and it is based on a pseudo-random-proportional action rule [110] that explicitly balances the exploration and exploitation abilities of the algorithm to look for a suitable path. In this function (3.2), the control parameters are β and $r_0 \in [0, 1]$. When r_0 approaches 1, exploration is neglected, and only the exploitation of present SRC values is taken into account.

$$u = \begin{cases} \arg\max_{j \in \mathcal{N}_n^m(t)} \{\tau_{ijk}(t)\eta_{nj}^{\beta}(t)\} & \text{if } r \leq r_0 \\ J & \text{if } r > r_0 \end{cases} \tag{3.2}$$

Again, the β value is a user-specified parameter that controls the potential benefit of choosing the output link j with a desirability value of η_{nj}.

Fig. 3.4 shows the evaluation of $\tau_{ijk}(t)\eta_{nj}^{\beta}(t)$ when $\beta = 1.0$ and $\beta = 2.0$. We can see that for a fixed SRC value, the shorter the path length through a given output port, the greater the value of the function, specially for large values of τ_{ijk}. As well as this, the greater the value of β, the transition rule in (3.2) creates a bias toward the nodes which are members of the shortest paths with a large SRC value. On selecting the next hop, care must be taken not to select an output port that would eventually create a route loop or without a feasible route to the destination.

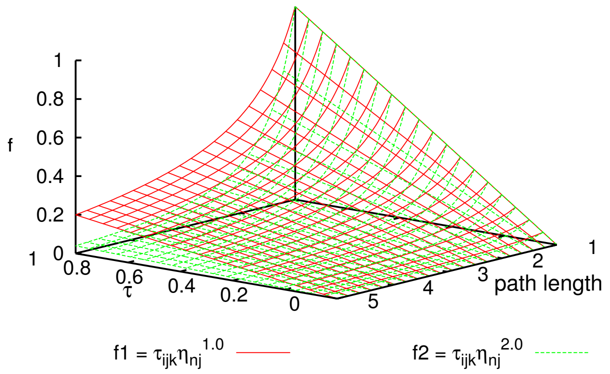


Fig. 3.4: Evaluation of (3.2) with $\beta = 1.0$ and $\beta = 2.0$.

As it has been introduced, the r_0 parameter defines the exploitation/exploration factor. The value r is a uniform random variable. If $r \leq r_0$, the algorithm exploits favoring the best output port from the current SRC values; otherwise, if $r > r_0$, the algorithm explores. In the latter case, the next output port is computed using an empirical probability distribution with probability mass function

$$f_J(j) = \Pr(J = j) = \Pr\{u \in \mathcal{N}_n^m(t) : J(u) = j\} = p_{iuk}(t), \tag{3.3}$$

where

$$p_{iuk}(t) = \frac{\tau_{iuk}(t)\eta^{\beta}_{nu}(t)}{\sum_{j \in \mathcal{N}^m_n} \tau_{ijk}(t)\eta^{\beta}_{nj}(t)}. \tag{3.4}$$

3.5.3 Local Updating Rule

The *local updating rule*, see (3.5), is used to emphasize the good switching selections performed by non-blocked bursts, and it is only applied after a successful switching reservation and before forwarding the burst to the next node towards destination. The parameter α_1 is used to control the update rate of the SRC value. The value $\Delta\tau_{ijk}$ will be explained later in the global updating rule.

$$\tau_{ijk}(t+1) = \tau_{ijk}(t) + \alpha_1 \Delta\tau_{ijk} \tag{3.5}$$

3.5.4 Global Updating Rule

The last function shown in (3.6) is the *global updating rule* and it is run by the returning BCP-ACK along the reverse path used by the corresponding burst. The aim of the rule is to reinforce or weaken the values of the switching tables of each node along the route. The parameter used for controlling the update of the switching table is again a user-specified value ρ_1. If it is small, then the values are increased/decreased slowly, and if it is greater, the previous experience, i.e. old values, is neglected in favor of more recent experiences, i.e. increments/decrements triggered by recent BCP-ACKs, as in

$$\tau_{ijk}(t+1) = (1-\rho_1)\tau_{ijk}(t) + \rho_1\gamma_{ij}\Delta\tau_{ijk}. \tag{3.6}$$

The γ_{ij} value depends on whether the burst is successfully received by the destination node and the path, $x^m(t)$, traveled by the burst. The value is compute as shown in (3.7). For the link belonging to $x^m(t)$, if the burst transmission is successful then $\gamma_{ij} = 1$, or $\gamma_{ij} = -1$ otherwise, while for the rest of links that do not belong to $x^m(t)$, $\gamma_{ij} = 0$.

$$\gamma_{ij} = \begin{cases} +1 & \text{if } link(n,j) \in x^m(t) \text{ and } success = true \\ -1 & \text{if } link(n,j) \in x^m(t) \text{ and } success = false \\ 0 & \text{if } link(n,j) \notin x^m(t) \end{cases} \tag{3.7}$$

On both, the local and global updating rules, the new SRC value is calculated subject to $\Delta\tau_{ijk}$. This value is an exponential decay function of the path length followed by the burst

$$\Delta\tau_{ijk} = e^{-\omega\Delta l}, \tag{3.8}$$

where $\Delta l = |x^r_n(t)| - |x^+_n(t)|$. The first term of the substraction, $x^r_n(t)$, denotes the shortest path to the destination using the chosen/used output port, whereas the second defines the length of the shortest path overall from the current node to destination whichever the output port in use. Thus, for instance, if the route followed to reach the destination is the shortest (or equivalent) and the reception of the burst is correct, the added extra SRC value is maximum as $\Delta l = 0$.

The parameter ω is the decay constant; the greater its value, the faster the exponential vanishes. Therefore, ω can be used to control the reinforcement in both, the local and global updating rules. Fig. 3.5 shows two examples of how the value of Δl strongly affects the final updated pheromone value as a function of ρ_1 for (3.6). Thus, if ρ_1 is large and the previous pheromone value is also large, a bad solution using a longer path than the shortest (or equivalent) can decrease the value of the pheromone concentration considerably.

3.5.5 Evaluation of the Updating Functions

As it can be inferred from the forwarding and updating rules, the concentration value throughout time greatly depends by the formulation itself. Next, we derive some conclusions about the switching rule concentration limits and expected values under some "ideal" long-term scenarios, i.e., when an output port is particularly favored and used to reach a specified destination.

Let by now derive the value of the switching value throughout time when both, local and global updating rules are applied. Without loss of generality, assume the starting time is (t). From (3.5) and (3.6), the first SRC value update is

$$\tau_{ijk}(t+1) = (1-\rho_1)(\tau_{ijk}(t) + \alpha_1 \Delta\tau_{ijk}) + \rho_1 \gamma_{ij} \Delta\tau_{ijk}. \tag{3.9}$$

If we iterate over the second rule update, the new value is

$$\tau_{ijk}(t+2) = (1-\rho_1)[\tau_{ijk}(t+1) + \alpha_1 \Delta\tau_{ijk}] + \rho_1 \gamma_{ij} \Delta\tau_{ijk}. \tag{3.10}$$

If we write (3.10) as a function of the initial rule value at time (t), $\tau_{ijk}(t)$, we get

$$\begin{aligned} \tau_{ijk}(t+2) = {} & (1-\rho_1)^2 \tau_{ijk}(t) + \\ & + (1-\rho_1)^2 \alpha_1 \Delta\tau_{ijk} + (1-\rho_1)\alpha_1 \Delta\tau_{ijk} + \\ & + (1-\rho_1)\rho_1 \gamma_{ij} \Delta\tau_{ijk} + \rho_1 \gamma_{ij} \Delta\tau_{ijk}. \end{aligned} \tag{3.11}$$

Moreover, the third iteration can be expressed as,

$$\begin{aligned} \tau_{ijk}(t+3) = {} & (1-\rho_1)^3 \tau_{ijk}(t) + \\ & + (1-\rho_1)^3 \alpha_1 \Delta\tau_{ijk} + (1-\rho_1)^2 \alpha_1 \Delta\tau_{ijk} + (1-\rho_1)\alpha_1 \Delta\tau_{ijk} + \\ & + (1-\rho_1)^2 \rho_1 \gamma_{ij} \Delta\tau_{ijk} + (1-\rho_1)\rho_1 \gamma_{ij} \Delta\tau_{ijk} + \rho_1 \gamma_{ij} \Delta\tau_{ijk}. \end{aligned} \tag{3.12}$$

From (3.11) and (3.12) we can see that a pattern is followed, so for the $n-th$ term of

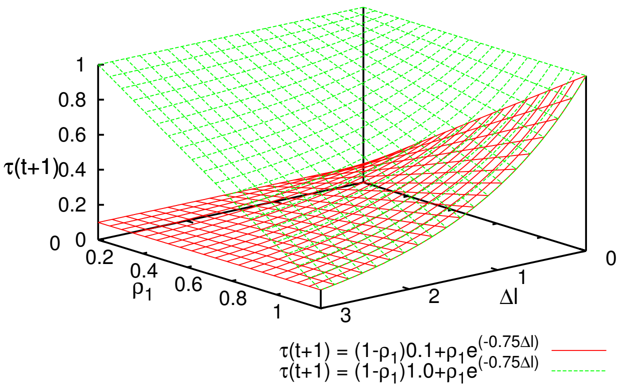


Fig. 3.5: Global updating rule analysis with two different SRC values (0.1 and 1.0).

the SRC value we have

$$\tau_{ijk}(t+n) = (1-\rho_1)[\tau_{ijk}(t+n-1) + \alpha_1\Delta\tau_{ijk}] + \rho_1\gamma_{ij}\Delta\tau_{ijk}, \tag{3.13}$$

which in terms of the initial $\tau_{ijk}(t)$, the SRC value follows the expression

$$\begin{aligned}\tau_{ijk}(t+n) = {} & (1-\rho_1)^n\tau_{ijk}(t) + (1-\rho_1)^n\alpha_1\Delta\tau_{ijk}+ \\ & + (1-\rho_1)^{n-1}\alpha_1\Delta\tau_{ijk} + \ldots + (1-\rho_1)\alpha_1\Delta\tau_{ijk}+ \\ & + (1-\rho_1)^{n-1}\rho_1\gamma_{ij}\Delta\tau_{ijk} + \ldots + \rho_1\gamma_{ij}\Delta\tau_{ijk}.\end{aligned} \tag{3.14}$$

This expression (3.14) can be further simplified, as in

$$\tau_{ijk}(t+n) = (1+\rho_1)^n\tau_{ijk}(t) + \alpha_1\Delta\tau_{ijk}\sum_{m=1}^{n}(1-\rho_1)^m + \rho_1\gamma_{ij}\Delta\tau_{ijk}\sum_{m=0}^{n-1}(1-\rho_1)^m. \tag{3.15}$$

Then, taking the infinity of the geometric series of both summation terms, that is $n \to \infty$, and due to the fact that $0 < \rho_1 < 1$, then

$$\sum_{m=1}^{\infty}(1-\rho_1)^m = \frac{1-\rho_1}{1-(1-\rho_1)} = \frac{1-\rho_1}{\rho_1}, \tag{3.16}$$

and

$$\sum_{m=0}^{\infty}(1-\rho_1)^m = \frac{1}{1-(1-\rho_1)} = \frac{1}{\rho_1}. \tag{3.17}$$

Furthermore, since $0 < \rho_1 < 1$, the first term $(1-\rho_1)^n$ from (3.15) also converges,

$$\lim_{n\to\infty}\bigg|_{0<\rho_1<1}(1-\rho_1)^n = 0. \tag{3.18}$$

Therefore, after a long time (or number of iterations), the switching concentration value would have an approximate value of

$$\begin{aligned}\lim_{n\to\infty}\tau_{ijk}(t+n) &= 0+\alpha_1\Delta\tau_{ijk}\frac{1-\rho_1}{\rho_1}+\rho_1\gamma_{ij}\Delta\tau_{ijk}\frac{1}{\rho_1}= \\ &= \alpha_1\Delta\tau_{ijk}\frac{1-\rho_1}{\rho_1}+\gamma_{ij}\Delta\tau_{ijk}= \\ &= \Delta\tau_{ijk}\left(\frac{\alpha_1}{\rho_1}-\alpha_1+\gamma_{ij}\right). \end{aligned} \tag{3.19}$$

As shown in (3.19), the switching concentration values will approximate to a value that depends on four main values/parameters: $\Delta\tau_{ijk}$, which depends on the routes followed by the bursts; γ_{ij}, which depends on the routes followed by the burst as seen by the network node and its input/output ports of the optical switch; and two user parameters, α_1 and ρ_1.

Regarding the γ_{ij} value, that is, depending on how the burst traffic flows make use of the available paths and output ports in the switches, we can get:

- If a specific switching configuration is not actively used, then $\gamma_{ij}=0$. From (3.19) we get $\tau_{ijk}=\Delta\tau_{ijk}\left(\frac{\alpha_1}{\rho_1}-\alpha_1\right)$. Assuming $\alpha_1\sim\rho_1$ and these take small values, then $\tau_{ijk}\sim\Delta\tau_{ijk}$.
- If the use of a switch configuration is actively used and produces good results, then $\gamma_{ij}=+1$ and assuming the same scenario as in the previous case, consequently $\tau_{ijk}\sim 2\Delta\tau_{ijk}$.
- And finally, if the use of a configuration is wrong over a long period, $\gamma_{ij}=-1$ and $\tau_{ijk}=0$, as $\tau_{ijk}\not<0$.

Furthermore, recall that the value $\Delta\tau_{ijk}=e^{-\omega\Delta l}$. If $\Delta l=0$, then $\Delta\tau_{ijk}=1$. However, if the path followed by the burst is not the shortest, then its value will be smaller than 1. In conclusion, the switching configurations that are actively used will move around $\Delta\tau_{ijk}\left(\frac{\alpha_1}{\rho_1}-\alpha_1+1\right)$ with its value being function of the length of the end-to-end path and with a tendency to follow load-balanced shortest paths or equivalents.

Fig. 3.6 shows the evaluation of the ρ_1 and α_1 parameters together. It is worth noting that these can have a great impact on the results. If we evaluate them together, we can come up to the following conclusions:

- If $\rho_1\lesssim 1$, then $\Delta\tau_{ijk}\left(\frac{\alpha_1}{\rho_1}-\alpha_1+\gamma_{ij}\right)\geq\Delta\tau_{ijk}\gamma_{ij}$. Therefore, recent experience (i.e., recent burst transmissions) is fostered and provokes high dynamism. Actually, the dynamism is primarily contributed by successful transmissions which can greatly change whether the node routes the burst to one output port or another.
- If $\alpha_1=0$, then $\Delta\tau_{ijk}\left(0-0+\gamma_{ij}\right)=\Delta\tau_{ijk}\gamma_{ijk}$, and the same behavior as in the previous point applies.
- If $\alpha_1=\rho_1$, then the approximate SRC has a value of $\Delta\tau_{ijk}\left(1-\alpha_1+\gamma_{ij}\right)$. Since $0\leq\alpha_1<1$, then SRC is $\geq\Delta\tau_{ijk}\gamma_{ij}$, thus having a similar performance as in the

previous point.

- Finally, if $0 \lesssim \rho_1 \ll 1$, then past knowledge prevails, hence good stable conditions are maintained and fostered in the long-term.

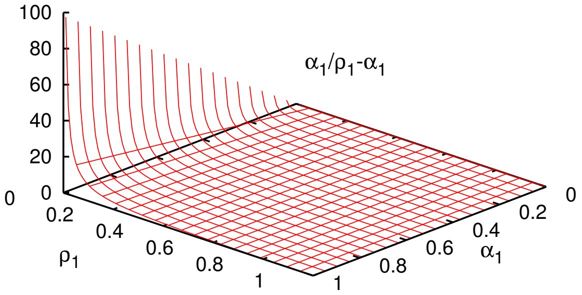


Fig. 3.6: Function evaluation of (3.19).

3.6 Simulation Results

The evaluation of the RWA protocol proposed in this chapter has been made through simulation. We divide the results section into two main sub sections. The objective is two-fold: first, to test in a simple and easy-to-manage network the operation of the algorithm, and second, to evaluate it on a more realistic scenario, even on different network topologies.

In order to extensively compare the proposed RWA, ALBA, three other RWA protocols are also evaluated, namely: shortest path routing with random wavelength assignment (SR), random path routing and wavelength selection (RR) and shortest path routing with first-fit Traffic Engineering wavelength assignment (FFTE). The former selects the lambda by means of a uniform probability distribution among the pool of wavelengths available on the network. In RR, every node calculates three (when possible) different routes from each possible source node to every remaining destination. Among these, it uniformly selects one route and one wavelength. Additionally, in FFTE [113], every node has an ordered list of assignable wavelengths, but each one using a different order. The node selects the first one *(first-fit)* available to transmit the burst.

In all the scenarios, a wavelength capacity of 10 Gbps is assumed, however the number of wavelengths on the network vary between the two chosen scenarios, as it will be explained later. Regarding the setup of the hardware devices, the control packet processing time and the non-blocking matrix switching setup are set to 10 μs and 5 μs, respectively. Moreover, an initial burst query processing of 1 μs is added between the generation of the burst and its transmission scheduling. Following one of the motivations of this present work, we devise a network without wavelength conversion, thus, the data transmissions must fulfill the wavelength-continuity constraint.

With regard to the traffic characteristics, a size-based algorithm [114] is used in the burst assembly with a packet arrival distribution following a Poisson process. The average packet size is set to 485 bytes. Bursts are of fixed size and equal to 100,000 bytes, which on a 10 Gbps channel is equivalent to 80 μs. With such short burst size, the network runs on a very dynamic environment, so that the dynamic properties of the ALBA protocol can be evaluated in detail. Many different works [85] deal with OBS network scenarios with burst sizes in this order of magnitude. Nevertheless, it is worth mentioning that there are also a number of works that devise burst sizes in the order of ms, especially on Grid over OBS [79].

3.6.1 Results on the Fish-Like Network

The scenario we consider here is a fish-like network composed of eight OBS network nodes, as shown in Fig. 3.7. Our aim is to specifically validate the operation of the routing and wavelength assignment of ALBA and its performance throughout time. The number of lambdas is intentionally reduced to only two in order to make easier the analysis of results. Three heavy hitter traffic flows of 0.66 Er. each (adding up 2 Er. in total) are transmitted from three different traffic sources (nodes N1, N2 and N3) to the same destination node (N8). In this scenario, all the traffic has to be switched/routed at a common node (N4), hence this node becomes a possible bottleneck. The simulation starts with the network empty, and at time 0.5 ms the three flows are inserted from their correspondent source nodes.

On the fish-like network, results have been gathered by repetition (i.e., simulation runs) and averaged over at least 30 measures in order to obtain 95% confidence intervals for all the results. However, the intervals are so narrow that they are omitted in order to improve the readability of the figures.

Fig. 3.8(a) shows a comparison between ALBA, SR and RR in terms of burst blocking probability (BP). The x-axis represents the time elapsed from the beginning of the simulation run. A clear differentiation can be observed regarding the blocking performance between the three protocols. As expected, SR gets the worst BP due to the fact that the three traffic source nodes (N1, N2 and N3) use the shortest path through N5 for delivering the bursts so that nearly half of them get blocked at N4.

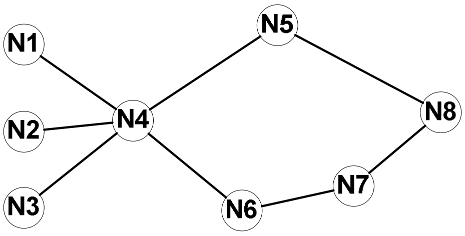
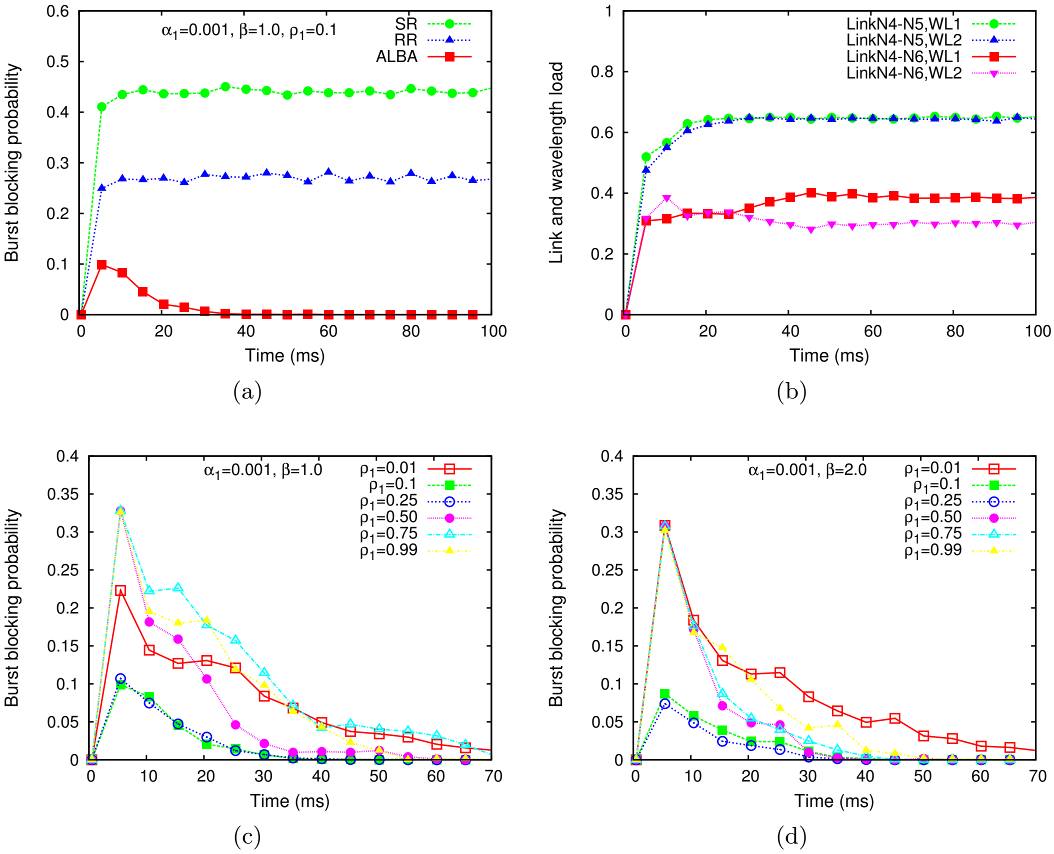


Fig. 3.7: Fish-like network with 8 nodes.



(a)

(b)

(c)

(d)

Fig. 3.8: Simulations on the fish-like network with 2 wavelengths: (a) burst blocking comparison of different RWA protocols, (b) link and wavelength utilization of the ALBA protocol, (c) burst blocking of ALBA with $\alpha_1 = 0.001$ and $\beta = 1.0$, and (d) burst blocking of ALBA with $\alpha_1 = 0.001$ and $\beta = 2.0$.

In a similar way, RR does not overcome the congestion state at N4 though making use of the two possible routes between the sources and node N8. On the contrary, in ALBA we have some burst losses at the beginning of the simulation, but then, the protocol overcomes the situation canceling the contentions for the rest of the simulation run. ALBA balances the traffic load between the two possible routes and wavelengths according to the SRC values and the state transition rules. Hence, after an initial period of time within the BCPs forage for the best route-lambda, the protocol stabilizes the SRCs that permit to cancel the contentions.

In ALBA, bursts are no longer transmitted over the shortest path, but balanced, not only over different routes, but also over different wavelengths. Fig. 3.8(b) shows the link and wavelength utilization for the ALBA case in the same network simulation scenario. The link between N4 and N5, which belongs to the shortest path is more loaded than the link N4-N6. The SRC scoring method takes into account the length of the path and the congestion level, and for that reason, in front of two possible paths of different lengths and the same congestion level, it takes the shortest.

The next two graphics in Fig. 3.8(c) and Fig. 3.8(d), show the performance of the protocol under different parametrization of β (1.0 and 2.0) and $0 < \rho_1 < 1.0$. We are interested in evaluating the convergence time that ALBA takes to cancel further burst collisions. From these two figures, it can be seen that either large or small values of ρ_1 tend to prolong the contention period. When ρ_1 is small, the SRC value is updated in small steps, hence more positive/negative burst deliveries are needed in order to foster a certain switching configuration at N4. Interestingly, we can also see that increasing the value of β, as in Fig. 3.8(d), the convergence time also decreases faster. A large value of β triggers, especially at the beginning of the simulation, many transmissions through the shortest path, which increases the number of burst collisions and negative updates of the SRC for the shortest route, hence switching to the other path more rapidly.

As we have seen, the parametrization of ALBA has a great impact on the BP and the delay incurred in canceling it. Next we provide a deeper analysis about the parametrization of ALBA. Table 3.3 shows the values given to ρ_1, α_1 and β for the fish-like scenario resulting in 90 different simulation cases.

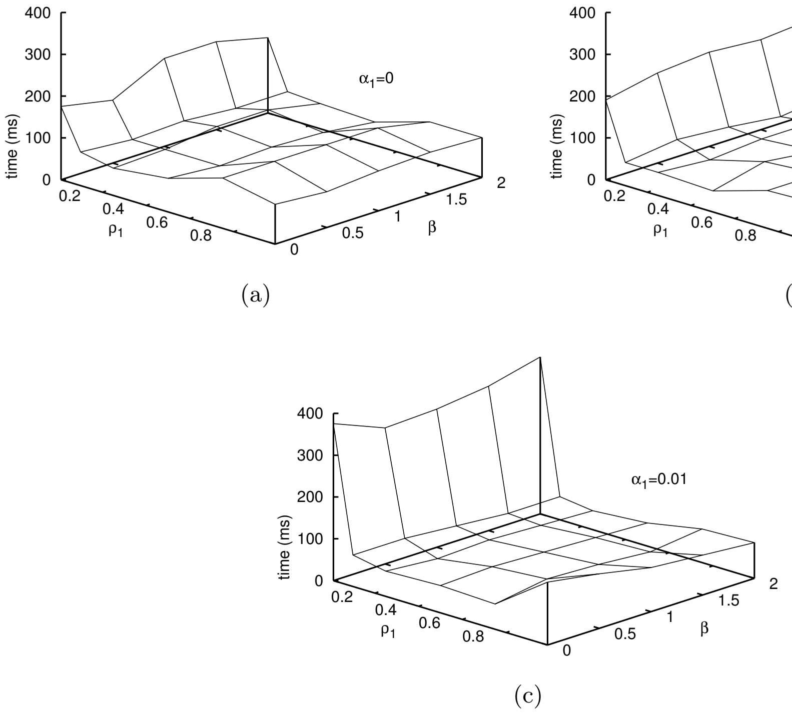
Table 3.3: Parameters values used in the simulations.

Parameter	**Fish-Like**	**NSFNET**
$r0$	0.8	0.9
ω and ϕ	0.75	0.75
ρ_1	0.01, 0.1, 0.25, 0.5, 0.75, 0.99	0.001, 0.01, 0.1
α_1	0, 0.005, 0.01, 0.1	0.001
β	0, 0.5, 1.0, 1.5, 2.0	1.0, 2.0, 3.0

Due to this parameter and value diversity the results are now shown in 3D-format (refer to Fig. 3.9). The z-axis represents the converge time to cancel further burst collisions within a time-frame of 400 ms. Figs. 3.9(a), 3.9(b), and 3.9(c) show a comparison of the parametrization of α_1 as a function of ρ_1 and β. We can see how, in general, the smaller the value of ρ_1, the more time the protocol needs to cancel the contentions. This behavior was also seen in Fig. 3.8(c). Likewise, in the comparison among the three graphs we can see that the greater the value of α_1, the shorter the convergence time for medium values of ρ_1. So, when using large values of α_1 in the local updating rule, successful burst switchings update the SRC values with greater increments, and as a consequence they remark more rapidly the link-wave that the upcoming bursts shall take.

3.6.2 Results on the NSFNET

In addition to the previous results, in this section the performance of the protocol under a more realistic scenario, such as the NSFNET network, is also evaluated. In



(a)

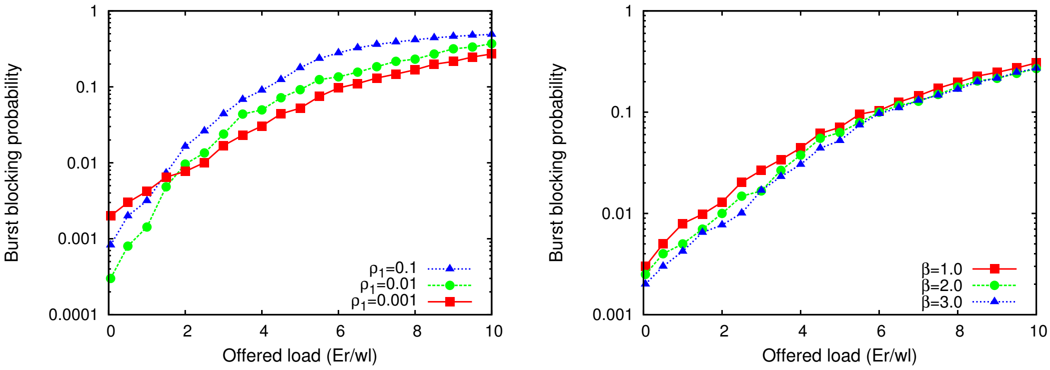
(b)

(c)

Fig. 3.9: Convergence time to cancel burst collisions on the fish-like network with 8 nodes and 2 wavelengths using ALBA: (a) as a function of ρ_1 and β with $\alpha_1 = 0$, (b) as a function of ρ_1 and β with $\alpha_1 = 0.001$, and (c) as a function of ρ_1 and β with $\alpha_1 = 0.01$.

this specific case, there are 16 wavelengths on the network with a capacity of 10 Gbps per channel. This number of wavelengths is already a representative value for the deployment of partially meshed WDM networks. As in the previous subsection, the burst length is limited to 100,000 bytes. Now the simulation uses the batch means method for gathering the results about burst loss probability and mean route length.

In Fig. 3.10 the burst blocking probability is represented for different ALBA parameterizations as a function of the total offered traffic load in Erlangs per wavelength. Fig. 3.10(a) depicts the results for different values of ρ_1 when $\alpha_1 = 0.001$ and $\beta = 3.0$. Interestingly, at low loads, when the value of ρ_1 is very small, the burst contention rate remains higher with respect to greater values of ρ_1. Due to the lesser number of bursts to transmit at low offered loads, with small values of ρ_1 the SRC values change over time at a lower rate which causes the switching configurations between different alternatives to be less emphasized. A similar SRC value in the state transition rule (3.2) with exploration ($r > r_0$) does not strenuously boost the use of a certain routing decision using the empirical distribution, so that BCPs wander at a higher rate, which can cause more burst collisions. However, at high loads, with a greater number of bursts transmissions, and thus, on a more dynamic scenario, a smaller value of ρ_1 provides



(a)

(b)

Fig. 3.10: ALBA burst blocking probability evaluation on the NSFNET with 16 wavelengths: (a) as a function of ρ_1 with $\alpha_1 = 0.001$ and $\beta = 3.0$; and (b) as a function of β with $\alpha_1 = 0.001$ and $\rho_1 = 0.001$.

better results in comparison with $\rho_1 = 0.01$ or $\rho_1 = 0.1$. In this case, the exploitation of the SRC is fostered instead of the exploration done by the BCPs, and therefore the past good knowledge of certain routing/switching configurations is kept.

With the value of ρ_1 that provides the best results, Fig. 3.10(b) shows the blocking probability for different values of β. Although at high loads, the plots almost converge to the same point, at low and medium loads we can see that the best performance is achieved for $\beta = 3.0$. When β is increased, shorter routes are boosted with respect to others and as the network load is not very high, short routes can still provide the best results. As a summary, we can conclude the protocol parametrization has an impact on the performance perceived under different network scenarios.

Table 3.4 summarizes some of the main conclusions described so far regarding the NSFNET results. The two fixed parameters are given in the first column, whereas in the second the best performance is described when the remaining parameter from the triple $\{\alpha_1, \beta, \rho_1\}$ is changed. The uparrow, $\uparrow$, indicates the best result is achieved when the parameter is increased.

Table 3.4: Parameter performance on NSFNET.

Fixed parameters	**Best performance if parameter**
α and ρ_1	$\beta \uparrow$.
α and β	Depending on network load: - if low load $\Rightarrow \rho_1 \uparrow$, - if high load $\Rightarrow \rho_1 \downarrow$

3.6.3 Network Topology Performance Comparison

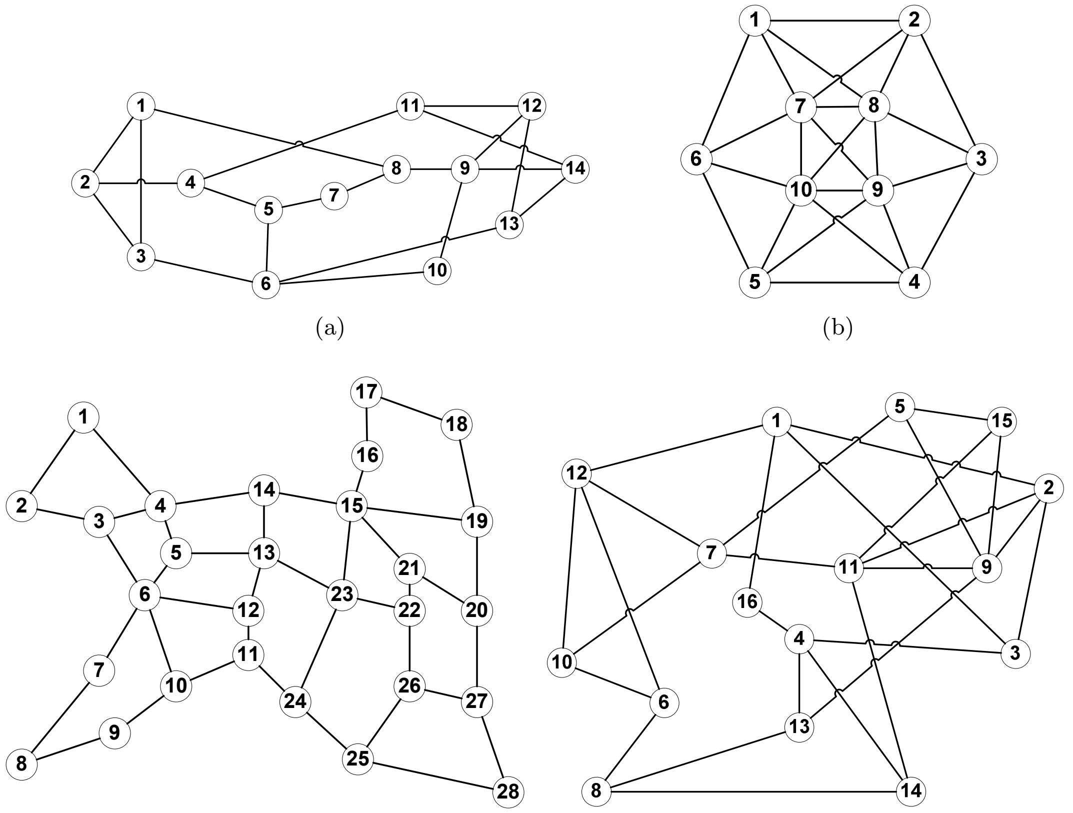
In this section, the objective is to evaluate the properties and performance of the protocol as a function of the underlying network topology. To this end, four different transport network topologies (NSFNET, extended European Optical Network (EON), RANDOM and the slightly modified SmallNet [115] have been used. A representation of each topology is available in Fig. 3.11. These networks have been chosen so that they provide different values in terms of connectivity degree, number of nodes and links, etc. Further details about their characteristics are shown in Table 3.5. In particular, the random network in Fig. 3.11(d) was generated using a Waxman algorithm [116] with the following parameters values: $\lambda = 0.000015$, $\alpha = 0.8$, $\beta = 0.35$ on a 1000x800 domain region. Other topologies like bus, ring or star are not likely to be representative for a RWA due to their simple connectivity and limited routing possibilities, and therefore they have not been considered in this analysis.

Table 3.5: Network topologies.

Feature	**NSFNET**	**EON**	**RANDOM**	**SMALLNET**
Nodes	14	28	16	10
Links	21	41	29	24
Mean degree ($\overline{deg}(G)$)	3.00	2.93	3.625	4.8
Mean shortest path	2.14	3.56	2.17	1.53

In all the network scenarios the number of wavelength is 16, without wavelength conversion and with a channel capacity of 10 Gbps. Moreover, neither fiber delay lines, nor WC are used on the network for contention resolution purposes. Additionally, the control packet processing time and the non-blocking matrix switching setup are set to 10 μs and 5 μs, respectively. Finally, regarding the traffic characteristics, a size-based assembly algorithm is used using a Poisson process packet arrival. The burst length is then fixed to 100,000 bytes, equivalent to 80 μs on a 10 Gbps channel.

Fig. 3.12 gives the burst blocking probability results as a function of the offered load to the network in Erlang per wavelength (Er/wl), and for each topology under analysis. At very low loads, the FFTE protocol gets the best results, mainly due to the fact that at this stage almost every node on the network uses a lambda that none of the remaining nodes use. Nevertheless, taking into account the whole load range, we can state that in the four cases ALBA outperforms the rest of algorithms, although with different margins. In this respect, the effect of the greater mean connectivity degree of the network ($\overline{deg}(G)$), as well as the smaller mean route length (MRL), largely benefits the performance of ALBA, so that, especially at high loads, the improvement margin in comparison to the basic SR protocol is more stressed. Therefore, the greater $\overline{deg}(G)$, the better the improvement provided by ALBA. Following this last statement, the extended EON topology strongly penalizes the BP performance, not only of ALBA,



(a) (b)

(c) (d)

Fig. 3.11: Set of network topologies under test: (a) NSFNET, (b) SmallNET, (c) EON, and (d) Waxman Random.

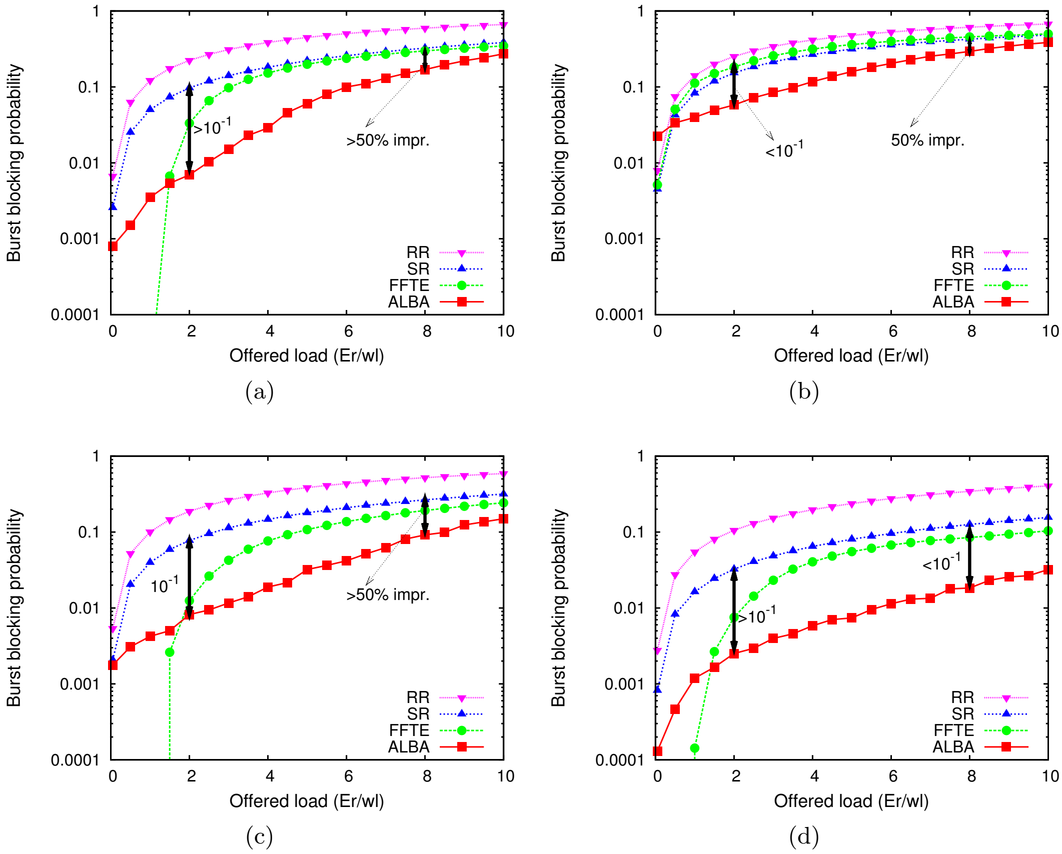
but also of the rest of tested algorithms due to the low connectivity degree of the nodes from the topology periphery and the greater MRL.

To further gain insight into the performance of ALBA on different topologies, Fig. 3.13(a) shows the BP but now as a function of the network load calculated as

$$\rho = \frac{\sum_{i=1}^{n} M_i}{\sum_{i=1}^{n} C_i} \tag{3.20}$$

where M_i is the traffic carried by link i, C_i is the capacity of the same link, and n is the number of links (bidirectional) on the network. In this way, we can get real comparable results whatever the underlying topology in use as we measure the effective carried load on the network.

At first sight, the BP growth rate between medium and higher loads is similar among the four topologies. Furthermore, we can see three main BP trends very related to the mean route length characteristic of each network. So, the shorter the mean route length is, the lower the BP. With regard to this, on the RANDOM and NSFNET topologies, which both have a very similar MRL (2.17 and 2.14, respectively), there is not

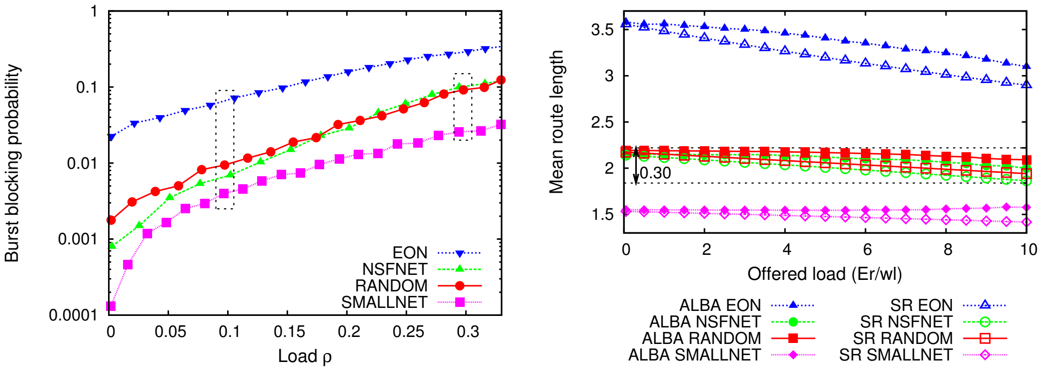


(a) (b) (c) (d)

Fig. 3.12: Burst blocking probability on different network topologies: (a) NSFNET, (b) EON, (c) Waxman Random, and (d) SmallNet.

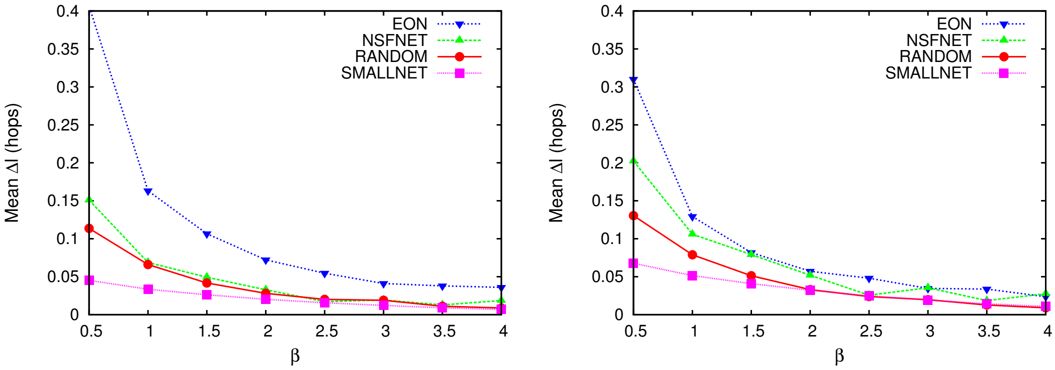
a definitive topology where ALBA outperforms despite $\overline{deg}(nsfnet) < \overline{deg}(random)$. Therefore, we can conclude that on similar network load conditions, the MRL is the main factor of influence on the ALBA BP performance.

Fig. 3.13(b) show the results of the mean route length as a function of the offered load on the four network topologies under study. To improve the readability of the graph, we compare ALBA only with shortest path routing (SR), which as the name indicates, it shall always provide the shortest path. We can see on the graph that almost in all the topologies, increasing the offered traffic provokes a decrease of the MRL, as in average, the bursts that survive take shorter paths consuming less resources per transaction. However, this decrease is less sharp for ALBA, or even in the case of the SmallNet does not decrease at all but remains nearly constant. As for ALBA, the burst transmissions tend to be balanced over different routes that are not necessarily the shortest ones, although comparable to the ones provided by the SR algorithm. As a result, ALBA is able to enhance the overall blocking probability without penalizing the burst end-to-end delay due to the increase of the MRL. This characteristic is important in order to meet the QoS of certain multimedia/streaming applications.



(a) (b)

Fig. 3.13: Comparison among topologies: (a) blocking probability vs. network load, and (b) mean route length (MRL).



(a) (b)

Fig. 3.14: Evaluation of the average Δl value: (a) in global updating rule with load = 5.0 Er/wl, and (b) in local updating rule with load = 5.0 Er/wl.

As it has been seen in the algorithm formulation section, the local (3.5) and global updating rules (3.6) are updated with a value that depends on Δl, the length difference between the shortest path using the chosen (used) output port and the shortest path overall from the current node processing the BCP (BCP-ACK). At the same time, Δl can depend on the value of β. We want to check here the influence of β on Δl. Fig. 3.14 shows the mean Δl value at load 5.0 Er/Wl for both global and local updating rules and for different topologies. In this case the rest of parameters remain the same and only β is modified. If we focus on the global updating in Fig. 3.14(a) we can see that, as expected, the average Δl depends on the size of the network and its mean shortest path value. The greater the network size, the greater Δl. Moreover, increasing the value of β drops Δl at an exponential rate that depends on the size of the network. Recall that β is used to emphasize the use of output links from the optical switch that are part of shorter paths. In conclusion, the β value can also control the load-balancing behavior of the algorithm boosting the use of alternative hop-by-hop paths that are equivalent in length to the shortest one.

3.7 Summary

In connectionless sub-wavelength optical networks, packet or burst contentions are one of the main issues. Even at very low loads, due to the use of one-way provisioning protocols and the lack of efficient buffering mechanisms in the optical domain, contentions can occur. In this chapter, we have presented an auto load-balancing distributed RWA algorithm for optical burst-switched networks with wavelength-continuity constraint as a proactive mechanism to decrease the number of contentions in the network.

The RWA and burst forwarding are based on the exploitation and exploration facilities using switching rule concentration values that incorporate contention and forwarding desirability information for every wavelength and port in the optical switch. To support such architecture, forward and backward control packets are used in the burst forwarding and updating rule processes, respectively.

In order to extensively analyze the benefits of the new algorithm, four different network topology scenarios have been used, proving from the results that the proposed method outperforms the rest of tested RWA algorithms at different margins depending on the characteristics of the topology without penalizing other parameters such as end-to-end delay.

In spite of these promising results, we can also summarize that cost-effective mechanisms like the one proposed in this chapter still present problems to cope with the blocking probability cause mainly in the core of the network, even at very low offered traffic loads. For this reason, we will introduce in the next chapter a mechanism that guarantees the deliver for in-transit bursts thereby overcoming such undesirable blocking behavior. A very interesting property of the new protocol will be its hybrid connectionless/connection-oriented bandwidth provisioning capabilities.

Chapter 4

Hybrid Connectionless and Connection-Oriented MAC-based Sub-Wavelength Optical Networks

The emergence of a broad range of network-driven applications (e.g. multimedia, online gaming) brings in the need for a network environment able to provide multi-service capabilities with diverse QoS guarantees. Even in the case when providing high-priority treatment in connectionless sub-wavelength optical networks, data delivery cannot be guaranteed. On the contrary, by making use of connection-oriented bandwidth provisioning, the network can guarantee the services that have been successfully allocated in the reservation stage. In this chapter, we propose a hybrid connectionless/connection-oriented (CL/CO) architecture based on a medium access control protocol to support multiple services and QoS levels in sub-wavelength optical burst-switched networks.

4.1 Introduction

Triple-play services (i.e. data, voice and video) and the new deployment of web-based multimedia applications have increased the amount of bursty traffic on the Internet. Such services may benefit from the utilization of packet/burst-switched networks. However, certain applications such as IPTV, together with Grid and Cloud computing (e.g. PC virtualization, etc.) can benefit from time-division multiplexing connection-oriented transport networks. In this sense, a desirable requirement is to provide and guarantee a great variety of services over the same optical network infrastructure.

Medium access control protocols can efficiently manage the huge optical bandwidth by providing contention avoidance schemes to further improve, or even guarantee, the packet or burst delivery on the network. This characteristic becomes of special interest for all-optical networks with limited contention resolution. However, its use in sub-wavelength all-optical networks, and in particular in OBS, has not been intensively analyzed. As introduced in Chapter 2, most present MAC solutions have focused on metro-ring architectures leaving out of their scope other common topologies, i.e. mesh.

In this chapter, we give an insight into the performance of a lossless (in the core of the network) enhanced multi-service OBS MAC protocol with QoS. The MAC protocol is an adaptation of the IEEE 802.6 Distributed Queue Dual Bus (DQDB) [117]. DQDB defines a queue-arbitrated access to the channel guaranteeing zero losses for transmitted frames. The original protocol is enhanced by adapting it over mesh network topologies and integrating it as a wavelength-aware MAC for wavelength continuity constraint slotted OBS networks. The protocol is referred as DAOBS, which stands for distributed access for optical burst switching.

A differential contribution with respect to past approaches is the use of the MAC on more complex topologies such as mesh. The protocol supports diverse services by means of a dual access scheme: a queue arbitrated (QA) and a pre-arbitrated (PA) burst transmission method, for connectionless and connection-oriented TDM sub-wavelength services, respectively. Differentiation among connectionless classes is provided by a multi-queue priority system along with a distributed queue-arbitrated channel access module.

In the queue-arbitrated mechanism, a set of counters keep track of the requests for free slots and their availability using request and burst control packets. With this information, bursts are only transmitted when there are free resources; hence burst losses due to overlapping are avoided. Therefore, this architecture guarantees the delivery for all bursts transmitted and in transit. Moreover, this system also permits higher priority bursts from a node to preempt lower priority ones from other nodes on the ingress queues and be placed and transmitted ahead in time, even among different nodes.

Connection-oriented services use the pre-arbitrated channel access which can incorporate different slot scheduling algorithms. The proposed algorithms deliver diverse results depending on the network state and traffic distributions. Hence, a dynamic decision making protocol may enhance the scheduler according to the type of service request, the present network utilization and the current traffic distribution on the network.

In spite of the good results obtained with the proposed MAC protocol for hybrid CL/CO, there are two main issues to solve. On the one hand, the queue-arbitrated MAC operation slightly increases the access delay for connectionless bursts. On the other hand, the architecture requires setting up and allocating the MAC entities as an overlay of lightpaths and light-trees. Both issues are tackled in the course of this chapter, first by quantifying the access delay using an analytical model, and second, by formulating mathematically the light-tree overlay and its optimization.

The remainder of the chapter is as follows. In Section 4.2, the proposed MAC protocol is introduced and its QoS enhancements and access modes are described in detail. Results through simulations are analyzed in section 4.3. The access delay analysis and the light-tree optimization are presented in Section 4.4 and Section 4.5, respectively. And finally, Section 4.6 concludes this chapter with the main contributions and results.

4.2 Hybrid CL/CO QoS-enabled MAC for OBS

In line with the previous connectionless RWA presented in Chapter 3, we propose using an optical burst-switched network without wavelength conversion as the sub-wavelength provisioning network substrate. Furthermore, the network data channel is time-sliced, e.g. slotted OBS [118], and a constant-based offset scheme is applied as in [119]. This last approach allows fixed offsets to be used by means of input delay lines at each input burst data port of a length equivalent to the maximum delay incurred in the processing of the burst control packet.

One of the main differences of DAOBS compared to other MAC-based schemes for OBS networks, is that in the former case, the protocol runs on mesh network topologies. To realize this, the network is logically partitioned as an overlay network of mono-color light-trees, wherein every node, depending on the overlay setup, may access different light-trees to reach some specific destination nodes.

Conceptually, the DAOBS optical tree has some similarities with a light-trail [94]. Both are network architectures that can span over ring and mesh topologies using optical buses or lightpaths enabling the participating nodes to share the capacity of the channel in a spatial reuse time-shared approach. Moreover, both also define the figure of a controller, the Head of Bus (HoB) in DAOBS and the *arbitrator* in the light-trail. However, unlike in [94], DAOBS is also a burst MAC protocol with collision avoidance in the core of the network that ensures the delivery of the burst without using electronic buffers. Furthermore, the request-granting process works differently. In DAOBS' QA access mode, this process works as a very simple algorithm based on a counting and monitoring process, whereas in the light-trail solution an explicit request-grant is established between the client node (the node requesting slots) and the *arbitrator*, which decides what node is granted in a per slot basis. Finally, DAOBS also includes a wavelength assignment module since a node can belong to more than a single light-tree and, as a result, multiple path/wavelength options are available to the nodes while transmitting data.

In this thesis we use a greedy meta-heuristic graph coloring algorithm to generate the MAC overlay network. The coloring of the graph is based on minimum-spanning trees. The remaining uncolored links and the minimum-spanning trees conform together a superset of the set of lightpath virtual topologies [14] able to improve the utilization capacity of the network. This scheme features two important characteristics: (a) it eliminates loops amongst the nodes belonging to the light-tree, and (b) the coloring of the rest of links avoids contending wavelengths on the same output port of the optical switch. As a result, one of the advantages is that with this tree-based instantiation, collisions from two different input ports are avoided. This is especially useful in the case of dealing with OBS networks without wavelength conversion. Furthermore, light-tree based topologies are useful for delivering multicast or groupcast services, or merely for multimedia broadcasting. Besides, optical musticasting can be more efficient than electronic multicasting since "splitting light" is conceptually easier than copying a packet in an electronic buffer [14]. For the static case, that is, all bandwidth demands among source and destination nodes are known in advance, we can optimize the establishment of the light-trees by formulating the problem as an integer linear programming optimization. This will be further developed in Section 4.5.

4.2.1 Architecture and Basic Operation

Fig. 4.1 shows a high-level diagram of the hybrid connectionless/connection-oriented multi-service aware DAOBS architecture. The most important entity in the architecture is the OBS MAC. The MAC is the key-enabler to support the service differentiation. It is composed of four main sub-modules:

- A QA burst access module. This module includes all the logic necessary to arbitrate the burst transmission based on the QA access mode operation.
- A PA burst access module. This module includes the logic to process the connection-oriented services. It generates the connection IDs (CID) and implements the reservation and provisioning algorithms for TDM bursts.
- A slot scheduling (SS) algorithm module. This module is active only in the HoB, which is responsible for managing the connection-oriented sub-wavelength channels within its light-tree.
- A burst assembly, classification and scheduling module. This last module includes the burst assembly algorithms and classifies the new burst into the QA and the

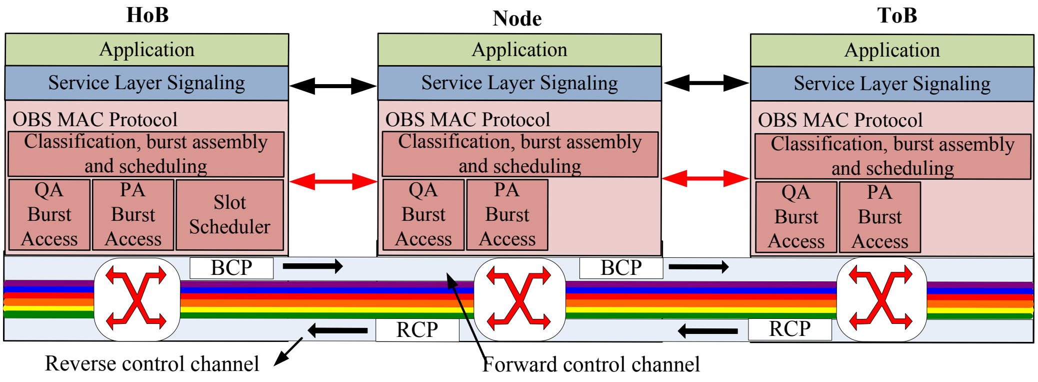


Fig. 4.1: DAOBS protocol and network architecture.

PA modules according to its service type. It also enqueues the bursts into its corresponding QA or PA ingress queues.

The root or top-most node of the light-tree is the Head of Bus node, and all leaf nodes are Tail of Bus (ToB) nodes (see Fig. 4.2). We consider two unidirectional control channels, which can be in-fiber (i.e. using a specific wavelength): the downstream or forward channel, which goes from the HoB node to the ToB nodes, and the upstream or reverse channel, on which QA access request packets are forwarded from the ToBs to the HoB. The HoB node is responsible for generating and forwarding the BCPs in the DAOBS light-tree at time-slot boundaries. The light-tree is usually composed of other nodes between the HoB and ToB. According to the operation of the two MAC access modes, all the nodes can transmit bursts to the rest of downstream nodes making use of the multiplexing capabilities within the light-tree.

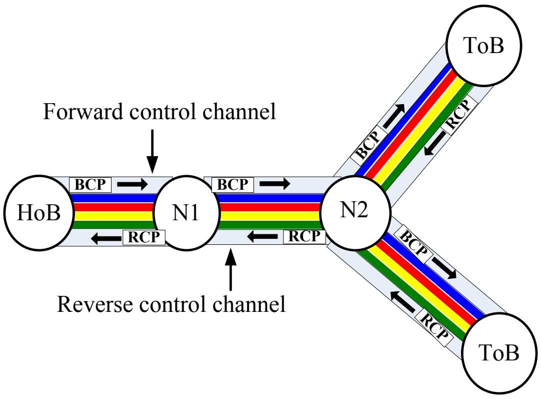


Fig. 4.2: Example of a DAOBS light-tree. If node N1 needs to transmit to any ToB, it requests for free slots to HoB, or eventually waits for a free slot coming on the downstream direction.

Based on the light-tree overlay, every node can belong to manifold light-trees. As a consequence, a node may have to manage multiple DAOBS entities, one for each accessible light-tree. In the proposed architecture, a DAOBS entity is identified by its *input port + output port + wavelength*, and a DAOBS light-tree by its *HoB + wavelength* on the network.

DAOBS provides two channel access mechanisms: queue arbitrated and pre-arbitrated. Each access scheme uses a different slot type. The former is devised for connectionless sub-wavelength burst transport services, that is, the transmission of bursts that have not been explicitly acknowledged the availability of a certain channel capacity. On the contrary, PA is used by services or applications that need a guaranteed reserved bandwidth. For a more detailed application usage example we refer to Table 4.3 in the results section.

Fig. 4.3 illustrates an example where bursts of different service types are transported on three wavelength channels. As seen in the figure, connectionless bursts are transmitted in QA slots, sharing the available channel capacity together with connection-oriented applications making use of PA slots. The number of PA reserved slots depends on the application bandwidth requirements. For instance, on channel $\lambda3$ two consecutive slots (every 6) are used by a single connection which gives a third of the wavelength capacity.

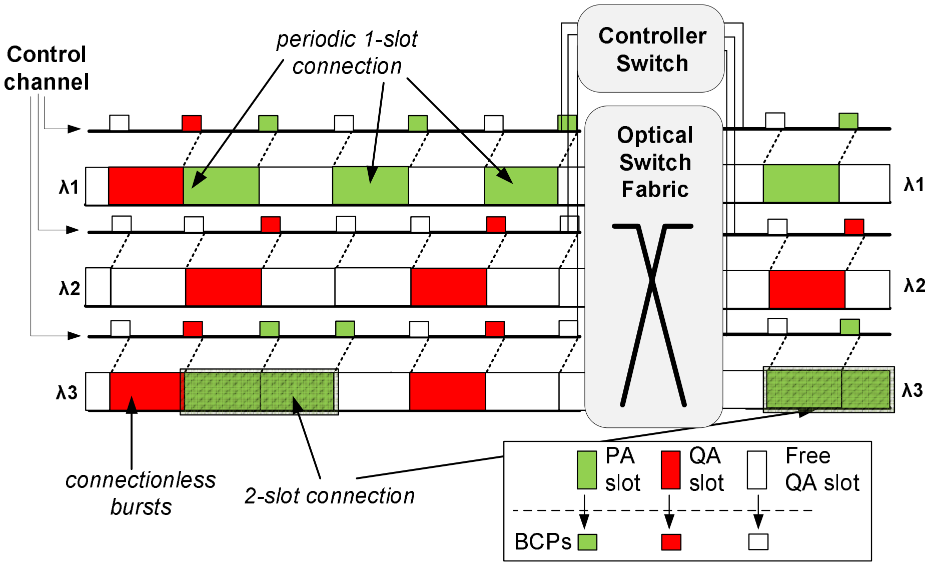


Fig. 4.3: Data channel PA/QA slot example.

In order to support such architecture of services, the burst control packet has a 1-bit field (PA/QA) to announce the upcoming slot type as shown in Fig. 4.4(a). If this field is equal to 1, the network node currently processing the BCP executes the PA access mode. Otherwise, the node runs the QA access mode. All BCPs are required to announce two important parameters in two separate fields, the HoB ID and Lambda. As described previously, each light-tree is identified on the network by its HoB and wavelength. Moreover, the BUSY bit marks whether the upcoming burst slot is occupied or free. If it is occupied, then the rest of fields, like source, destination, burst ID and burst size are meaningful.

BCPs are created by the HoB nodes every time slot, transmitted over the control channel and processed by the core nodes in advance to the reception of the upcoming slot, which can eventually be taken by a data burst. The PA/QA type is assigned according to the flow diagram from Fig. 4.5. At every slot interval, the HoB creates QA slots unless a PA slot has been reserved on the super-cycle slot window by the service layer. In such a case, a new PA slot is created by setting the PA/QA field to 1. Otherwise, a QA slot is forwarded to the next node downstream on the light-tree. In both cases, the HoB checks if it is indeed able to fill the next slot by transmitting one of its enqueued bursts. If so, then the node sets the BCP's BUSY bit to true, and transmits the burst together with the BCP.

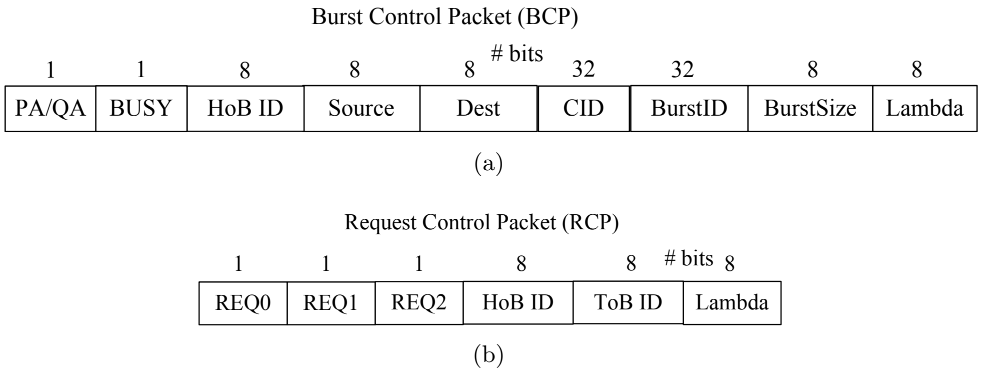


Fig. 4.4: Format of the control packets: (a) burst control packet, and (b) request control packet.

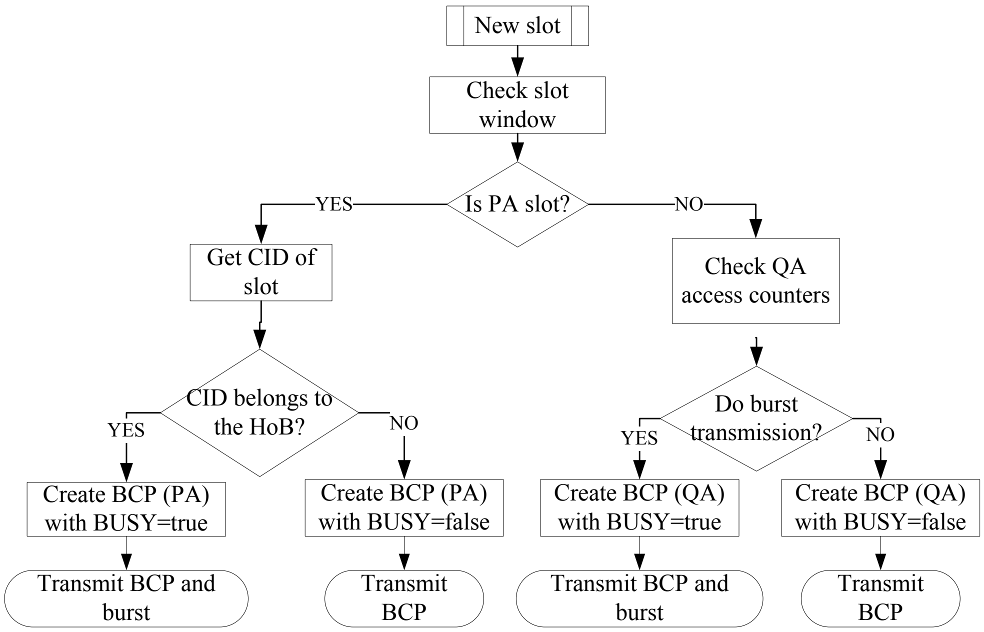


Fig. 4.5: Flow diagram for the HoB slot processing.

4.2.2 Queue Arbitrated Access

The queue arbitrated access mode relies on a bidirectional control channel communication between the HoB node and the remaining nodes in the light-tree. Two are the control packets used to enable such bidirectional communication: the burst control packet (refer to Fig. 4.4(a)) and the request control packet (RCP). On the downstream or forward channel, the BCPs are forwarded from the HoB node to the ToB nodes, and on the upstream or reverse channel, the RCPs travel from the ToBs to the HoB.

One of the advantages of the proposed protocol is its low complexity. In the QoS-enabled DAOBS QA access mode, the hardware complexity is easily bearable. Furthermore, the scheduling complexity is again virtually null as the QA mode just works as a set of counters which can easily be monitored and updated.

4.2.2.1 Queue Arbitrated Operation Mode

Request control packets are used by the downstream nodes on the light-tree to request for free slots, and as the BCPs, RCPs are sent every time slot. This is necessary in order to let all the nodes participating in the light-tree to check whether free slots are coming or not and to permit them to request for free slots periodically. Despite the volume of control packets to process could become cumbersome, we assume that bursts are much longer than the average packet length; hence the electronic processing requirements can easily be met.

The format of the RCP is shown in Fig. 4.4(b). The packet has four main fields: the *HoB ID*, the *ToB ID*, *Lambda* and a number of request bits (*REQ*). As in the BCP, the HoB and Lambda identifiers are unique for each light-tree, so as to identify the DAOBS light-tree instance referred by the RCP. The ToB ID identifies one of the tail of the bus nodes. ToBs are the generators of RCPs. Finally, the number of request bits depends on the number of connectionless traffic classes in use. For instance, in the aforementioned figure, the packet has three different REQ bits, one for each class under consideration.

As introduced before, BCPs have a BUSY bit for announcing whether or not the upcoming slot is free. Recall that a QA access slot is announced by the PA/QA field set to 0. In the case that the BUSY bit is equal to 1, then the rest of fields in the BCP (*BurstID*, *Source*, *Dest.*, etc.) are meaningful since the upcoming slot is occupied.

All the nodes in the light-tree can transmit bursts to downstream nodes according to the operation of the QA access, that is, a node requests for free slots to the upstream and HoB nodes using the REQ bits of the RCP to subsequently transmit bursts by taking the upcoming free slots on the downstream direction. A set of counters for each priority keep track of the requests and free slots coming from downstream and upstream nodes, respectively. Connectionless burst losses happen only at the edge of the network because of queue blocking. Offering more load than the acceptable causes bursts to be dropped due to having the ingress queue fully loaded.

4.2.2.2 Queue Arbitrated Module Architecture

In a network node, the QA access module of the DAOBS MAC entity is composed of the following components for each queue-arbitrated priority class i:

- a distributed access state machine (DASM),
- a request control machine (RQM),
- a local queue (LQ), and,
- a distributed queue (DQ).

Fig. 4.6 displays how all these elements are interconnected in a DAOBS entity on wavelength λ_k with three connectionless burst priority classes. The DASM and

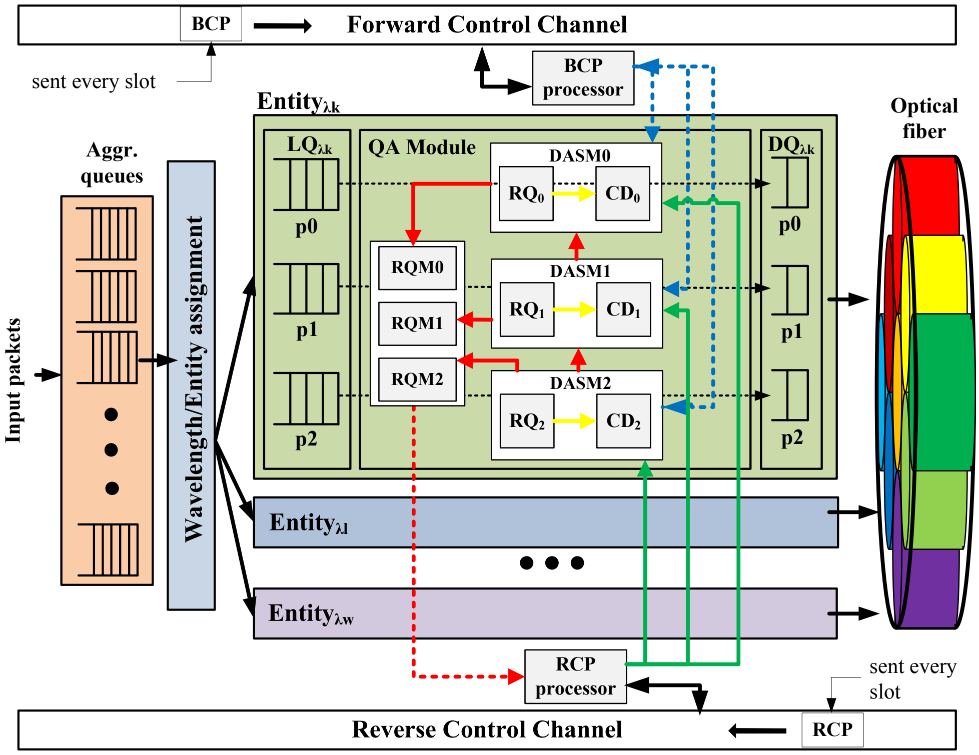


Fig. 4.6: DAOBS QA modular architecture. Entity with three burst priorities on wavelength λ_k.

RQM state machines are also interconnected to the BCP and RCP packet processor, respectively.

The LQ temporarily stores bursts coming from the wavelength/entity assignment module while waiting to gain access to the optical channel. The DQ is a one-position queue that stores the next burst to transmit. The set of DQs from the nodes that belong to the DAOBS light-tree form the so-called virtual distributed queue. Thus, when a burst at a certain node gets into one of its DQs is similar to access a FIFO queue distributed among the nodes from the light-tree. Each node can only have a single burst at a time in the virtual distributed queue, which is realized by limiting the size of each node's DQ to enqueue only one burst at a time. The queue arbitrated naming origins from this operational mode.

There is a DASM for each priority, and in the example given in Fig. 4.6 these are: DASM0, DASM1 and DASM2, wherein the greater the number, the higher its priority. DASMs responsible for monitoring and managing the busy/free and request counting processes of the protocol, and as such, it requires to access the RCP and BCP processing units in order to read the contents of the respective control packets. Also, as it is shown in the figure, a high priority DASM can signal and preempt to lower priority ones in order to ensure that higher class connectionless bursts receive a better service level. The counter preemption among classes will be described in detail later.

Furthermore, for each priority there is a request control machine. RQMs process the requests triggered by the QA entity in the network node. Later, these same RQMs communicate with the RCP processor to set the REQ bit of its corresponding priority whenever the RQM determines so and the upcoming RCP packet has a REQ bit equal to 0, that is, none of the downstream nodes have requested slots on that priority.

4.2.2.3 Queue Arbitrated State Machine

Each DASM can be in two different states as shown in Fig. 4.7. In the *idle* state, the node, for a certain DAOBS light-tree and priority, has nothing to transmit, whereas in the *active* state, the node has successfully placed a request for a free slot and it is waiting to transmit a burst. Additionally, each DASM has two counters: a request counter (RQ) and a count-down counter (CD). On the one hand, the RQ monitors the number of requests made by downstream nodes on the optical light-tree and by higher priority DASMs in the same DAOBS entity. On the other hand, the CD counts the number of free slots the current node is not allowed to use before being given access to transmit the burst. All entity counters are reset to zero upon initializing the QA DASM.

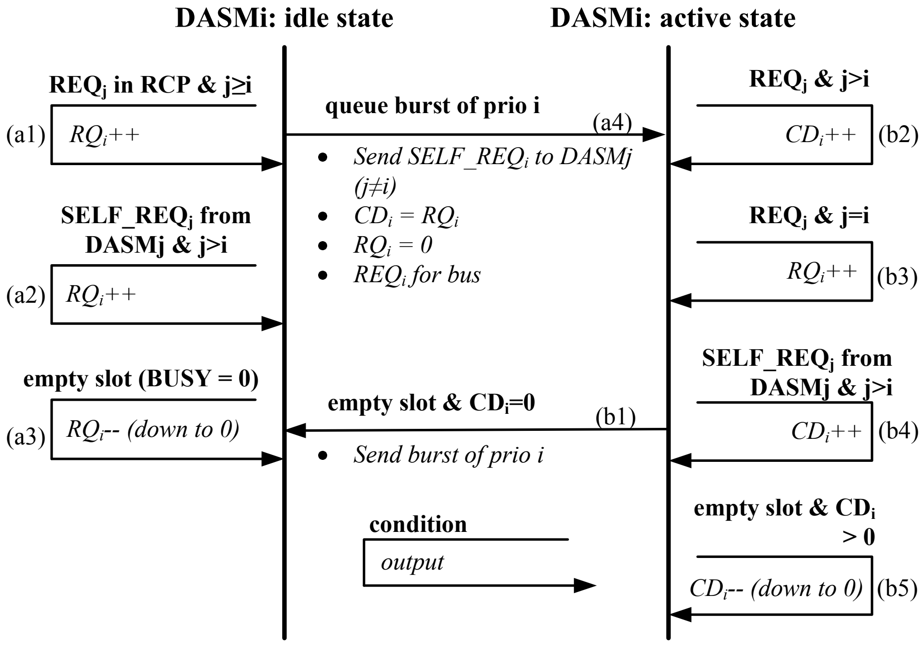


Fig. 4.7: QA DASM flow at priority i.

While a DASM at priority i is *idle*, it monitors the RCPs on the reverse control channel and BCPs on the forward control channel and increases (see Fig. 4.7 step (a1)) or decreases (a3) the RQ_i for every $REQ_j = 1$ in the RCP of a priority $j \geq i$, and for every $BUSY = 0$ in the BCP, respectively. Similarly, if the DASMi receives a $SELF_REQ_j$ signal from a higher priority DASMj ($j > i$) within the same entity (a2), then the RQ_i is also increased.

As soon as a burst is enqueued into the LQ of a certain DAOBS entity and priority i, if the DASMi is *idle* it switches to the *active* state following the transition (a4) shown in Fig. 4.7. The same happens if after returning from a successful burst transmission there are more bursts to transmit in the LQ, i.e., as soon as transmitting the burst and returning to the *idle* state. This state transition (a4) triggers the following events: first, a $SELF\_REQ_i$ signal is sent to the remaining DASMj ($j \neq i$) in the entity; then the value of the RQ_i is dumped to the CD_i ($RQ_i \leftarrow CD_i$), after which the RQ_i is reset to 0; and finally a REQ_i signal is sent to the RQM of that same priority in order to set the REQ bit of that priority to 1 in the next upcoming free RCP received from the downstream node.

In the *active* state, the DASMi continues monitoring the BCPs and RCPs. For any REQ bit of priority $j \geq i$ (b2) or $SELF\_REQ_j$ signal from a DASMj with $j \geq i$ (b4), CD_i is increased by 1. These two steps let higher priority bursts on the DAOBS light-tree to be placed ahead and thus transmitted before, even between bursts from two different nodes of the same light-tree. In (b3), for every REQ bit of priority $j = i$, the RQ_i is increased by 1. Likewise, for every empty slot on the forward control channel (b5), and provided that the $CD_i > 0$, the CD_i is decreased by 1 (down to 0). Finally, the transmission of the burst from the DQ at priority i happens when CD_i reaches 0 and an empty slot comes from the upstream link. This last step involves the transition (b1) of DASMi from *active* to *idle*.

4.2.2.4 Queue Arbitrated Wavelength Assignment

The wavelength assignment module is responsible for allocating bursts transmissions to a specific DAOBS light-tree. As stated before, each DAOBS tree is identified by its HoB and wavelength. Therefore, the wavelength selection is in fact a DAOBS entity assignment (see Fig. 4.6). In the QA access, the wavelength assignment is controlled by an algorithm that takes into account the values of the counters RQ and CD and the number of bursts ahead in the LQ in the assignment. The value of these counters determines the position of the node and its burst transmission in the distributed FIFO queue at a certain priority class. The number of bursts ahead in the LQ gives information about the number of transmissions not processed yet in the current node. Depending on these values, the node will take longer to transmit a burst on that specific light-tree, thus increasing or decreasing the channel access delay, and consequently, the end-to-end delay of the burst. In order to overcome such scenario, the DAOBS light-tree that is expected to provide the lowest access delay is always selected. That is, for priority i, we select the wavelength from the set $\mathcal{W}_n^m$ that guarantees,

$$\min\nolimits_{\lambda_j \in \mathcal{W}_n^m} (RQ_{i,\lambda_j} + CD_{i,\lambda_j} + size(LQ_{i,\lambda_j})) \tag{4.1}$$

where $\mathcal{W}_n^m$ is the wavelength/DAOBS candidates list for transmitting a burst from node n to m. The list needs to be pre-computed using the routes and wavelengths available from the light-tree overlays, which are recorded in each node when these are established.

4.2.3 Pre-arbitrated Access

In the PA access, HoB nodes are responsible for scheduling connection requests for connection-oriented TDM channels. Whenever an application requests this type of transmission mode from a downstream (core) node –or also from the HoB itself–, the node explicitly requests to the light-tree HoB the allocation of sufficient time-slots based on the amount of bandwidth requested and the specific type of service. In turn, the HoB will acknowledge the core node about the success of the connection. Although this feature is supported by a higher service layer that monitors applications' service requests, is the MAC layer which implements the scheduling, allocation and differentiation of slots.

The connection request and acknowledgement use a specific set of service layer messages distinct to the BCP and RCP introduced for the queue-arbitrated access mode. These messages are transmitted and forwarded through the control channel together with the BCPs and RCPs.

4.2.3.1 Pre-arbitrated Operation Mode

Fig. 4.8 shows an example of an explicit multi-layer setup and release of a connection-oriented TDM channel and the slot allocation at the MAC level. The process is composed of the following steps:

(1) Let a user/application request a channel between node N1 and a downstream node on the light-tree, e.g., the ToB.

(2) The service layer computes the connection requirements of the application call. For instance, let the service layer at node N1 map the connection to use 1 out of 2 slots, as shown in Fig. 4.8. At this step, the service layer also sorts the list of possible HoBs that can handle this application.

(3) Using this algorithm, the setup message is transmitted to the chosen HoB which is in charge of scheduling all the PA slots for this DAOBS light-tree.

(4) Upon receiving the setup message, the HoB allocates, if possible, the slots for the connection from N1. After processing the slot scheduling, the HoB acknowledges the node whether the connection has been scheduled or not by means of an acknowledgment message.

(5) The service layer at N1 is notified of the availability and successful setup of the connection, in which case, the application can start transmitting its data

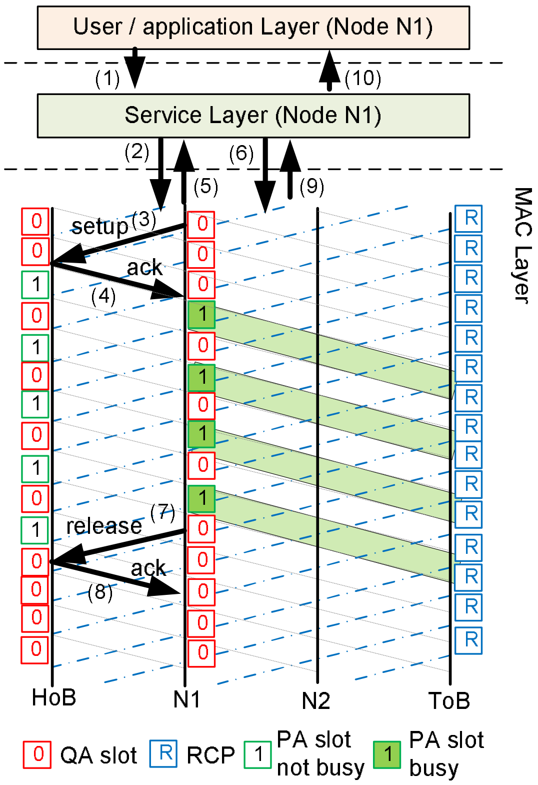


Fig. 4.8: Example of a multi-layer PA service setup and allocation of slots in the MAC layer.

packets. The transmission of bursts will be made according to the PA reserved slots allocated to node N1.

(6) When the service finishes, the service layer initiates the disconnection.

(7) A release message is sent to the HoB node, which frees the allocated slots for the connection.

(8) An acknowledgement message is then transmitted back to node N1.

(9) And the service layer is acknowledged about the successful connection release.

(10) Finally, an explicit acknowledgment signal can also be sent to the application layer.

One of the advantages of the DAOBS PA slot allocation is that even using an explicit release, eventual upcoming PA slots which would remain unused can be changed by the designated core node and announced as a QA slot to the rest of downstream nodes. For instance, in Fig. 4.8, the node N1 initiates the release of the connection sending a message to the HoB. Meanwhile, a PA slot has already been created and scheduled at the HoB following the information stored in the connection table, which is not updated yet due to the message propagation delay. In such a case, if the PA slot belongs to a connection from N1, and the connection is in the release state, the node can change the type of slot to be used by other downstream nodes, or even reuse it to transmit one of its bursts through the QA access.

In order to reduce the connection establishment delay, a cross-layer algorithm in the service module sorts the list of HoBs available to serve the TDM connection so that light-trees for which the own service request node is actually a HoB are given preference. As a result, the setup can be processed between the service and MAC layers in the node itself avoiding a two-way connection establishment, hence reducing the connection setup delay.

4.2.3.2 PA/QA Slot Scheduling

It is worth noting that many different scheduling algorithms can be implemented in order to satisfy a variety of service requirements, such as, jitter, delay, bandwidth, call blocking probability, etc. Together with the PA and QA channel access modules, a slot scheduler is integrated within the MAC layer of the HoB nodes. The scheduler allocates the time-slots based on the bandwidth and connection request types. As introduced before, both downstream nodes and the HoB itself can request for TDM channels on the specific light-tree governed by this same HoB.

The allocation of slots is made along a super-cycle slot window of a size that depends on the minimum and maximum possible bandwidth request. For instance, with a 10 Gbps channel and a minimum connection request of 155 Mbps, equivalent to an OC-3 of SONET, the size of the super-cycle window will be,

$$W = \left\lfloor \frac{10 \cdot 10^9}{155 \cdot 10^6} \right\rfloor = 64\ slots \tag{4.2}$$

Regarding the types of scheduling that can be realized, Fig. 4.9 shows three different policies. For instance, let assume we need to process a 4-slot TDM connection request; we can:

1. schedule the request over first-fit continuous slots (FF),
2. allocate a “pure” 4-slot TDM over a periodic number of slots (SPFF), or
3. schedule the 4-slot bandwidth guaranteed service over a number of slots neither periodic nor continuous, but random (NCR).

FF-based algorithms need to be considered in order to support not only applications with guaranteed bandwidth requirements, but also with minimum delay variation between super-cycle periods. FF and NCR algorithms are fairly straightforward, hence we do not describe any particular algorithm implementation.

Periodic scheduling is necessary to provide jitter-controlled characteristics for multimedia services with periodicity within the own super-cycle. The periodic scheme is implemented by the sliding periodic first-fit (SPFF) algorithm. The sliding feature increases the probability for other future requests to be allocated according to the periodic constraint by creating gaps along the super-cycle window.

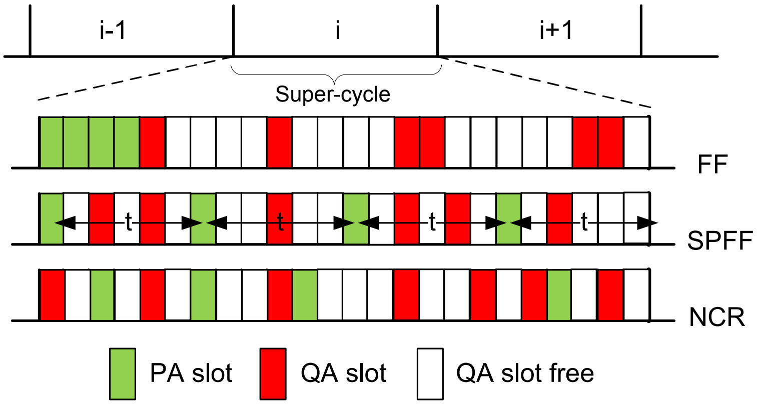


Fig. 4.9: Pre-arbitrated slot scheduling algorithms.

In SPFF, the number of requested slots N must obey that,

$$size(W) \equiv 0\,(\mod N)\,, \tag{4.3}$$

where $size(W)$ stands for the size of the slot window W (in number of slots), also referred as *super-cycle*. Equation (4.4) defines the mathematical constraints the set of slots must guarantee when using the SPFF algorithm. The scheduling is successful if a group of slots S_i can be found such that each slot s_i is free ($v(s_i) = 0$) in the window fulfilling the periodicity constraint, $t = W/N$,

$$S_i = \{s_i \in W \mid v(s_i)=0 \;\forall\; s_{i+1}=s_i+t,\; t=\frac{W}{N},\; 1\leq i\leq N\} \tag{4.4}$$

Algorithm 3 shows in detail the SPFF implemented in this Thesis. In the algorithm pseudocode, the input S_w stands for the predefined amount of sliding slots for every execution of the algorithm, S_0 is the last reserved slot from the previous algorithm invocation, and s_p, s_s and s_j are auxiliary pointers used by the algorithm to check the free slots that fulfill the constraints (4.4).

The algorithm starts (see lines 4-5) by checking whether the amount of requested slots obeys the arithmetic module constraint (4.3). If the constraint is met (line 5), then in line 6, the last reserved slot used in the previous algorithm invocation is loaded. From this slot index, a number of slots (S_w) is added, so that the initial slot to be checked moves ahead (slot sliding), i.e., it shifts S_w positions. From that point, the availability of free slots is checked for the number of slots N passed as a parameter to the algorithm until the group of slots (lines 11-19), S_i, is found. In line 9, the number of iterations (t) in the *while* loop depends on N. As a result, the smaller the number of requested slots N, the greater the possibilities to find S_i. If the N requested slots of capacity are allocated (line 20), then the algorithm updates the last allocated slot (line 22). This value will be taken in the next algorithm execution to update the slot sliding. The

Algorithm 3 SPFF scheduling algorithm.

```
1:  input: $N$, $W$, $S_w$, $S_0$
2:  output: $S_i \leftarrow \varnothing$
3:  variables: $t$, $Res$, $Rem$, $Count \leftarrow 0$, $i \leftarrow 0$, $j$, $s_j$, $s_p$, $s_s$
4:  $Rem \leftarrow W \mod N$
5:  if $Rem$ is 0 then
6:      $s_s \leftarrow (S_0 + S_w) \mod W$ {Load last reserved slot}
7:      $Res \leftarrow$ false
8:      $t \leftarrow W/N$
9:      while ($Res$ is false) and ($i < t$) do
10:         $s_p \leftarrow s_s + i$
11:         for $j = 0$ to $N - 1$ do
12:             $s_j \leftarrow (s_p + j \cdot t) \mod W$ {Compute next periodic slot}
13:             if $value(s_j)$ is 0 then
14:                 $Count \leftarrow Count + 1$
15:                 $S_i \leftarrow S_i \cup s_j$
16:             else
17:                 $S_i \leftarrow \varnothing$
18:             end if
19:         end for
20:         if $Count$ is $N$ then
21:             $Res \leftarrow$ true {Return true}
22:             $S_0 \leftarrow s_j$ {Update last reserved slot}
23:         end if
24:         $i \leftarrow i + 1$
25:     end while
26: end if
27: return $S_i$
```

algorithm finishes by returning the set of assigned slots S_i (line 27). Finally, if the algorithm cannot find an available set of slots, it will return the empty set ($S_i = \varnothing$).

One of the benefits of this algorithm is that it does not only accomplish the periodicity for the slot scheduling within the slot window, but also enhances the success rate for n-slot connections. Connections demanding very few number of slots, m (where $n > m$), are allocated along the slot window reducing the collision with the required space demanded by other n-slot connections.

4.3 Simulation Results

In this section we assess the performance of the multi-service QoS-enabled MAC protocol proposed in this chapter. To this end, simulations are conducted on the well-known NSFNET network composed of 14 nodes and 21 bidirectional links. In such a scenario, there are 16 wavelengths and 10 Gbps per channel. We assume in all the examples that the network neither has wavelength converters nor FDLs for contention resolution; hence burst transmissions are subject to the wavelength continuity constraint. Regarding the setup of hardware devices, the control packet processing time and the non-blocking matrix switching time are set to 10 μs and 5 μs respectively. These two values define the offset-emulated delay between the control packet and the data burst.

With respect to the traffic characteristics, bursts are created at each node by using a volume-based algorithm [114] with an input Poisson packet arrival process and fixed size per burst of 100,000 bytes. For simplicity, the burst destination is uniformly distributed to all the remaining nodes, so that the probability of a burst to be sent to any other node in the network is the same. The heuristic used to set up the DAOBS light-tree overlays tries to equally provide to all nodes the same capacity to reach the remaining nodes.

Results were gathered using the batch means method. 95% confidence intervals were also obtained, but since they are quite narrow, they have been omitted in order to improve the readability of the graphs.

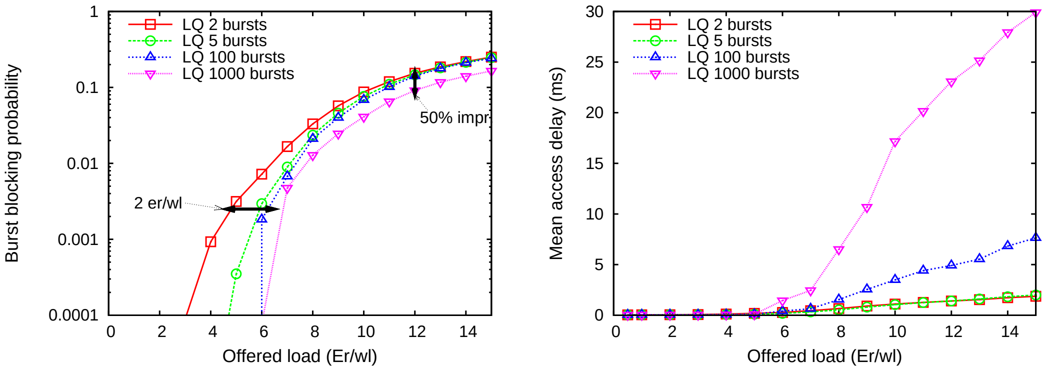
Results are divided in four main subsections. Firstly, we evaluate the performance of the QA access mode as a function of the local queue size and without QoS. Secondly, we analyze the QA access mode with QoS. Next subsection evaluates the performance of the whole MAC architecture with both, connection-oriented and connectionless sub-wavelength burst services. And finally, we present a critical analysis about the proposed service classification.

4.3.1 Queue Arbitrated Access Mode Without QoS

Initial results deal with the protocol performance of the QA access mode when different local queue (LQ) sizes are used and only one connectionless traffic class is transmitted on the network. Fig. 4.10(a) shows the burst blocking probability (BP) as a function of the total offered load to the network in Erlang per wavelength (Er/wl). The LQ sizes (in number of bursts) used across the simulations are: 2, 5, 100 and 1000 bursts. Intuitively, the smaller the size of the LQ, the sooner the blocking probability starts rising. For sizes between 2 and 100 bursts, the results at high loads asymptotically converge to the same value. Only for the case in which the LQ size is equivalent to 1000 bursts we can see an improvement ($\sim 50\%$) of the blocking probability but at the expense of increasing two orders its size.

Furthermore, Fig. 4.10(b) represents the mean access delay (in ms) for the same group of LQ lengths. At the expense of decreasing the mean blocking probability, the delay experienced when the LQ size is of 1000 bursts rises up to nearly 30 ms at very high loads. Hence, a trade-off between BP and access delay comes up. At low-to-medium offered loads, the BP improvement for larger LQ is considerable with more than 2 Er/wl of accepted offered load at the same blocking rate. However, at high loads, an average access delay 15 times greater may not justify a BP improvement of about 50%.

As we can see, the access delay goes up as the offered load to the network increases. Although bursts transmissions in QA are connectionless, nodes are not able to transmit



(a) (b)

Fig. 4.10: QA access mode without QoS for different LQ lengths: (a) burst blocking probability, and (b) mean access delay.

as soon as the burst is ready due to the operation of the request-grant and counting process. Later in the chapter, in Section 4.4, we will model this access delay, so that we can quantify the expected delay and act upon it.

4.3.2 Queue Arbitrated Access Mode With QoS

After evaluating the performance of a single traffic class using the connectionless queue-arbitrated access, in this section we give an insight into the results of the DAOBS QA access mode when a number of different burst traffic classes are transmitted on the network. In this experiment the LQ size is configured to store up to 5 bursts. The following notation is used in the graphs: burst priority classes are numbered from 0 to 2, being *class 2* the highest priority traffic. To demonstrate the performance of the protocol on different traffic scenarios, two traffic class distributions were tested, as shown in Table 4.1. In the table, we represent the traffic percentages assigned to each class.

Table 4.1: Traffic distribution configurations.

Distribution	**Class 0**	**Class 1**	**Class 2**
Dist. 1	50%	30%	20%
Dist. 2	70%	20%	10%

Fig. 4.11(a) shows the BP as a function of the offered load under the two traffic distributions. *Class 2* blocking probability is not plotted on the graph as in both cases it is null for the whole load range. The rest of classes get different results depending on the traffic configuration. Besides, we can see that in the second distribution, when the higher priority traffic load is decreased with respect to the total, the burst blocking probability decreases for both *class 0* and *class 1*. Intuitively, the lower the *class 2* (i.e., the highest priority) traffic load is, the more resources available for the rest of classes.

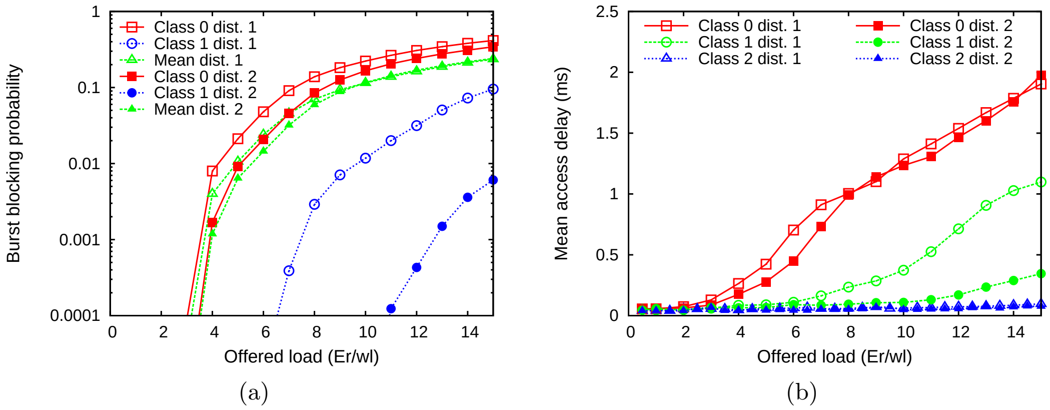


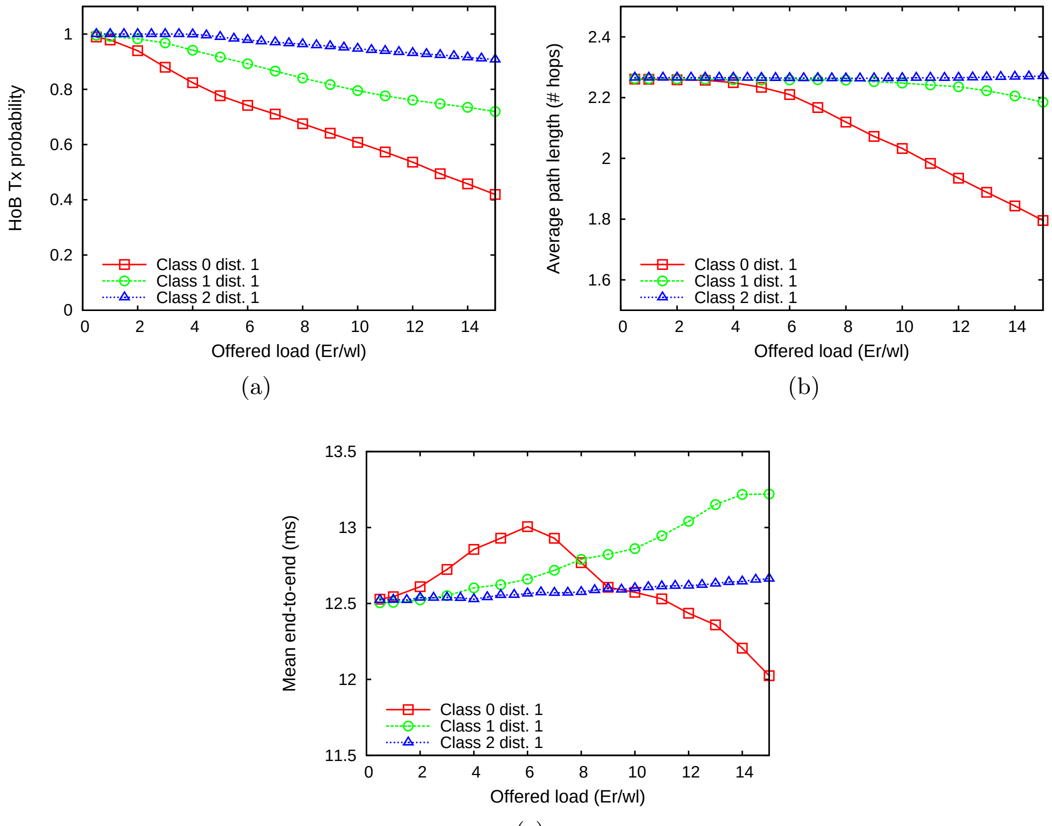
Fig. 4.11: QA access mode with QoS and LQ=5 bursts for the two distributions in Table 4.1: (a) burst blocking probability, and (b) access delay.

Consequently, less high priority traffic preempts over lower priority traffic; hence the low priority traffic loss probability decreases. Nevertheless, at high loads the mean BP plots of both traffic distributions tend to converge to similar values, thus ensuring a predictable average blocking performance of the protocol whatever traffic distribution is present on the network at high loads.

Fig. 4.11(b) compares the mean access delay against the offered load. In both traffic configurations, *class 2* bursts not only have the lowest access delay, but also get a delay nearly constant for the whole load range. *Class 0* bursts have a similar delay trend for both traffic configurations and finally, *class 1* bursts (i.e. the intermediate class) get a different access delay depending on the traffic distribution. When *class 2* proportion is 10% and *class 1* represents the 20% of the total traffic on the network, the access delay for *class 1* resembles more the delay of *class 2*. So, it can be concluded that the delay experienced by a burst traffic class depends on the aggregate traffic between itself and all its higher priority traffics. This is due to the operation of the multi-priority QA access mode which allows higher priority bursts to increment the counters of other lower priority state machines. The priority QA system ultimately puts ahead the former bursts (high-priority) in the queue and delays the transmission of the latter.

Fig. 4.12(a) shows a different performance parameter. In the graph, the probability that a certain class of traffic is being transmitted from a Head of the Bus node is counted. In this case we only show the results obtained from the first traffic distribution. Clearly, we can observe that *class 2* bursts are mainly transmitted from HoBs, whereas the transmission rate for *class 0* and *class 1* decreases almost linearly as the load is increased. High priority burst traffic is more often transmitted from a HoB because this node tends to see greater available channel capacity on the light-tree while the requests from downstream nodes are not yet received due to the propagation delay.

Fig. 4.12(b) shows the average route length (in number of hops) for the three burst



(a)

(b)

(c)

Fig. 4.12: QA access mode with QoS and LQ=5 bursts for the first distribution in Table 4.1: (a) HoB transmission probability, (b) average path length, and (c) end-to-end delay.

priorities. As it can be seen in the graph, highest priority bursts almost have a constant route length for the entire load range. Nonetheless, lowest priority bursts experience a drop of the path length which is directly related to the fact pointed in Fig. 4.12(a). By moving *class 0* burst transmissions from the HoB to inside the DAOBS light-tree, the origin node gets closer to the destination, which at the most can be one of the ToBs. As a result, the average route length is shortened.

With respect to the burst end-to-end delay (in ms), Fig. 4.12(c) shows a comparison between the three classes. In this specific scenario, an average end-to-end delay between 1.5 ms is guaranteed between the three traffic classes still providing a clear differentiation of burst loss probability. As concluded from Fig. 4.11(a), *class 0* bursts see increased their end-to-end delay as the offered load rises, according to the access delay increase also seen in Fig. 4.11(b). However, we can see a change of the trend from an offered load of 6 Er/wl onwards where the delay starts falling. Following the reasoning from previous figures, this fact can be explained as follows: at high loads, *class 0* bursts get a higher blocking probability and those that are able to get to destination tend to follow a shorter path, i.e. fewer number of hops, hence decreasing the mean delay even though the access delay goes up as shown in Fig. 4.11(b). The rest of traffic classes (*class 1* and *class 2*) experience an end-to-end delay only affected by the access delay increase since both of them show almost a constant average route length within the load range under consideration (see Fig. 4.12(b)).

In order to evaluate, not only the performance in terms of the offered traffic load, but also as a function of the normalized carried traffic load per link, Fig. 4.13 shows the blocking probability as a function of the normalized link load ratio as

$$\rho = \frac{\sum_{i=1}^{n} M_i}{\sum_{i=1}^{n} C_i}\ , \tag{4.5}$$

where M_i is the traffic carried by link i, C_i is the capacity of link i, and n is the number of links (bidirectional) on the network. The network load is represented between 0 and 0.5 and computed using the link utilization results from the same simulation runs.

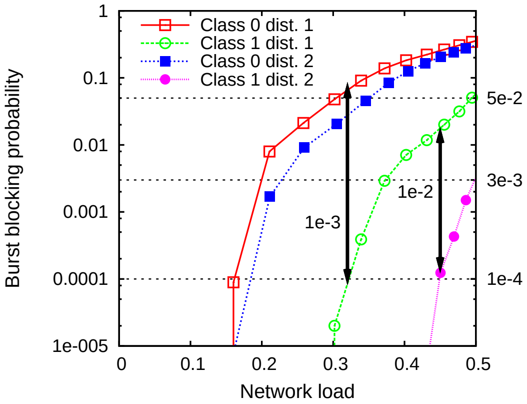


Fig. 4.13: QA access mode burst blocking probability as a function of the network load.

As it is shown in Fig. 4.11(a), *class 2* is lossless for the network load under test, hence it is not represented in the graph. For the rest of classes, the protocol provides differentiation of up to three orders of magnitude between *class 0* and *class 1* at an average load of 0.3, and an improvement of two orders for *class 1* traffic between the two distributions at load 0.45.

Based on the QoS requirements shown in Table 4.2 and providing that the packet loss rate (PLR) can be approximated by the burst blocking probability for fixed size bursts [120], the protocol can guarantee the QoS of a diverse number of applications. High or very high loss sensitive traffic (e.g. Grid applications or live video broadcasting) can be mapped as *class 1* for a wide network load range (up to 45% for the second traffic distribution), or even be mapped as *class 2* traffic, which is lossless.

Table 4.2: Applications' QoS requirements [5].

Application	**Delay**	**Jitter**	**Loss sensibility (PLR)**
Interactive audio/video	<150 ms	<75 ms	High (<1e-3)
Inter. transaction data	<50 ms	<10 ms	High (game <5e-2) to very high (grid <1e-4)
Video/audio streaming	<2 s	<40 ms	High (<3e-3) to very high (live video <1e-4)
Legacy applications	Not spec.	Not spec.	Low

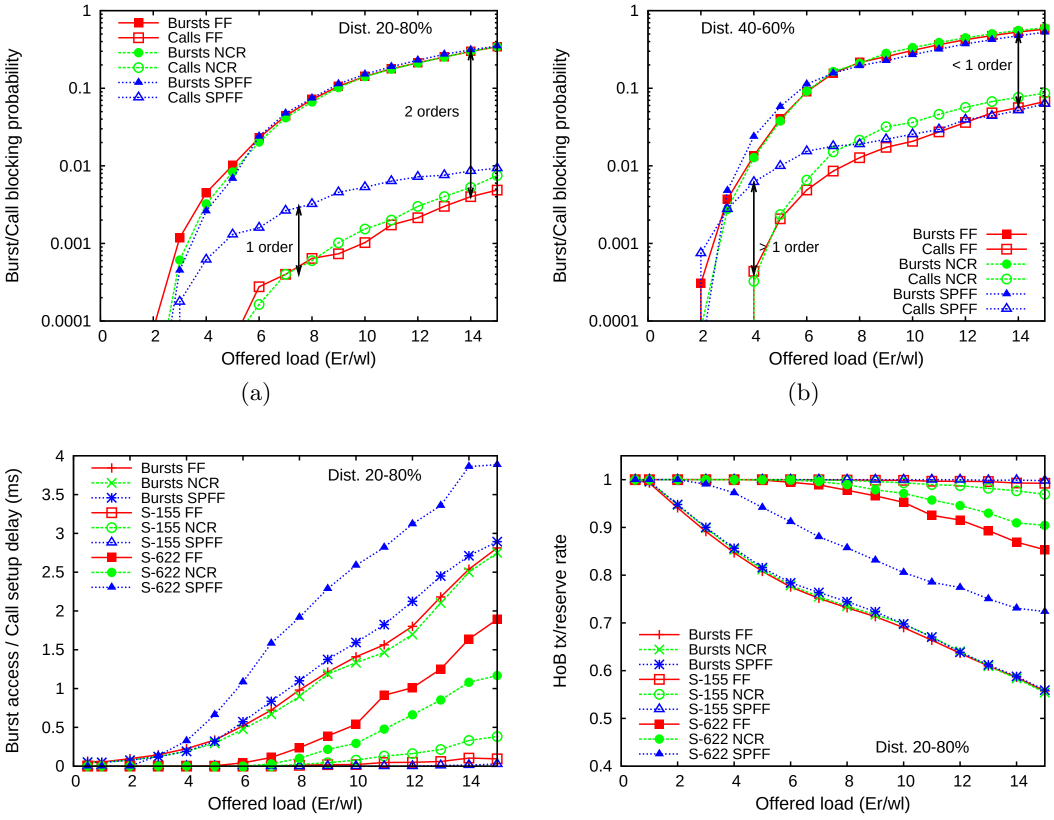
4.3.3 Results With Mixed Connectionless and Connection-Oriented Burst Services

We evaluate in this section the performance of the proposed OBS MAC protocol using the three aforementioned slot scheduling algorithms in a multi-service scenario where connectionless bursts share the capacity of the channel with connection-oriented guaranteed bandwidth services (i.e., TDM). Regarding the traffic characteristics, bursts are of fixed size for both types of service (connection-oriented and connectionless) and equal to the slot size (100,000 bytes). For the TDM services, we consider two different connection types of 155 Mbps (S-155) and 622 Mbps (S-622), equivalent to a bandwidth capacity of OC-3 and OC-12, respectively. The former is mapped to use 1 slot every 64, and the second to use 4 out of 64 slots. In both cases, the connection arrivals follow a Poisson process with an exponentially distributed holding time of 100 ms and 200 ms, respectively (intentionally short to speed up simulations). The average load generated by the connection-oriented calls is calculated as follows,

$$A_{CO} = \frac{1}{C}\sum_{i=1}^{n}\lambda_i\frac{1}{\mu_i}b_i \tag{4.6}$$

where C denotes the channel capacity in bps, λ_i the call arrival rate of connection service i, $1/\mu_i$ is the mean holding time and, finally, b_i is the demanded bandwidth of service i in bps. Two different traffic distributions were also considered; firstly, a distribution in which TDM services represent 20% of the offered load and connectionless bursts the remaining 80% (Dist. 20-80%), and secondly, a distribution with 40% and 60% (Dist. 40-60%), respectively.

Fig. 4.14(a) and Fig. 4.14(b) show the mean burst and call blocking probability of both service types (bursts and TDM) as a function of the offered load to the network in Erlang per wavelength and distributed by scheduling algorithm. In Fig. 4.14(a) we can see that the burst BP is very similar among the three implemented scheduling algorithms, specially at high loads, where the three converge to same values. Only at low loads, burst blocking in SPFF and NCR gets a slight improvement in comparison with the first-fit algorithm. Therefore, we can conclude that under this particular testing scenario, the PA scheduling algorithms have little effect on the blocking probability of



(a) (b)

(c) (d)

Fig. 4.14: Dual PA/QA access modes: (a) Mean burst/call blocking probability (BP) in traffic distribution 20-80%, (b) BP with traffic distribution 40-60%, (c) Mean access delay for connectionless bursts and mean connection setup delay in traffic distribution 20-80%, and (d) HoB connectionless burst transmission and HoB connection setup rates in traffic distribution 20-80%.

connectionless bursts.

Regarding the call blocking probability of the TDM services, this remains always below the burst BP, as the MAC protocol gives precedence to the scheduling of PA slots. That is, once a connection-oriented service is assigned its requested slots, these remain assigned and cannot be preempted by connectionless bursts. Moreover, the mean connection BP of SPFF is the greatest among the three algorithms. Because of the slot periodicity constraint imposed by (4.4), the number of schedule options is reduced in comparison with the other two PA scheduling algorithms. Although SPFF does not outperform, this type of algorithm provides extra capabilities for services that require restricted jitter and delay variations, so that it needs to be considered. Also, we can see that the connection blocking in FF is slightly lower than in NCR. We believe that in this case, the non-continuous policy improves the allocation of higher bandwidth requests, but at the same time, increases the blocking for the rest of connections. On the contrary, the FF policy is more stringent with high capacity requests due to the slot continuity, thus freeing more slots on the super-cycle windows for lower bandwidth connection requests.

If we compare Fig. 4.14(a) and Fig. 4.14(b) side by side, we can see that the greater the connection-oriented offered traffic, the higher the burst BP for the connectionless burst traffic and the higher the connection BP for the TDM services. For instance, increasing the TDM load from 20% to 40% reduces the blocking probability difference between connectionless and connection-oriented calls to less than 1 order at high loads mainly due to the considerable increase of the connection blocking probability (by 1 order of magnitude). As a result, doubling the TDM traffic load increases both burst BP and call BP, hence a decrease of the carried traffic load is also expected.

In Fig. 4.14(c) the mean access delay for connectionless bursts and the mean connection setup delay for the connection-oriented services in the first traffic distribution scenario (20-80%) is shown. In general, connectionless bursts see increased their access delay due to the decrease of channel capacity left by the allocation of TDM services. Although the results are very similar using the three scheduling algorithms, a slight shorter delay in the NCR can be observed. The connectionless burst access delay of the SPFF is slightly greater than the other two algorithms. In this case a greater number of connections are required by the core nodes in the light-tree, hence less network capacity is left to other upstream nodes between the HoB and the origin core nodes.

With respect to the connection-oriented services, as expected, S-622 connections have a higher average connection setup delay as a greater number of setups are initiated by core nodes on the light-tree due to the slot requirements of n-slot connections. Moreover, we can see that the more restrictive the scheduling constraints are, the longer the setup delay is. For instance, because the SPFF algorithm constraints are the most restrictive, the setup delay for the n-slot connections is the largest. Unlike SPFF, the NCR algorithm allows the required bandwidth to be allocated randomly using the void slots on the slot window, hence it is less restrictive and as a result, easier for the n-slot connections to be allocated.

Fig. 4.14(d) shows the HoB burst transmission rate for the connectionless services and the connection setup rate when the origin node of the connection-oriented services is HoB of the light-tree. This is an interesting performance parameter because it explains many of the results from previous paragraphs. In general, the higher the offered load to the network, the lower the HoB connectionless burst transmission rate. Besides, we can also see that the reservation rate of S-622 connections using the SPFF algorithm is the lowest in comparison with the other two algorithms, which corroborates the performance of the mean connection setup delay seen in Fig. 4.14(c).

Finally, Fig. 4.15 shows in more detail the call BP distributed by algorithm and connection type, but now as a function of the average normalized link load as computed in (4.5). S-622 connections using the SPFF get the worst BP performance. Despite enhancing the scheduling algorithm with the *slot sliding* capability, the fact that the

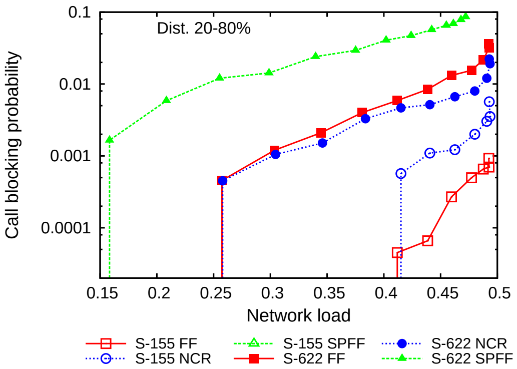


Fig. 4.15: BP for different connection types in traffic distribution 20-80%.

scheduling of slots must be done in accordance with (4.4) considerably reduces the slot allocation search space for n-slot connections. On the contrary, the S-155 connections that only require 1-slot can easily be allocated, and in fact, the blocking probability of S-155 turns to be null for the whole load range. Moreover, the allocation of 1-slot connections decrements even more the search space for the n-slot connections, hence diminishing the setup successfulness ratio of these second ones. Regarding the other two scheduling algorithms, the BP of S-622 connections using FF is higher than NCR as the slot allocation is continuous in the FF service type, whereas NCR takes advantage of randomly scheduling the connections among the void slots, so that a greater number of options are available given the same slot window size. However, S-155 connections get a slightly lower blocking probability using the FF, merely due to the fact that S-622 connections get a higher one, thus more free bandwidth is available to allocate the remaining 1-slot connections.

Each algorithm delivers its optimum performance under specific traffic conditions. For this reason, by introducing a scheduling machine able to take dynamic decisions and autonomously select the optimum algorithm based on the type of service and traffic distributions we may easily improve the overall PA slot reservation performance. The selection of the algorithm would also need to take into account the network status, e.g., resource utilization, to make sure that the slot reservations can be efficiently accommodated in the spare channel capacity. Moreover, in order not to disrupt the service of live connections (e.g., already set up), the algorithm optimization would need to be processed only for new connection requests.

4.3.4 PA/QA Access Mode Transmission Decision

In view of the results from previous analysis, in this subsection, a service and access mode assignment is proposed using the two DAOBS access schemes (PA and QA) and the QoS differentiation within the QA access mode itself.

First of all, it is of interest to evaluate the burst blocking performance under similar

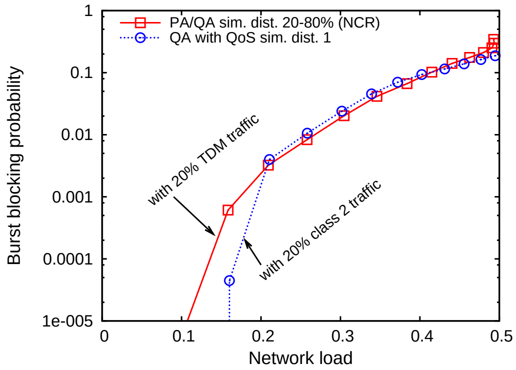


Fig. 4.16: BBP comparison between standalone QA access with QoS and PA/QA dual access modes.

scenarios. Fig. 4.16 gives the burst BP as a function of the mean offered load for the connectionless bursts (80% of the load) when these are transmitted together with other connection-oriented TDM services (20% of the load). Likewise, the mean BP of the connectionless burst with only QA access supporting multiple QoS with the first traffic distribution from Table 4.1 is also represented. We have taken this two plots as in both cases the highest priority traffic (connection-oriented traffic with PA access in the first case, and *class 2* traffic in the second case) represents 20% of the offered load. As we can see, at high loads, the two plots lay within the same value range, with very slight differences between them. Therefore, it is expected that the class differentiation between the two lowest priority traffics in the QA access mode will perform similarly even in the case that the channel is also shared with connection-oriented traffic using PA access, as long as the load of this traffic class is similar to *class 2* traffic.

Turning to the performance of the highest priority traffic in the circumstances mentioned in Fig. 4.16, the decision making of which traffic should be assigned as *class 2* connectionless, or as connection-oriented, depends specifically on the type of traffic and service required. In Section 4.3.2, the QoS differentiation provided by the QA access mode has been analyzed concluding that for both traffic distributions *class 2* is lossless for the whole load range. On the contrary, from Section 4.3.3 the connection-oriented traffic, which is given preference using the PA access mode, gets a non zero connection BP starting from medium offered loads, but still lower than the blocking probability of connectionless bursts. In view of this, *class 2* QA traffic is envisioned for very high priority connectionless traffic such as control traffic without explicit jitter control guarantees, whereas PA access mode can be more suitable for services demanding a guaranteed bandwidth for a long period (in comparison to the burst size) that under some circumstances may also demand for jitter and delay guarantees.

As a concluding remark, Table 4.3 summarizes the application use of each class of service provided by the multi-service DAOBS architecture.

Table 4.3: DAOBS Classes of Service (ConnectionLess (CL), Connection-Oriented (CO)).

Mode	CoS	Type	Use	App
QA	class 0	CL	Best-effort traffic without loss guarantees or delay.	Transactional data
	class 1	CL	Priority traffic that tolerates some losses without guaranteed delay.	Interactive audio-video
	class 2	CL	Very high priority traffic that does not tolerate losses at all.	Interactive data, control
PA	class 3	CO	Long-life TDM traffic with guaranteed bandwidth and delay.	IPTV, cloud computing, Grid

4.4 Delay Model of a DAOBS Lightpath

In previous sections we observed that DAOBS provides the mechanisms and flexibility to allow hybrid sub-wavelength connectionless and connection-oriented services to share the same wavelength capacity. Another interesting result was DAOBS' achievement to improve the burst blocking probability on the network avoiding blockings for in-transit bursts while using the queue arbitrated access mechanism. This turned to be a nice feature to guarantee the delivery of connectionless traffic once enqueued in the local and distributed queue tandem. However, one of the disadvantages of this mechanism is that it introduces an extra access delay for transmitting bursts. As such, bursts are enqueued in the local queues waiting to gain access to the distributed queue from which they can be transmitted into the optical channel. As a matter of fact, it would be interesting to quantify this delay, not only through simulation, but also mathematically, so as to enhance the wavelength or DAOBS entity assignment and reduce the burst blocking and end-to-end delay.

The objective in this section is to obtain a network-wide upper and lower approximation of the average channel access delay on a DAOBS lightpath or light-tree. Approximations will be established by differentiating between the case when the system is modeled with a finite number of sources and a worst-case scenario where the arrival rate to the DQ is treated as generated by an infinite population. Recall that in a pure finite-source system, call congestion is smaller than time congestion because the arrival rate is state-dependent [121].

The present analysis can be achieved by assuming that burst arrivals into the system are random and equally distributed amongst the nodes that belong to the lightpath (or light-tree). The burst arrival process is assumed to be Poisson. As concluded in [122], self-similarity is only relevant for optical packet buffering into bursts and negligible for burst blocking probability, thus confirming the Poisson burst arrival assumption widely used in the literature. Moreover, in the simulation versus model performance comparison, nodes are interspaced a propagation delay corresponding to one slot. Under these ideal conditions, the bursts in the queue system of a single DAOBS lightpath or light-tree, are served in a manner that closely approximates to a cyclic queue system (i.e., the LQs, one for each node) and a M/G/1/-/N (or M/G/1 if we consider an infinite population) with constant service time (i.e., the lightpath/light-tree virtual DQ) as shown in Fig. 4.17.

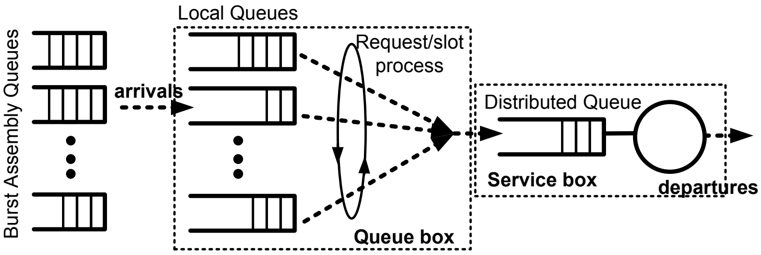


Fig. 4.17: DAOBS queue model.

In this model, the burst access delay comprises the time between the burst arrival into the LQ and the use of a free slot on the forward or downstream channel to transmit the burst. Therefore, this delay can be seen as: 1) the time spent in the LQ, W_{LQ}, and 2) the time spent in the DQ, W_{DQ}. As a result, the total access delay is,

$$W_{AC} = W_{LQ} + W_{DQ} \tag{4.7}$$

Following this approach, the calculation of W_{AC} is performed in two stages. First, we will derive the mean delay in the local queue. With this result, we will be able to compute later the mean delay in the distributed queue.

4.4.1 Mean Delay in the Local Queue

The time a tagged burst spends in the LQ depends on both the number of bursts ahead in the LQ and the time the first burst remains in the DQ. We note that the latter can only be occupied by a single burst from each node. The LQ delay is composed of three variables, as shown in Fig. 4.18:

- the slot arrival synchronization,
- the queue waiting time of bursts ahead in the queue, and
- the delay in queue of the actual burst previous to be enqueued into the DQ.

Bursts can arrive in between time slots and since the arrival process is assumed to be Poisson, the remaining slot length, α, can be considered to be uniformly distributed within the slot length, x. Therefore, the average slot synchronization delay is $\overline{\alpha} = x/2$.

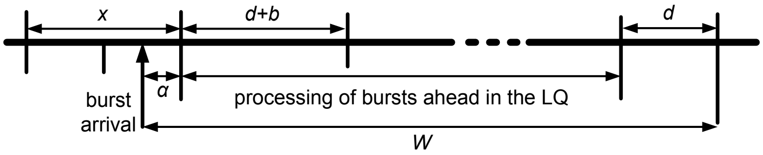


Fig. 4.18: Waiting delay faced by an arriving burst.

Next, the arriving burst has to wait for the processing of other bursts ahead of it in the queue. The average number of bursts in the queue seen by an arriving burst is identical to the average queue length $E[Q]$ which by Little's law is

$$E[Q] = \lambda E[W] \ . \tag{4.8}$$

Each burst in the LQ requires an average time $\overline{d}+\overline{b}$ to finish its processing, where $\overline{d}$ is the delay experienced by the burst, which is related to the mean delay experienced in the DQ, and $\overline{b}$ is the mean processing time that models the request propagation delay. Due to the memoryless property of the exponential interarrival distribution, the total waiting time in the queue faced by the arriving burst is $E[W] = E[Q](\overline{d}+\overline{b}) + \overline{\alpha} + \overline{d}$, which after some calculation using Little's Law we get

$$E[W] = \frac{\overline{\alpha} + \overline{d}}{1 - \lambda(\overline{d} + \overline{b})}, \tag{4.9}$$

where λ is the offered traffic rate per node. Let $x = 1$ slot, and the propagation delay be $\overline{b} = 1$ slot (i.e., as used in the simulation evaluation), we finally have

$$W_{LQ} = \frac{1/2 + W_{DQ}}{1 - \lambda(W_{DQ} + 1)}, \tag{4.10}$$

wherein W_{DQ} is the mean waiting delay in the DQ. If we want to get the lower approximation of W_{LQ}, then $W_{DQ} = W_{DQ,LB}$, and for the upper one, $W_{DQ} = W_{DQ,UB}$.

Next, we derive the mean delay in the distributed queue.

4.4.2 Mean Delay in the Distributed Queue

To calculate the lower approximation of the mean burst delay in the DQ we will consider the use of an M/G/1/-/N queue. This model assumes there are N sources generating jobs, i.e., bursts, which are buffered in the system's queue if their service cannot start immediately. Moreover, a source that has already a pending job cannot generate a new job until the previous one is served, hence the arrival rate is state-dependent. This behavior is similar to the blocking behavior of the DQ in DAOBS, since the DQ can only be populated by a single burst from each source node in the lightpath or light-tree.

The Markovian assumption of the burst arrival into the DQ is justified based on

[123], where the use of a Poisson process to model the request arrival generated by downstream nodes and the slot-occupation provided the best accuracy. In our case, this request/slot-occupation process defines the arrival process in the so-called virtual distributed queue. In this analysis, we follow the M/G/1/-/N analysis first presented by Bose [121]. The M/G/1/-/N is a generalization of the M(n)/G(n)/1/K queue that was first analyzed by Courtois and Georges [124] and further discussed by Takine et al. in [125].

Because now the service times are not memoryless, to analyze such system the easiest method is to embed the Markov-chain at the departure instants of the bursts leaving the system. That is, the number of bursts in the DQ (i.e., the system state) at the imbedded points corresponding to the time instants after a burst departure will form a Markovian Chain. Let $p_{d,k}$ be the probability of state k in the embedded Markov-chain, so

$$p_{d,k} = P\{\text{system in state k}\} \text{ for } k = 0, 1, \ldots, (N-1). \tag{4.11}$$

The state transition probability at equilibrium can be represented as

$$p_{d,jk} = P\{n_{i+1} = k | n_j = j\} \text{ for } 0 \leq j, k \leq (N-1). \tag{4.12}$$

The corresponding state probabilities $p_{d,k}$ can be found by solving the following set of N balance equations

$$p_{d,k} = \sum_{j=0}^{N-1} p_{d,j} \cdot p_{d,jk} \text{ for } k = 0, 1, \ldots, (N-1), \tag{4.13}$$

and the normalization condition $\sum_{k=0}^{N-1} p_{d,k} = 1$.

We need to compute $p_{d,k}$ from these equations in order to find the equilibrium state probabilities $\{p_k\}$, for $k = 0, 1, \ldots, N$, at an arbitrary time instant. Actually, the best way to solve the above equations is by using the generating function $P_d(z)$ defined as

$$P_d(z) = \sum_{k=0}^{N-1} p_{d,k} \cdot z^{N-k-1}, \tag{4.14}$$

and using (4.13), we get

$$P_d(z) = \sum_{j=0}^{N-1} p_{d,j} \sum_{k=0}^{N-1} p_{d,jk} \cdot z^{N-k-1}. \tag{4.15}$$

After several transforms and simplifications [121] we can obtain $p_{d,0}$ which is the probability of finding the server idle at the burst departure instant. The former can be written as

$$p_{d,0} = \frac{1}{\sum_{i=0}^{N-1} \binom{N-1}{i} (\beta_i)^{-1}}, \tag{4.16}$$

where $\beta_0 = 1$ for $i = 0$, and

$$\beta_i = \prod_{j=1}^{i} \frac{B_i^*(j\lambda)}{1 - B_i^*(j\lambda)} \text{ for } i = 1, ..., (N-1), \tag{4.17}$$

in which $B_i^*(s)$ is the Laplace-Stieltjes Transform (LST) of the service time probability density function (pdf) $b(t)$. Taking into account that the mean service time is $E[X] = 1/\mu$ and in our case this time is deterministic (1 slot), the LST is

$$B_i^*(s) = e^{-s/\mu}. \tag{4.18}$$

To find the waiting delay for this distributed queue we consider bursts generated by one of the N sources in the system. A burst spends a mean time W in the system, and after it is serviced, the source can generate another burst. Assuming that the burst inter-arrival from the request/slot counting process follows an exponential distribution, the mean inter-arrival time is $1/\lambda$. The throughput rate of an individual source, γ_i, will then be

$$\gamma_i = \frac{1}{W + \frac{1}{\lambda}}\ , \tag{4.19}$$

and the system throughput γ will be the sum of the throughputs of all the N sources, as in

$$\gamma = \frac{N}{W + \frac{1}{\lambda}}\ . \tag{4.20}$$

The overall throughput rate γ with the mean service time $1/\mu$ can be interpreted as the probability that the queue server is not idle. Let p_0 be the equilibrium probability that the server is idle at an arbitrary time instant, then we have

$$1 - p_0 = \frac{\gamma}{\mu}\ . \tag{4.21}$$

From (4.20) and (4.21), if we eliminate γ, we also have

$$W = \frac{N}{(1-p_0)\mu} - \frac{1}{\lambda}. \tag{4.22}$$

The only unknown value from (4.22) is p_0. To get p_0 we can use the idle (IP) and busy period (BP) analysis, as in

$$p_0 = \frac{E[IP]}{E[IP] + E[BP]}. \tag{4.23}$$

When the server (plus queue) is idle, the arrival process to the queue is Poisson with rate $\gamma_0 = (N-0)\lambda = N\lambda$, and therefore $E[IP] = 1/(N\lambda)$. Moreover, since the busy period terminates with a burst departure instant which leaves the system empty, then the mean length of the busy period will be $E[BP] = 1/(\mu p_{d,0})$. Using these in (4.23), we get p_0 as

$$p_0 = \frac{1/(N\lambda)}{1/(N\lambda) + 1/(\mu p_{d,0})} = \frac{p_{d,0}}{p_{d,0} + N\lambda/\mu}. \tag{4.24}$$

Substituting this result in (4.22) we get the mean time spent in the system (waiting and in service)

$$W = \frac{N}{\mu} - \frac{1-p_{d,0}}{\lambda}, \tag{4.25}$$

in which all the values are known and can be calculated from previous equations, i.e., $p_{d,0}$ from (4.16).

In this analysis, we are interested in the mean waiting time in the queue; hence, the lower approximation ($W_{DQ,LB}$) of the mean waiting delay in the distributed queue can be written as

$$W_{DQ,LB} = W - \frac{1}{\mu} = \frac{N-1}{\mu} - \frac{1-p_{d,0}}{\lambda}\ . \tag{4.26}$$

To calculate the upper approximation we consider the worst-case scenario where the request/slot-occupation is modeled by an infinite population. Thus, the mean delay will be calculated using an M/G/1 with blocking probability

$$P_B = P\{\text{arrival finds burst in DQ}\}, \tag{4.27}$$

that is the blocking of an upcoming burst when the DQ is already taken by a node's burst.

Taking into account this approximation, and noting now λ as the overall offered traffic rate to the lightpath, and $\rho = \lambda/\mu$ the offered traffic load, the effective throughput of the queue is then $\rho_c = \rho\cdot(1-P_B)$. The carried traffic load is related to the probability of finding the system empty, hence

$$p_0 = 1 - \rho_c = 1 - \rho\cdot(1-P_B). \tag{4.28}$$

P_B can be found knowing that the equilibrium state probability p_k for $k = 0, 1, \ldots, K$ at an arbitrary time instant is

$$p_k = p_{a,k} = (1-P_B)\cdot p_{d,k}, \tag{4.29}$$

where $p_{d,k}$ are the state probabilities at departure instants (as derived in (4.13)), and $p_{a,k}$ are the state probabilities at arrival instants regardless of whether the burst joins the queue or is blocked. We can then rewrite P_B as

$$P_B = p_{a,K} = 1 - \sum_{k=0}^{K-1} p_{a,k}. \tag{4.30}$$

Moreover, from (4.29), if $k=0$, then the probability to find the system empty is

$$p_0 = (1-P_B) \cdot p_{d,0}, \tag{4.31}$$

and using (4.28), then we can write $1-\rho\cdot(1-P_B) = (1-P_B)\cdot p_{d,0}$. Isolating P_B from the previous equation, we get

$$P_B = 1 - \frac{1}{p_{d,0}+\rho}. \tag{4.32}$$

Also, from the well-known M/G/1 queue waiting delay [121],

$$W_{M/G/1} = \frac{\lambda\overline{X^2}}{2(1-\rho)}. \tag{4.33}$$

In our system, due to the slotted operation and transmission of bursts from the DQ, we can set the service time as deterministic. If so, the second moment is equal to the square of the mean, hence $\overline{X^2} = \overline{X}^2 = (1/\mu)^2$. Finally, taking into account the P_B calculated above and the effective throughput of the queue ρ_c we can compute the upper approximation of the DQ mean delay as

$$W_{DQ,UB} = \frac{\rho_c \frac{1}{\mu}}{2(1-\rho_c)} = \frac{\rho(1-P_B)\frac{1}{\mu}}{2(1-\rho(1-P_B))} = \frac{\frac{\rho}{\mu(p_{d,0}+\rho)}}{2\left(1-\frac{\rho}{p_{d,0}+\rho}\right)}. \tag{4.34}$$

Summing up, the overall access delay of a burst, taking into account the delay experienced in both the local queue and the distributed queue is, for the lower approximation

$$W_{AC,LB} = \frac{1/2 + \frac{N-1}{\mu} - \frac{1-p_{d,0}}{\lambda}}{1-\lambda\left(\frac{N-1}{\mu} - \frac{1-p_{d,0}}{\lambda} + 1\right)} + \frac{N-1}{\mu} - \frac{1-p_{d,0}}{\lambda}, \tag{4.35}$$

and for the upper one,

$$W_{AC,UB} = \frac{1/2 + \dfrac{\rho/(\mu(p_{d,0}+\rho))}{2(1-\rho/(p_{d,0}+\rho))}}{1-\lambda\left(\dfrac{\rho/(\mu(p_{d,0}+\rho))}{2(1-\rho/(p_{d,0}+\rho))}+1\right)} + \frac{\dfrac{\rho}{\mu(p_{d,0}+\rho)}}{2\left(1-\dfrac{\rho}{p_{d,0}+\rho}\right)} \tag{4.36}$$

4.4.3 Numerical Results

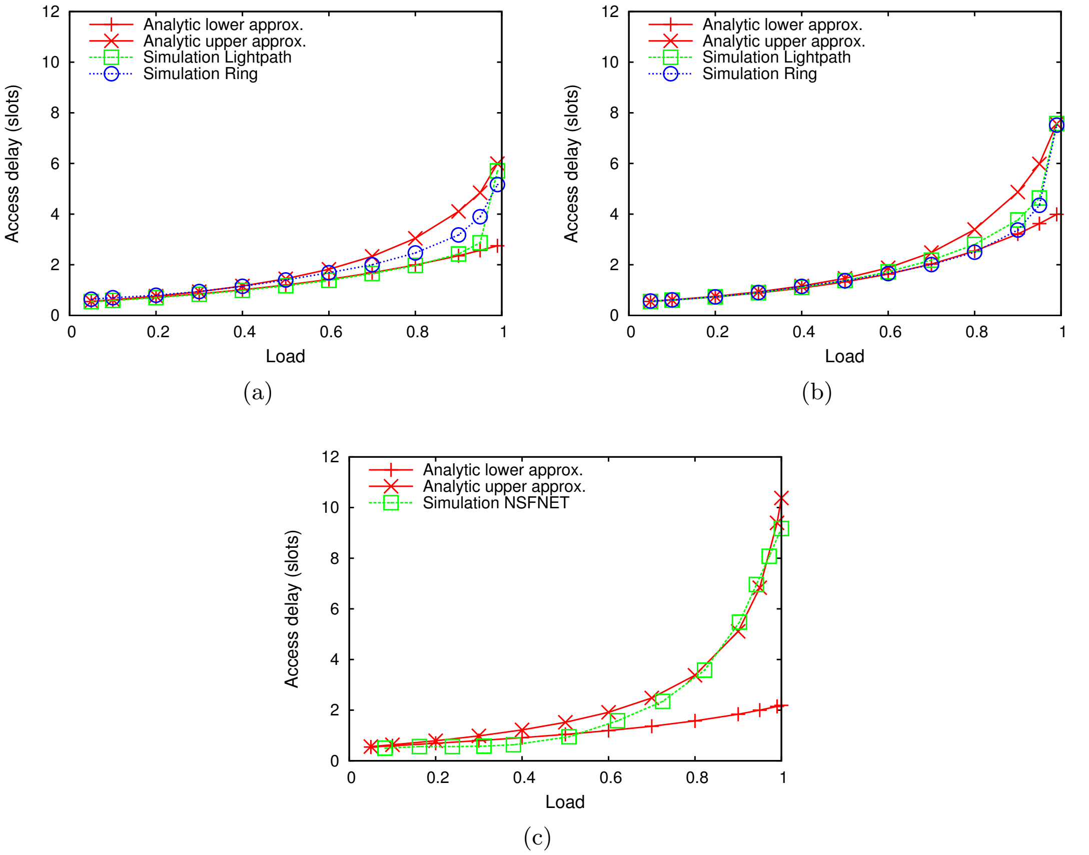
The validation of the proposed model was made through simulation. We considered three different network topologies: lightpaths of 8 and 16 nodes, rings of 8 and 16 nodes with 4 and 8 wavelengths, respectively; and the NSF network topology with 16 wavelengths. DAOBS expands over ring and mesh networks as a set of lightpaths and light-trees, respectively. In all cases, nodes are interspaced one slot. The traffic load was distributed equally amongst network nodes and burst lengths were equal to 1 slot.

Fig. 4.19(a) and 4.19(b) show the access delay results for the 8 and 16-node lightpath and ring networks as a function of the normalized load per lightpath. As we can see, the simulation results stand well between the lower and upper approximations for both topologies. In particular, in the lightpath topology, we can observe that shortening the length of the lightpath in number of nodes, the simulation results fit better the lower approximation. With fewer number of nodes, the state-dependent arrival from the finite population gives a better alikeness.

We also compared the analytical results with those obtained in the NSFNET simulation scenario. In this case, we considered the average shortest path length of this network (2.14) as the number of actives nodes into the model, and computed the delay under similar loads. As shown in Fig. 4.19(c), simulation results resemble better our upper approximation overall. This is due to the overlay required to extend DAOBS on mesh topologies which generates a more irregular composition of LQ/DQ queues across the network with respect to the normalized load per link. This results in, at low loads about every light-tree is used by a single node, thus performing closer to the lower approximation, whereas at higher loads other nodes make use of same light-tree resources increasing the load into the DQ and consequently fitting to the M/G/1 approximation better.

4.5 Light-Tree Topology Optimization Without Wavelength Conversion

As we introduced at the beginning of this chapter, DAOBS uses a light-tree-based virtual topology or overlay. Each of these light-trees have a single HoB node that generates the burst control packets necessary for the DAOBS mechanisms to operate. Moreover, on the light-tree establishment, trees sharing the same output port using two different input links in the switch are not allowed. This allows avoiding collisions among



(a)

(b)

(c)

Fig. 4.19: Access delay: (a) lightpath and ring of 8 nodes and 4 wavelengths, (b) lightpath and ring of 16 nodes with 8 wavelengths, and (c) lightpaths setup on NSFNET.

traffic flows not belonging to the same light-tree. Then, the DAOBS MAC protocol orders the transmissions among the nodes belonging to the same light-tree avoiding burst overlapping contentions. All this guarantees the delivery for in-transit bursts.

An issue we may question ourselves is, “how can we generate this virtual topology so as to enhance and provide more capacity and improve the efficiency of the protocol?” We must note that, without wavelength conversion, each light-tree uses only one lambda, and due to the effect of the input port restrictions, some optical capacity can be lost and not effectively assigned to the rest of given light-trees. Thus, an optimization of this process is required.

The dynamic approach of the light-tree computation is usually made using greedy algorithms. Due to the dynamic nature, the establishment of current light-trees may block the establishment of upcoming ones, even though there are resources to set up both using alternative paths or wavelengths. Thus, in static routing, when several sessions are known in advance, a combined one-step computation of all sessions rather than a step-by-step is advantageous. In such a case, the computation can be derived to a mathematical formulation, which can be further solved using integer linear programming (ILP).

4.5.1 General Problem Statement and Formulation

In this section, the problem of setting up a group of light-trees at an overall minimum cost on a given topology is formally stated. In the problem statement we are given the following input information.

- A physical topology represented by $G = (V, E)$ and consisting of a weighted undirected graph. V is the set of network nodes and E is the set of links between nodes. Each link is assigned a weight (e.g., the physical distance between the corresponding node pair or the number of used wavelengths on the link).
- The number of wavelengths on each fiber. This number is symbolized by W.
- A group of k light-tree sessions defined by a root node and a set of destination nodes that must be reached by the light-tree.

The goal is to set up all k light-tree sessions on the given physical topology while minimizing the overall cost. In this case, the cost of a light-tree session is represented by the sum of the weights of the physical links used by it. As pointed by [126], the problem of establishing multiple directed trees at a minimum cost is an NP-complete problem. The problem is a generalized version of the directed Steiner minimum tree (DSMT).

Without wavelength converters in the switch architectures, the entire light-tree is on a common wavelength. The problem formulation on a wavelength-continuity constraint optical network is similar to the problem with wavelength converters. However, in the former case, there are some additional variables and constraints. It is worth noting that proper wavelength assignment in this cases is very important in order to minimize the overall cost. Furthermore, we are also interested on applying certain constraints on the light-tree topology generation to avoid contentions from two different input ports. Namely, the constraint imposes that two different light-trees cannot be established through a core node using two different input links and the same output port over the same wavelength. In this way, contentions in the optical switch from two different input ports are avoided.

The notation which shall be used throughout is the following:

- s and d refer to the source and destination node in the light-tree session.
- m and n are the endpoints of the certain physical link on a light-tree.
- i is an index for the light-tree session number, so $i = 1, 2, \dots, k$.
- c is the index of the wavelength used on the light-tree.
- mnp is used to define a certain switching entry at node n, between ports m and p.

Next two sections describe the formulation with and without fractional capacities, i.e., sub-wavelength capacity demands, while meeting the requirements established by the light-tree virtual topology.

4.5.2 Problem Formulation Without Fractional Capacities

In this first problem formulation we introduce the optimization formulation proposed in [127]. In this case, all light-tree sessions demand the whole capacity of the wavelength. As a result, once a light-tree is established on a specific link, no other light-tree using different input links can make use of that same output link; hence we do not incur in any collision among bursts in the way we have previously mentioned, i.e., light-trees using two different input ports. This is not a very realistic scenario in our case; as expressed throughout this thesis, main contributions have to do with sub-wavelength provisioning. However, this formulation will serve us well for presenting the nomenclature and establishing the background for further extensions in the next section.

Given the following input parameters:

- N is the number of nodes on the network.
- W is the number of wavelengths per fiber.
- P_{mn} is the physical topology, where the fiber links are assumed to be bidirectional, that is $P_{mn} = P_{nm} = f$, and f is the number of fibers per link. Hence, if there is no connection between nodes m and n, $P_{mn} = P_{nm} = 0$. For simplicity, we will assume there is only one fiber per link, so, $P_{mn} \leq 1$.
- $D_p(m)$ = number of physical output ports (links) from node m.
- w_{mn} = weight of the link between nodes m and n.
- C = capacity of the channel (e.g. 10 Gbps).
- A group of k light-tree sessions S_i for $i = 1, 2, 3, \ldots, k$. Each light-tree has a source or root node and a set of destination nodes that must be covered by the entire tree. The group of nodes is denoted by $\{s_i, d_{i_1}, d_{i_2}, \ldots\}$. The number of destination nodes of light-tree i is denoted by D_i.
- The nodes do not have wavelength converters; hence the light-trees must be set up using only one wavelength.

The variables of the problem are:

- M_{mn}^{ic} is a boolean variable to denote if light-tree i uses link between nodes m and n on wavelength c, otherwise, $M_{mn}^{ic} = 0$.
- V_p^i is another boolean variable, which is equal to one if node p belongs to light-tree i; otherwise, $V_p^i = 0$. A node belongs to a light-tree if it is either source, destination or an intermediate node.
- F_{mn}^i is an integer commodity-flow variable. Each destination node of the tree needs one unit of commodity. So, D_i units of commodity flow out of source

node s_i on light-tree i. F^i_{mn} is the number of units of commodity flowing on the link from node m to n for light-tree i. This value also represents the number of destination nodes in light-tree i downstream of the link between nodes m and n.

- C^i_c is a boolean variable equal to one if light-tree i is on wavelength c; otherwise, $C^i_c = 0$. Since we do not have wavelength converters on the network, the light-tree can occupy only one wavelength.

The objective of the problem is to optimize:

- The total cost of all light-trees, which in this case the purpose is to minimize such cost:

$$\text{Minimize} \sum_{i=1}^{k}\sum_{c=1}^{W}\sum_{m,n} w_{mn} \cdot M^{ic}_{mn} \tag{4.37}$$

To solve the problem we have the follow constraints:

- Light-tree creation constraints:

$$\forall i, \forall n \neq s_i : \sum_{m,c} M^{ic}_{mn} = V^i_n \tag{4.38}$$

$$\forall i : \sum_{m,c} M^{ic}_{ms_i} = 0 \tag{4.39}$$

$$\forall i, \forall j \in S_i : V^i_j = 1 \tag{4.40}$$

$$\forall i, \forall m \neq d_{i_j}, j = 1, \ldots, D_i : \sum_{n,c} M^{ic}_{mn} \geq V^i_m \tag{4.41}$$

$$\forall i, m : \sum_{n,c} M^{ic}_{mn} \leq D_p(m) \cdot V^i_m \tag{4.42}$$

$$\forall m, n : \sum_{i,c} M^{ic}_{mn} \leq P_{mn} \cdot W \tag{4.43}$$

$$\forall m, n, c : \sum_{i} M^{ic}_{mn} \leq P_{mn} \tag{4.44}$$

- Commodity-flow constraints:

$$\forall i, \forall m \notin S_i : \sum_{n} F^i_{nm} = \sum_{n} F^i_{mn} \tag{4.45}$$

$$\forall i, \forall m = s_i : \sum_{n} F^i_{s_i n} = D_i \tag{4.46}$$

$$\forall i, \forall m = s_i : \sum_{n} F^i_{ns_i} = 0 \tag{4.47}$$

$$\forall i, \forall m = d_{i_j}, j = 1, \ldots, D_i : \sum_{n} F^i_{nm} = \sum_{n} F^i_{mn} + 1 \tag{4.48}$$

$$\forall i, m, n : \sum_c M_{mn}^{ic} \leq F_{mn}^i \tag{4.49}$$

$$\forall i, m, n : F_{mn}^i \leq N \cdot \sum_c M_{mn}^{ic} \tag{4.50}$$

$$\forall i, m, n : F_{mn}^i \leq D_i \tag{4.51}$$

- Wavelength-related constraints and continuity constraint:

$$\forall i : \sum_c C_c^i = 1 \tag{4.52}$$

$$\forall m, n(n > m), \forall i, c : M_{mn}^{ic} + M_{nm}^{ic} \leq C_c^i \tag{4.53}$$

Next, we explain the meaning of all these constraint equations. Equation (4.38) ensures that every node that belongs to a light-tree (except the source) has an incoming edge/link. Equations (4.39) and (4.40) ensure that the source node has no incoming edge, since it is the root of the light-tree, and that every source node and the destination nodes belong to the light-tree i. Moreover, (4.41) checks that every node belonging to the tree, except the destinations, have at least one outgoing link on the light-tree. Equation (4.42) complements the previous one ensuring that every node with at least one outgoing edge belongs to the light-tree. Equations (4.43) and (4.44) restrict the number of light-trees to the capacities available on the link, namely, number of fibers and wavelengths. Specifically, the former restricts the number of light-tree segments between nodes m and n, while the second one restricts the number of trees between the pair of nodes and wavelength c to the number of fibers between these two nodes.

The commodity-flow constraints are related to the problem of multiple commodities (goods) flowing through the network between source and destination nodes [128, 129]. These constraints are used to create a connected tree with end-to-end connectivity between the source and every destination on the light-tree. In this sense, (4.45) ensures that any intermediate node, which is neither a source nor a destination, the incoming flow is the same as the outgoing flow value. Equation (4.46) also checks that the number of outgoing flows from the source is the number of destinations of the tree. Moreover, (4.47) restricts that no flow of the tree i has as a destination the root node, whereas (4.48) ensures that the total outgoing flow at a destination node is one less than the incoming flow units. Therefore, a destination node, even in the core of the light-tree is a sink of one flow unit. The equation works for both, final leaf destination nodes and core destination nodes. The next two commodity-flow equations (4.49) and (4.50) check that every link used by a light-tree has a positive flow and the rest of links not used by this tree have no flow. Finally, (4.51) limits the flow through a link to the number of destinations of the tree at the most.

The last group of constraints are linked with the wavelength-continuity constraint. The former (4.52) restricts a light-tree to only use one wavelength. Equation (4.53) counts that all the links used by a light-tree are on the same wavelength. Because light-trees use the whole capacity of the channel, no collisions occur between light-trees using two different input ports and a common output port. Therefore, we do not need further constraints.

Finally, it is of interest to approximate the complexity of the problem formulation in number of variables and constraints. On the one hand, the number of unknown variables in this problem is $O(kW + kN + kN^2 + kWN^2) \sim O(kWN^2)$, which comes from: (a) M_{mn}^{ic} adding kWN^2 variables, (b) V_p^i which denotes if node p belongs to light-tree i results in kN variables, (c) the commodity flow variable sums kN^2 variables more, and finally (d) the boolean variable C_c^i adds kW to the sum. Thus, the number of variables increases linearly as a function of the number of wavelengths and light-trees on the network, and quadratically with the number of nodes. On the other hand, the constraint complexity is bounded by $O(kWN^2)$ counting the number of constraints to satisfy for instance in (4.53).

4.5.3 Problem Formulation With Sub-wavelength Light-tree Capacity Demands

In this new reformulation of the problem, light-trees use a fractional capacity, i.e., sub-wavelength. Therefore, if two light-trees do not incur in input link restriction incompatibilities and its bandwidth request is ensured by the channel capacity, then they must be allowed to share the same wavelength. However, if an output port would be able to allocate the demanded capacity through a common output port, but the two light-tree sessions use different input links at a specific node, this can not be permitted. To ensure that this is satisfied, new constraint variables and equations will need to be added to the original problem detailed in the previous section. Note that this problem is similar to one introduced in [127]; however, in our case, we assume an all-optical switching of data. Therefore, no O/E/O conversions are present. This implies that electrical grooming is not realized. For other IP over WDM grooming ILP solutions, we also refer to [130].

Furthermore, we can differentiate two specific sub-cases: (a) every root or HoB node defined in the connection request list creates and maintains its own light-tree (it will be referred as *case (a)*), or (b) light-tree connection requests are aggregated with other conforming light-trees, as long as the wavelength capacity and input link restriction is guaranteed (referred as *case (b)*). Obviously, in the first case, the total number of light-tree instances will be greater. We will evaluate both cases.

We should note that optimization techniques similar to the problem we address in this section have been proposed in the literature. For instance, the authors in [131] developed an ILP formulation to optimize the light-trail setup in WDM ring networks. Obviously, our proposed formulation differs from it since we assume in our case a mesh network topology. A previous work for the light-trail optimization in mesh networks was proposed in [132]. However, light-trails differ from light-trees because the former are simply an extension of a lightpath wherein the downstream nodes are allowed to groom traffic all-optically. Light-trees enable one-to-many communication (multicast), and a difference of our proposed solution with respect to others like [133] is the absence of O/E/O grooming.

Given the following parameters:

- N, W, P_{mn}, and the rest of parameters used in the previous formulation remain valid.
- The capacity demanded by each light-tree i with respect to the total capacity of the optical channel (C), is f_i (i.e., the sub-wavelength capacity).

The variables of the problem are:

- Variables M_{mn}^{ic}, V_p^i, F_{mn}^i and C_c^i remain valid and have the same meaning as in the previous formulation.
- S_{mnp}^{ic} is a boolean variable to denote if switching entry at node n between ports m and p is set on wavelength c for light-tree i, otherwise $S_{mnp}^{ic} = 0$.

Now, the problem takes into account the sub-wavelength demands of the light-trees and the objective is to optimize:

- The total cost of all light-trees by minimizing this value:

$$\text{Minimize} \sum_{i=1}^{k} f_i \cdot \sum_{c=1}^{W} \sum_{m,n} w_{mn} \cdot M_{mn}^{ic} \tag{4.54}$$

To solve the first case, *case (a)*, we have to add/substitute to the previous problem formulation the following constraints:

- Light-tree creation constraints: the following two equations substitute (4.43) and (4.44), respectively.

$$\forall m, n : \sum_{i,c} f_i \cdot M_{mn}^{ic} \leq P_{mn} \cdot W \tag{4.55}$$

$$\forall m, n, c : \sum_{i} f_i \cdot M_{mn}^{ic} \leq P_{mn} \tag{4.56}$$

- And add the following equations to the switching-related light-tree constraint group.

$$\forall c, n, p, m : \sum_{i} S_{mnp}^{ic} \leq 1 \tag{4.57}$$

$$\forall c, n, p : \sum_{m,i} S_{mnp}^{ic} \leq P_{np} \tag{4.58}$$

$$\forall c, i, m, p, n \neq s_i : M^{ic}_{mn} + M^{ic}_{np} - V^i_n \leq S^{ic}_{mnp} \tag{4.59}$$

$$\forall c, i, p, n = s_i : M^{ic}_{mn} \leq \sum_m S^{ic}_{mnp} \tag{4.60}$$

Equations (4.55) and (4.56) restrict the number of light-trees to the capacities available on the channel, but now taking into account that the capacity demands can be fractional. Specifically, the former restricts the number of light-tree segments between nodes m and n summing up the sub-wavelength capacities. Equation (4.56) restricts the number of trees between any pair of nodes and wavelength c to the number of fibers between these two nodes.

The new four light-tree and switching-related constraints serve the purpose to avoid the collisions between light-trees even though the capacity demands per link would be met, yet allowing each one of them to handle its own light-tree session. Firstly, (4.57) ensures that for a given switching entry at node n between nodes m and p and on wavelength c, the sum of all light-tree sessions using this same entry is at the most only 1. Secondly, (4.58) checks that given an output link at a switching node, the number of light-tree sessions using the same wavelength on any other input port is also, at the most the number of fibers on the output link, which for simplicity was set to $P_{np} = 1$. Basically, these two first equations restrict the switching connectivity of the light-trees. Finally, (4.59) sets the value of the switching entry S^{ic}_{mnp} at a node different to the root on session i, and (4.60) does the same but for the root node of the light-tree.

Due to the addition of the new variables, now the problem complexity increases in an order of magnitude with respect to the size of the network. Specifically, the number of unknown variables in this new problem formulation is $O(kW + kN + kN^2 + kWN^2 + kWN^3) \sim O(kWN^3)$. In comparison with the previous formulation in Section 4.5.2, the switching entry variables add kWN^3 new variables to the system. Therefore, now the number of variables increases linearly as a function of the number of wavelengths and light-trees on the network, and increases to the cube with the number of nodes. On the other hand, the constraint complexity is bounded by $O(kWN^3)$, mainly from (4.59) and (4.60).

Regarding the second case (*case (b)*), i.e., we allow aggregation of conforming light-trees, we need to substitute the switching-related light-tree constraints as follow:

- And add the following equations to the switching-related light-tree constraint group.

$$\forall c, n, p, m : \sum_i f_i \cdot S^{ic}_{mnp} \leq 1 \tag{4.61}$$

$$\forall c, n, p : \sum_{m,i} f_i \cdot S^{ic}_{mnp} \leq P_{np} \tag{4.62}$$

$$\forall c, i, m, p, n \neq s_i : M^{ic}_{mn} + M^{ic}_{np} - V^i_n \leq S^{ic}_{mnp} \tag{4.63}$$

$$\forall c, i, m, p, n = s_i : S^{ic}_{mnp} = 0 \tag{4.64}$$

Constraints (4.61), (4.62), (4.63) and (4.64), substitute (4.57), (4.58), (4.59) and (4.60), respectively. Because now the light-trees can be aggregated, we include the sub-wavelength (or fractional) capacity demand in the switching-related light-tree constraints. Therefore, (4.61) and (4.62) retain the same meaning as before, but now allow to include original light-tree requests that use fractional capacity on the same input port. Equation (4.63) is the same one as defined in the previous case (4.59). And finally, (4.64) defines that origins of original light-tree requests do not need to manage their own light-tree as long as the capacity demand is guaranteed.

4.5.4 ILP Results

In this section, some results are presented for illustrative purposes. To this end, we use the 6-node network topology in Fig. 4.20, as it is also used in [127]. In such a scenario, we define six different light-tree sessions, one from every node on the network: $T_1 = F \rightarrow \{A, B, C\}$, $T_2 = E \rightarrow \{A, B\}$, $T_3 = A \rightarrow \{C, D\}$, $T_4 = B \rightarrow \{A, C, D, E\}$, $T_5 = C \rightarrow \{A, E, F\}$, and $T_6 = D \rightarrow \{A, B, F\}$. Every light-tree is defined by a root node and a group of destination nodes that have to be reached on the same light-tree. The capacities demanded by these light-trees vary between the two problem solutions. In the former case, demands are assigned the whole capacity of the wavelength. For simplicity the weight cost of every link on the network is set as $w_{mn} = 1, \forall m, n$.

Table 4.4: Light-tree optimization with full-wavelength capacity demands.

Source	**Dests**	**Route**	**Cost**	λ
F	A,B,C	F-A,A-B,B-C	3	λ_1
E	A,B	E-A,A-B	2	λ_2
A	C,D	A-E,E-D,D-C	3	λ_2
B	A,C,D,E	B-A,B-D,D-C,D-E	4	λ_1
C	A,E,F	C-D,D-E,E-F,F-A	4	λ_2
D	A,B,F	D-B,B-A,A-F	3	λ_2

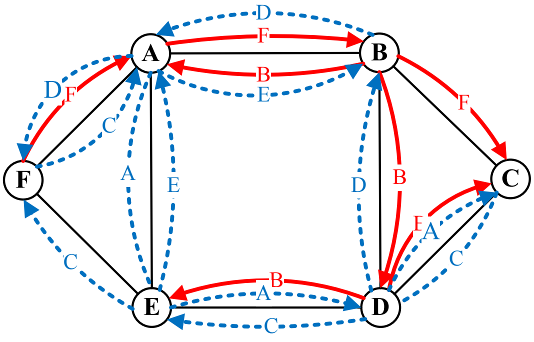


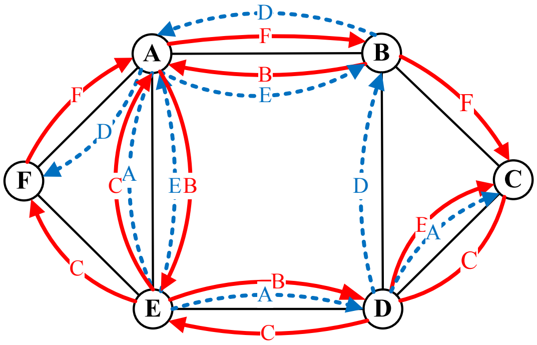
Fig. 4.20: Optimization results with full-wavelength capacity demands.

Table 4.4 and Fig. 4.20 show the results for the full-wavelength formulation problem.

As the link weight cost is 1, the total cost of this solution is the number of links used by all the light-trees. Therefore, the total cost is 19. The minimum number of wavelengths that lets establish all the light-trees is two.

Table 4.5: Light-tree optimization with sub-wavelength capacity demands and switching constraints for *case (a)*.

Source	**Dests**	f_i	**Route**	**Cost**	λ
F	A,B,C	0.5	F-A,A-B,B-C	1.5	λ_1
E	A,B	0.5	E-A,A-B	1	λ_2
A	C,D	0.5	A-E,E-D,D-C	1.5	λ_2
B	A,C,D,E	0.5	B-A,A-E,E-D,D-C	2	λ_1
C	A,E,F	0.5	C-D,D-E,E-A,E-F	2	λ_1
D	A,B,F	0.5	D-B,B-A,A-F	1.5	λ_2



Sub-wavelength case (a)

Fig. 4.21: Optimization results with sub-wavelength demands and switching constraints for *case (a)*.

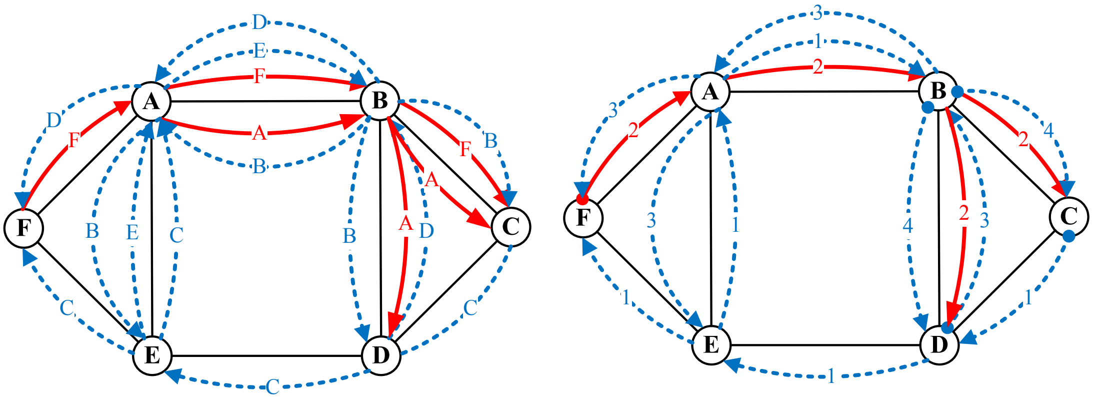
In the second formulation for *case (a)*, the solution provided by the ILP solver is very similar to the non-fractional example. The results can be seen in Table 4.5 and Fig. 4.21. In this specific case, the capacities demanded by the light-trees are in every case 0.5, which is half the capacity of a wavelength, i.e., sub-wavelength granularity. The number of links occupied by the light-trees is the same as in Fig. 4.20 which results in exactly half of the total cost in comparison with the previous full-wavelength problem. However, the distribution of the wavelengths among the set of light-trees is a bit different. It is worth noting that using fractional capacities constraints in the first problem formulation would have given light-trees sharing output links and using different input links at the switching node by assuming the capability to groom traffic in the electrical domain. Nonetheless, this is not allowed in our DAOBS architecture due to the inability of O/E/O conversion for in-transit bursts and the use of the light-tree input link constraints to avoid burst collisions.

Finally, Table 4.6 and Fig. 4.22 show the results for the same sub-wavelength light-tree demand, but now for *case (b)* constraints. As we can see, in this case the solution provided by the ILP solver is quite different from the previous two solutions.

Table 4.6: Light-tree optimization with sub-wavelength capacity demands and switching constraints for *case (b)*.

Source	Dests	f_i	Route	Cost	λ	Aggregated light-tree
F	A,B,C	0.5	F-A,A-B,B-C	1.5	λ_1	LT 2
E	A,B	0.5	E-A,A-B	1	λ_2	LT 1
A	C,D	0.5	A-B,B-C,B-D	1.5	λ_1	LT 2
B	A,C,D,E	0.5	B-C,B-D,B-A,A-E	2	λ_2	LT 3 & 4
C	A,E,F	0.5	C-D,D-E,E-A,E-F	2	λ_2	LT 1
D	A,B,F	0.5	D-B,B-A,A-F	1.5	λ_2	LT 3

Should we took a look at Fig. 4.22(a), we would see that now there are original light-tree demands which are groomed on common output links. For instance, connection request from node B and D share link B-A. As pointed in the formulation, this behavior is now permitted because we allow light-trees to be aggregated as long as the capacity demands are guaranteed and the same wavelength is not used from two different input links over the same output link. The aggregated light-trees that would be set up on the network are shown in Fig. 4.22(b).



Sub-wavelength case (b)

(a)

Sub-wavelength case (b), aggregated light-trees

(b)

Fig. 4.22: Optimization results with sub-wavelength demands and switching constraints for *case (b)*: (a) original light-trees generated, and (b) aggregated light-trees.

4.6 Summary

The provisioning of hybrid connectionless and connection-oriented services over the same optical network substrate is not straightforward. A possibility to realize this, as stated in the background chapter, is to physically separate the network resources into two groups, one for each service type, respectively. Another option is to virtually separate the resources, but still maintaining resources and operation independence. In this thesis, we have followed a different approach by enabling true resources sharing between CL and CO services using a MAC protocol.

This chapter has introduced and extensively analyzed a novel multi-service MAC protocol for OBS mesh networks. An advantage of this protocol is its integrated design that permits to potentially serve a broad range of applications with diverse QoS requirements. The MAC provides two main access methods: queue arbitrated for connectionless bursts and pre-arbitrated for TDM connection-oriented services. On the one hand, the queue arbitrated access is based on a counting and monitoring process of burst and request control packets traveling in opposite control channel directions and a distributed preemption-based scheme in a multi-queue access priority system. On the other hand, the pre-arbitrated mode is based on the pre-reservation of slots supported by a higher service layer module. A benefit of the MAC protocol is that contentions for in-transit bursts are avoided, thus guaranteeing the data delivery even for connectionless applications.

Results evaluated through simulation have shown that highest priority bursts are guaranteed zero losses and very low access latencies in the QA access mode. Even for the intermediate traffic class, the protocol can guarantee an acceptable burst blocking probability for a diverse number of applications. Regarding the PA access mode, results showed that doubling the offered TDM traffic load increased in more than one order their connection blocking probability, slightly affecting the blocking of connectionless bursts. Moreover, three different slot scheduling algorithms for allocating TDM connections have also been evaluated, yielding to diverse results depending on the requested bandwidth. The overall results demonstrate the suitability of the proposed architecture for future integrated multi-service optical networks.

In this chapter we have also tackled two of the issues related with the DAOBS architecture and its operation. Firstly, we have mathematically modeled the access delay resulting from the connectionless queue arbitrated access. This model may be used to quantify the expected delay in order to improve the wavelength assignment and reduce the end-to-end delay. Secondly, we have mathematically formulated the generation of the virtual light-tree overlay topology for the static traffic case, so as to optimize the resources usage on the network, typically by minimizing the overall cost. In this case, we have contributed with some simple constraint extensions to generate the virtual topology when light-tree demands are fractional to the wavelength capacity.

We have seen that a hybrid CL/CO sub-wavelength optical network architecture is possible. However, such approach has leveraged the necessity to accurately plan the virtual light-tree overlay in the static case. In the dynamic traffic scenario, some network capacity could be lost due to setting up greedily the light-trees in a decentralized manner. In the next chapter, we will pursue the sub-wavelength bandwidth provisioning from another perspective, that is, using a centralized connection-oriented approach for guaranteed bandwidth provisioning.

Chapter 5

Guaranteed Sub-Wavelength Provisioning on Time-Shared Optical Network

In Chapters 3 and 4 we have reviewed and proposed connectionless and hybrid connectionless and connection-oriented architectures for sub-wavelength optical networks. In particular, we have seen that full connectionless transmission cannot guarantee data delivery. Usually, in this case, blockings are produced on in-transit packets or bursts (i.e., in the core of the network). Alternatively, the use of a hybrid CL/CO MAC-based architecture enabled the provisioning of guaranteed data transport, for both CL and CO applications, as long as bursts are accepted into the MAC transmission queues.

In this chapter, we look further into the provisioning of guaranteed sub-wavelength channels, but now more from the network operator's perspective. We consider as our scenario an optical metropolitan network. We explore the benefits of a PCE-enabled time-shared optical network (TSON) that uses a centralized sub-lambda assignment element to schedule the connections avoiding collision within the network and guaranteeing the bandwidth availability for the accepted connections. Five different assignment policies in increasing constriction order are also analyzed and compared against their integer linear programming formulation for a static example case. For the dynamic case, simple heuristics on a tandem path and sub-wavelength computation engine are implemented and assessed through simulation.

5.1 Introduction

Network centric services enable new business opportunities for network operators and media content providers by combining network and IT (computing and content storage) resources. Examples of such services include PC virtualization, video-on-demand, storage area networking, etc. Many of these applications show a connection-oriented behavior, for instance, when a client sets up an IPTV program for a certain duration of time like minutes or hours. Moreover, many of these applications cannot justify the reservation of the whole wavelength capacity. Actually, many of them may just request a finer wavelength granularity an order or two below the full-wavelength capacity. Therefore, connection-oriented sub-wavelength provisioning needs to also be realized.

The finer sub-wavelength granularities, together with the continuous growth in Internet traffic (mainly driven by video and peer-to-peer applications), make a huge impact in the metropolitan area network. As a matter of fact, network costs of current metro architectures depend significantly on traffic growth; the higher the traffic is, the higher the network costs. Moreover, metro optical networks need to evolve to bridge the gap between the finer granularity traffic flows in the access networks, likely based on PON technologies, and the coarse lightpath connections used in the backbone or core networks, based on wavelength-switching or OCS (see Fig. 5.1). Thereby, new all-optical sub-wavelength architectural solutions fully controlled by advance control planes are needed to deliver the huge expected increase in traffic in a cost-effective way and ensure low cost broadband Internet access and increased bandwidth transparency.

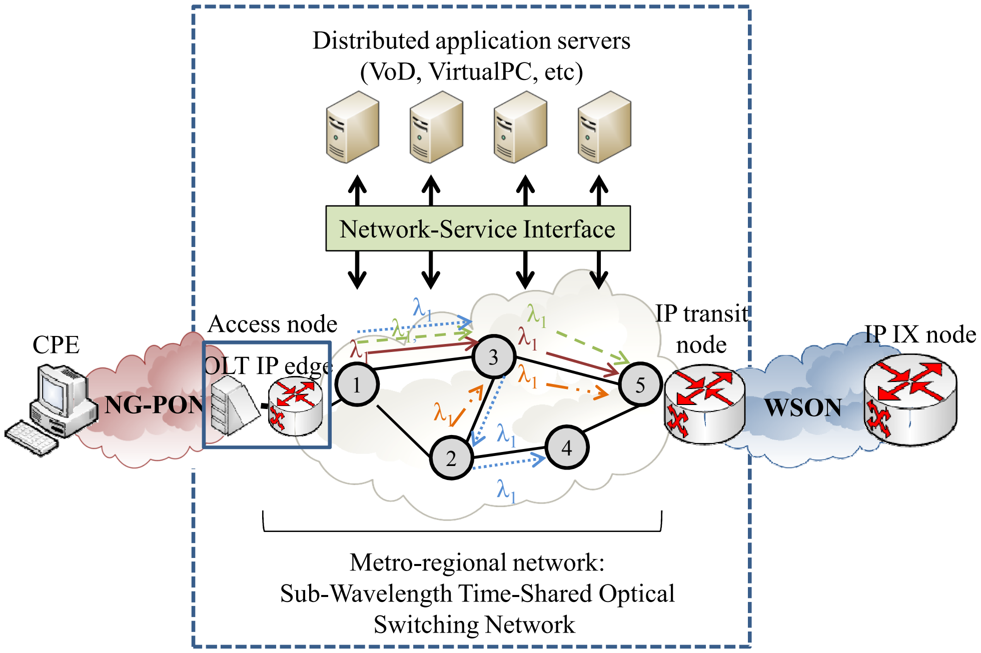


Fig. 5.1: Time-shared optical network application scope.

An option to upgrade the metro network capacity and reduce the operational and investment cost is the provisioning of sub-wavelength switching (i.e. the time-shared utilization of a single wavelength by optical bursts, packets or slots). The introduction of sub-wavelength granular all-optical switching technologies in metro-regional networks is motivated by many studies on the evolutionary trend of network traffic and emerging technologies [134, 135, 136]. As such, we can identify some advantages like:

- CAPEX minimization through optimization of expensive high-capacity optoelectronic IP ports.
- OPEX minimization by saving space and power consumption making use of optical instead of electronic switching.
- Metro regional network granularity fulfillment, since most network connections between IP edge nodes are typically in the sub-wavelength range.

The all-optical sub-wavelength benefits may not only come from the data plane optimization, but also from the benefits on achieving and extending standardized control planes. As such, the control plane needs to simplify the operational tasks by dynamically establish, restore and reallocate connection tunnels across the network. Moreover, multi-domain interoperability is a very important issue, and by using standardized control plane implementations, we are able to ensure interoperability, not only between domains, but also between network technologies and provisioning granularities.

In this chapter, we explore the benefits of a PCE-enabled time-shared [101] sub-wavelength optical network (TSON) that uses a centralized sub-lambda assignment element to schedule the connections avoiding collision within the network. As opposed to other optical network architectures based on the connectionless provisioning of sub-wavelength bandwidth introduced in the previous chapters, in this case, our approach is connection-oriented.

In TSON, the lower layer bandwidth provisioning is realized by time-sharing the wavelength capacity. In this way, the wavelength bandwidth is partitioned into time-slots of fixed size, which are further organized along a fixed-length time frame [103]. Such a mechanism enables incoming connections to request sub-wavelength bandwidth by mapping such capacity demand into a fixed number of time slots, which are then allocated for the entire connection holding time. Therefore, different connection requests can share the wavelength capacity, enabling the provisioning of sub-wavelength connectivity. The diverse wavelength granularity comes at the expense of, not only realizing the route and wavelength selection for the connection, but also determining the time-slot assignment. As introduced in the background chapter, this problem is also known as routing, wavelength and time-slot assignment (RWTA), and its primary goal is to maximize the network resources usage (i.e., wavelength bandwidth) while minimizing the overall connection blocking probability.

The contributions of this work are threefold: first, we introduce a novel PCE-enabled architecture for the provisioning of sub-wavelength services. Second, we formally define the routing, wavelength and slot assignment optimization problem through a detailed integer linear formulation. As such, five different sub-wavelength assignment policies are analyzed and compared against their integer linear programming. Due to the complexity and unscalable runtime of ILP, even on a small network and for a very few number of connection requests, we also propose two heuristic algorithms for two of the policies. The algorithms are evaluated in a static traffic scenario wherein all logical connections are known in advance. And third, we propose a practical implementation of the sub-wavelength assignment. The assignment is fulfilled from coarser to finer granularity by means of a dual stage PCE and time-slot computation engine. Finally, simulation results are conducted on two different network topologies to assess the performance of the algorithms under dynamic traffic, i.e., connections arrive into the system based on a stochastic process. Throughout the analysis, we assume the wavelength-continuity constraint on the network, hence connection requests, when established, make use of the same wavelength for all the links along the path from the source to the destination node.

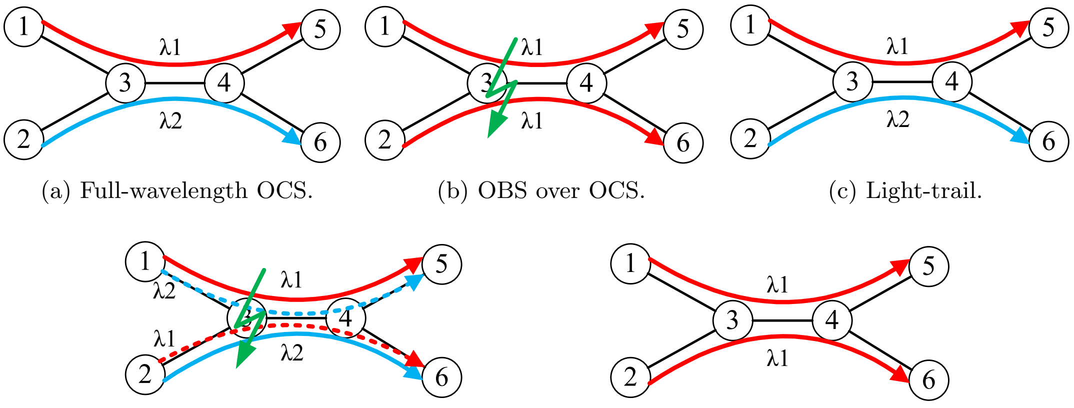
The remainder of this chapter is as follows: Section 5.2 reports on some of the work related with this chapter contributions. Section 5.3 presents an evaluation of different architectures to realize the time-shared optical network and highlights the advantages of the chosen one. In Section 5.4 we formally establish the sub-wavelength time-shared scheduling problem and derive its mathematical formulation for different scheduling sub-problems. Section 5.5 introduces two of the policy heuristics and evaluates their performance for a static traffic scenario. Section 5.6 presents the sub-wavelength RWA heuristics used within the proposed network architecture and Section 5.7 shows their simulation results for the dynamic traffic case. Finally, Section 5.8 summarizes the main findings from this chapter.

5.2 Related Work on Sub-Wavelength Network Provisioning

The provisioning of sub-wavelength optical network services is not new. We highlighted in Chapter 2 some of the solutions proposed in the literature. It is our purpose to review more in detail some of these architectures in order to emphasize the benefits of the proposed architecture.

Let assume two different sub-wavelength connections requesting for 4 Gbps of bandwidth each go from nodes 1 to 5 and from node 2 to node 6. 10 Gbps is the total wavelength capacity. Fig. 5.2(a) shows the full-wavelength OCS case for this particular example. In such a case, to insure the connectivity, two different wavelengths are required (λ_1 and λ_2). Full-wavelength OCS has the advantages that it is a proven and mature technology with available standardized equipment. However, it performs poorly when dealing with sub-wavelength connections. Also, electrical grooming requires to increase investment on equipment which raises the operational cost per port.

One of the first solutions for the provisioning of sub-wavelength on hybrid OBS/OCS scenarios is OBS over OCS [89, 91]. The OCS control plane is used to define the optical logical topology, and on top of that, sub-wavelength provisioning by means of OBS is



(a) Full-wavelength OCS.

(b) OBS over OCS.

(c) Light-trail.

(d) LOBS-H.

(e) TSON.

Fig. 5.2: Alternative architectures.

enabled. One of the advantages of this approach is the joint benefit of allowing the network managers to do a TE network planning at the OCS layer while at the same time having the flexibility of the short-live connectionless nature OBS. However, it also presents some disadvantages, namely: higher OPEX due to management of the two networks (i.e., OBS + OCS), and the absence of guaranteed burst transport, as contentions can still occur depending on how the virtual mesh topology is created. As shown in Fig. 5.2(b), a single wavelength would suffice to provision the required channels. However, collisions may still happen due to sharing the same wavelength between nodes 3 and 4.

Light-trails are another option for provisioning sub-wavelength channels across the network [94]. This architecture establishes time-shared wavelength buses with passive add/drop elements which provide spatial grooming. One of the benefits is the low-cost implementation with current optical enabling devices. Connections do not require nodal reconfiguration and the sub-wavelength provisioning is done multipoint-to-multipoint along the unidirectional wavelength bus. Nevertheless, we can name three main disadvantages. Firstly, this approach only provides time-shared wavelengths for a single lightpath. Secondly, it also exposes higher node architecture complexity when upgrading to a light-mesh topology. And finally, light-trails also require the arbitration and slot allocation for each burst transmission. As is, this architecture requires to use two different wavelengths to provide the same level of connectivity as introduced in the example (refer to Fig. 5.2(c)).

Finally, LOBS-H [87, 88] allows sharing wavelengths for in-home traffic and off-home-circuit connections. That is, lightpath home-circuits are established across the network and the bursts transmitted over them (in-profile traffic) are guaranteed no collision. Other traffic from other source-destination pairs can make use of the same wavelength. However, in this case bursts are marked as out-profile traffic, and upon contention with in-profile traffic, the former bursts are dropped. Therefore, this architecture cannot guarantee the burst delivery for off-home-circuit traffic. For full-guaranteed bandwidth provisioning, this architecture would require to set up two different LOBS-home circuits, as depicted in Fig. 5.2(d).

The most desirable architecture would be the solution seen in Fig. 5.2(e). Since the bandwidth requested by the two connections is still achievable on the common link between nodes 3 and 4, these could share the same wavelength. Enabling efficient sub-wavelength provisioning may improve the network resources utilization. This is the objective to address throughout this chapter.

We note that the proposed TSON architecture has similarities with the so-called time-driven switching (TDS) networks [99]. TDS applies both time division multiplexing and frame structures to provide end-to-end sub-wavelength circuits. Time is divided into time frames of equal size grouped in time cycles. On setting up a connection between a source and destination node, a free time frame is searched in the cycles associated to each link along the path. If found, the time frames are reserved, and this sequence form a synchronous virtual pipe. To make this possible, the whole network must be synchronized to recognize the time division, which in the TDS case, the authors assume a global system provides absolute time reference with high accuracy, e.g., form a GPS system.

An example of TDS networks are fractional lambda switching (FλS). In FλS [100], the capacity of an optical carrier is divided into a larger number of sub-channels based on time frames and the data content of each time frame is independently switched using TDS techniques.

A key contribution of this chapter from TDS is the formal definition of the RWTA scheduling policies and their integration in a realistic and standardized path and sub-wavelength computation architecture.

5.3 PCE/TSON Architectures

The aim of the proposed network architecture is to provide sub-lambda services on PCE/TSON control plane interworked networks with lossless data transmission. This means that blocking on the network is limited to the sub-wavelength connection provisioning phase instead of the actual data burst transport. This is possible because the data transmission is allocated the time-slots in a way that do not collide with any of the pre-established sub-wavelength connections. Moreover, by enabling all-optical time-shared sub-wavelength allocation, complexity and cost per port of the network nodes is decreased considerably, which results in attractive CAPEX cost reduction benefits for the network operator.

We propose an augmented control plane interworking approach. In this design, different control plane instances run on each layer but some information is exchanged between them aiming at improving the network bandwidth allocation. It ensures an easier compatibility with different underlying transport technologies, which is especially beneficial in the case the TSON network would be interconnected with other networks, either circuit or packet-based. In addition, it allows for multi-domain interoperability since the upper-layer path computation (e.g. PCE) is standardized. Also, in two of the proposed architectures, a full implementation of the GMPLS stack runs on top of the TSON network to signal the sub-wavelength lightpaths on the network.

It is very important to differentiate the tasks assigned to each control plane and the level of cross-layer information in order not to overburden the network with duplicated functions. As stated in [89], tasks can be classified between sub-wavelength long and short time-scale, and between sub-wavelength signaling and routing.

In order to enable interoperability across domains, we envision the use of a two-tier sub-lambda assignment. On the top level, the management-based PCE [137] is responsible for computing the routes and wavelengths (or only routes, depending on the specific sub-wavelength policy) that provide enough bandwidth to allocate the IP client service. Alongside, OSPF-TE manages the network topology and resources availability information dissemination for high-level path computations. Later, the sub-lambda assignment element (SLAE) processes the list of routes/wavelengths from the PCE to finely compute and assign the time-slot labels in a contention-free basis. The availability of sub-lambda resources is aggregated and abstracted per port and wavelength to the traffic engineering databases (TED) periodically or asynchronously when processing the sub-wavelength connection setup.

5.3.1 Time-shared Optical Network Data Plane

In TSON, the optical channel is time-shared, i.e., the raw wavelength bandwidth per link is divided in time-slots, similarly as defined in [99]. These slots are then allocated to different connections based on the operation of the RWTA implemented by the PCE-SLAE sub-wavelength assignment engine.

The sub-wavelength allocation is composed of a two-tier slot-frame structure as shown in Fig. 5.3. The minimum assignable piece of bandwidth is the time-slot. Depending on the time-slot allocation, a set of continuous slots can be concatenated to form a bigger fragment (i.e., burst). The top framing tier is the actual frame. A frame is composed of a predefined number of time-slots. This number depends on the wavelength capacity and the minimum channel sub-wavelength allocation defined in TSON.

In the selection of the time-slot size, the actual requirements of the optical network equipment (e.g., optical transceivers, optical cross-connects, etc.) must be met. As a matter of fact, we cannot define a shorter time-slot (in time) than the OXC switching configuration set up time. For instance, throughout the performance evaluation we

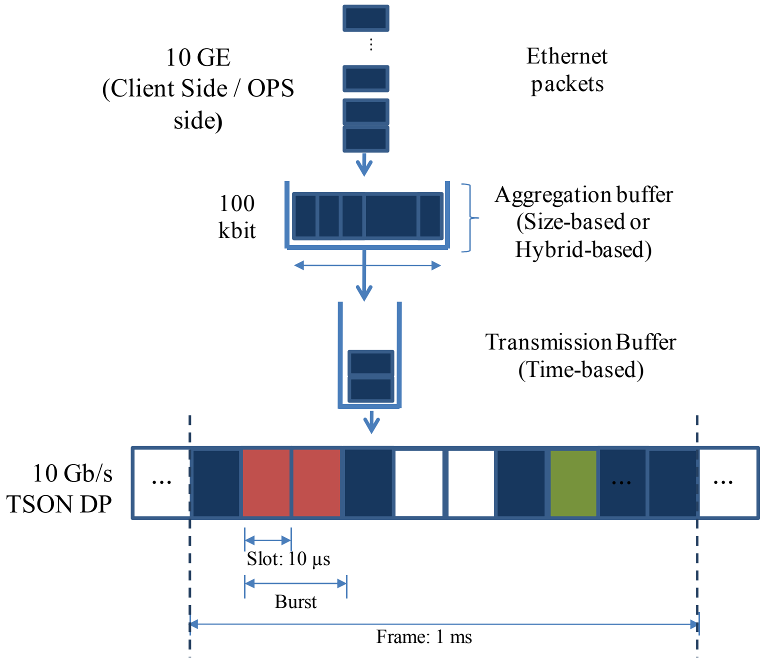


Fig. 5.3: Client frame aggregation.

have assumed a wavelength capacity of 10 Gb/s and a minimum sub-wavelength channel capacity of 100 Mb/s, with a time-slot length of 10 μs assuming the availability of OXCs with a configuration and switching time in the order of 200 ns. With this configuration, the number of slots that form the frame is $F = \frac{10 \text{ Gb/s}}{100 \text{ Mb/s}} = 100$. The frame structure repeats over time unless new connections are dynamically removed or allocated.

Fig. 5.3 also shows the aggregation of the client packets for its transmission into the allocated slots. In this particular example, we assume a 10 GE interface from the client port. Packets are aggregated into a buffer. The aggregation or assembling of packets can be size-driven or hybrid-based. In the latter, data is released from the buffer when a timer expires even if the intermediate buffer is not full. The assembled data bursts are transferred to the intermediate transmission buffers. From these buffers, bursts are dropped into the corresponding time-slots in the final transmission stage.

Next, we describe the main characteristics of the proposed sub-wavelength network architectures for guaranteed bandwidth provisioning. The first two architectures are based on an augmented GMPLS/PCE/TSON control plane interworking. The latter is simpler and only involves the interworking of PCE/TSON control planes.

5.3.2 Daisy-chain Centralized SLAE Time-shared Optical Network

The daisy-chain centralized SLAE Time-shared Optical Network (DC-TSON) architecture (see Fig. 5.4) is characterized by having a centralized PCE on the GMPLS layer with visibility of the per-wavelength utilization and a centralized SLAE on the TSON layer. The SLAE has a complete view of the time-slot utilization on all the links across the network. Using this information, the SLAE is able to compute non-blocking slot allocation for the bursts, hence subsequently avoiding burst contention in the core of the network. As a result, blockings are only produced at the sub-lambda connection phase.

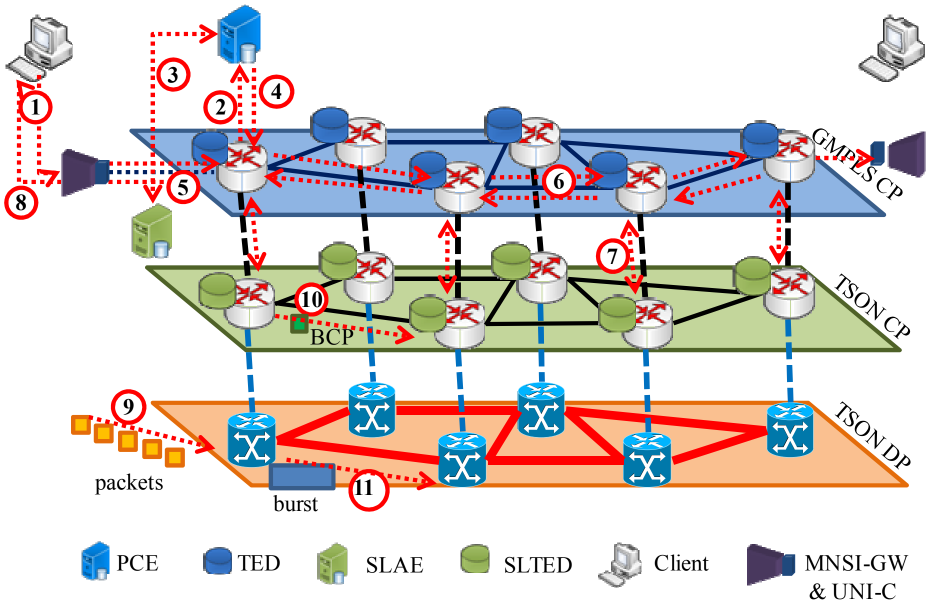


Fig. 5.4: DC-TSON architecture: connection request setup.

As shown in Fig. 5.4, the connection request setup involves the following steps. The service gateway (MNSI-GW & UNI-C)[1] redirects the connection to the ingress GMPLS node (step (1)) which requires the computation of the RWA and label assignment (2). The latter is realized jointly by the PCE, which computes the initial list of routes and wavelengths, and the SLAE, which is responsible for making the final finer time-slot and label assignment (steps (3) and (4)). With this information, the sub-lambda lightpath two-way signaling (i.e. for unidirectional lightpath) can be initiated along the chosen path (steps (5) and (6)). This process involves the cross-layer communication between the GMPLS and TSON CP at every node (7) to verify the label slot availability with the local SL-TED. Later, after the MNSI-GW acknowledges the client interface (8), the packet traffic flow (9) is properly assembled at the ingress TSON node and transmitted in the network (11) in the form of bursts for the assigned amount of time-slots (eventually also with their corresponding burst control packet (10) if automatic signaling at intermediate nodes is not active). The reason of operating RWA and slot assignment functions independently is for deploying sub-wavelength services with an upper standardized GMPLS/PCE-based control layer that is able to interoperate with other domains (technological, operational and/or geographical).

[1] The description of the service gateway is out of the scope of this Thesis. For more information, please refer to project MAINS [138].

5.3.3 Daisy-chain Distributed Time-shared Optical Network

The main difference of the daisy-chain distributed SLAE Time-shared Optical Network (DD-TSON) (see Fig. 5.5) with respect to the previous one shown in Fig. 5.4 is the relocation of an SLAE into each TSON node with local information stored in the SL-TED. Therefore, the computation of the sub-wavelength assignment (time-slots) is carried out distributively from the list of suggested labels computed by the ingress node.

The SLAE at the ingress node computes a list of suggested/candidate labels (see step (3) in Fig. 5.5) based on PCE pre-computed paths and wavelengths that fulfill the sub-lambda requirements. Later, the candidate labels on the RSVP-TE PATH message are validated at each node along the lightpath (5). It is worth noting that this solution can be susceptible to a higher blocking rate since the suggested/candidate label usage from the source node may not be possible on some of the remaining path network nodes.

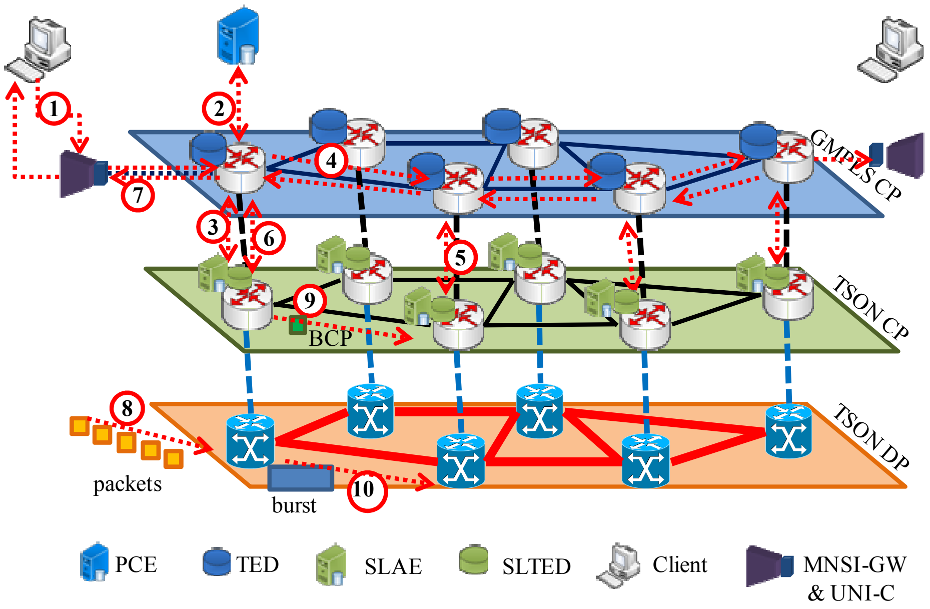


Fig. 5.5: DD-TSON architecture: connection request setup.

5.3.4 Tree-based Time-shared Optical Network

The last proposed architecture is the tree-based centralized Time-shared Optical Network (TC-TSON) (see Fig. 5.6). TC-TSON is characterized by having a centralized PCE on the top layer, with visibility of the per-wavelength utilization, and a centralized sub-lambda assignment engine on the TSON layer with a complete view of the time-slot utilization on all the links of the network. Using this information, the connection can be established avoiding contentions in the core of the network since the SLAE is able to compute non-blocking time-slot allocation.

In this architecture, the TSON CP and in particular the SL-TED are responsible for updating the local TED of the PCE entity with aggregated info of the sub-wavelength utilization. The local SL-TEDs keep track of all the time-slot available resources per port and wavelength in the network node. The centralized SLAE also has a network-wide global SL-TED updated from its own scheduling, which is also cross-checked with the label information from lightpath setup messages and TSON node SL-TED periodic updates.

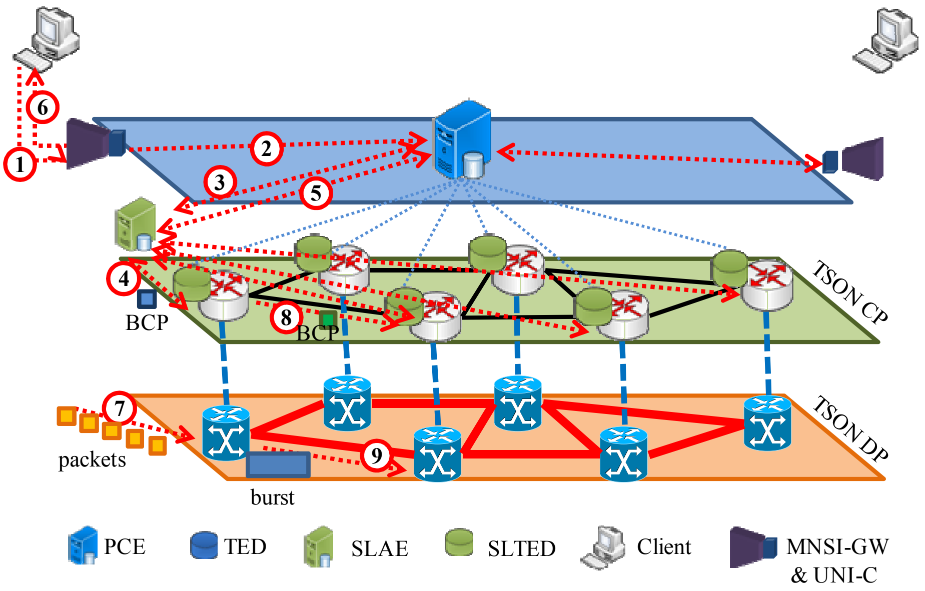


Fig. 5.6: TC-TSON architecture: connection request setup.

Fig. 5.6 shows the connection request and sub-lambda lightpath set up (one-way, no bidirectional). The client connection request is directed through the network to the service interface gateway (MNSI-GW) (see step (1) in Fig. 5.6). From the MNSI-GW, the connection is forwarded to the PCE which is responsible for the computation of a list of paths and wavelengths. Specifically, the latter is realized together by the PCE and SLAE (steps (2) and (3)). The SLAE generates the labels of the computed sub-wavelength LSP.

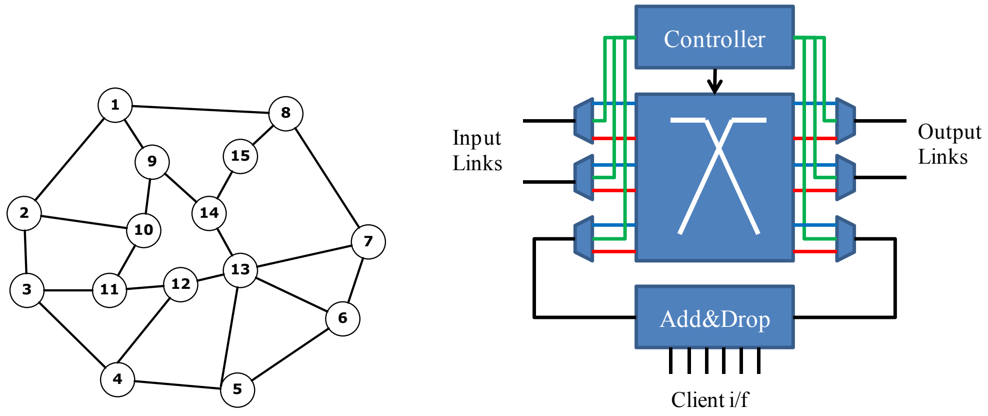
In the tree-based centralized architecture, a master TSON node is responsible for signaling the lightpath to the rest of nodes involved in the reservation. For simplicity, this node can be hold together with the SLAE and even be shared in one of the TSON nodes. In fact, this is the approach we have utilized in the simulation implementation. Provided that, the SLAE signals to the sub-wavelength-capable network node control plane the assigned time-slots for the upcoming connection (4), i.e., the previously computed labels. Two reservation signaling mechanisms are possible: one-way or two-way. Using one-way signaling, setup delay can be decreased to the longest propagation delay between the master node and the farthest node that belongs to the lightpath. Two-way signaling doubles the setup delay, but it adds an extra layer of reliability by acknowledging the master node the successful label reservation and configuration from each of the nodes along the lightpath. Once the setup is finished, the PCE is acknowledged about the connection (5), and through the MNSI-GW, the client node gets the connection response (6). At this stage, the client node is able to start transmitting the packets, which are enqueued and assembled at the ingress node and transmitted on the allocated slots.

5.3.5 Simulation Performance GMPLS/PCE/TSON Architectures

We analyze in this section the performance of the proposed GMPLS/PCE/TSON and PCE/TSON interworking architectures. An event-driven JAVA-based simulator was implemented for testing purposes. Fig. 5.7(a) shows the topology we considered in this performance evaluation. It is based on Madrid's regional metro network and composed of 15 nodes, 23 bidirectional links, with a nodal degree connectivity of $\overline{deg(G)} = 3.07$, an average link length $L = 56.87$ km and length standard deviation $\sigma_L = 41.70$ (i.e., some links are much longer than others).

In such a simulation scenario, we consider 16 wavelengths with 10 Gb/s per channel. All nodes on the network are both edge and core, hence apart from forwarding traffic from other network nodes they can also add and drop traffic for themselves. Fig. 5.7(b) shows the node architecture. We assume in all the examples that sub-lambda lightpaths are subject to the wavelength and time-slot continuity constraint. A slot and frame size of 10 μs and 1 ms is used, respectively. Moreover, for the sub-wavelength connection requests we use a single-wavelength first-fit continuous heuristic running on the two-tier PCE-SLAE routing, wavelength and time-slot assignment engine. These algorithms will be introduced more in detail later in Section 5.6. Finally, regarding the setup of the devices, in the daisy-chain signalling architectures (GMPLS/PCE/TSON, noted as DC-TSON and DD-TSON) we assume a message processing time of 10 ms or 5 ms (both results are plotted) for the RSVP-TE messages, while in the third one (PCE/TSON, TC-TSON), due to the simplicity of the tree-based signaling system and use of hardware-based solutions (FPGA and embedded processors), the message processing delay is lowered to 10 μs.

With respect to the traffic characteristics, connection arrivals are based on a Poisson process with rate λ and an exponential holding time with mean $1/\mu = 60$ s. For simplicity, the destination of the connection is uniformly distributed to all the remaining nodes on the network. In order to assess the sub-lambda assignment performance for different connection demands, in the simulations we have considered two different sub-lambda traffic scenarios. In the former, connection requests are of a granularity equal to 1 Gb/s (i.e. sub-lambda LPs request this amount of bandwidth), while in the second,



(a) (b)

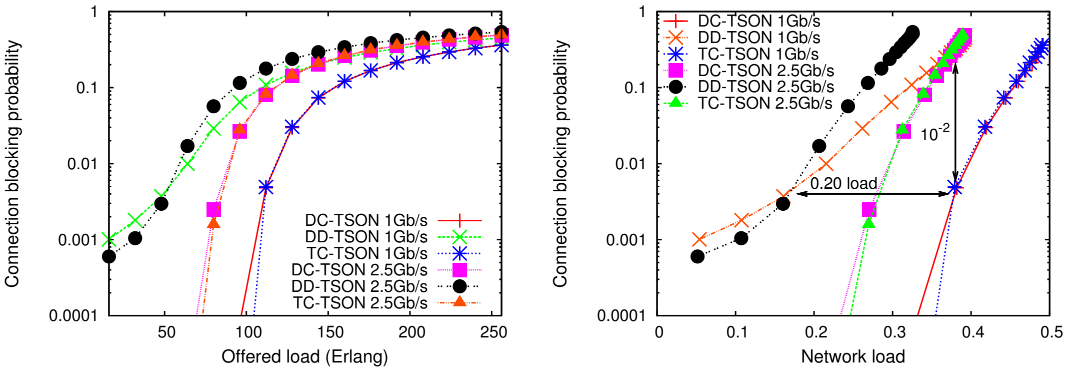
Fig. 5.7: (a) Madrid's network topology, and (b) network node architecture used throughout simulation.

the sub-wavelength requests are of 2.5 Gb/s. Results are gathered using the batch means method over simulation runs of 6.0E5 sub-lambda connection requests. 95% confidence intervals were also obtained, but since they are quite narrow, they have been omitted in order to improve the readability of the graphs.

Fig. 5.8 shows the connection blocking probability: Fig. 5.8(a) as a function of the offered load to the network in Erlang and Fig. 5.8(b) as a function of the network load (computed per link and wavelength). As expected, the GMPLS/PCE/TSON architecture with distributed SLAE (DD-TSON) shows the worst performance due to its distributed sub-lambda assignment with localized information (e.g. two orders worse at the same network load). The other two architectures, GMPLS/PCE/TSON with centralized SLAE and PCE/TSON, get almost identical results since both use the same sub-lambda algorithms running on the centralized SLAE.

Recall that in the DC-TSON, the SLAE marks the given time-slots as booked when the labels are assigned, hence avoiding upcoming connections to book over the same slots and generate backward lightpath setup blockage. This also explains the almost non-existent differences when the RSVP-TE message processing changes from 10 ms to 5 ms, as we will see later on. Also, when the sub-lambda connection bandwidth is increased from 1 Gb/s to 2.5 Gb/s, the successful allocation of such connections becomes harder and saturation points are reached sooner as shown in Fig. 5.8(a). In particular, at low loads, 1 Gb/s connections present for DD-TSON a slightly higher blocking probability in comparison with 2.5 Gb/s due to the higher arrival rate necessary to generate the same offered load, which penalizes the distributed SLAE assignment. It has to be noted that if cranckback techniques are applied to DD-TSON the blocking probability will possibly improve at the expense of longer time-to-service delay.

To complement the previous analysis, Fig. 5.8(b) shows the same blocking probability but now as a function of the actual network load. In this case we can easily



(a)

(b)

Fig. 5.8: Performance results comparison on a metro network with 16 wavelengths: (a) mean connection blocking probability vs. offered load, and (b) mean connection blocking probability vs. network load.

notice that increasing the size of the connection request –from 1 to 2.5 Gb/s– results in a lower network utilization ratio at the same offered load due to the less chances to allocate resources to higher-bandwidth demanding requests. At a network load over 0.4, the blocking probability is greater than 30 absolute percentage points. Only at low-medium loads the performance for the tree-based signaling architecture is slightly better, which is directly related to its better time-to-connection performance (as shown later).

Fig. 5.9 shows the results of the mean lightpath setup delay. As we introduced, the message processing delay is different between the GMPLS-based and PCE-only-based architectures. The graph reveals that by reducing the message processing delay from 10 to 5 ms, this time becomes the main factor for the RSVP-TE connection setup. However, it is worth noting that for the TC-TSON, this delay is almost fixed whichever the offered load to the network, which ensures a bounded time-to-service delay. As a consequence, the TC-TSON approach not only can deliver the same blocking probability as the DC-TSON, and better compared to DD-TSON, but also an almost constant time-to-service delay. In the worst case scenario of increasing the processing delay in TC-TSON an order of magnitude, the setup delay would still remain lower than in the rest of architectures.

Another interesting value to represent is the mean lightpath length (in number of hops) as shown in Fig. 5.10(a). All architectures manage to provide similar values when dealing with the same type of connection requests. Moreover, when the connection bandwidth request is increased, the mean lightpath length is smaller in comparison to the case 1 Gb/s scenario. This value is related to the fact that the allocation of resources over shorter lightpaths has more chances to be successful. Also, DD-TSON experiments shorter paths compared to the other two architectures. This is due to the higher blocking as a result from the distributed SLAE computation and signaling. The

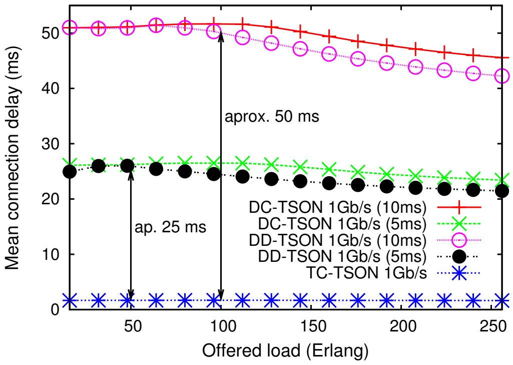


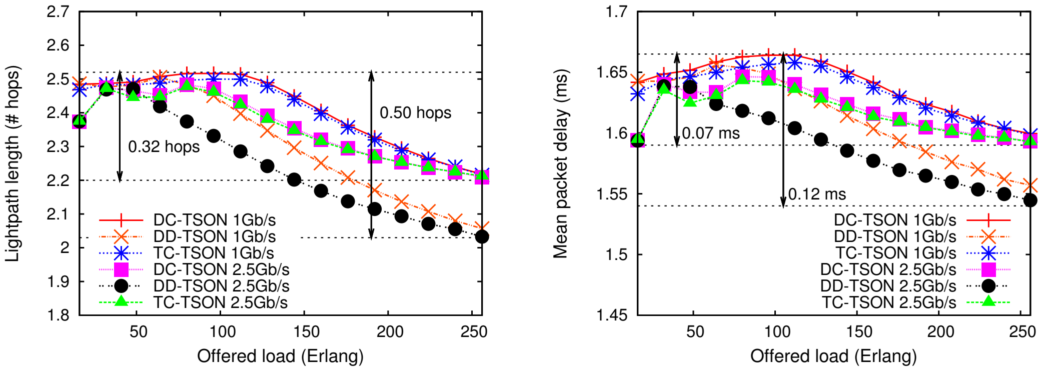
Fig. 5.9: Performance results comparison on a metro network with 16 wavelengths: mean connection setup (time-to-service) delay.

number of hops can help explain the mean packet end-to-end delay (see Fig. 5.10(b)), which follows the same patterns and is bounded to 1.85 ms for the network topology under test. As the blocking probability is increased at higher loads, the lightpaths that require less number of hops get a higher success setup ratio, hence the average packet delay decreases, e.g., 0.07 ms decrease for DC-TSON.

Finally, the add port utilization vs. offered load of each TSON node is illustrated in Fig. 5.11. As in the blocking probability results both the type of architecture and the BW amount per request are the main factors that affect the throughput of the node. For 1 Gb/s connection requests, DC-TSON and TC-TSON deliver the highest node normalized port utilization ratio, about 70% of the add port capacity at 250 Erlang. DD-TSON can deliver ∼ 60%, 10% less than centralized SLAE architectures. The node throughput in case of having 2.5 Gb/s connection requests for DC-TSON and TC-TSON is ∼ 55% of the add port capacity.

In summary, results show that although the GMPLS/PCE/TSON architectures (DC-TSON and DD-TSON) can be more complex in terms of control plane implementation, they also meet the requirements of a standardized control plane for the provisioning of end-to-end lightpaths across multiple domains and switching technologies. On the other hand, the PCE/TSON architecture (TC-TSON) not only achieves the same blocking performance as the centralized DC-TSON, but also guarantees a smaller time-to-service delay due to its lower complex tree-based signaling.

Hereafter, we will gain more insight into the TC-TSON architecture and assess several other parameters of importance like the time-slot assignment to ensure guaranteed end-to-end data delivery for accepted connection requests. As such, we will first formulate the routing, wavelength and time-slot assignment problem optimization. Later, we will propose several heuristics to cope with the time-consuming ILP optimization and evaluate these on both static and dynamic traffic scenarios.



(a) (b)

Fig. 5.10: Performance results comparison on a metro network with 16 wavelengths: (a) mean lightpath length vs. offered load, and (b) mean packet end-to-end delay vs. offered load.

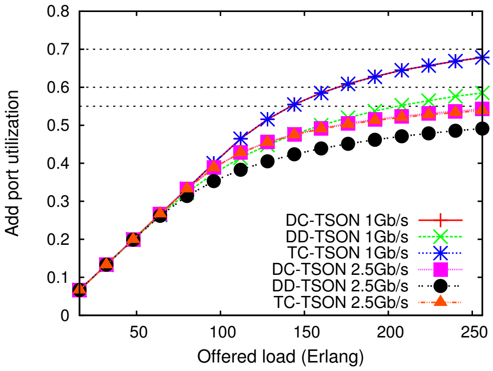


Fig. 5.11: Node's normalized add port utilization.

5.4 Sub-Wavelength Assignment General Problem Statement

In this section, we formally define the sub-wavelength allocation problem for a given set of unicast connections from an origin to a destination node. Let assume the following problem scenario:

- A physical topology represented by a weighted undirected graph $G = (V, E)$. V is the set of network nodes and E is the set of link between nodes. Each link is assigned a weight (e.g., the physical distance between the corresponding node pair, or some other parameter).
- The set of wavelengths on the network, which is symbolized by W.
- A group of c sub-wavelength lightpaths, B, sessions defined by their index and burst/capacity demand (i.e., number of slots).
- A matrix/frame with a specific sub-wavelength horizon that represents the pool

of resources available on each link. This is represented by M_{ij}. The number of columns of the matrix is given by the sub-wavelength frame horizon (s slots) and the number of rows equals the number of wavelengths, the cardinality of W, $|W|$.

The goal is to set up all c sub-wavelength lightpaths on the given sub-wavelength network resource topology G, while minimizing the overall cost –a function cost to be defined– or maximizing other parameters such as number of scheduled connections or overall add port throughput. This problem can be considered a particular case of graph coloring problem, and therefore, it is of type combinatorial NP-complete [139].

Several policies can be followed while scheduling the burst connections within the network. Some of these may constrain the scheduling options of the sub-wavelength lightpath, hence decreasing the availability of resources and consequently diminishing their successful scheduling completeness. Five are the main constraints we can consider: (1) multi-wavelength scheduling, (2) single wavelength scheduling, (3) wavelength tunable[2] scheduling, (4) non-continuous slot/time assignment, and (5) continuous slot/time assignment. Provided that, we can infer five main policies. Fig. 5.12 exemplifies each one of the policies considered in this analysis. We assume an arriving connection requests 5 time-slots. The slots assigned by each policy are marked by a grid shading pattern.

1. Multi-wavelength non-continuous time-slot assignment (MW-NC). Any allocation of time-slots is possible, on any wavelength, at any time-slot within M_{ij} (refer to Fig. 5.12(a)).
2. Single wavelength non-continuous time-slot assignment (SW-NC). The connection can only use a single wavelength, but assigned time-slots do not need to be continuous (Fig. 5.12(b)).
3. Single wavelength continuous time-slot assignment (SW-C). The same as SW-NC, but time-slot continuity is enforced (Fig. 5.12(c)).
4. Tunable wavelength non-continuous time-slot assignment (T-NC). It works similarly as MW-NC, but the same slot index cannot be used on different wavelengths (Fig. 5.12(d)).
5. Tunable wavelength continuous time-slot assignment (T-C). It applies the same T-NC concept, but assigned slots have to be continuous in time (Fig. 5.12(e)).

For simplicity, we will take the case where the network resources are time-sliced in slots and time-shared connections have to be allocated within the matrix/frame horizon. A burst can take one or more slots depending on the requested sub-wavelength granularity.

[2] In this case, tunable refers to the capability to allocate time-slots on different wavelengths along the frame horizon. Recall that tunable can also refer to the physical transceiver capabilities to tune the laser to different lambdas within the wavelength grid.

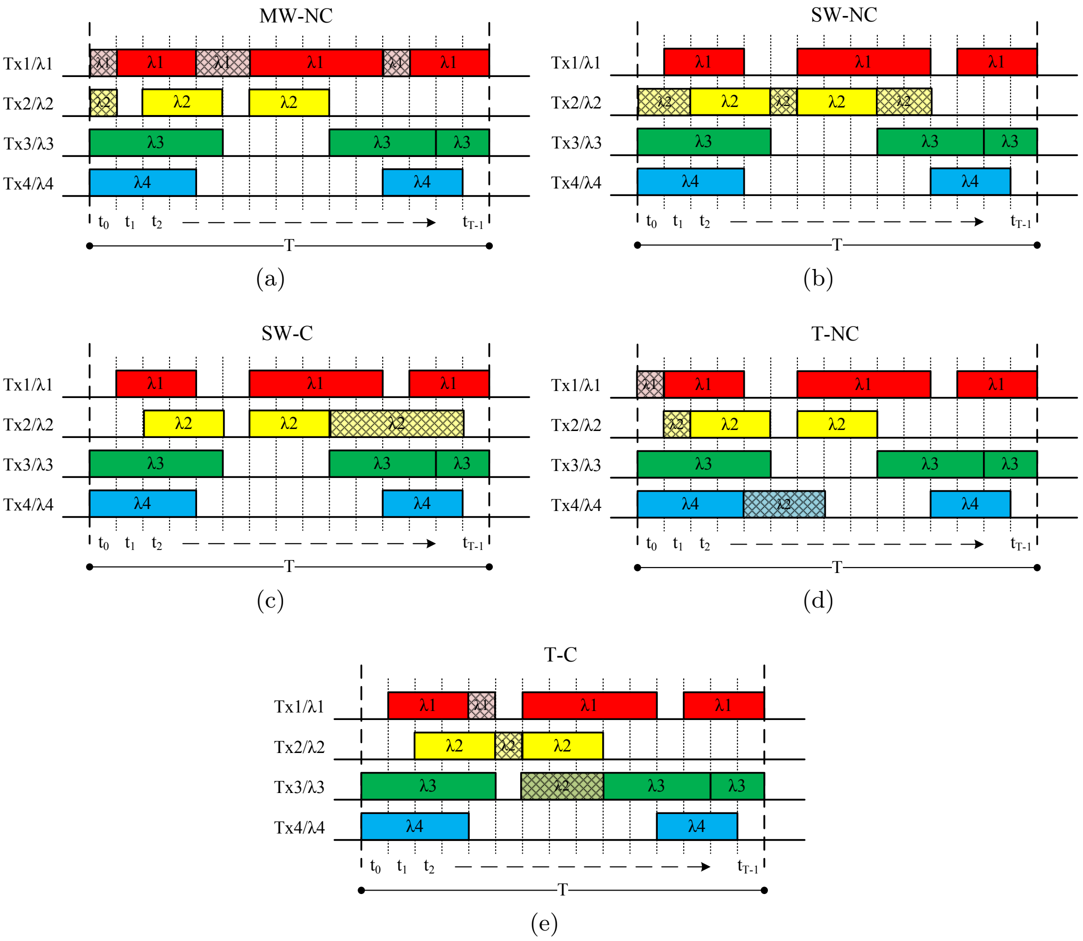


Fig. 5.12: Time-slot assignment policies: (a) MW-NC, (b) SW-NC, (c) SW-C, (d) T-NC, and (e) T-C.

5.4.1 Problem Formulation

We start by defining the variables, parameters and general constraints that are used in all the assignment problems. To such end, we will make use of the commodity-flow formulation to represent the establishment of the lightpaths and their sub-wavelength capacity demand as commodities. The basis of the present work is [127], which serves the purpose of presenting the nomenclature, which is further extended for the sub-wavelength assignment policies.

Given the following input parameters:

- V is the set of nodes on the network. The number of nodes on the network is $|V|$.
- W is the set of wavelengths per fiber. The number of wavelengths is its cardinality, $|W| = w$.
- P_{mn} is the physical topology, where the fiber links are assumed to be bidirectional, that is $P_{mn} = P_{nm} = f$, where f is the number of fibers per link. For simplicity, we will assume $f = 1$ if there is connectivity. If there is no connection between

nodes m and n, $P_{mn}=P_{nm}=0$.

- s = number of slots in the matrix/frame.
- $M = [m_{ij}]_{|W|\times s}$ is the matrix that represents the link resources. Variables $1 \leq i \leq |W|$ and $1 \leq j \leq s$ are the matrix indices.
- $B = \{b_1, b_2, \ldots, b_c\}$ represents the group of sub-lambda connections. Each connection has a source, s_k, and a destination node, d_k. The number of connection requests to schedule is the cardinality of B, $|B| = c$. Variable $1 \leq k \leq |B|$ is the connection index.
- L_k denotes the length (number of slots) requested by connection k.
- Nodes do not have wavelength converters, hence connections, and their assigned slots, must follow the wavelength-continuity constraint.

The common variables of the problem are:

- C_k is a boolean variable denoting whether connection k is scheduled. If scheduled, $C_k = 1$, 0 otherwise.
- m_{mn}^{ijk} is a boolean variable used to denote if slot m_{ij} of link P_{mn} is used by connection k. If so, then $m_{mn}^{ijk} = 1$, 0 otherwise.
- V_p^k is also a boolean variable to denote if node p belongs to lightpath connection k. If it is part of the lightpath, then $V_p^k = 1$, 0 otherwise. A node belongs to a connections if it is either source, destination or an intermediate node.
- L_{mn}^k is another boolean variable to denote whether link between nodes m and n is used by connection k. Since we restrict in this problem to setup unicast connections, the specific connection flow cannot be balanced over two different links sharing the same origin link node.
- F_{mn}^k defines the commodity-flow variable. Specifically, it represents the number of commodities of connection k flowing between nodes m and n.

Different optimization objectives can be defined for the problem. For instance, to minimize the total cost for the lightpath setup on the network,

$$\text{minimize} \sum_k \sum_i \sum_j \sum_{m,n} w_{mn} \cdot m_{mn}^{ijk} \tag{5.1}$$

Another option can be to optimize the number of scheduled time-shared lightpaths regardless of their length and route followed, so,

$$\text{maximize} \sum_k C_k \tag{5.2}$$

or in order words, minimize the blocking probability given both network resources and connections.

The following sections describe the formulation for the specific assignment policies we have considered (see Fig. 5.12): multi-wavelength non-continuous (MW-NC), Single-wavelength non-continuous (SW-NC), single-wavelength continuous (SW-C), tunable non-continuous (T-NC) and tunable continuous (T-C). As the reader will see, the formulation of the policies is realized from lower to higher level of constriction. Specifically, we will see that the simplest case is the MW-NC scheduling policy, which requires the least number of constraints and variables[3].

5.4.2 Multi-Wavelength Non-Continuous Policy

The multi-wavelength non-continuous problem is the least restrictive time-slot assignment policy. In MW-NC, the slots pertaining to a sub-wavelength lightpath can be scheduled all over the link resources matrix. Different wavelengths and not necessarily within a continuous slot set can be used, including wavelengths on the same slot index (multi-wavelength).

In MW-NC, if we focus on optimizing the number of connections (minimize the global blocking probability), the solution of the problem is subject to the following general constraints:

- Lightpath routing and general constraints:

$$\forall k : C_k \leq 1 \tag{5.3}$$

$$\forall k, m, n : L^k_{mn} \leq C_k \tag{5.4}$$

$$\forall k, \forall n \notin \{s_k, d_k\} : \sum_{m,i,j} m^{ijk}_{mn} = L_k \cdot V^k_n \tag{5.5}$$

$$\forall k, \forall m \notin \{s_k, d_k\} : \sum_{n,i,j} m^{ijk}_{mn} \leq L_k \cdot V^k_m \tag{5.6}$$

$$\forall k, \forall n \notin \{s_k, d_k\} : V^k_n \leq C_k \tag{5.7}$$

$$\forall k, \forall n \in \{s_k, d_k\} : V^k_n = 1 \tag{5.8}$$

$$\forall k, \forall n = s_k : \sum_{m,i,j} m^{ijk}_{mn} = 0 \tag{5.9}$$

$$\forall k, \forall m = d_k : \sum_{n,i,j} m^{ijk}_{mn} = 0 \tag{5.10}$$

$$\forall k, \forall m = s_k : \sum_{n,i,j} m^{ijk}_{mn} = L_k \cdot C_k \tag{5.11}$$

$$\forall k, \forall n = d_k : \sum_{m,i,j} m^{ijk}_{mn} = L_k \cdot C_k \tag{5.12}$$

[3] We have opted for keeping uniformity throughout the formulation of the different policies. In doing so, it may appear the number of variables and constrains is too large in some cases.

$$\forall m, n : \sum_{i,j,k} m_{mn}^{ijk} \leq P_{mn} \cdot w \cdot s \tag{5.13}$$

$$\forall m, n, i, j : \sum_{k} m_{mn}^{ijk} \leq P_{mn} \tag{5.14}$$

$$\forall m, n(n > m), \forall k : \sum_{i,j} m_{mn}^{ijk} + \sum_{i,j} m_{nm}^{ijk} \leq L_k \cdot C_k \tag{5.15}$$

Next, we explain the meaning of all these constraint equations. Equation (5.3) constrains the connection scheduling variable to be either 0 or 1, while (5.4) makes sure that the L_{mn}^{k} value of link between m and n used by a lightpath is constraint by the scheduling of the actual lightpath. From (5.5) and (5.6), a node different from source and destination belongs to the lightpath if the sum of incoming and outgoing slots is equal to the sub-wavelength lightpath capacity demand. Moreover, as stated by (5.7) all intermediate nodes can or cannot be part of the lightpath, while source s_k and destination d_k nodes always are part of it (5.8). Equations (5.9) and (5.10) ensure that no slots are flowing into a source node or out from the destination, respectively. In addition, (5.11) and (5.12) ensure that if the connection is scheduled, the number of slots on outgoing and incoming links is the connection slot capacity. By (5.13) we restrict the total number of scheduled slots per link to its maximum capacity, which is $P_{mn} \cdot w \cdot s$. With (5.14) we restrict the maximum number of scheduled sessions per slot to be P_{mn}, that is, 1 if there is link connectivity between two nodes. Finally, (5.15) ensures that the number of scheduled slots per bidirectional link for a single lightpath is the connection length (in number of slots).

In order to control the scheduling of the sub-wavelength connections and their corresponding number of slots, we define the following set of commodity-flow constraints.

- Commodity-flow constraints:

$$\forall k, \forall m \notin \{s_k, d_k\} : \sum_{n} F_{nm}^{k} = \sum_{n} F_{mn}^{k} \tag{5.16}$$

$$\forall k, \forall m = s_k : \sum_{n} F_{mn}^{k} = L_k \cdot C_k \tag{5.17}$$

$$\forall k, \forall n = s_k : \sum_{m} F_{mn}^{k} = 0 \tag{5.18}$$

$$\forall k, \forall n = d_k : \sum_{m} F_{mn}^{k} = L_k \cdot C_k \tag{5.19}$$

$$\forall k, m, n : \sum_{i,j} m_{mn}^{ijk} \leq F_{mn}^{k} \tag{5.20}$$

$$\forall k, m, n : F_{mn}^{k} \leq \sum_{i,j} m_{mn}^{ijk} \tag{5.21}$$

$$\forall k, m, n : F^k_{mn} \leq L_k \cdot C_k \tag{5.22}$$

$$\forall k, n, \forall m \neq d_k : F^k_{mn} = L_k \cdot L^k_{mn} \tag{5.23}$$

Equation (5.16) constrains that at an intermediate node, all flows that enter the node, must be forwarded to some output link, thus, intermediate nodes cannot hold commodities since they are not the final destination. Next two equations (5.17) and (5.18) define for each source node the number of commodities flowing out from the node must be equal to the length of the connection k if this is successfully established, and that no commodities can get back into the source node. Likewise, (5.19) controls that all commodities flow into the destination node, also if this connection is established. The next two commodity-flow equations (5.20) and (5.21) check that every link used by a lightpath connection has a positive flow and the rest of links not used by the lightpath have no flow. Additionally, (5.22) limits the flow through a link to the number of slots of such burst sub-wavelength connection k. Finally, (5.23) controls that all slots from a connection only take one output link, therefore, flows cannot be load distributed over different routes.

As for constraining the slot continuity along the lightpath, we need to set the following constraint. Recall that this characteristic is necessary since we assume that nodes do not have fiber delay lines (FDL) to delay the arrival of the slots. For simplicity, we will also assume that propagation delays are null, hence the same slot on the previous link is also used on the output link.

- Link slot continuity constraint:

$$\forall k, i, j, \forall n \notin \{s_k, d_k\} : \sum_m m^{ijk}_{mn} = \sum_m m^{ijk}_{nm} \tag{5.24}$$

In order to approximate the complexity of the problem formulation in number of variables and constraints we refer to: on the one hand, the number of unknown variables in this problem is $O(|B|+s|W||B||V|^2+|B||V|+2|B||V|^2) \sim O(s|W||B||V|^2)$, which comes from, principally, the size of the sub-wavelength link resources matrix, the number of connections and the network size. Thus, the number of variables increases linearly as a function of the number of wavelengths, connection requests and matrix slot length, and quadratically as a function of the network topology. On the other hand, the constraint complexity is bounded also by $O(s|W||B||V|^2)$ counting the number of constraints to satisfy, for instance, in (5.5) or (5.6).

5.4.3 Tunable Non-Continuous Policy

The T-NC policy introduces a new constraint: connections are restricted to use different time-slot indices, either over a single or multiple wavelengths, i.e., two simultaneous slots on different wavelengths are not allowed. Therefore, basically we have to add a new constraint to account for this behavior.

- Restricted wavelength-slot constraint:

$$\forall m, n, k, j : \sum_{i} m_{mn}^{ijk} \leq P_{mn} \tag{5.25}$$

Indeed, (5.25) constrains each connection to not use more than one slot at the same time-slot index, $j \in M$, on the same link between nodes m and n.

5.4.4 Single-Wavelength Non-Continuous Policy

The SW-NC policy is similar to T-NC but with the added constraint to schedule the slots for a sub-wavelength connection on a single wavelength. To this end, a new variable and a pair of constraints are added to the formulation. The extra variable is,

- C_k^i, of boolean type denotes whether sub-wavelength lightpath is scheduled on wavelength i.

And the single wavelength constraints related to this policy are the following.

- Single wavelength constraints:

$$\forall k : \sum_{i} C_k^i = C_k \tag{5.26}$$

$$\forall m, n(n > m), \forall k, i : \sum_{j} m_{mn}^{ijk} + \sum_{j} m_{nm}^{ijk} \leq L_k \cdot C_k^i \tag{5.27}$$

On the one hand, (5.26) ensures that if a connection is scheduled then it will only be on a single wavelength. On the other hand, (5.27) constraints all the slots of sub-wavelength connection k to be established on a single wavelength. Equation (5.25) is also used in this problem.

5.4.5 Single-Wavelength Continuous Policy

For the SW-C policy, a new constraint is introduced: the slot allocation for connection k needs to be continuous, hence the sub-wavelength (requested bandwidth) cannot be split along the matrix/frame horizon. To exemplify such approach, Fig. 5.13 shows four possible cases to consider when dealing with time-slot continuity. In the diagrams, the dashed arrows define the first assigned slot and the continuous arrow the ending slot. Two frames are shown to also account for the continuity across different frames. As we introduced, the frame structure is repeated over time unless new connections are set up or removed dynamically.

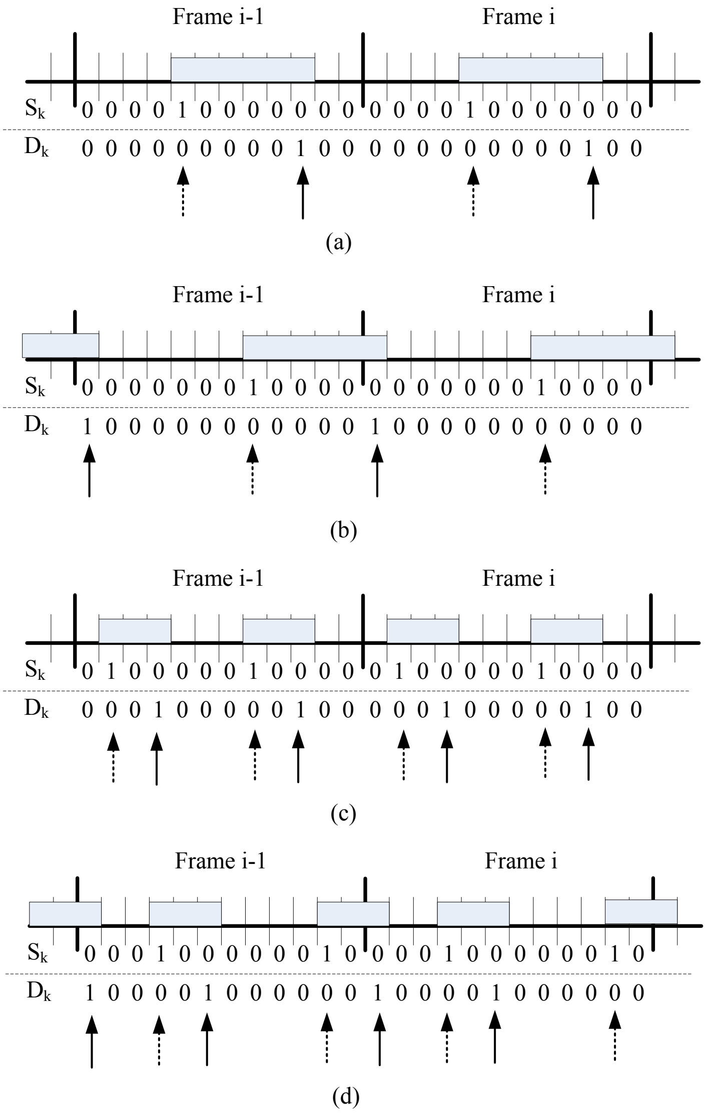

Frame i-1
Frame i
Sk 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
Dk 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0
(a)
Frame i-1
Frame i
Sk 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
Dk 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
(b)
Frame i-1
Frame i
Sk 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0
Dk 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0
(c)
Frame i-1
Frame i
Sk 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0
Dk 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0
(d)

Fig. 5.13: Continuous scheduling constraint.

Cases (a) and (b) illustrate a continuous slot scheduling. Even in case (b), the scheduling is continuous although it covers two different frames as these are time-merged (i.e., continuous in time). These two cases are allowed by the SW-C policy. On the contrary, cases (c) and (d) split the sub-wavelength connection on two different bursts (or group of slots).

To check the continuity of a sub-wavelength connection, we will look at the scheduling output, whether or not a slot is actually set as a function of the input time-slot index, that is, the derivative of the time-slot assignment. Taking a look at Fig. 5.13, we can see that, whenever the connection is split on more than one burst or continuous set of time-slots, at least two start and end slots occur for each allocated chunk within each link resources frame horizon. To track a start or end slot we apply the differential, Δm_{ij}, between two consecutive slot values. Therefore, whenever a change of slot occupation occurs, the slope will be different to 0,

$$m = \frac{\Delta m_{ij}}{\Delta x} = \frac{m_{i,x2} - m_{i,x1}}{x_2 - x_1} \neq 0. \tag{5.28}$$

where x_1 and x_2 are the indices of two consecutive slots.

To track the starting/ending slots of the allocated connections we add two more variables:

- S_{mn}^{jk}, where $1 \leq k \leq |B|$ represents the connection index and j denotes the slot, is an integer array to track the starting slot of the connection allocation on link between nodes m and n.
- D_{mn}^{jk}, is also an integer array, but now for tracking the ending slot of the connection allocation.

Upon this, the extra constraints needed to ensure the burst slot allocation continuity are:

- Continuity constraints:

$$\begin{aligned} \forall m, n, k, \quad & \forall j_1, j_2 \in S, j_1 < s, j_2 = j_1 + 1 : \\ & S_{mn}^{j_2 k} \geq \sum_i m_{mn}^{ij_2 k} - \sum_i m_{mn}^{ij_1 k} \end{aligned} \tag{5.29}$$

$$\begin{aligned} \forall m, n, k, \quad & \forall j_1, j_2 \in S, j_1 < s, j_2 = j_1 + 1 : \\ & D_{mn}^{j_1 k} \geq \sum_i m_{mn}^{ij_1 k} - \sum_i m_{mn}^{ij_2 k} \end{aligned} \tag{5.30}$$

$$\forall m, n, k, j : S_{mn}^{jk} \geq 0 \tag{5.31}$$

$$\forall m, n, k, j : D_{mn}^{jk} \geq 0 \tag{5.32}$$

$$\forall m, n, k : \sum_j S_{mn}^{jk} \leq 1 \tag{5.33}$$

$$\forall m, n, k : \sum_j D_{mn}^{jk} \leq 1 \tag{5.34}$$

Equations (5.29)-(5.34) basically generate the starting/ending slot array for the burst allocation, just as shown in Fig. 5.13. While (5.29) and (5.30) are defined to track the starting and ending slots, respectively, (5.31) and (5.32) make sure that the rest of unallocated slots for such connection are equal to 0. Finally, (5.33) and (5.34) ensure that if the slot is scheduled, only one start/end slot is present.

SW-C is the most restricted policy, hence it introduces the largest number of constraints and complexity. However, if we approximate the variable complexity as we did for the previous policies, we get the same approximation. Specifically, the number of unknown variables in this new problem formulation is $O(|B| + s|W||B||V|^2 + |B||V| + 2|B||V|^2 + |W||B| + 2s|B||V|^2) \sim O(s|W||B||V|^2)$. Thus, the number of variables increases linearly as a function of the number of wavelengths, connection requests and matrix slot length, and quadratically as a function of the network topology size. The complexity in terms of constraints is also increased in order to allow for the continuity inequalities. However, the approximate constraint complexity remains the same, $O(s|W||B||V|^2)$, since the slot index used to check the gradient on consecutive slots is run once per link, wavelength and connection.

5.4.6 Tunable Continuous Policy

The latest policy restricts the assignment along a continuous set of time-slots, but allows to switch to another wavelength from slot to slot index. It reuses partially some of the variables and constraints introduced hereafter.

Table 5.1 summarizes the list of variables and constraints needed by each assignment policy and their approximate complexity in terms of number of variables and constraint inequalities.

Table 5.1: Summary of variables and constraints needed by each scheduling policy.

Policy	**Variables**	**Equation constraints**	**Variables complexity**	**Approximate constraints complexity**
MW-NC	m^{ijk}_{mn}, V^k_p, F^k_{mn}, L^k_{mn}, C_k	(5.3)-(5.15), (5.16)-(5.23) and (5.24)	$O(\|B\| + s\|W\|\|B\|\|V\|^2 + \|B\|\|V\|+2\|B\|\|V\|^2) \sim O(s\|W\|\|B\|\|V\|^2)$	$O(s\|W\|\|B\|\|V\|^2)$
T-NC	m^{ijk}_{mn}, V^k_p, F^k_{mn}, L^k_{mn}, C_k	(5.3)-(5.15), (5.16)-(5.23), (5.24) and (5.25)	$O(\|B\| + s\|W\|\|B\|\|V\|^2 + \|B\|\|V\|+2\|B\|\|V\|^2) \sim O(s\|W\|\|B\|\|V\|^2)$	$O(s\|W\|\|B\|\|V\|^2)$
SW-NC	m^{ijk}_{mn}, V^k_p, F^k_{mn}, L^k_{mn}, C_k, C^i_k	(5.3)-(5.15), (5.16)-(5.23), (5.24), (5.25), (5.26) and (5.27)	$O(\|B\| + s\|W\|\|B\|\|V\|^2 + \|B\|\|V\| + 2\|B\|\|V\|^2 + \|B\|\|W\|) \sim O(s\|W\|\|B\|\|V\|^2)$	$O(s\|W\|\|B\|\|V\|^2)$
T-C	m^{ijk}_{mn}, V^k_p, F^k_{mn}, L^k_{mn}, C_k, S^{jk}_{mn}, D^{jk}_{mn}	(5.3)-(5.15), (5.16)-(5.23), (5.24), (5.25), and (5.29)-(5.34)	$O(\|B\| + s\|W\|\|B\|\|V\|^2 + \|B\|\|V\| + 2\|B\|\|V\|^2 + 2s\|B\|\|V\|^2) \sim O(s\|W\|\|B\|\|V\|^2)$	$O(s\|W\|\|B\|\|V\|^2)$
SW-C	m^{ijk}_{mn}, V^k_p, F^k_{mn}, L^k_{mn}, C_k, C^i_k, S^{jk}_{mn}, D^{jk}_{mn}	(5.3)-(5.15), (5.16)-(5.23), (5.24), (5.25),(5.26), (5.27), and (5.29)-(5.34)	$O(\|B\| + s\|W\|\|B\|\|V\|^2 + \|B\|\|V\| + 2\|B\|\|V\|^2 + \|B\|\|W\| + 2s\|B\|\|V\|^2) \sim O(s\|W\|\|B\|\|V\|^2)$	$O(s\|W\|\|B\|\|V\|^2)$

5.4.7 ILP Numerical Results

In this section, we present some results about the proposed ILP formulations for illustrative purposes. To this end, we have considered the small network topology shown in Fig. 5.14, with 2 wavelengths per link and a link resources frame/matrix length of 6 slots. Such values are intentionally small to speed up the runtime of the ILP formulation. In the experiments, we used the GNU Linear Programming Kit (GLPK) [140] has been used as the ILP solver.

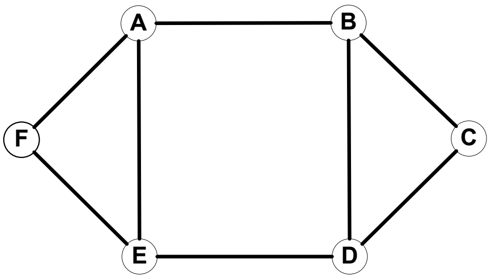


Fig. 5.14: 6-node topology.

To check the correctness of the ILP formulation for the five proposed sub-wavelength scheduling policies, 10 unicast sub-lambda connections of various granularities (BW) were established on the network between a source S and a destination node D. Tables 5.2 and 5.3 show the list of connections and the results obtained from each one of the scheduling policies formulated in the ILP. Labels are defined as $<lambda>$:$<slot>$, where$<lambda>\in \{1,2\}$ and $<slot>\in \{1,2,\ldots,6\}$. As we can see from the tables, the scheduled slots follow the restrictions imposed by each policy. For instance, in the SW-C case, all scheduled slots follow the continuous group set on a single wavelength, while in the T-C case, more than one wavelength can be used. On the contrary, the non-continuous policies allow slots of any kind to be used as long as they satisfy the non multi-wavelength constraint. An example follows; connection 4 from E to B on route E–A–B uses slots $\{2:2, 2:4, 1:6\}$ under policy T-NC, while under the continuous policy T-C, the slots are $\{1:1, 2:2, 2:3\}$, i.e., in the second case the slots are continuous over two different wavelengths. It is also worth noting that in the MW-NC case, allocation on simultaneous slots is allowed; see for instance connection from A to C over route A–B–C, wherein the assigned labels are $\{1:2, 1:5, 1:6, 2:6\}$; hence, slot index 6 is used on both wavelengths.

Furthermore, we provide in Fig. 5.15 the performance results of various network parameters, such as blocking probability, average lightpath cost and average lightpath length. In order to minimize the runtime complexity, we set 30 different ILP runs for each scheduling policy with 6 sub-wavelength unidirectional lightpath connections. To ensure that connection collisions can be produced, we uniformly select the source and destination of each connection k to be $S_k \in \{A, F, E\}$ and $D_k \in \{B, C, D\}$. Therefore,

Table 5.2: Sub-wavelength connections and results for the MW-NC, T-NC and SW-NC policies.

S	D	BW	MW-NC		T-NC		SW-NC	
			Route	**Label**	**Route**	**Label**	**Route**	**Label**
A	C	4	A-B-C	1:2,1:5,1:6,2:6	A-B-C	1:1,1:2,2:3,2:6	A-B-C	2:1,2:3,2:4,2:6
A	D	4	A-E-D	1:1,1:2,1:3,2:6	A-E-D	1:1,2:2,1:3,1:6	A-E-D	2:3,2:4,2:5,2:6
F	C	2	F-E-D-C	2:2,2:3	F-A-B-C	1:3,1:4	F-E-D-C	1:1,1:3
E	B	3	E-A-B	1:1,1:3,1:4	E-A-B	2:2,2:4,1:6	E-A-B	1:1,1:2,1:4
D	A	4	D-B-A	1:3,2:3,1:4,1:5	D-E-A	1:2,2:3,1:4,2:5	D-E-A	2:3,2:4,2:5,2:6
C	F	3	C-D-E-F	1:2,1:6,2:6	C-D-E-F	1:3,2:4,2:6	C-D-E-F	1:1,1:2,1:3
B	F	3	B-A-F	1:1,2:4,2:6	B-A-F	1:1,1:3,2:4	B-A-F	1:3,1:4,1:6
E	C	2	E-D-C	1:5,1:6	E-D-C	2:1,1:5	E-D-C	1:5,1:6
C	A	3	C-B-A	2:1,2:2,1:6	C-B-A	2:1,2:2,1:4	C-B-A	2:1,2:3,2:5
B	F	2	B-A-F	1:2,2:5	B-A-F	2:3,2:5	B-A-F	2:2,2:4

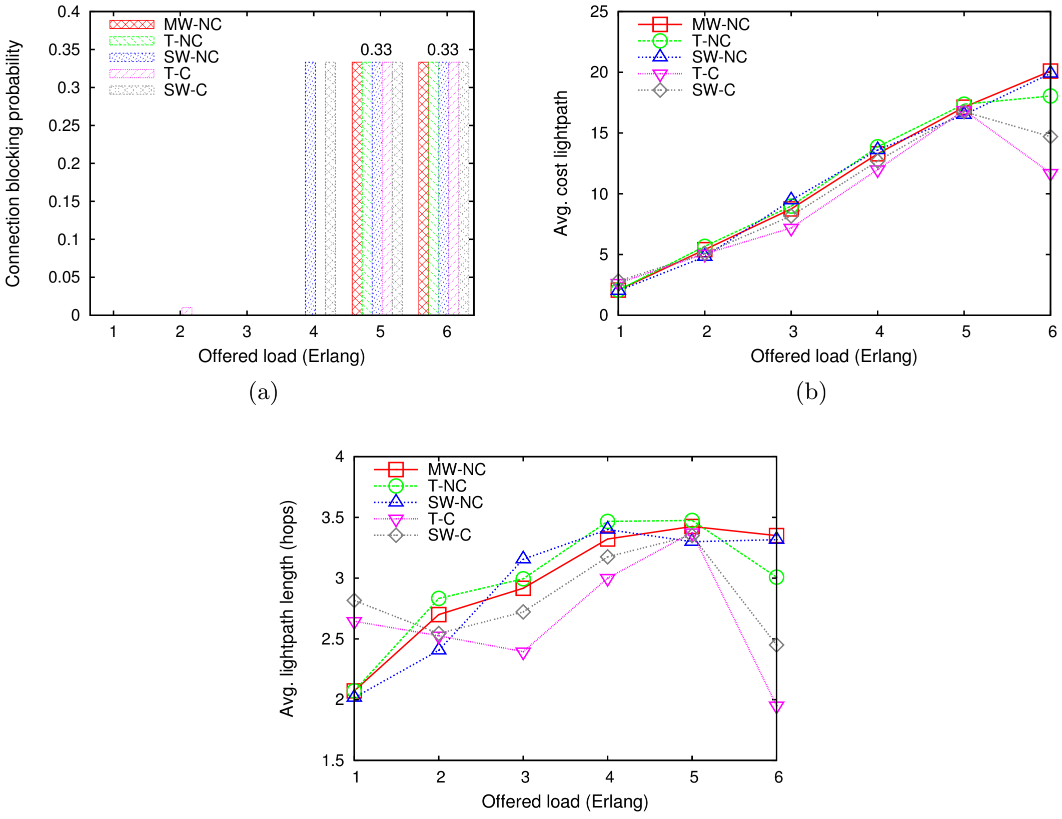
Table 5.3: Sub-wavelength connections and results for the T-C and SW-C policies.

S	D	BW	T-C		SW-C	
			Route	**Label**	**Route**	**Label**
A	C	4	A-B-C	2:1,1:2,1:5,2:6	A-B-C	1:2,1:3,1:4,1:5
A	D	4	A-E-D	2:1,2:4,1:5,1:6	A-E-D	1:1,1:2,1:3,1:4
F	C	2	F-A-B-C	1:3,1:4	F-A-B-C	2:4,2:5
E	B	3	E-A-B	1:1,2:2,2:3	E-A-B	2:1,2:2,2:6
D	A	4	D-E-A	1:3,1:4,2:5,1:6	D-E-A	1:1,1:2,1:3,1:4
C	F	3	C-D-E-F	2:2,2:3;2:4	C-D-E-F	2:1,2:2,2:3
B	F	3	B-A-F	2:4,2:5,1:6	B-A-F	1:2,1:3,1:4
E	C	2	E-D-C	1:1,1:2	E-D-C	2:1,2:6
C	A	3	C-B-A	2:2,1:3,1:4	C-B-A	2:4,2:5,2:6
B	F	2	B-A-F	1:5,2:6	B-A-F	2:1,2:2

links A–B and E–D will be possible congestion bottlenecks.

Fig. 5.15(a) shows the connection blocking probability as a function of the total offered load (in Erlang). Loads are defined by the number of time-slots requested by the connections. For instance, when load is equal to 1 Erlang, each one of the six connections requests for 1 slot, whereas when the load is 6 Erlang, they request 6 slots.

As expected, the blocking probability stabilizes at 0.33 for loads 5 and 6 Erlang for all the scheduling policies. In this respect, 6 Erlang imply that connections request the whole capacity of the wavelength. Therefore, all 6 slots available per link and wavelength are requested. Obviously, if only 2 wavelengths are available per link, and two possible congestion links are also available (as mentioned previously), only 4 out of 6 connections will be possible. Moreover, the single wavelength scheduling policies (SW-NC and SW-C) also produce losses even at a load of 4 Erlang due to their inability to split the connection slots over different wavelengths, thus not making possible to reuse the two slots per wavelength that remain unused after setting up the other 4 sub-wavelength connections. Finally, we can see in the T-C case that at 2 Erlang blocking



(a)

(b)

(c)

Fig. 5.15: ILP performance results: (a) connection blocking probability, (b) average lightpath cost, and (c) average lightpath length.

probability is not zero. This is due to a single case where not all the connections were possible due to the expiration of the ILP runtime.

In spite of having formulated the ILP to optimize (minimize) the blocking probability among the connection requests at the input, we can also gain some insight on which are the actual paths and network cost of the successfully established connections. Recall that these values might not be optimized globally. As we have introduced, the ILP formulation optimizes the number of accepted connections based on (5.2). However, because the setup of the ILP is common for all the sub-wavelength assignment policies, we will be able to establish a comparative analysis among them. Fig. 5.15(b) shows the network cost per successful lightpath connection. The cost per lightpath is calculated as,

$$\sum_{k}\sum_{i}\sum_{j}\sum_{m,n} w_{mn}\cdot m_{mn}^{ijk} \tag{5.35}$$

for all k such $C_k = 1$. For simplicity, all link weights are equal to 1, hence $\forall m, n$ with $P_{mn} = 1 \Rightarrow w_{mn} = 1$. As we can see, the average cost lightpath is similar in the five cases. Only when the connections demand the whole wavelength capacity, the average cost varies among the five policies. Related to this result, we can also calculate the average path length (in number of hops) of the established lightpaths. These results are shown in Fig. 5.15(c). In general, we can see that the T-C policy provides the shortest paths, specially at loads over 3 Erlang. Moreover, we can see two different trends: on the one hand, the policies with slot continuity constraints produce the longest paths for an offered load of 1 Erlang whereas at a load of 3 Erlang and over, they generate the shortest paths. On the other hand, the non-continuous policies present a smoother trend, increasing its path length as the offered load rises. Consequently, slot continuity seems to be the main constraint to address and the one that demands the network to achieve a more accurate resources allocation.

5.5 Heuristics for the Static Scheduling Scenario

As usual, the complexity and runtime of the mathematical formulation and optimization is not scalable. As a result, the ILP is not useful for the optimization of large problem scenarios. This is specially true in the present case due to growing number of variables and constraints as a consequence of the sub-wavelength (time-slot) assignment. Note that RWA is already a very complex problem, thus adding a new level of granularity (time-slot) to the RWA grows the feasible solution set by an order of two or three, depending on the time-frame size.

Quasi-optimal heuristics are required in order to solve in a timely manner the static scheduling scenario proposed in the previous section. Next, we describe two of the most representative heuristic algorithms. The former computes the set of sub-wavelength lightpaths following the SW-C scheduling policy, and the latter does the same but applying the T-C approach. Similar heuristics can also be defined for the remaining three policies, but they have been omitted in this chapter.

Both algorithms are based on a two-tier path, wavelength and time-slot computation. As introduced before, this strategy fits the PCE-SLAE computation approach devised for the three proposed TSON architectures. In particular, the mechanism fits perfectly the PCE-managed and centralized SLAE used in TC-TSON.

5.5.1 SWFFC: Heuristic for the SW-C Policy

Algorithm 4 shows the least-used lightpath with single-wavelength first-fit continuous heuristic algorithm, which implements the SW-C scheduling policy for the static connection request scenario.

Given the set of sub-wavelength connection requests $B = \{b_1, b_2, \ldots, b_c\}$, the network graph and resources $G = (V, E, W, S)$ with V nodes, E links, W wavelengths and S resource slots per wavelength (i.e., minimum allocation sub-wavelength granularity), and K number of possible shortest paths, the objective of the algorithm is to compute

Algorithm 4 Least-used lightpath with single-wavelength first-fit continuous (SWFFC) scheduling heuristic.

```
 1: Inputs: $B = \{b_1, b_2, \ldots, b_c\}$, $G=(V,E,W,S)$, $K$
 2: Output: $[lp_m]_c$
 3: Initialize K-shortest paths for each source-destination pair and order by increasing length
 4: $B_s \leftarrow sortConnectionIncCost(B)$
 5: $k \leftarrow 0$
 6: while $k < K$ and $isEmpty(B_s)$ is not true do
 7:    for all $m$ in $|B|$ do
 8:       $b_x \leftarrow heap(B_s)$
 9:       $(s_x, d_x) \leftarrow getSourceAndDest(b_x)$
10:       $r_x \leftarrow getRoute(s_x, d_x, k)$
11:       $bw_{max} \leftarrow 0$, $\lambda_{max} \leftarrow null$
12:       for all $\lambda_l$ in $W$ do
13:          $bw_{r_x,\lambda_l} \leftarrow computeFreeBW(r_x, \lambda_l)$
14:          if $bw_{r_x,\lambda_l} > bw_{max}$ and $bw(b_x) < bw_{r_x,\lambda_l}$ then
15:             $bw_{max} \leftarrow bw_{r_x,\lambda_l}$, $\lambda_{max} \leftarrow \lambda_l$
16:          end if
17:       end for
18:       if $\lambda_{max}$ is not null then
19:          $[m_j] \leftarrow computeResources(r_x, \lambda_{max})$
20:          $[label_x] \leftarrow \emptyset, j \leftarrow 0$
21:          while $j < (|S| + bw(b_x))$ and $countLabel([label_x]) < bw(b_x)$ do
22:             $s_j \leftarrow getSlot(j, [m_j])$
23:             if $isSlotFree(s_j)$ is true then
24:                $[label_x] \leftarrow addSlot(s_j)$
25:             else
26:                $[label_x] \leftarrow \emptyset$
27:             end if
28:             $j \leftarrow j + 1$
29:          end while
30:          if $countLabel([label_x])$ is $bw(b_x)$ then
31:             $[lp_m]_c \leftarrow addLightpath(rw_x, [label_x])$
32:             Update network resources on $G$
33:             Goto line (7)
34:          end if
35:       end if
36:       $B_s \leftarrow push(b_x)$
37:    end for
38: end while
```

and return the list of sub-wavelength lightpaths $[lp_m]_c$ with their corresponding assigned time-slots (or labels).

The algorithm starts by initializing the k-shortest paths between each source and destination pair on the network, and sorting them in increasing number of hops. Eventually, this step can be pre-computed in advance and passed as a parameter to the algorithm. As usual, we need to first sort the list of connection requests, which in this specific case are ordered in increasing order based on the network cost, $Cost_x$, to schedule the connection over the shortest path, $x^+_{s,d}$, as given by (5.36)

$$Cost_x = \sum_{l \in x^+_{s,d}} bw(b_x), \tag{5.36}$$

where $l \in x^+_{s,d}$ denotes the set of links that are part of the shortest path, and $bw(b_x)$ the sub-wavelength capacity demanded by connection b_x (in number of slots). This process is realized by function *sortConnectionIncCost*() in line 4.

In loop from lines 6-38, the list of connections is evaluated. As stated within the *while* clause, each connection is inspected a maximum of K times if not scheduled wherein. For each connection, the algorithm gets the indexed k-shortest path, r_x (lines 9-10). Next, the algorithm seeks the wavelength along this path with the greatest available bandwidth that provides enough free capacity to allocate the current connection demand (lines 12-17). If a wavelength is found with enough bandwidth, $bw(b_x) < bw_{r_x,\lambda_l}$, an array of the resources along the route is computed, $[m_j]$, as shown in line 19. The inner *while* loop (lines 21-29) does the lookup of a continuous set of free (available) slots. If the number of slots found is equal to the requested bandwidth (line 30), then the lightpath and sub-wavelength time-slots are allocated to the correspondent connection request and added to the return list $[lp_m]_c$ (line 31). This process requires to update the network resources from G (line 32). In the case the current route index does not successfully provision the sub-wavelength bandwidth for the connection, the request is pushed again into the pending connection list (line 36). Recall that if the connection is successfully established, it it not pushed again into the list, hence its processing is skipped in the following loop iteration carried at line 6.

Regarding the complexity of Algorithm 4, assume the network is represented by graph $G = (V, E, W)$, with V the set of nodes and E the unidirectional links. Moreover, W are the wavelengths per link, and $|S|$ granularity slots per wavelength. The number of connection requests is $|B|$. Assuming that the k-shortest paths for each source-destination pair are processed in advance, we can derive the complexity as follows: the initial sorting of the connections (line 4) can be processed in $O(|B| \log |B|)$. After that, we get into the main loop, which is bounded by the number of shortest paths k (loop 6-38). For each one of them, we run through the set of connections B to process (loop 7-37), and for each of these the following steps are executed: compute the free bandwidth along the path for each of the wavelengths (lines 12-17), which contributes with $O(|W||V||S|)$. Upon that, the resources matrix is computed for all the links of the route (line 19), $O(|W||V||S|)$, and along it, all slots are checked looking for the free continuous set (lines 18-35), which is in the order of $O(|S|)$. If the slots are assigned, then the network resources are also updated, which is also doable in $O(|W||V||S|)$ (line 32). As a result of all these steps, the runtime complexity is in the order of $O(|B| \log |B| + k|B|(3|W||V||S| + |S|)) \sim O(k|B||W||V||S|)$, if $|B| \ll k|W||V||S|$.

5.5.2 LUSCT: Heuristic for the T-C Policy

Algorithm 5 Least-used-slot continuous tunable (LUSCT) scheduling heuristic.

```
 1: Inputs: $B = \{b_1, b_2, \ldots, b_c\}$, $K$
 2: Output: $[lp_m]_c$
 3: Initialize K-shortest paths for each source-dest pair and order by increasing length
 4: $B_s \leftarrow sortConnectionIncCost(B)$
 5: $k \leftarrow 0$
 6: while $k < K$ and $isEmpty(B_s)$ is not true do
 7:   for all $m$ in $|B|$ do
 8:     $b_x \leftarrow heap(B_s)$
 9:     $(s_x, d_x) \leftarrow getSourceAndDest(b_x)$
10:     $r_x \leftarrow getRoute(s_x, d_x, k)$
11:     $bw_{r_x} \leftarrow computeFreeLinkBW(r_x)$
12:     if $bw_{r_x} > bw(b_x)$ then
13:       $[m_{ij}] \leftarrow computeResources(r_x)$
14:       $[label_x] \leftarrow \emptyset, i \leftarrow 0, s_{max} \leftarrow maxFreeSlotIndex([m_{ij}])$
15:       while $j < (|S| + bw(b_x))$ and $countLabel([label_x]) < bw(b_x)$ do
16:         $s_{ij} \leftarrow getSlot(i, s_{max}, [m_{ij}])$
17:         if $isSlotFree(s_{ij})$ is true then
18:           $[label_x] \leftarrow addSlot(s_{ij})$
19:         else
20:           $index \leftarrow 0$
21:           repeat
22:             $i \leftarrow i + 1$, $index \leftarrow index + 1$
23:             if $i \geq |W|$ then
24:               $i \leftarrow 0$
25:             end if
26:           until $index < |W|$ and $isSlotFree(s_{ij})$ is false
27:           if $isSlotFree(s_{ij})$ is true then
28:             $[label_x] \leftarrow addSlot(s_{ij})$
29:           else
30:             $[label_x] \leftarrow \emptyset$
31:           end if
32:         end if
33:         $j \leftarrow j + 1$, $s_{max} \leftarrow s_{max} + 1$
34:         if $s_{max}$ is $|S|$ then
35:           $s_{max} \leftarrow 0$
36:         end if
37:       end while
38:       if $countLabel([label_x])$ is $bw(b_x)$ then
39:         $[lp_m]_c \leftarrow addLightpath([lp_m]_c, rw_x, [label_x])$
40:         Update network resources on $G$
41:         Go to line 7
42:       end if
43:     end if
44:     $B_s \leftarrow push(b_x)$
45:   end for
46: end while
```

Having in mind the tunable capabilities of the T-C policy, we need to further optimize the time-slot assignment to improve the link resources utilization. Because now it is not a first-fit slot approach, we refer to the heuristic as least-used-slot continuous tunable

(LUSCT).

Algorithm 5 shows the heuristic for the static connection request problem. The heuristic, like in SWFFC, allocates the list of sub-wavelength connections on the available network resources with the objective to maximize the maximum number of accepted connections.

The algorithm starts by ordering the input list of connection in increasing cost computed with (5.36) (line 4). The main loop runs for each k-shortest path available for each connection source and destination pair. In comparison with the SWFFC approach, in this case we compute the free link capacity for all the wavelength channels on every link (line 11). If the available bandwidth is enough to satisfy the demanded bandwidth, $bw_r > bw(b_x)$ (line 12), then the matrix of network resources is computed, $[m_{ij}]$ (line 13). Using the latter, a new function is executed, $maxFreeSlotIndex([m_{ij}])$ (line 14) which aims at improving the slot assignment for the tunable case. The function takes as a parameter the network resource matrix computed previously and its purpose is to track the slot index, s_{max}, that provides the greatest amount of free bandwidth across all the wavelengths. Once this is computed, the *while* loop (lines 15-37) is responsible for finding the free slots that comply with the continuous and tunable constraints. Essentially, if the next slot on the same wavelength is not free, the algorithms runs a loop (lines 21-26) to find another wavelength conforming the slot continuity. If the number of slots required by the connection are found (line 38), then the algorithm proceeds with the allocation of the sub-wavelength lightpath and updates the network resources. If after running for all the slots and wavelengths from the network resources matrix, the algorithm does not find enough free slots, the connection request is pushed again into the list (line 44) in order to be processed later in the next *while* iteration (line 6).

Similarly as in the previous algorithm, the complexity of Algorithm 5 is also in the order of $O(k|B||W||V||S|)$. As in Algorithm 4, in this case the main loop also inspects for all the slots and wavelengths along the selected route, although in a different way to support the wavelength tunability during the slot allocation.

5.5.3 Heuristic Numerical Results Under the Static Traffic Scenario

The purpose of defining heuristics is to speed up the runtime of the optimization problem. In any case, the heuristics should provide a performance as close as possible to the ILP optimization. Although we have only described thoroughly the SWFFC and LUSCT heuristic implementations for their corresponding SW-C and T-C policies presented in the previous subsections, we actually implemented all of them to evaluate their performance. Table 5.4 shows the correspondence between the scheduling policies and the heuristic implementation.

Table 5.4: Scheduling policies and heuristic implementation correspondence.

Policy	**Heuristic**
MW-NC	Multi-Wavelength First-Fit (MWFF)
T-NC	Least-Used-Slot Tunable (LUST)
SW-NC	Single-Wavelength First-Fit (SWFF)
T-C	Least-Used-Slot Continuous Tunable (LUSCT)
SW-C	Single-Wavelength First-Fit Continuous (SWFFC)

5.5.3.1 ILP and Heuristics Performance Comparison

Fig. 5.16 shows the results for the same network and connection request scenario analyzed in Section 5.4.7. The connection blocking probability is represented in Fig. 5.16(a). Although the number of connection requests is small, we can see for this specific case that the performance is the same as the one shown in Fig. 5.15(a) for the ILP.

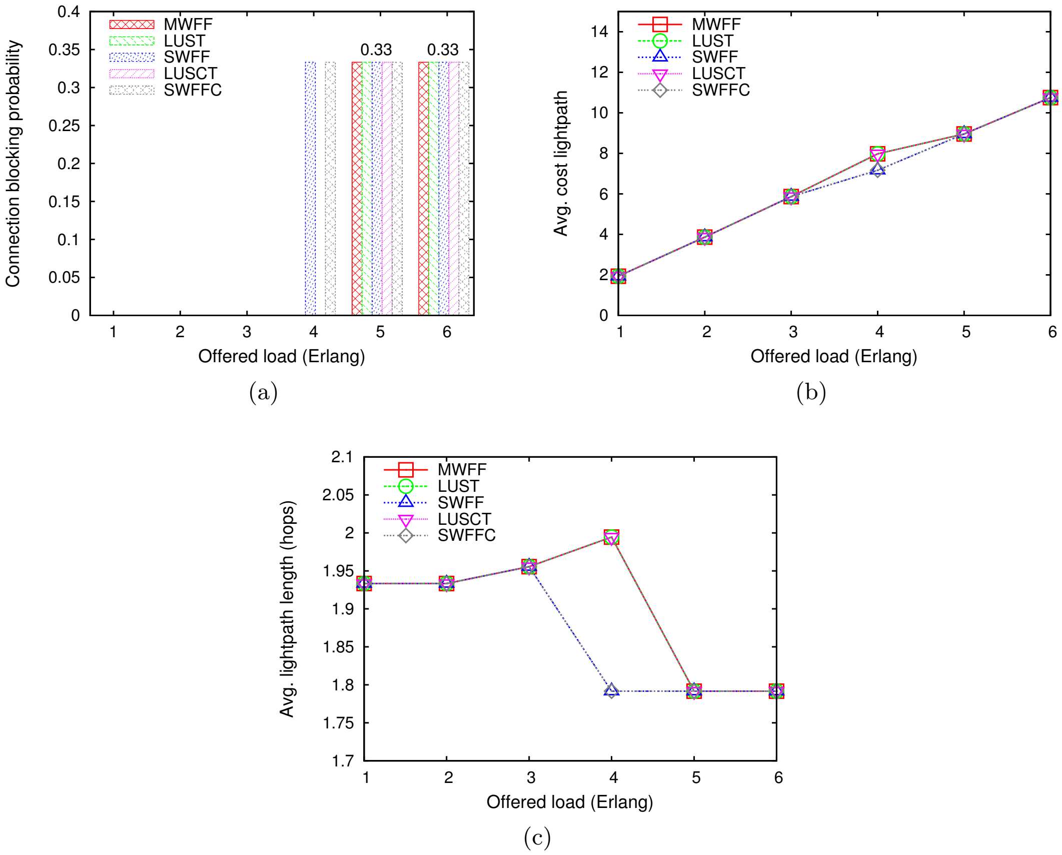
Fig. 5.16(b) shows the average cost per lightpath as calculated in (5.36). We can see that all the heuristic implementations have a similar performance. Only at a load of 4 Erlang, the tunable or multi-wavelength heuristics have a slightly greater lightpath cost, primarily due to the fact that in those three cases, a greater number of connections are established (actually, all 6 connections). The explanation behind such performance is related with the ordering of the connection request set in one of the first steps of the heuristics algorithm (refer to line 4 in Algorithm 5). In this case, it is worth noting that in all 5 implementations, the connections are ordered in the same manner, from smaller to greater cost over the shortest path.

Also, Fig. 5.16(c) depicts the average lightpath length (in number of hops). At a load of 4 Erlang, all the connections are successfully established in the tunable and multi-wavelength heuristic algorithms. As a result, some of these connections are allocated over one of the non-shortest paths, and therefore, the average length increases. For the rest of load scenarios, the results are all the same among the heuristics for the same reason introduced above.

In comparison with the ILP results, we can see that the heuristics provide a much shorter average lightpath length and smaller cost. This is due to the fact that to pursue for the optimization of the scheduled number of connections, we ordered the input list of connections from smaller to greater lightpath cost. We note as well that the ILP was optimized to maximize the number of accepted connections (see (5.2)), and for this reason, the ILP execution did not account for the length of the scheduled lightpath as long as it could find a feasible solution.

5.5.3.2 Heuristics Performance for Larger Scenario

The performance assessment between the ILP and the proposed heuristics is limited by the actual ILP formulation and its runtime complexity. Needless to say that run-



(a)

(b)

(c)

Fig. 5.16: Heuristics performance results: (a) connection blocking probability, (b) average lightpath cost, and (c) average lightpath length.

ning a simulation case scenario with dozens of connections and with a fairly small frame/matrix length can take days. Since we also have the heuristics, and these are able to run in the order of milliseconds for a single request input set, following we show some results on larger simulation scenarios.

For comparison purposes, we have implemented two other sub-wavelength algorithms from the literature. We adopt the algorithms presented in [103] to set the baseline for our comparison. In this work, the authors present some algorithms for the disjoint RWTA problem (i.e., the problem is divided disjointly into routing, wavelength and time-slot selection subproblems [104]). A new least resistance weight (LRW) function, which incorporates link load and path-length information, is used for computing the path cost, along with a time-slot-allocation algorithm based on the least loaded (LL) wavelength. A difference with respect our specific algorithm implementation is that we do not assume links have multiple fibers; hence, we avoid such assignment step from the algorithms.

The LRW function is given by

$$w_{ij} = \frac{C_{max}^{T}}{C_{ij}^{A}}, \; \forall(i,j) \in E, \tag{5.37}$$

where C_{max}^{T} is the maximum link total free capacity in the network as defined by,

$$C_{max}^{T} = \max_{i,j}(C_{ij}^{T}), \; \forall(i,j) \in E, \tag{5.38}$$

and C_{ij}^{T} is the amount of total capacity on the link $(i,j) \in E$. E denotes the set of link/edges on the network. $1/C_{ij}^{A}$ is the measure of resistance a link offers for establishing a connection. For time-sharing scenarios with the wavelength capacity divided in time-slots, the LRW weight function becomes $w_{ij} = S_{max}^{T}/S_{ij}^{A}$, where S_{max}^{T} is the maximum value of total free slot capacity per link, over all network links, and S_{ij}^{A} is the number of available slots on the (i,j) link.

The authors define two algorithms: least-loaded time-slot assignment (LLT) and the LLT with alternate wavelength (LLTAW). Our equivalent to LLT is SWFF, and to LLTAW is LUST. For more details, readers are referred to [103].

As in previous examples, we make use of the 6-node network (see Fig. 5.14). The number of wavelengths per link is now increased to 8 and the size of the frame/matrix (capacity of the wavelength) is 100 slots. The results are provided as a function of the number of connection requests into the system for a static traffic scenario. That is, all connection requests are known at the beginning of the simulation (off-line traffic scenario). For simplicity, all connections request the sub-wavelength capacity equivalent to 8 slots (8/100 the wavelength capacity), and the sources-destinations are computed from a uniform distribution. For each connection request set, we average the results over 30 different runs with different seed. 95% confidence intervals were also obtained, but since they are very narrow, these are omitted in order to improve the readability of the graphs.

Fig. 5.17(a) shows the connection blocking probability as a function of the number of connections offered to the network. The number of alternate routes or k-shortest paths is $k = 1$. As seen in previous analysis, the algorithms corresponding to the policies that permit the time-slot assignment on different wavelengths (LUSCT, LUST and MWFF) perform better than their non-tunable counterparts. In comparison with the other baseline algorithms, LLT and LLTAW have a lower blocking probability than the rest of proposed heuristics. This behavior is due to the on-demand and dynamic routing of LLT and LLTAW. These two algorithms are capable to recompute the link weights after processing successfully a connection request. With the new link weights, the algorithm runs the shortest weighted routing to compute the least loaded path. Therefore, instead of having a single route as in the rest of heuristics ($k = 1$), LLT and LLTAW are able to recompute a new one each time. LLTAW also performs better than

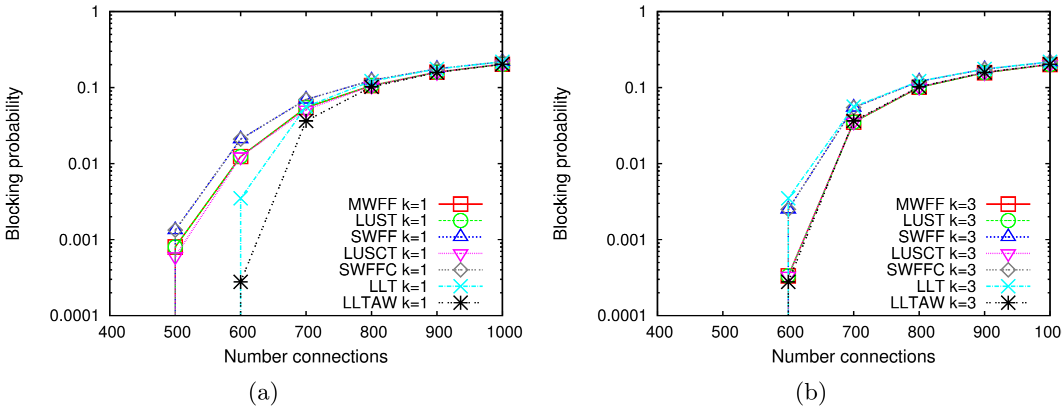


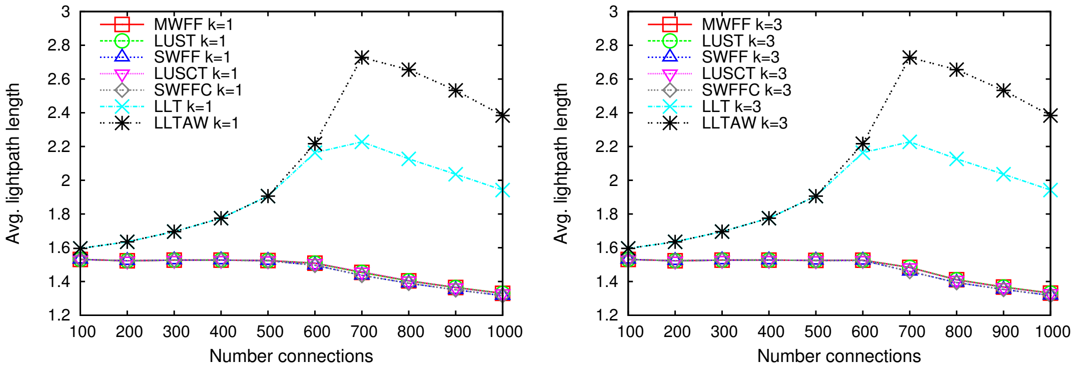
Fig. 5.17: Heuristic results on 6-node network with 8 wavelengths: (a) BP vs. number of connections for k=1, and (b) for k=3.

LLT due to the alternate wavelength allocation, i.e., with tunability.

When we increase the number of k-shortest routes (see Fig. 5.17(b)), the performance of LUSCT, LUST and MWFF become almost the same as LLTAW. Therefore, we can state that the benefits of the LLTAW in terms of blocking probability can be realized by offering off-line source-based k shortest route alternatives, that is, pre-computed shortest paths. This is interesting, since in the LLTAW, the routing protocol needs to run and recompute the routes for each incoming connection request, which is time consuming. Recall that in our heuristics, the number of shortest paths (see line (3) in Algorithm 4 or 5) determines the number of attempts each connection has in order to find free time-slots (see line (6)). We can also see that the LUSCT performs slightly better than SWFFC in both cases thanks to the unrestricted use of more than one wavelength per lightpath.

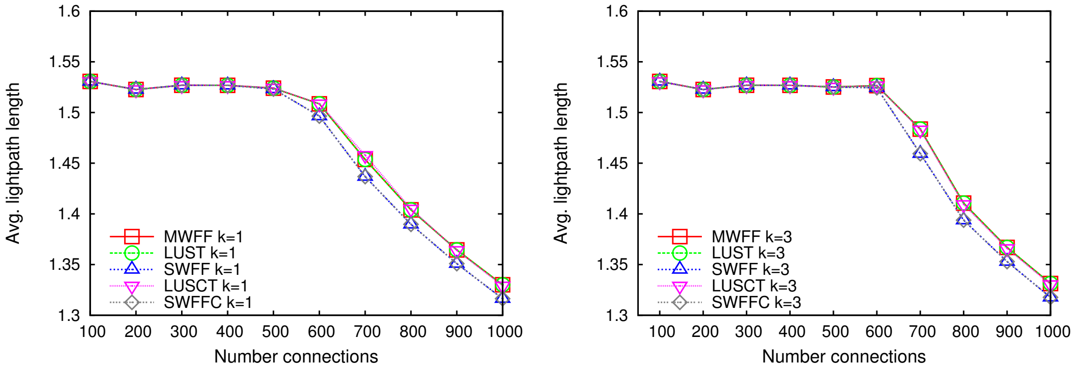
Fig. 5.18(a) and Fig. 5.18(b) show a comparison of the average lightpath length for successful connection requests and for k-shortest paths with $k = 1$ and $k = 3$, respectively. Basically, we can see that the proposed heuristics, which make use of the source-based pre-computed shortest paths (in length), show a much shorter path length than LLT and LLTAW. The latter compute the route based on the LRW function. This actually increases on average the length of the routes. At 700 connection requests, the lightpath length in the LLTAW nearly doubles the length seen by the rest of proposed algorithms. Therefore, we can expect that the compared baseline protocols cannot deliver the best end-to-end packet delay across the network. Also, we note that LLTAW and LLT yield no difference at all between Fig. 5.18(a) and 5.18(b), as both algorithms recompute the best path online based on the weight function.

If we focus on the proposed heuristic algorithms, we can observe that Fig. 5.19(a) and Fig. 5.19(b) show two different performance trends. Basically, the most flexible policies not only provide the best blocking performance as shown in Fig. 5.17(a), but



(a) (b)

Fig. 5.18: Heuristic results on 6-node network with 8 wavelengths: (a) average lightpath length (in number of hops) vs. number of connections for k=1, and (b) for k=3.



(a) (b)

Fig. 5.19: Heuristic results on 6-node network with 8 wavelengths among the proposed heuristics: (a) average lightpath length (in number of hops) vs. number of connections for k=1, and (b) for k=3.

also manage to do so on slightly longer lightpaths, especially when the number of connections is great, e.g., from 500 requests and over. By providing more bandwidth assignment flexibility, network resources (i.e., time-slots) can be used more efficiently, specially when connections start to get blocked on the overall shortest path.

Two other important performance parameters to address are the average fragmentation and number of wavelengths used per successful connection. As defined in previous sections, when it comes to RWTA, several policies are possible, such as restricting time-slot continuity or tunability (i.e., use of different wavelengths per connection). The first becomes the main factor when computing the number of fragments per connection, whereas the tunability determines the number of wavelengths that can be used per connection.

We define a fragment as the continuous set of allocated time-slots. To count the number of fragments of a time-shared lightpath we use the same idea reported behind

Fig. 5.13 and (5.28).

Fig. 5.20 illustrates two examples concerning fragmentation and number of lambdas used per connection: (i) lightpath 1 (LP1) uses a single wavelength but it is composed of two fragments; in contrast, (ii) LP2 uses three wavelengths but the set of time-slots allocated to it are continuous in time conforming a single fragment.

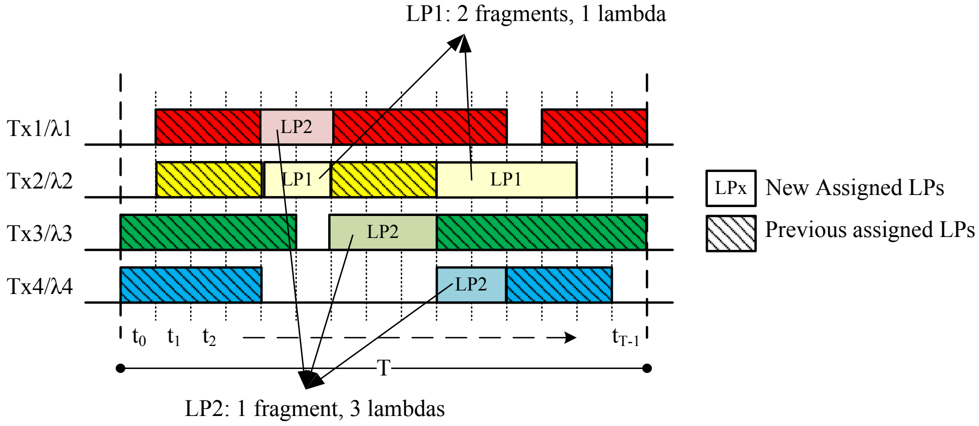
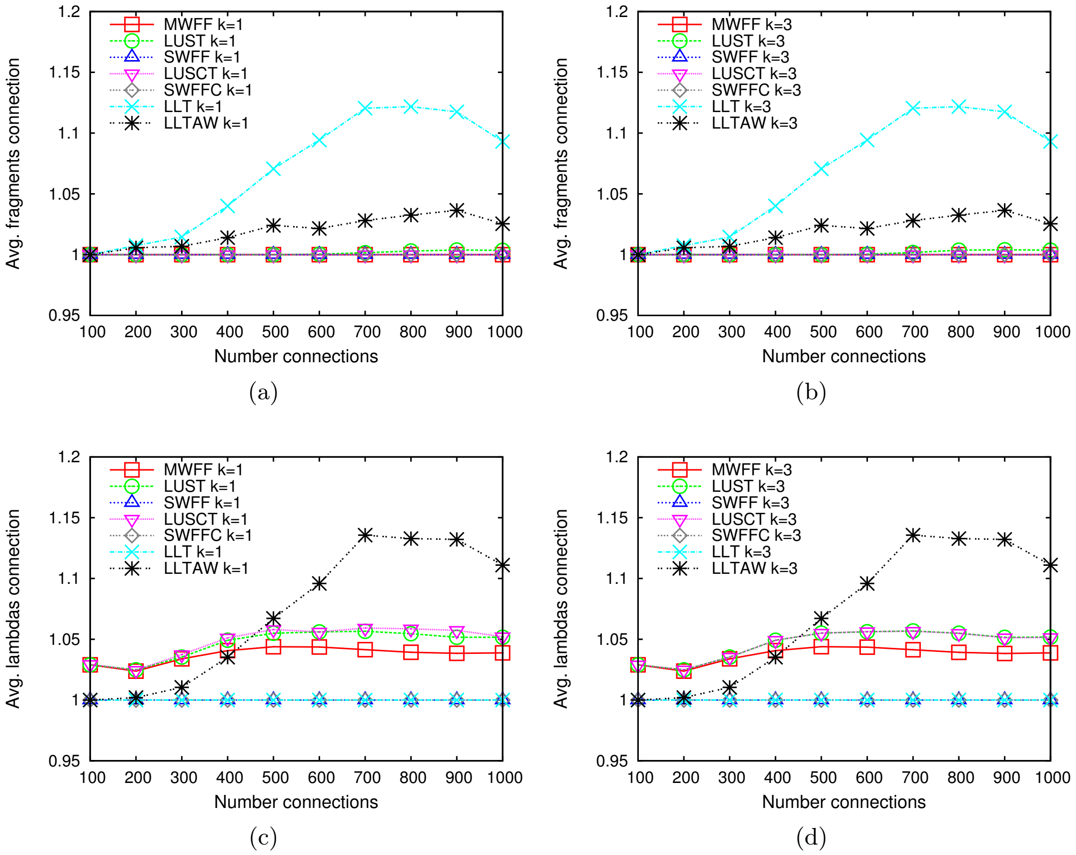


Fig. 5.20: Fragmentation vs. number of used wavelengths.

Fig. 5.21(a) shows the number of fragments that each successful connection request generates on average as a function of the input number of connection requests. In the static traffic scenario, as is this case, we can see that the proposed heuristics slightly differ from creating a single fragment. We can identify some variations, but these are barely noticeable. Obviously, those heuristics following the continuity constraint (SWFFC and LUSCT) do always generate a single fragment.

Moreover, the two baseline algorithms, LLTAW and LLT, generate a greater amount of fragments. In both cases, the algorithm tracks the least-used and constrained slot to start allocating the connection. Also, when adding the alternate wavelength flexibility as defined in LLTAW, the number of fragments is decreased in comparison with LLT. The results with $k = 3$ shortest paths shown in Fig. 5.21(b) are very similar to the $k = 1$ case. Particularly, recall that in this static traffic scenario, connection requests are processed sequentially over the predefined k-shortest paths. This provokes that connections are assigned in most cases a continuous groups of time-slots. This is not the case for LLTAW and LLT, which are more dynamically-driven by the least resistance weight function.

The final set of results deal with the average number of wavelengths used by successful connections. Fig. 5.21(c) and Fig. 5.21(d) show this performance parameter for $k = 1$ and $k = 3$ shortest paths, respectively. Again, all the heuristics that have the non-tunability or multi-wavelength constraint shall only use a single wavelength. Therefore, LLT, SWFFC and SWFF remain at a value of 1 for the whole x-axis. Also, we can see that the proposed tunable heuristics do increase the average number of



(a) (b)

(c) (d)

Fig. 5.21: Heuristic results on 6-node network with 8 wavelengths: (a) average number of fragments per connection vs. number of connections for k=1, and (b) for k=3; and (c) average number of wavelengths used per connection vs. number of connections for k=1, and (d) for k=3.

wavelengths used per connection, but not to the extend the LLTAW does, especially at high loads or number of connections. This implies that the LLTAW would require to enable the transmitter to switch among a greater number of wavelengths, which might degrade the performance, produce to much signaling in the network and stress the node transceivers.

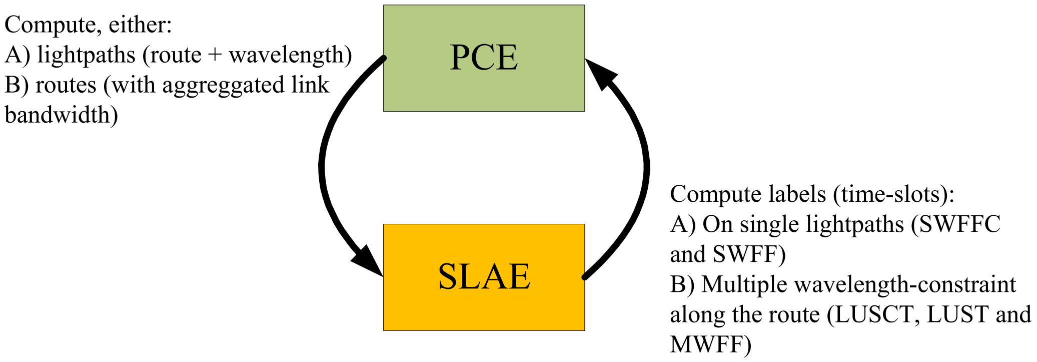
In summary, we can conclude that the proposed heuristics do present some benefits in front of other algorithms proposed in the literature, especially in terms of time-slot connection fragmentation and average number of wavelengths used per connection.

5.6 PCE and SLAE Routing, Wavelength and Time-slot Assignment in TC-TSON

As we saw in Section 5.3.4, the TC-TSON network architecture uses a centralized PCE and SLAE elements which are responsible for computing and generating the labels for the arriving sub-lambda lightpath requests. As a result of this, the performance greatly depends on the resources scheduling algorithms.

Furthermore, as shown in the previous section, the sub-wavelength routing, wavelength and slot assignment is very complex and no solution can be found in polynomial time for big and complex networks. Hence, heuristics are necessary in order to pursue for sub-optimized solutions able to run, either locally or distributively, the dynamic nature of the sub-wavelength assignment.

As it has been introduced, the PCE/TSON sub-lambda architecture is based on an augmented model, so both control planes in the network are involved in the resources allocation. As such, we devise a dual-assignment computation method as shown in Fig. 5.22; one carrying a coarser bandwidth route and wavelength assignment pre-processing, and the other computing finer sub-wavelength label assignment. More specifically, the PCE is responsible for, based on the aggregated resources information available in the TED, computing an initial list of candidate lightpaths for the arriving client connection. Subsequently, the SLAE produces the finer time-slot assignment using the list of pre-computed candidate lightpaths provided by the PCE.



Compute, either:
A) lightpaths (route + wavelength)
B) routes (with aggreggated link bandwidth)

Compute labels (time-slots):
A) On single lightpaths (SWFFC and SWFF)
B) Multiple wavelength-constraint along the route (LUSCT, LUST and MWFF)

Fig. 5.22: PCE and SLAE tandem time-shared bandwidth computation.

In this section we exemplify two different PCE-SLAE assignment approaches for two of the proposed heuristics introduced in the previous section: SWFFC and LUSCT. These two sub-wavelength policies require different PCE route computations so as to enable or not the tunable time-slot assignment by the SLAE. In the former case, the PCE has to provide both route and wavelengths, whereas in the LUSCT case, the PCE only needs to compute routes and aggregate link bandwidth availability. We start by describing the SWFFC solution.

5.6.1 SWFFC Algorithm

Algorithm 6 shows a simple sub-wavelength-unaware RWA algorithm for the PCE. The algorithm generates a list of n candidate lightpaths, RW (both routes and wavelengths), from node s to node d for a specific sub-wavelength capacity bw_c, based on the available

Algorithm 6 Least-used path and wavelength PCE used by SWFFC.

```
 1: Inputs: $n$, $s$, $d$, $bw_c$
 2: Output: $[lp_i]_n$
 3: $\mathcal{R} = [r_i]_k = kShortestPath(s, d)$
 4: $\mathcal{RW} = [rw_j]_l \leftarrow \emptyset$
 5: for all $r_i$ in $\mathcal{R}$ do
 6:     for all $\lambda_i$ in $\mathcal{W}$ do
 7:         $b_i = getMinFreeBw(r_i, \lambda_i)$
 8:         if $b_i \geq bw_c$ then
 9:             $[rw_j]_l \leftarrow [rw_j]_l + \{r_i, \lambda_i\}$
10:         end if
11:     end for
12: end for
13: if $\mathcal{RW} \neq \emptyset$ then
14:     $\mathcal{RW} \leftarrow orderLeastUsed(\mathcal{RW})$
15:     $c_0 \leftarrow connectionParameter(bw_c)$
16:     $[lp_k]_n \leftarrow slae(n, c_0, \mathcal{RW})$
17:     if $[lp_k]_n$ is not $\emptyset$ then
18:         return $[lp_k]_n$
19:     end if
20: end if
21: return $\emptyset$
```

Algorithm 7 Single-wavelength first-fit continuous (SWFFC) SLAE.

```
 1: Function: implements $slae(n, c, \mathcal{RW})$
 2: Input: $n$, $c$, $\mathcal{RW}$
 3: Output: $[lp_k]_n$
 4: $counter \leftarrow 0$
 5: repeat
 6:     $rw_k \leftarrow heap(\mathcal{RW})$
 7:     $[m_k] \leftarrow routeFrame(rw_k)$
 8:     $[label_k] \leftarrow \emptyset, i \leftarrow 0$
 9:     while $i < (num\_slots + c)$ and $countLabel([label_k]) < c$ do
10:         $s_i \leftarrow getSlot(i, [m_k])$
11:         if $isSlotFree(s_i)$ is true then
12:             $[label_k] \leftarrow [label_k] + s_i$
13:         else
14:             $[label_k] \leftarrow \emptyset$
15:         end if
16:         $i \leftarrow i + 1$
17:     end while
18:     if $countLabel([label_k])$ is $c$ then
19:         $[lp_k]_n \leftarrow [lp_k]_n + rw_k, [label_k]$
20:         $counter \leftarrow counter + 1$
21:     end if
22: until $counter$ is $n$ or $\mathcal{RW}$ is $\emptyset$
23: return $[lp_k]_n$
```

bandwidth per wavelength, b_i.

The first step is the computation of the k-shortest paths between s and d. Recall that these can be computed in advance. Next, for every route r_i and wavelength λ_i in W, the minimum available bandwidth available along the route is calculated using $getMinFreeBw(r_i, \lambda_i)$ (line 7). Only those lightpaths that satisfy the required connection bandwidth, $b_i \geq bw_c$, are kept in the list (line 8-9). After calculating the list of possible routes and wavelengths RW, if this is not empty, the finer sub-lambda assignment is processed. To this end, the PCE calls the function $slae(n, c_0, RW)$ implemented by the SLAE. Previously, the PCE orders the list of lightpaths RW from maximum to minimum available bandwidth (line 14).

The finer sub-wavelength assignment is processed by the SLAE, which is responsible for allocating the time-slots required by the connection request and assign the corresponding labels. Algorithm 7 shows the implementation of the single-wavelength first-fit continuous assignment policy. The input to the algorithm is the list of n lightpaths RW pre-computed by the PCE and the number of slots (or time length) required by the connection, c. Essentially, the SWFFC algorithm tries to allocate a set of continuous slots for the connection over a specific route and wavelength. Within the loop (lines 5-22), the algorithm gets the first lightpath from list RW, $heap(RW)$. For this route and wavelength, a frame matrix is calculated (line 7) by time-shifting and joining the link resources information along the lightpath. The join function takes into account the time-slot propagation delay from each path hop in order to satisfy the slot-continuity constraint along the path.

In a real network implementation, propagation delay is produced along the path interconnecting the nodes. Let assume a link l_i produces a delay d_i and the length of the frame is T slots. Moreover, let t_0 be the first slot allocation on link l_0 along the path, x_m, traversing links $\{l_0, \ldots, l_n\}$. The time-slot shifting for slot m_{t_i} on link $l_i \in x_m$ is computed as

$$m_{t_i} = \left(m_{t_0} + \sum_{j=0}^{i-1} d_j \right) \bmod T. \tag{5.39}$$

The join function, implemented by $routeFrame(rw_k)$ at line 7 in Algorithm 7 also removes the unusable resources across the lightpath and gives the results to the sub-wavelength assignment element which is responsible for allocating the finer granularity slots. This frame matrix is computed using the SL-TED information available at the SLAE. When using a time-slotted frame matrix approach, and OR operation is used for computing the join function of the resources for each of the links along the route. An example follows: assume a path x_m traversing links $\{l_0, \ldots, l_n\}$. The bit-representation of the resource frame/matrix for lightpath on wavelength λ_k on any of the links of path x_m is $[m_{t_i}]_{l_j}$. In such a case, the $routeFrame(rw_k)$ computes the $[m_k]$ as

$$[m_k] = [m_{t_0}] \text{ OR } [m_{t_1}] \text{ OR } \ldots \text{ OR } [m_{t_n}]. \tag{5.40}$$

In other words, the join function $routeFrame$ applies the time-shifting to each of the resource frame/matrix slots of a particular link and computes the bitwise OR operation with the rest of the resource frame/matrix (also time-shifted) from the rest of links.

The purpose of the inner loop (lines 9-17) is to find the set of slots that satisfy the continuous constraint. If so, and the number of slots within the $[label_k]$ object is equal to c (line 18), then the lightpath and labels $rw_k, [label_k]$ are inserted into the output list $[lp_k]_n$ (line 19) and the $counter$ of found labels increased (line 20). The algorithm finalizes by returning the list of lightpaths and labels (i.e., time-slots) found (line 23).

Regarding the complexity of this PCE-SLAE tandem SWFFC assignment, let the network of V nodes, E edges and W wavelengths be modeled by the graph $G{=}(V, E, W)$ and consider a time-slot frame of size L and k shortest paths between s and d. The computation cost of the PCE algorithm is in the order of $O(k|W||V| + (k|W|)^2) \sim O((k|W|)^2)$, if we assume the k-shortest paths are already computed in advance; otherwise we would have to add $O(k|V|^2 \log |V|))$. The first loop adds $O(k|W||V|)$, while the sorting in line 14 adds $O((k|W|)^2)$ if we use a bubble sort algorithm. Likewise, the complexity of the SLAE depends on the size of the frame and the list of candidate lightpaths passed as a parameter RW. Approximately, it can be determined to be

$O(k|W|(|V|L+L)) \sim O(k|W||V|L)$. It is worth noting that the given complexity of the SLAE algorithm is the worst case and it can be further reduced depending on the number of candidate lightpaths to process.

5.6.2 LUSCT Algorithm

The continuous tunable policy requires the PCE to only calculate the aggregate free bandwidth per route. Therefore, instead of computing the group RW as in the previous example, we just need to calculate a list of routes R, as shown in Algorithm 8. Actually, because we assume the k-shortest paths are already computed, we only need to remove the routes from the list that do not fulfill the minimum aggregated bandwidth requested by the connection. This process is carried out within the loop from lines 4-15.

The list of routes is then processed by the LUSCT SLAE algorithm (see Algorithm 9). The heuristic applies the same idea shown in Algorithm 5 to track the slot index with the maximum number of free wavelengths. This is implemented by the function $maxFreeSlotIndex([m_{kl}])$.

Within the loop between lines 5-36, the routes are checked for finer time-slot assignment until the list is empty or a predefined number of labels is found, n. Basically, for each route, a time continuous set of slots needs to be found, but not necessarily on the same wavelength. This implies the use of two inner loops to check on both, slots and wavelengths. To this end, the indices j and i are used for slots and wavelengths, respectively. As long as the next slot s_{ij} on the same wavelength is free, this is assigned into the label set. Otherwise, the algorithm tries to find a time-continuous slot on another wavelength (see lines 14-19). If after using all the wavelengths the algorithm is not able to find a free one, then the previous set of found slots is reset (line 24). The $routeMatrix(r_k)$ function in line 7 is equivalent to the $routeFrame(rw_k)$ used previously in the SWFFC algorithm. However, now, instead of being given a lightpath (i.e., path + lambda), we are given a path, and the function has to compute the join operation for all the slots and lambdas per link. Therefore, instead of having an array of bits $[m_k]$ representing the slots resources, we have a matrix of bits $[m_{kl}]$.

The computation cost of the PCE algorithm in LUSCT is in the order of $O(k|V|+k^2) \sim O(k|V|)$, if we assume the k-shortest paths are already computed in advance (otherwise we would have to add $O(k|V|^2 \log |V|)$) and the mean path length is greater than the number of shortest paths. Likewise, the complexity of the SLAE depends on the size of the frame, the initial list of candidate paths and the number of wavelengths per link. Hence, it can be determined to be $O(k(|V|L|W|+L|W|)) \sim O(k|W||V|L)$. As a result, the complexity of the SLAE LUSCT is the same as SWFFC.

In both algorithms, either the PCE is not able to provide an initial list of route–wavelengths with aggregated bandwidth availability or the SLAE cannot find a list of time-slots based on a concrete policy, the connection request is blocked.

Algorithm 8 Least-used path PCE used by LUSCT.

```
Inputs: $n$, $s$, $d$, $bw_c$
Output: $[lp_i]_n$
$\mathcal{R} = [r_i]_k = kShortestPath(s, d)$
for all $r_i$ in $\mathcal{R}$ do
    $b_i \leftarrow bw$
    for all $l_i$ in $r_i$ do
        $bl_i \leftarrow getFreeBw(l_i)$
        if $bl_i < b_i$ then
            $b_i \leftarrow bl_i$
        end if
    end for
    if $b_i < bw_c$ then
        $[r_i]_k \leftarrow [r_i]_k - r_i$
    end if
end for
if $\mathcal{R} \neq \emptyset$ then
    $\mathcal{R} \leftarrow orderLeastUsed(\mathcal{R})$
    $c_0 \leftarrow connectionParameter(bw_c)$
    $[lp_k]_n \leftarrow slae(n, c_0, \mathcal{R})$
    if $[lp_k]_n$ is not $\emptyset$ then
        return $[lp_k]_n$
    end if
end if
return $\emptyset$
```

Algorithm 9 Least-used-slot continuous tunable (LUSCT) SLAE.

```
Function: implements $slae(n, c, \mathcal{R})$
Input: $n$, $c$, $\mathcal{R}$
Output: $[lp_k]_n$
$counter \leftarrow 0$
repeat
    $r_k \leftarrow heap(\mathcal{R})$
    $[m_{kl}] \leftarrow routeMatrix(r_k)$
    $i \leftarrow maxFreeSlotIndex([m_{kl}]$
    $[label_k] \leftarrow \emptyset, j \leftarrow 0, aux \leftarrow 0$
    while $j < (num\_slots + c)$ and $count\_label([label_k]) < c$ do
        $s_{ij} \leftarrow getSlot(i, j, [m_{kl}])$
        if $isSlotFree(s_{ij})$ is true then
            $[label_k] \leftarrow [label_k] + s_{ij}$
        else
            repeat
                $i \leftarrow i + 1$
                if $i \geq num\_waves$ then
                    $i \leftarrow 0$
                end if
            until $index < num\_waves$ and $isSlotFree(s_{ij})$ is false
            if $isSlotFree(s_{ij})$ is true then
                $[label_k] \leftarrow [label_k] + s_{ij}$
            else
                $[label_k] \leftarrow \emptyset$
            end if
        end if
        $aux \leftarrow aux + 1, j \leftarrow j + 1$
        if $j \geq num\_slots$ then
            $j \leftarrow 0$
        end if
    end while
    if $countLabel([label_k])$ is $c$ then
        $[lp_k]_n \leftarrow [lp_k]_n + r_k, [label_k]$
        $counter \leftarrow counter + 1$
    end if
until $counter$ is $n$ or $\mathcal{R}$ is $\emptyset$
return $[lp_k]_n$
```

5.7 Simulation Results

This section reports on the performance of the proposed PCE/TSON interworking control plane scenario defined in the TC-TSON architecture for a dynamic traffic scenario. We must note that this case is different to the ILP scenario wherein all traffics were statically known in advance. The simulation scenario is the same as the one defined in Section 5.3.5. However, now we only use the TC-TSON architecture, which provided the best results, to test all the heuristics defined in the previous sections.

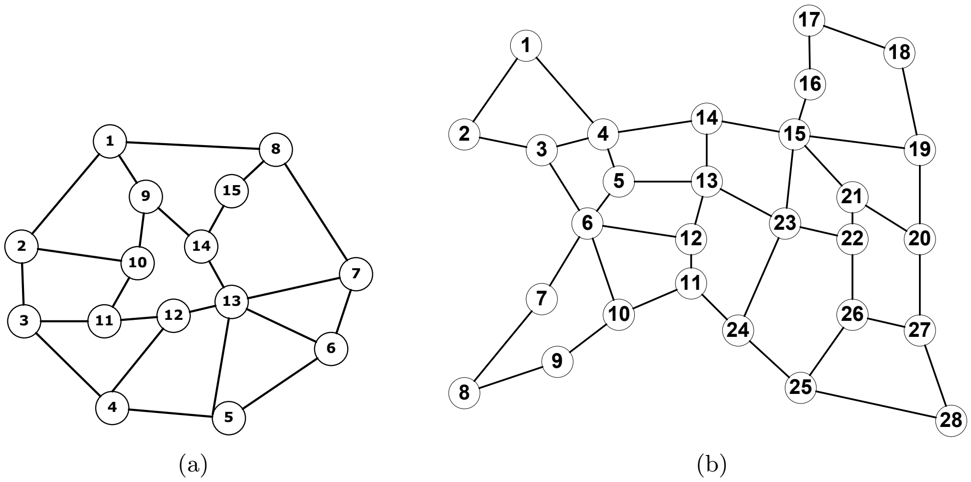


Fig. 5.23: Simulation topologies for the dynamic scenario: (a) TID network, and (b) EON.

Fig. 5.23(a) and Fig. 5.23(b) show the topologies considered in this evaluation. As introduced, Fig. 5.23(a) is based on Madrid's regional metro network. This topology is composed of 15 nodes, 23 bidirectional links, with a nodal degree connectivity of $\overline{deg}(G_{TID}) = 3.07$, an average link length $\overline{L_{TID}} = 56.87$ km and length standard deviation $\sigma_{L_{TID}} = 41.70$ (i.e. some links are much longer than others).

The EON (see Fig. 5.23(b)), composed of 28 nodes and 41 bidirectional links, has a mean route length of 3.56 hops and $\overline{deg}(G_{EON}) = 2.93$, an average link length $\overline{L_{EON}} = 625.37$ km and length standard deviation $\sigma_{L_{EON}} = 264.64$.

We have chosen these two network topologies for being representative to their respective scenarios. The TID network is based on Madrid's metropolitan network, whereas EON is a wide area network based across Europe with longer average link distances. In both cases, node 13 is the master TSON node that holds the centralized SLAE and from which the signaling of the sub-wavelength lightpath and its corresponding labels (time-slots and wavelengths) is initiated.

The number of wavelengths in each scenario changes from test to test. We assume in all the examples that sub-wavelength lightpaths are subject to the wavelength and time-slot continuity constraint along the path. Thus, nodes do not have fiber delay lines to delay the arrival of specific slots when switching them between the input and output port.

For simplicity, all sub-wavelength connections request for 1 Gbps bandwidth. Following the dynamic traffic premise, traffic arrival is modeled by a Poisson process with an exponential connection holding time with average $1/\mu = 60$ s. Moreover, the results are based on the time-slotted approach with a slot and a resources link frame size of 10 μs and 1 ms, respectively. Unless noted, the number of shortest paths available is $k = 3$.

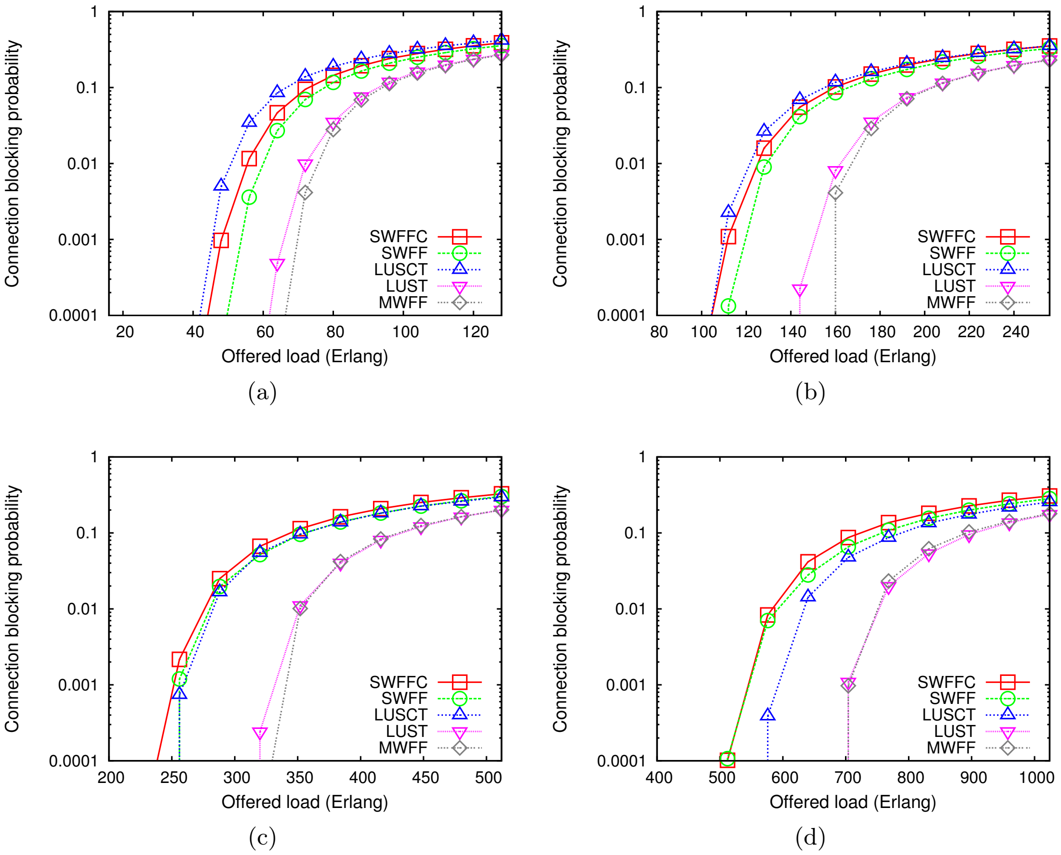
5.7.1 Blocking Probability and Throughput Performance

We will first assess the performance of the five scheduling policies analyzed in the ILP formulation section as a function of the offered load to the network and the number of wavelengths available per link, from 8 to 64. Fig. 5.24 shows the connection blocking probability for each of the policies obtained in the TID network with 8, 15, 32 and 64 wavelengths. As expected and seen previously in Fig. 5.15(a) and Fig. 5.16(a), the multi-wavelength first-fit tunable (MWFF) provides the best blocking performance among the five policies under test. This is due to its flexibility of assigning sub-wavelength slots on different wavelengths and without slot continuity, thus truly taking advantage of the free capacity available on the lightpath. Also, the LUST heuristic performs almost at the same level as MWFF.

Surprisingly, the least-used-slot continuous tunable (LUSCT) does not provide better results across all the different wavelength scenarios than its non-tunable counterpart SWFFC. For a small number of wavelengths on the network, LUSCT blocking probability is worse than SWFFC. However, as the number of wavelengths per link increases (see transition from Fig. 5.24(a)-(d)), so does the performance of LUSCT. This is related to the fact that, when the number of wavelengths is small, the use of the tunable relaxation produces more fragmentation on the slot continuity across the network by using many wavelengths per connection with respect to the total number of wavelengths per link. On the other side, when more wavelengths are available, such fragmentation among connections is more unlikely to occur due to the greater number of wavelengths. These results will be corroborated later (in Section 5.7.3) when assessing the number of wavelengths used per connection by each of the heuristics.

To further illustrate this behavior, Fig. 5.25(a) and Fig. 5.25(b) show the connection blocking probability as a function of the offered load as Erlangs per wavelength, so as to make a fairer comparison among scenarios with various wavelengths. We only show the results for the two algorithms we introduced in Section 5.6, SWFFC and LUSCT. As seen from the figures, the wavelength improvement of the LUSCT is sharper than in SWFFC. Hence, it can be suggested that the former performs better as the number of wavelengths per link increases. Basically, this has to do with the benefits that the greater number of wavelengths on the network provides to the second algorithm. As introduced, LUSCT is allowed to switch wavelengths, so that by having more wavelengths to switch amongst, the algorithm yields to a better resource allocation and utilization. Nonetheless, it is worth noting that as we show in Fig. 5.24(a) and 5.24(b), the blocking performance of LUSCT with fewer wavelengths is actually worse than in SWFFC.

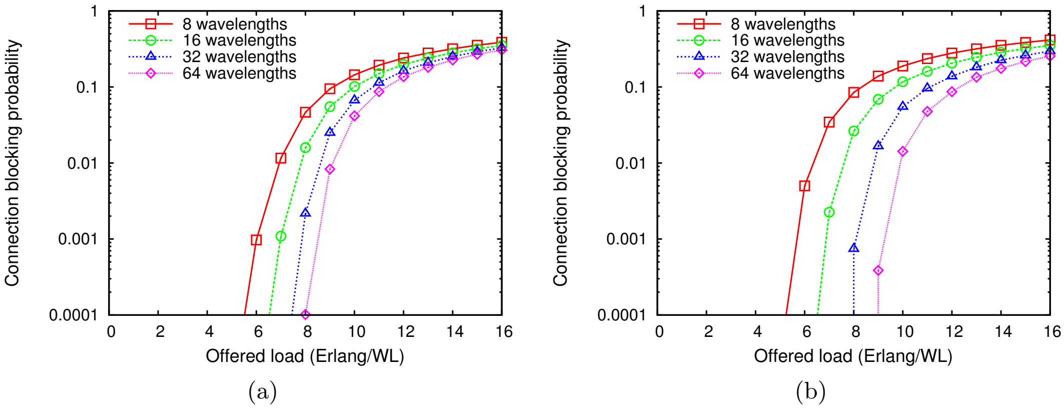
As for the EON is concerned, the graphs in Fig. 5.26 show the blocking probability results also for 8 to 64 wavelengths. As it can be observed, the performance comparison among the heuristics is very similar to the TID network scenario. So is the performance



(a) (b)

(c) (d)

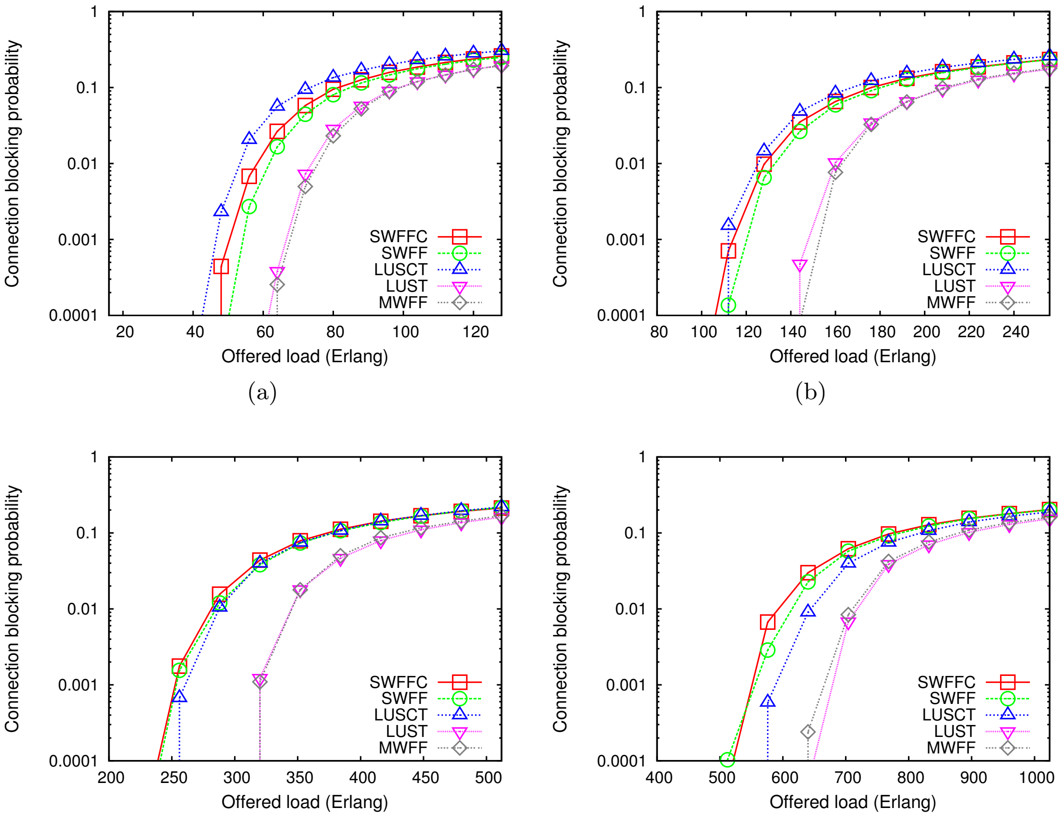
Fig. 5.24: Connection blocking probability results in TID: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths.



(a) (b)

Fig. 5.25: Blocking probability wavelength comparison of (a) SWFFC and (b) LUSCT in TID network.

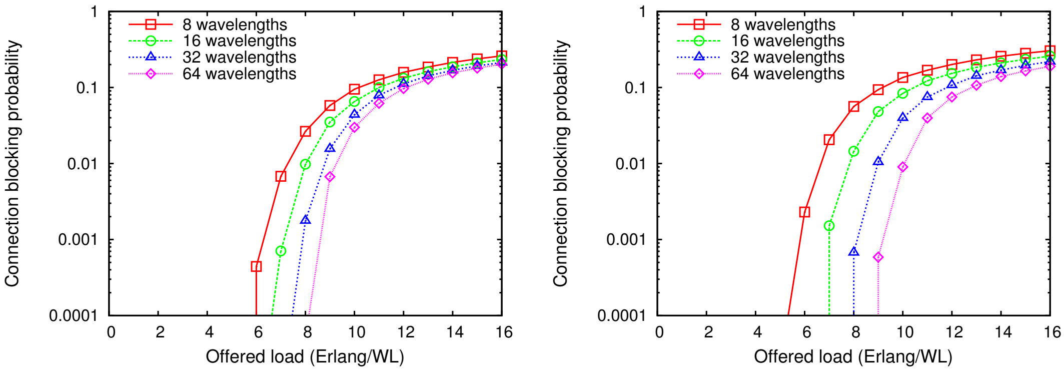
when comparing the SWFFC and LUSCT across the wavelength set, as depicted in Fig. 5.27(a) and Fig. 5.27(b), respectively. It is worth noting that although qualitatively these results are similar with those obtained from the TID network, quantitatively



(a) (b)

(c) (d)

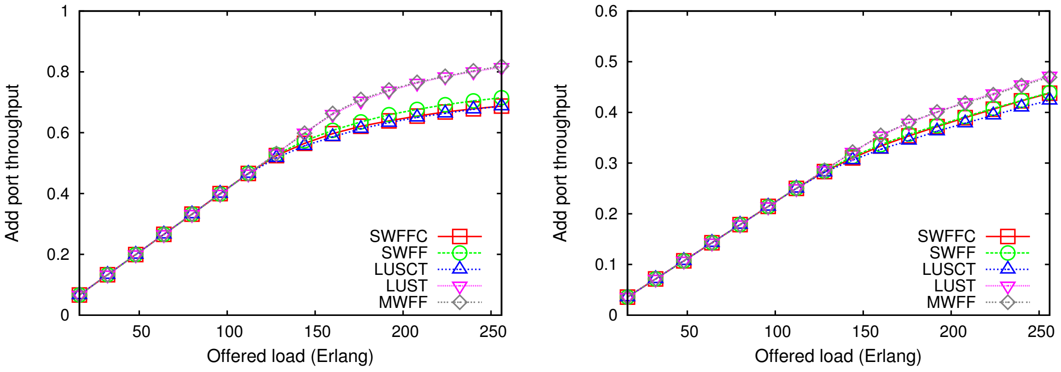
Fig. 5.26: Connection blocking probability results in EON: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths.



(a) (b)

Fig. 5.27: Blocking probability wavelength comparison of (a) SWFFC and (b) LUSCT in EON.

these are different. If we, for instance, compare Fig. 5.24(c) and Fig. 5.26(c) with 32 wavelengths, we can see that in EON the single-wavelength and continuity heuristic (SWFFC) experiences a slightly lower blocking probability. These variations can be



(a) (b)

Fig. 5.28: Normalized throughput per add port with 16 fixed wavelength transmitters: (a) TID network and (b) EON.

explained by the network size difference between the two topologies. Although in EON the average shortest path is longer than on the TID network, at the same offered load and due to the greater number of nodes in EON, the connection requests are more distributed in the former case. This behavior is also related to using a centralized PCE-SLAE architecture which is able to balance connections over the available k-shortest paths having full network resources utilization information.

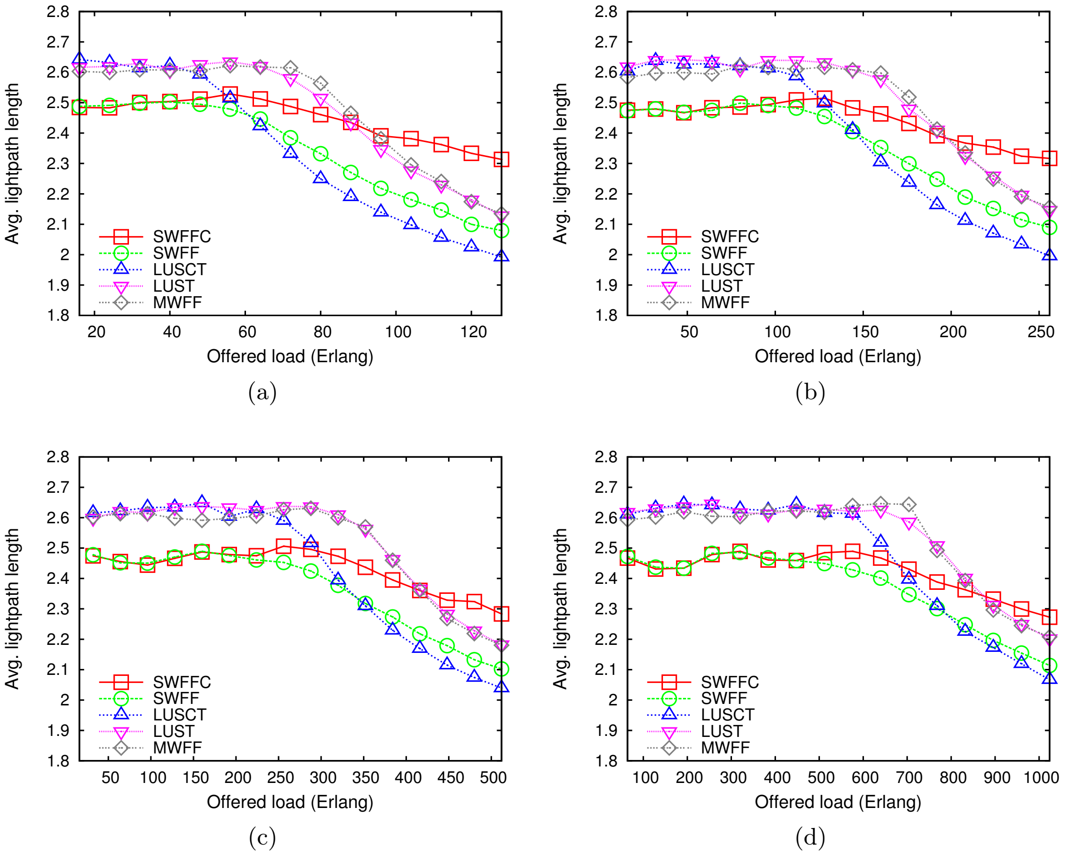
With the throughput performance we are able to assess the ability of each network node to add traffic to the network. Fig. 5.28(a) shows the average add port throughput per node as a function of the offered load in the TID network. Values are normalized by the total add port capacity of the node, which for this specific case with 16 wavelengths is equivalent to 160 Gbps. We can see that the two least constraint policies (LUST and MWFF) achieve a higher throughput at higher offered loads due to their greater ability to allocate time-slots non-continuously and on different wavelengths.

As for the EON topology results shown in Fig. 5.28(b), the performance comparison is similar to the TID network. However, in this case, due to the greater number of nodes on the network in comparison with TID, the averaged normalized throughput per node is lower at the same offered traffic load.

5.7.2 Average Lightpath Length

The graphs in Fig. 5.29 and Fig. 5.30 show the average path length (in number of hops) as a function of the offered traffic load to the network (in Erlang) for the TID network and EON, respectively. As usual, while increasing the offered load, the routes followed by the sub-wavelength successful connections tends to shorten. This is due to the higher load on the network and that those connections using shorter paths tend to be accepted with higher probability.

If we focus first on the TID scenario, we can observe four main trends across all the



(a) (b) (c) (d)

Fig. 5.29: Average path length in TID: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths.

graphs from 16 to 64 wavelengths: (1) the trend followed by SWFFC, (2) the one of SWFF, (3) another one for LUSCT, and finally (4) the trend followed by both LUST and MWFF. Among all, the heuristics that enable the tunable allocation of the time-slots (i.e., not a single wavelength) show a sharper decrease of the lightpath length from half the offered load onwards. Recall that all these algorithms make use of the tandem PCE-SLAE route, wavelength and time-slot assignment.

In the case of LUST, LUSCT and MWFF, the PCE is only responsible for providing routes whose link aggregated free bandwidth provides enough bandwidth for the upcoming connection request. Such a list is ordered in increasing order of used bandwidth, or what is the same in decreasing order of free bandwidth. Later, the SLAE takes this input list and computes the finer time-slot granularity based on the heuristic policy. This approach is different to the single-wavelength policies wherein the PCE computes a list of lightpaths which is further processed by the SLAE to compute the finer sub-wavelength slot allocation.

In general, the lightpaths allocated to the connections at low loads are shorter in the SWFFC and SWFF algorithms. We refer in this case to the PCE which seems to

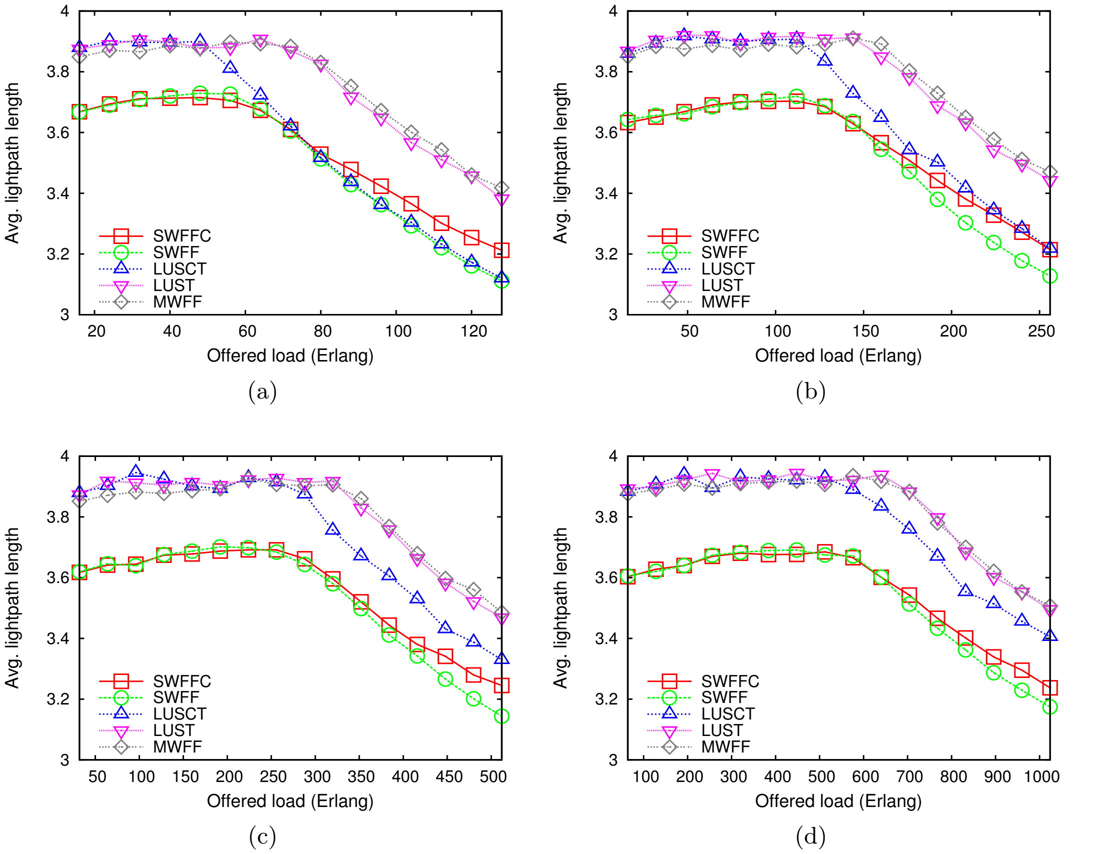


Fig. 5.30: Average path length in EON: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths.

filter out shorter lightpaths when computing these using the aggregated free bandwidth per link and wavelength, instead of only per link. We can explain this from the fact that by restricting to single-wavelength scheduling decreases the chances to find greater lightpath free bandwidth on longer paths. As a result, on average, the paths followed by SWFF and SWFFC are shorter at low loads. Also, at higher loads, the blocking probability of the single-wavelength policies is always higher than for their tunable alternatives. As a result of this, fewer connections are accepted contributing to the network resources utilization. This, in turn, avoids these single-wavelength policies experiencing a big drop in terms of average lightpath length.

Among the tunable policies, the LUSCT results in shorter lightpaths when the offered load increases. The continuity constraint imposed by the policy allows for lesser chances to schedule the time-shared connection request.

With regards to EON, the heuristics perform similarly as in the TID network as we can see in Fig. 5.30(a)-(d). Obviously, because the mean route length in EON is longer than in TID, on average, the graphs also show a longer lightpath length. Finally, we note the path length difference between the tunable and single-wavelength policies is

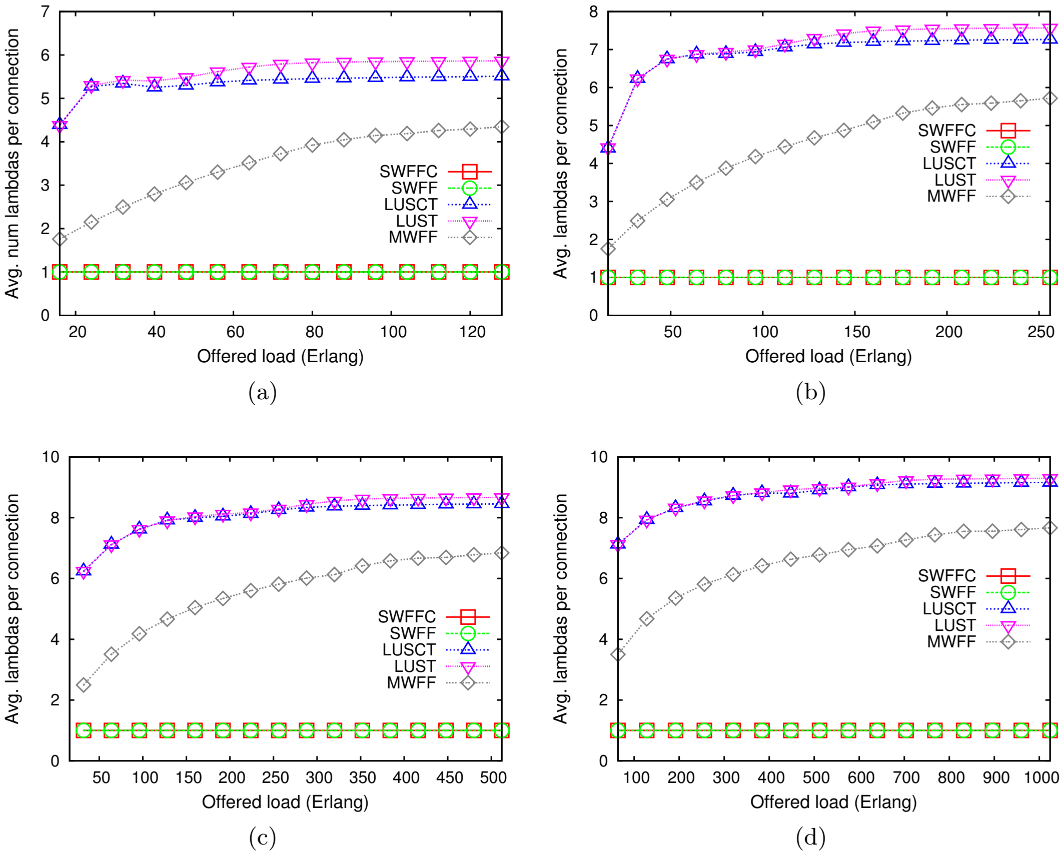


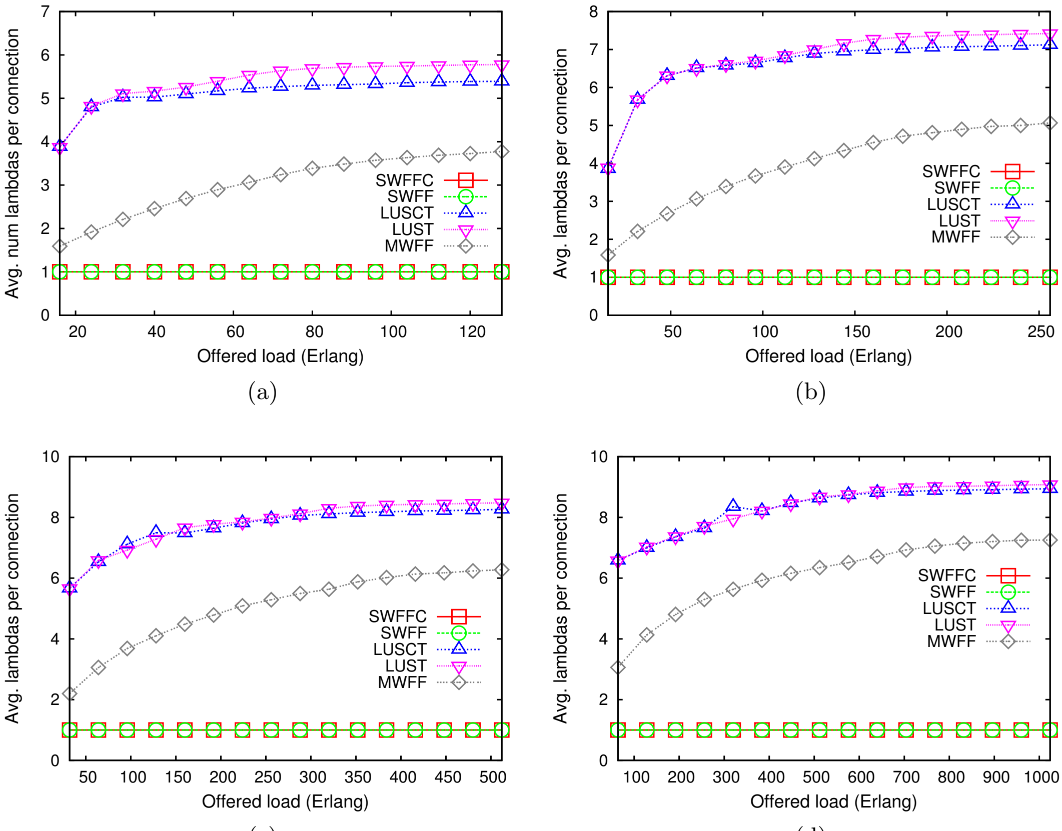
Fig. 5.31: Average number of wavelengths used per connection in TID: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths.

greater in EON.

5.7.3 Wavelengths Used per Connection

Next parameter to assess is the average number of wavelengths used per successful connection request. Fig. 5.31(a)-(d) show the results while changing the number of wavelengths in the TID network. As expected, the heuristics SWFFC and SWFF can only use a single wavelength per connection, which is reflected on each graph. The other three heuristics are allowed to allocate time-slots on different wavelengths. Among these, the MWFF shows an increasing tendency from low to high loads. This is due to the particular implementation of the heuristic that tries to allocate first time-slots along one of the wavelengths before switching to the next one, so, it packs the allocated time-slots in as fewer wavelengths as possible.

Another interesting result is the wavelength utilization when increasing the number of wavelengths in the network, in particular for the LUSCT and LUST heuristics. In this case, when the links on the network have 8 wavelengths, the utilization per connection, almost for the entire offered load range, is about 5.5 wavelengths. As the



(a)

(b)

(c)

(d)

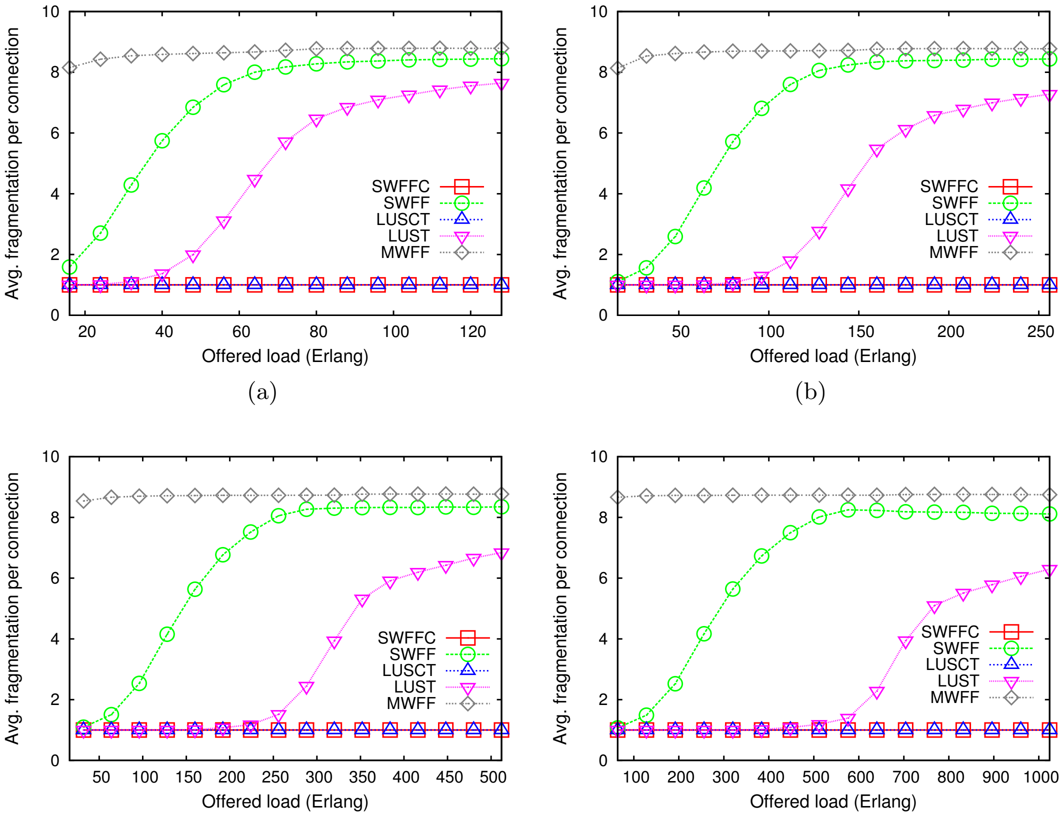
Fig. 5.32: Average number of wavelengths used per connection in EON: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths.

number of wavelengths per link is increased, so is the number of wavelengths used per successful connection, approaching the upper bound of 10, which in this particular case corresponds to the number of time-slots requested. A 1 Gbps connection request equals to 10 time-slots within a time-shared wavelength frame of 100 time-slots length. It is worth noting that the LUST and LUSCT implementations are not optimized to pack connections in as fewer number of wavelengths as possible. Therefore, by the heuristic operation itself, this value tends to be high.

As for the EON topology is concerned, the results shown in Fig. 5.32 are very similar to the ones obtained in TID and the same description applies.

5.7.4 Fragmentation per Connection

The last set of results have to do with the fragmentation per successful connection. That is, in how many pieces or continuous set of time-slots connections are allocated. As it has been pointed out, the arriving connections into the system request for a maximum of 10 time-slots, hence the maximum number of "fragments" cannot be greater than this value.



(a)

(b)

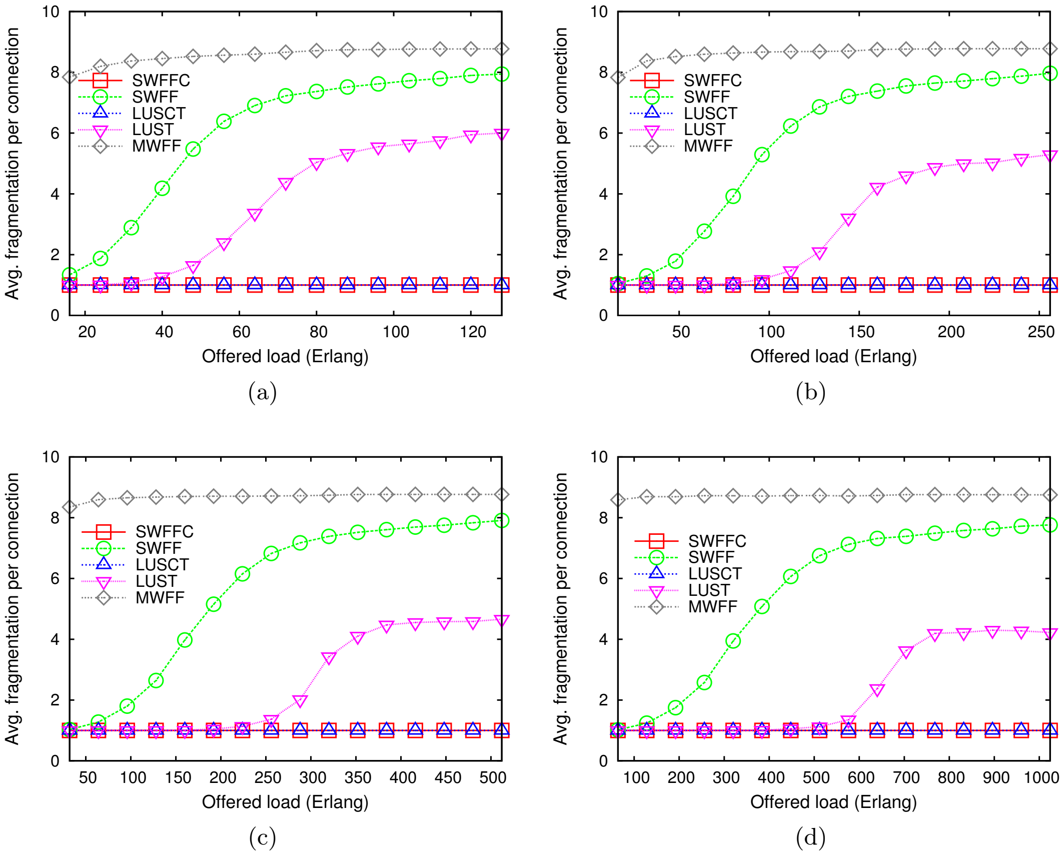
(c)

(d)

Fig. 5.33: Average fragmentation per connection in TID: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths.

Fig. 5.33(a)-(d) show the average fragmentation per successful connection as a function of the offered load (in Erlang) to the TID network. Obviously, the heuristics based on the continuous allocation of time-slots, namely SWFFC and LUSCT, have an average of 1.0, i.e., only one fragment is allowed by the policy constraints. The other three heuristics perform differently when increasing the offered load. MWFF is the heuristic that produces the highest level of fragmentation. Almost for the whole load under consideration in every wavelength scenario, fragmentation moves around 8-8.5 fragments on average. This is explained due to the fact that the implemented policy for MWFF is to pack connections into fewer number of wavelengths by exploring first all slots on a wavelength before switching to the next one. This produces more fragmentation within the same wavelength.

Another heuristic that shows the same performance across the different wavelength scenarios is SWFF whose fragmentation per connection stands in between the MWFF and LUST. This is mainly due to packing the whole connection within a single wavelength which yields the algorithm to produce more fragmentation, similarly as reported by MWFF. On the contrary, because LUST is allowed to use different wavelengths, even



(a) (b)

(c) (d)

Fig. 5.34: Average fragmentation per connection in EON: (a) 8 wavelengths, (b) 16 wavelengths, (c) 32 wavelengths, and (d) 64 wavelengths.

not continuously, it spreads the time-slot allocation on a fewer number of fragments. As we have seen in Fig. 5.31, LUST is one of the heuristics that uses more lambdas per connection on average. Finally, we can also see that when the number of wavelengths per link is increased, LUST and SWFF start rising later, especially for the former.

Again, as for the EON topology is concerned, the results shown in Fig. 5.34 are very similar to the ones obtained in TID and the same description applies.

5.8 Summary

Network operators and media content providers could achieve successful business opportunities by enabling new network centric services combining network and IT resources. However, some of these services require guaranteed transmission on the network at the sub-wavelength level. That is, after a connection request has successfully been allocated, it requires a guaranteed and loss-free transmission. Furthermore, most of these applications do not request for full-wavelength capacity.

In this chapter, we have proposed several time-shared optical network (TSON) architectures to accomplish the all-optical bandwidth provisioning. One of the objectives is to tie the architecture as much as possible to current control plane implementations looking further to achieve control plane interworking both, vertically among technologies and horizontally among network domains. To this end, we have reviewed three architectures based on GMPLS/PCE/TSON interworking. Among all, the PCE/TSON based architecture provides the best delay-to-service and connection blocking probability.

Using as an input the centralized sub-wavelength routing, wavelength and time-slot assignment from PCE/TSON, we have also formulated five different time-slot scheduling policies from lower to higher level of constriction. ILP formulations together with heuristics have been proposed and assessed on a small 6-node network topology with static traffic input. Our findings corroborate that blocking probability is highly influenced by the time-slot assignment flexibility.

Finally, in the last sections of the chapter we have implemented the aforementioned five heuristics into the dual PCE-SLAE for the dynamic assessment of the proposed heuristics. Two different topologies have been used: a metropolitan-based optical network and a wide-area European optical network. Several parameters have been evaluated showing the benefits in terms of blocking probability of the more flexible time-slot assignment algorithms at the expense of increasing the complexity in terms of connection fragmentation and number of wavelengths used per successful connection.

Chapter 6

Conclusion

Internet has experienced a tremendous evolution from its ARPANET origins. Nowadays, new video and other network centric applications push the underlying network infrastructures to provide the necessary bandwidth to accommodate the growing traffic demands. That is, the supporting network technologies need to keep up this technological revolution in order to deliver this traffic growth.

As it is well known, optical network technologies supply huge bandwidth thanks to the capabilities of the wavelength division multiplexing. WDM provides the ability to transmit on several optical channels through the same fiber. Current developments are able to achieve Terabit bandwidth in a single fiber.

It is also true that not all applications require full-wavelength capacity. Optical grooming using optical-to-electrical-to-optical conversion is currently the main choice with regards to sub-wavelength provisioning. However, such a traffic grooming cannot benefit from all-optical switching and protocol transparency.

In this Thesis we contributed with new sub-wavelength optical network architectures using different communication paradigms. Namely, we explored three different paradigms: connectionless (CL), connection-oriented (CO) and hybrid CL/CO.

In connectionless sub-wavelength optical networks, like packet or burst switching, data loss due to contention is one of the main issues. Even at very low network loads, due to the use of one-way provisioning protocols and the lack of efficient buffering methods in the optical domain, contentions can occur. Approaches to cope with contentions include reactive resolution schemes, such as fiber delay lines, deflection routing and wavelength conversion, and proactive schemes like access control protocols and efficient RWA. As for RWA is concerned, we introduced in Chapter 3 an auto load-balancing distributed RWA algorithm for optical burst-switched networks with wavelength-continuity constraint.

The RWA and burst forwarding are based on the exploitation and exploration facilities using switching rule concentration values that incorporate contention and forwarding desirability information for every wavelength and port belonging to the optical switch. The burst route is generated hop-by-hop based on the forwarding rules. To support such architecture, forward and backward control packets are used in the burst forwarding and updating rule processes, respectively.

The proposed RWA was extensively analyzed, first on a simple scenario, and later on four different network topologies. Results showed that the proposed method outperforms the rest of RWA algorithms under test. Performance improvements changed from one network to another and even with its auto-load balancing capabilities, the algorithm did not damage other parameters like end-to-end delay in comparison to shortest path routing. We checked that the size of the network (i.e., mean route length) is the main variable that penalizes the algorithm performance, mostly due to the wavelength-continuity constraint. Also, we derived mathematically the long-term switching concentration values under an ideal scenario so as to gain insight on the parameterizations setup involved in the protocol. In summary, despite the goodness of the RWA to proactively reduce burst contentions in the core of the network, these cannot be totally canceled.

Even in the case when providing high-priority treatment in connectionless sub-wavelength optical networks, data delivery cannot be guaranteed. Another way to cope with contentions is by controlling the burst transmission into the channel. Medium access control protocols can help reduce the burst collision and acknowledge the transmission only when the data channel is free. We proposed in Chapter 4 a MAC-based hybrid CL/CO sub-wavelength optical network architecture which is able to cancel burst contentions for in-transit bursts.

The provisioning of hybrid connectionless and connection-oriented services over the same optical network substrate is not straightforward. A possibility to realize this, as stated in the background chapter, is to physically separate the network resources into two groups, one for each service type, respectively. Another option is to virtually separate the resources, but still maintaining resources and operation independence. In this Thesis we followed a different approach by enabling true resources sharing between CL and CO services using the MAC protocol. One of the benefits of the architecture is that it provides multi-service capabilities with diverse QoS guarantees both through connectionless or connection-oriented access modes which can potentially serve a broad range of applications with diverse QoS requirements.

The MAC provides two main access methods: queue arbitrated (QA) for connectionless bursts and pre-arbitrated (PA) for TDM connection-oriented services. On the one hand, the queue arbitrated access is based on a counting and monitoring process of burst and request control packets traveling in opposite control channel directions and a distributed preemption-based scheme in a multi-queue access priority system. The QA multi-queue system allows different traffic classes to be defined providing relative QoS differentiation of burst blocking probability and end-to-end delay. On the other hand, the pre-arbitrated mode is based on the pre-reservation of slots supported by a higher service layer module. A benefit of the MAC protocol is that contentions for in-transit bursts are avoided, thus guaranteeing the data delivery even for connectionless applications.

Results evaluated through simulations showed that highest priority bursts are lossless and experiment the lowest access latencies in the QA access mode. Even for the intermediate traffic class, the protocol could guarantee an acceptable burst blocking probability for a diverse number of applications. We also found out that the performance of lower traffic classes is defined by the percentages of traffic of other higher priority classes.

Regarding the PA access mode, results showed that doubling the offered TDM traffic load increases in more than one order their connection blocking probability, slightly affecting the blocking of connectionless bursts transmitted through the QA access mode. Moreover, three different slot scheduling algorithms for allocating TDM connections along with connectionless bursts were also evaluated, providing diverse results depending on the requested bandwidth. The overall results demonstrated the suitability of the proposed architecture for future integrated multi-service optical networks.

With regards the hybrid CL/CO MAC protocol, we also tackled two important issues. The first dealt with the access delay due to the queue arbitrated mode. We represented the MAC as a tandem queue system composed of local queues and a virtual distributed queue. We then modeled mathematically the access delay with a lower and upper approximation. For the lower approximation we used an M/G/1/-/N queue analysis, and for the upper approximation an M/G/1. Overall, we detected that the model was able to compute approximately the access delay in a single priority class scenario. As a consequence, the model may be used to quantify the expected access delay in order improve the wavelength assignment and reduce the overall end-to-end delay.

The second issue we handled was the virtual light-tree overlay topology optimization under a static traffic demand. We formulated mathematically the generation of such topology with the objective to minimize the overall cost. As such, we contributed with some simple constraint extensions to generate the virtual topology when light-tree demands are fractional to the wavelength capacity, as it is the topic of this Thesis. We checked the correctness of the formulation under two simple static traffic scenarios.

After assessing the pure connectionless mode and the hybrid CL/CO MAC-based sub-wavelength optical network, we proposed in Chapter 5 several time-shared optical network (TSON) architectures to accomplish guaranteed all-optical bandwidth provisioning. From a network operator's perspective, we tied the architecture as much as possible to current control plane implementations. The objective was to achieve control plane interworking both vertically among technologies and horizontally among network domains. To this end, we reviewed three architectures based on GMPLS/PCE/TSON interworking. Among all, the PCE/TSON based architecture provisioned the best delay-to-service and connection blocking probability.

Using as an input the centralized sub-wavelength routing, wavelength and time-slot assignment from PCE/TSON, we also formulated posteriorly five different time-slot scheduling policies from lower to higher level of constriction. ILP formulations together with heuristics were proposed and assessed on a small 6-node network topology with static traffic input. Our findings corroborated that the blocking probability is highly influenced by the time-slot assignment flexibility. To illustrate this, the policy which allowed for non-continuity time-slot allocation of multiple wavelengths obtained the best blocking probability. However, we also observed that the greater the flexibility, the more is the implementation complexity of the node to queue and assemble the bursts due to bandwidth fragmentation.

To evaluate the performance of the heuristics under a more realistic scenario, we implemented the same five time-sharing sub-wavelength policies under a dynamic traffic setup. As such, and in order to pursue for a real path computation environment, we designed a dual RWTA: the coarser route and wavelength assignment was assigned to the standardized PCE, whereas the finer time-slot (sub-wavelength) assignment was committed to the SLAE. To assess the performance of this architecture, two different topologies were used: a metropolitan-based optical network and a wide-area European optical network. Several parameters were analyzed showing the benefits in terms of blocking probability of the more flexible time-slot assignment algorithms at the expense of increasing the complexity in terms of connection fragmentation and number of wavelengths used per successful connection, just as we concluded from the static scenario.

This Thesis has covered quite diverse methodologies and proposed different architectures. We conclude that none of the architectures is the best overall. Such a statement greatly depends on the applications and services we expect to support on the network. Each of the architectures had its advantages and drawbacks. Even the maybe more interesting hybrid CL/CO required to assess the extra access delay incurred in the connectionless access mode and the optimization of the light-tree virtual topology. Despite this, by allowing resources sharing among diverse CL and CO-based applications, we can utilize resources more efficiently. It is clear though that the limitations of current optical enabling devices, both technologically and economically, do not seem to offer the required QoS guarantees for which clients are willing to pay and network operators to support. Therefore, solutions like TSON can become a feasible solution in the mid-term while further research on pure optical connectionless provisioning makes true such option.

Chapter 7

Future Research and Work

In this Thesis we focused on some architectures we believe are of interest for subwavelength provisioning in future optical networking. Needless to say that the topic can be quite broad, even when only considering some specific parts, as we did. In this chapter we introduce some of the open issues not addressed by this Thesis and other findings worth to evaluate. We divide the future research into three parts, one for each main contributing chapter.

7.1 Chapter 3: ALBA

In this chapter we addressed the burst contention issue in connectionless optical burst-switched networks with wavelength-continuity constraint. We saw through simulation that the proposed RWA algorithm was able to decrease the burst blocking probability due to burst contention in comparison with the rest of protocols under test. However, we believe further work can be done in the following fields:

- In order to improve the simulation performance assessment it would be interesting to compare the proposed ALBA algorithm with other load-balancing RWA for OBS. Some of the options to consider are [72, 69, 141].
- Although wavelength conversion can be an expensive addition to the optical network, it is the most effective way to decrease the blocking probability. ALBA was initially conceived for wavelength-continuity constraint optical networks. Therefore, a possible way to improve the results obtained with ALBA is actually extend it for wavelength conversion capable networks. This would possibly imply to redefine the assignment, forwarding and updating rules of the protocol to take into account the greater flexibility in the wavelength domain.

7.2 Chapter 4: DAOBS

This chapter devoted to the hybrid CL/CO. To this end, a MAC-based protocol for OBS, named DAOBS, was proposed. We assessed the protocol under different traffic distributions concluding that this architecture could deliver a multi-service platform with low-complexity implementation. We also addressed a pair of issues related to the access delay in the queue arbitrated mode and the light-tree virtual topology optimization for static traffic demands. Possible ways to improve what has already been addressed in this Thesis:

- A mathematical model for the queue arbitrated access mode was proposed. However, this model only considered a single priority class, and as we described, the QA model can treat more than a single class (e.g., in the simulation performance evaluation we tested 3 priority classes). Therefore, a possible extension is to consider the multi-queue priority case in the mathematical model.
- Also, a model capable of measuring the blocking probability could turn to be useful while deciding the DAOBS entity/wavelength assignment.
- One of the limitations of the proposed model is that it requires to setup the light-tree virtual topology or lightpaths to instantiate the DAOBS entities on each of the nodes. This, in turn, requires to have a certain knowledge of the traffic demands or expected traffic throughout time. In the performance evaluation section we assumed a uniform traffic pattern among the nodes on the network, i.e., traffic to the remaining nodes is the same. As such, we used an heuristic to setup the virtual topology by making use of shortest spanning trees. Therefore, it would be useful to also assess the protocol having in mind other possible traffic matrices and develop the required virtual topology heuristics accordingly. As suggested in the external review process, a future/follow-up work of this contribution would be to consider mathematical approaches such as column generation (i.e., a column is a good candidate light-tree), instead of heuristic approaches to solve the dual-layer design problem to cope with the complexity of the ILP.
- Finally, another enhancement would be to improve the resources utilization within the virtual topology by enabling spatial reuse of slots within the light-tree itself. That is, to propose a mechanism for reusing free slots in the light-tree whose transmission would no content with others taking place in other links.

7.3 Chapter 5: TSON

The last contribution tackled the provisioning of guaranteed connection-oriented time-shared for sub-wavelength provisioning in optical networks. We proposed three architectures accomplishing different levels of control plane inter-working. We also contributed with the formulation for the optimization of five different time-slot assignment policies to combine with the connection-oriented approach. Their corresponding heuristics were also assessed on both static and dynamic traffic scenarios. Future work to consider include:

- As far as standardization is concerned, GMPLS lacks all-optical sub-wavelength provisioning. Extensive work is required to propose and adopt new ways to label such lightpaths and make it compatible with the rest of the GMPLS stack.
- Another task is to improve the distributed SLAE GMPLS/TSON architecture. New mechanisms to suggest labels distributively could help decrease the connection blocking probability. This issue is being currently under investigation in full-wavelength OCS networks, but when it comes to sub-wavelength provisioning in all-optical networks, this is yet an unexplored scenario.

Chapter 8

Publications

This is a chronological list of publications that I have produced and published throughout this Thesis.

8.1 Publications Related to this Thesis

Journals

- Chapter 4: **J. Triay** and C. Cervelló-Pastor, “Delay Analysis of Slotted OBS Networks Under DAOBS MAC Protocol,” *IEEE Communications Letters*, vol. 15, no. 8, pp. 778–780, Aug. 2011.
- Chapter 4: **J. Triay**, G. S. Zervas, C. Cervelló-Pastor, and D. Simeonidou, “Multi-service QoS-enabled MAC for Optical Burst Switching,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 2, no. 8, pp. 530–544, Aug. 2010.
- Chapter 3: **J. Triay** and C. Cervelló-Pastor, “An Ant-based Algorithm for Distributed Routing and Wavelength Assignment in Dynamic Optical Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 4, pp. 542–552, May 2010.

Conferences and Workshops

- Chapter 5: G. Zervas, **J. Triay**, N. Amaya, Y. Qin, C. Cervelló-Pastor, and D. Simeonidou, “Time Shared Optical Network (TSON): A Novel Metro Architecture for Flexible Multi-Granular Services,” in *Proc. of 37th European Conference on Optical Communication (ECOC) 2011*, Geneva, Switzerland, Sep. 2011, pp. 1–3.
- Chapter 3: **J. Triay** and C. Cervelló-Pastor, “Topology Analysis of Auto Load-Balancing RWA in Optical Burst-Switched Networks,” in *Proc. 2011 IEEE Wireless and Optical Communications Conference (WOCC)*, Newark, NJ, USA, Apr. 2011, pp. 1–6.

- Chapter 5: **J. Triay**, G. Zervas, C. Cervelló-Pastor, and D. Simeonidou, "GMPLS/PCE/OBST Architectures for Guaranteed Sub-Wavelength Mesh Metro Network Services," in *Proc. of Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC) 2011*, Los Angeles, CA, USA, Mar. 2011, pp. 1–3.
- Chapter 4: **J. Triay**, G. Zervas, C. Cervelló-Pastor, R. Nejabati, and D. Simeonidou, "Multi-service MAC-enabled OBS Networks," in *Proc. of Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC) 2010*, San Diego, CA, USA, Mar. 2010, pp. 1–3.
- Chapter 4: **J. Triay**, G. Zervas, C. Cervelló-Pastor, R. Nejabati, and D. Simeonidou, "QoS-enabled Distributed Access on Optical Burst-Switched Networks," in *Proc. 14th Conference on Optical Network Design and Modeling (ONDM) 2010*, Kyoto, Japan, Feb. 2010, pp. 1–6.
- Chapter 3 and 4: C. Cervelló-Pastor, **J. Triay**, and S. Sallent, "Distributed Resources Assignment for Optical Burst Switching (Invited Paper)," in *Proc. of 2009 International Workshop on Optical Burst/Packet Switching (WOBS), collocated with 6th Int. Conf. on Broadband Communications, Networks and Systems (BROADNETS)*, Madrid, Spain, Sep. 2009, pp. 1–8.
- Chapter 4: **J. Triay**, J. Perelló, C. Cervelló-Pastor, and S. Spadaro, "On Avoiding-Minimizing Burst Collisions in Optical Burst-Switched Networks without Wavelength Conversion," in *Proc. of 11th Int. Conf. on Transparent Optical Networks (ICTON)*, Azores, Portugal, Jun. 2009, pp. 1–4.
- Chapter 3: **J. Triay** and C. Cervelló-Pastor, "An Ant-based Algorithm for Distributed RWA in Optical Burst Switching," in *Proc. of 11th Int. Conf. on Transparent Optical Networks (ICTON)*, Azores, Portugal, Jun. 2009, pp. 1–4.
- Chapter 4: **J. Triay** and C. Cervelló-Pastor, "Distributed Contention Avoidance in Optical Burst-Switched Ring Networks," in *Proc. of 11th IEEE Singapore International Conference on Communication Systems (ICCS) 2008*, Guangzhou, China, Nov. 2008, pp. 715–720.
- Chapter 4: **J. Triay** and C. Cervelló-Pastor, "QoS Multi-Spanning Tree for Optical Burst Switching," in *Proc. of 4th IEEE International Telecommunication Networking Workshop on QoS in Multi-service IP Networks (IT-NEWS)*, Venice, Italy, Feb. 2008, pp. 89–94.
- Chapter 4: **J. Triay** and C. Cervelló-Pastor, "A Traffic Engineered Multi-Spanning Tree for Optical Burst Switching," in *Proc. of EuroFGI Workshop on IP QoS and Traffic Control*, Lisbon, Portugal, Dec. 2007.

8.2 Other Publications

Journals

- P. R. Pereira, A. Casaca, J. Rodrigues, V. N. Soares, **J. Triay**, and C. Cervelló-Pastor, "From Delay-Tolerant Networks to Vehicular Delay-Tolerant Networks," *IEEE Communications Surveys and Tutorials.* To appear.

Conferences and Workshops

- **J. Triay**, C. Cervelló-Pastor, and V. M. Vokkarane, "Analytical model for Hybrid Immediate and Advance Reservation in Optical WDM Networks," in *IEEE Global Communications Conference (GLOBECOM) 2011*, Houston, TX (USA), Dec. 2011. To appear.
- **J. Triay**, D. R. Rousseau, C. Cervelló-Pastor, and V. M. Vokkarane, "Dynamic Service-Aware Reservation Framework for Multi-Layer High-Speed Networks," in *Proc. 5th Workshop on Performance Modeling and Evaluation in Computer and Telecommunication Networks (PMECT) collocated with ICCCN 2011*, Maui, HI, USA, Jul. 2011, pp. 1–7.
- K. Bhaskaran, **J. Triay**, and V. M. Vokkarane, "Dynamic Anycast Routing and Wavelength Assignment in WDM Networks Using Ant Colony Optimization (ACO)," in *Proc. of IEEE International Conference on Communications (ICC) 2011*, Kyoto, Japan, Jun. 2011.
- E. Escalona, Y. Qin, G. Zervas, **J. Triay**, G. Zarris, N. Amaya-Gonzalez, R. Nejabati, C. Cervelló-Pastor, and D. Simeonidou, "Experimental Demonstration of a Novel QoS-Aware Hybrid Optical Network," in *Proc. 36th European Conference and Exhibition on Optical Communication (ECOC) 2010*, Torino, Italy, Set. 2010, pp. 1–3.
- G. S. Zervas, G. Zarris, N. Amaya-Gonzalez, **J. Triay**, E. Escalona, Y. Qin, C. Cervelló-Pastor, R. Nejabati, and D. Simeonidou, "Experimental Demonstration of a Self-Optimised Multi-Bit-Rate Optical Network," in *Proc. 36th European Conference and Exhibition on Optical Communication (ECOC) 2010*, Torino, Italy, Set. 2010, pp. 1–3.
- Y. Qin, K. Cheng, **J. Triay**, E. Escalona, G. S. Zervas, G. Zarris, N. Amaya-Gonzalez, C. Cervelló-Pastor, R. Nejabati, and D. Simeonidou, "Demonstration of a C/S based Hardware Accelerated QoT Estimation Tool in Dynamic Impairment-Aware Optical Network," in *Proc. 36th European Conference and Exhibition on Optical Communication (ECOC) 2010*, Torino, Italy, Set. 2010, pp. 1–3.
- D. Remondo, S. Sargento, M. Cesana, M. Nunes, I. Filipini, **J. Triay**, A. Agustí, M. De Andrade, L. Gutiérrez, S. Sallent, and C. Cervelló-Pastor, "Integration

of Optical and Wireless Technologies in the Metro-Access: QoS Support and Mobility Aspects," in *Proc. of 5th Euro-NGI Conference on Next Generation Internet Networks (NGI)*, Aveiro, Protugal, Jul. 2009, pp. 1–8.

- **J. Triay**, C. Cervelló-Pastor, M. Calderón, and P. J. Argibay, "Diseño e Implementación de un Prototipo de Red OBS," in *Actas Jornadas de Ingeniería Telemática (JITEL)*, Málaga, Spain, Sep. 2007, in Spanish.
- **J. Triay**, J. Rubio, and C. Cervelló-Pastor, "Design and Implementation of an OBS Control Plane Architecture," in *Proc. of 10th EUNICE Summer School 2006*, Stuttgart, Germany, Sep. 2006.
- J. Gonzalo, **J. Triay**, X. Hesselbach, and J. Abella, "ENIGMA: A Testbed for MPLS and QoS Integration on IP Networks," in *Proc. of 2nd Int. Conf. on Testbeds and Research Infrastructures for the Development of Networks and Communities (TRIDENTCOM) 2006*, Barcelona, Spain, Mar. 2006, pp. 558–563.

Bibliography

[1] Cisco, “Hyperconnectivity and the Approaching Zettabyte Era,” White Paper. [Online] Available at: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.pdf, Jun. 2010.

[2] S. Yao, B. Mukherjee, S. Yoo, and S. Dixit, “A unified study of contention-resolution schemes in optical packet-switched networks,” *IEEE/OSA Journal of Lightwave Technology*, vol. 21, no. 3, pp. 672–683, Mar. 2003.

[3] A. Gumaste, T. Das, A. Mathew, and A. Somani, “An autonomic virtual topology design and two-stage scheduling algorithm for light-trail WDM networks,” *IEEE/OSA Journal on Optical Communications and Networking*, vol. 3, no. 4, pp. 372–389, Apr. 2011.

[4] M. Klinkwoski, “Offset Time-Emulated Architecture for Optical Burst Switching - Modelling and Performance Evaluation,” Ph.D. dissertation, Dept. of Computer Architecture, Universitat Politècnica de Catalunya (UPC), Nov. 2007.

[5] IST IP Nobel Phase 2, “Project deliverable D1.1: Architectural vision of network evolution,” IST IP Nobel project consortium, Tech. Rep., Aug. 2006.

[6] S. Huang, R. Dutta, and G. Rouskas, “Traffic grooming in path, star, and tree networks: complexity, bounds, and algorithms,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 4, pp. 66–82, Apr. 2006.

[7] A. Jaekel, A. Bari, Y. Chen, and S. Bandyopadhyay, “New techniques for efficient traffic grooming in WDM mesh networks,” in *Proc. of 16th International Conference on Computer Communications and Networks (ICCCN) 2007*, Honolulu, HI, USA, Aug. 2007, pp. 303–308.

[8] K. Zhu and B. Mukherjee, “Traffic grooming in an optical WDM mesh network,” *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 1, pp. 122–133, Jan. 2002.

[9] W. Yao and B. Ramamurthy, “A link bundled auxiliary graph model for constrained dynamic traffic grooming in WDM mesh networks,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 8, pp. 1542–1555, Aug. 2005.

[10] Cisco, “Cisco Visual Networking Index: Forecast and Methodology, 20092014 ,” White Paper. [Online] Available at: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf, Jun. 2010.

[11] M. O’Mahony, C. Politi, D. Klonidis, R. Nejabati, and D. Simeonidou, “Future optical networks,” *IEEE/OSA Journal of Lightwave Technology*, vol. 24, no. 12, pp. 4684–96, Dec. 2006.

[12] S. Sarkar, S. Dixit, and B. Mukherjee, “Hybrid wireless-optical broadband-access network (WOBAN): A review of relevant challenges,” *IEEE/OSA Journal of Lightwave Technology*, vol. 25, no. 11, pp. 3329–3340, Nov. 2007.

[13] G. Kramer and G. Pesavento, “Ethernet passive optical network (EPON): building a next-generation optical access network,” *IEEE Communications Magazine*, vol. 40, no. 2, pp. 66–73, Feb. 2002.

[14] B. Mukherjee, *Optical WDM Networks*, 1st ed. Springer, 2006.

[15] S. Sygletos, I. Tomkos, and J. Leuthold, “Technological challenges on the road toward transparent networking,” *OSA Journal of Optical Networking*, vol. 7, no. 4, pp. 321–350, Apr. 2008.

[16] N. Ghani, S. Dixit, and T.-S. Wang, “On IP-over-WDM integration,” *IEEE Communications Magazine*, vol. 38, no. 3, pp. 72–84, Mar. 2000.

[17] International Telecommunication Union Standardization, “Generic framing procedure (GFP),” ITU-T Standard G.7041/Y.1303, Apr. 2011.

[18] ——, “Network node interface for the synchronous digital hierarchy (SDH),” ITU-T Standard G.707/Y.1322, Jan. 2007.

[19] ——, “link capacity adjustment scheme (LCAS) for virtual concatenated signals,” ITU-T Standard G.7042/Y.1305, Feb. 2004.

[20] S. Clavenna, “The Future of Sonet/SDH,” Heavy Reading, Tech. Rep., Nov. 2003.

[21] G. Bonaventura, G. Jones, and S. Trowbridge, “Optical transport network evolution: hot standardization topics in ITU-T including standards coordination aspects,” *IEEE Communications Magazine*, vol. 46, no. 10, pp. 124–131, Oct. 2008.

[22] S. Yao, B. Mukherjee, and S. Dixit, “Advances in photonic packet switching: An overview,” *IEEE Communications Magazine*, vol. 38, no. 2, pp. 84–94, Feb. 2000.

[23] S. Yao, S. Yoo, B. Mukherjee, and S. Dixit, “All-optical packet switching for metropolitan area networks: Opportunities and challenges,” *IEEE Communications Magazine*, vol. 39, no. 3, pp. 142–8, Mar. 2001.

[24] D. J. Blumenthal, J. E. Bowers, and C. Partridge, “Mapping a Future for Optical Networking and Communications,” GENI (Global Environment for Network Innovations), Tech. Rep. GDD-05-03, July 2005.

[25] M. Maier and M. Reisslein, “Trends in optical switching techniques: A short survey,” *IEEE Network*, vol. 22, no. 6, pp. 42–47, Nov. 2008.

[26] C. Qiao, W. Wei, and X. Liu, “Towards a polymorphous, agile and transparent optical network (PATON) based on polymorphous optical burst switching (POBS),” in *Proc. of 23th Conference of the IEEE Communications Society (INFOCOM) 2004*, Barcelona, Spain, Apr. 2006, pp. 1–5.

[27] I. T. U. Standardization, “Architecture for the Automatic Switched Optical Network (ASON),” ITU-T Standard G.8080/Y.1304, Jan. 2001.

[28] ——, “Definitions and Terminology for Automaticaly Switched Optical Networks (ASON),” ITU-T Standard G.8081/Y.1353, 2004.

[29] E. Mannie, “Generalized Multi-Protocol Label Switching (GMPLS) Architecture,” IETF RFC 3945, Oct. 2004.

[30] A. Zalesky, “To burst or circuit switch?” *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 305–318, feb 2009.

[31] V. Chan, “Editorial: Near-Term Future of the Optical Network in Question?” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 9, pp. 1–2, Dec. 2007.

[32] G. Zervas, M. De Leenheer, L. Sadeghioon, D. Klonidis, Y. Qin, R. Nejabati, D. Simeonidou, C. Develder, B. Dhoedt, and P. Demeester, “Multi-granular optical cross-connect: Design, analysis and demonstration,” *IEEE/OSA Journal on Optical Communications and Networking*, vol. 1, no. 1, pp. 69–84, Jun. 2009.

[33] F. Xue, S. J. B. Yoo, H. Yokoyama, and Y. Horiuchi, “Performance comparison of optical burst and circuit switched networks,” in *2005 Optical Fiber Communication Conference and Exposition and The National Fiber Optic Engineers Conference (OFC)*, Anaheim, CA, USA, Mar. 2005, p. OWC1.

[34] M. O'Mahony, D. Simeonidou, D. Hunter, and A. Tzanakaki, "The application of optical packet switching in future communication networks," *IEEE Communications Magazine*, vol. 39, no. 3, pp. 128–135, Mar. 2001.

[35] L. Lui, L. Xu, K. Lau, P. Wai, H. Tam, and M. Demokan, "All-optical packet switching with all-optical header processing and 2R regeneration," in *Conference on Lasers and Electro-Optics (CLEO) 2005*, vol. 1, May 2005, pp. 719–721.

[36] C. Qiao and M. Yoo, "Optical burst switching (OBS) - a new paradigm for an optical Internet," *Journal of High Speed Networks*, vol. 8, no. 1, pp. 69 – 84, Mar. 1999.

[37] Y. Chen, C. Qiao, and X. Yu, "Optical burst switching: A new area in optical networking research," *IEEE Network*, vol. 18, no. 3, pp. 16–23, May 2004.

[38] T. Coutelen, H. Elbiaze, and B. Jaumard, "Performance comparison of OCS and OBS switching paradigms," in *Proc. of 2005 7th International Conference on Transparent Optical Networks (ICTON)*, vol. 1, Jul. 2005, pp. 212–215.

[39] X. Liu, C. Qiao, X. Yu, and W. Gong, "A fair packet-level performance comparison of OBS and OCS," in *Optical Fiber Communication Conference and the National Fiber Optic Engineers Conference (OFC) 2006*, Mar. 2006, p. 3 pp.

[40] A. Agrawal, T. S. El-Bawab, and L. B. Sofman, "Comparative account of bandwidth efficiency in optical burst switching and optical circuit switching networks," *Photonic Network Communications*, vol. 9, no. 3, pp. 297–309, 2005.

[41] P. Molinero-Fernández and N. McKeown, "Performance of circuit switching in the Internet," *OSA Journal of Optical Networking*, vol. 2, no. 4, pp. 83–96, Apr. 2003.

[42] R. Parthiban, C. Leckie, A. Zalesky, M. Zukerman, and R. Tucker, "Cost comparison of optical circuit-switched and burst-switched networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 27, no. 13, pp. 2315–2329, Jul. 2009.

[43] G. Lee and J. Choi, "Flow classification for IP differentiated service in optical hybrid switching network," in *Convergence in Broadband and Mobile Networking Information Networking*, ser. Lecture Notes in Computer Science, vol. 3391, 2005, pp. 635–642.

[44] M. Duser and P. Bayvel, "Analysis of a dynamically wavelength-routed optical burst switched network architecture," *IEEE/OSA Journal of Lightwave Technology*, vol. 20, no. 4, pp. 82–89, Feb. 2002.

[45] J. Wei and R. McFarland, "Just-In-Time signaling for WDM optical burst switching networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 18, no. 12, pp. 2019–37, Dec. 2000.

[46] I. Baldine, G. Rouskas, H. Perros, and D. Stevenson, "Jumpstart: a Just-In-Time signaling architecture for WDM burst-switched networks."

[47] M. Yoo and C. Qiao, "Just-Enough-Time (JET): a high speed protocol for bursty traffic in optical networks," in *1997 Digest of the IEEE/LEOS Summer Topical Meetings: Vertical-Cavity Lasers/Technologies for a Global Information Infrastructure/WDM Components Technology/Advanced Semiconductor Lasers and Applications/Gallium Nitride Materials, Processing, and Devices*, Montreal, Que., Canada, Aug. 1997, pp. 26–27.

[48] K. Dolzer, C. Gauger, J. Spath, and S. Bodamer, "Evaluation of reservation mechanisms for Optical Burst Switching," *AEU-International Journal of Electronics and Communications*, vol. 55, no. 1, pp. 18–26, 2001.

[49] J.-B. Chang and C.-S. Park, "Efficient channel-scheduling algorithm in optical burst switching architecture," in *Workshop on High Performance Switching and Routing (HPSR) 2002*, Kobe, Japan, 2002, pp. 194–198.

[50] B.-C. Kim, Y.-Z. Cho, J.-H. Lee, Y.-S. Choi, and D. Montgomery, "Performance of optical burst switching techniques in multi-hop networks," in *Proc. of the 21st IEEE Global Telecommunications Conference (GLOBECOM) 2002*, vol. 3, Taipei, Taiwan, Nov. 2002, pp. 2772–2776.

[51] T. Battestilli and H. Perros, "An introduction to optical burst switching," *IEEE Communications Magazine*, vol. 41, no. 8, pp. S10–S15, Aug. 2003.

[52] Y. Chi, L. Zhengbin, and X. Anshi, "Parallel link server architecture: a novel contention avoidance mechanism in OBS networks," *Photonic Network Communications*, vol. 14, no. 3, pp. 297–305, 2007.

[53] M. Lévesque and H. Elbiaze, "Graphical probabilistic routing model for OBS networks with realistic traffic scenario," in *Proc. of the 28th IEEE Global Communications Conference (GLOBECOM) 2009*, Honolulu, Hawaii, USA, Nov. 2009, pp. 5459–5464.

[54] I. Chlamtac, A. Fumagalli, and C.-J. Suh, "Multibuffer delay line architectures for efficient contention resolution in optical switching nodes," *IEEE Transactions on Communications*, vol. 48, no. 12, pp. 2089–2098, Dec. 2000.

[55] C. Gauger, "Dimensioning of fdl buffers for optical burst switching nodes," in *Proc. of 6th Working Conference on Optical Network Design and Modelling (ONDM) 2002*, Torino, Italy, Feb. 2002, pp. 117–132.

[56] ——, “Performance of converter pools for contention resolution in optical burst switching,” in *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 4874, Boston, MA, USA, 2002, pp. 109–117.

[57] X. Wang, H. Morikawa, and T. Aoyama, “Burst optical deflection routing protocol for wavelength routing WDM networks,” in *Proc. of the SPIE - The International Society for Optical Engineering*, vol. 4233, Richardson, TX, USA, Oct. 2000, pp. 257–66.

[58] Y. Chen, H. Wu, D. Xu, and C. Qiao, “Performance analysis of optical burst switched node with deflection routing,” in *2003 IEEE International Conference on Communications (ICC)*, vol. 2, Anchorage, AK, USA, May 2003, pp. 1355–9.

[59] H. Cankaya, S. Charcranoon, and T. El-Bawab, “A preemptive scheduling technique for OBS networks with service differentiation,” in *Proc. of the 22nd IEEE Global Telecommunications Conference (GLOBECOM) 2003*, vol. 5, San Francisco, CA, USA, Dec. 2003, p. 2704.

[60] H. Overby, “Quality of Service differentation: Teletraffic analysis and network layer packet redundancy in optical packet switched networks,” Ph.D. dissertation, Dep. of Telematics, Norwegian University of Science and Technology, 2005.

[61] K. Chua, M. Gurusamy, Y. Liu, and M. Phung, *Quality of Service in Optical Burst Switched Networks*, ser. Optical Networks. Springer, 2007.

[62] M. Yoo, C. Qiao, and S. Dixit, “A novel fault detection and localization scheme for mesh all-optical networks based on monitoring-cycles,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, pp. 2062–2071, Oct. 2000.

[63] H. Overby, N. Stol, and M. Nord, “Evaluation of QoS differentiation mechanisms in asynchronous bufferless optical packet-switched networks,” *IEEE Communications Magazine*, vol. 44, no. 8, pp. 52–57, Aug. 2006.

[64] Q. Zhang, V. Vokkarane, J. Jue, and B. Chen, “Absolute QoS differentation in optical burst-switched networks,” *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 9, pp. 1781–1795, Nov. 2004.

[65] J. Teng and G. Rouskas, “Traffic engineering approach to path selection in optical burst switching networks,” *OSA Journal of Optical Networking*, vol. 4, no. 11, pp. 759–777, Nov. 2005.

[66] I. Baldine and G. Rouskas, “Reconfiguration and dynamic load balancing in broadcast WDM networks,” *Photonic Network Communications*, vol. 1, no. 1, pp. 49–64, Jun. 1999.

[67] A. Narula-Tam and E. Modiano, "Dynamic load balancing in WDM packet networks with and without wavelength constriants," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, pp. 1972–1979, Oct. 2000.

[68] M. Klinkowski, J. Pedro, D. Careglio, M. Pióro, J. Pires, P. Monteiro, and J. Sol-Pareta, "An overview of routing methods in optical burst switching networks," *Elsevier Optical Switching and Networking*, vol. 7, no. 2, pp. 41–53, Apr. 2010.

[69] M. González-Ortega, J. López-Ardao, R. Rodríguez-Rubio, C. López-García, M. Fernández-Veiga, and A. Suarez-González, "Performance analysis of adaptive multipath load balancing in WDM-LOBS networks," *Computer Communications*, vol. 30, no. 18, pp. 3460–3470, Dec. 2007.

[70] J. Lu, Y. Liu, G. Mohan, and K. Chua, "Gradient projection based multipath traffic routing in optical burst switching networks," in *Proc. of 2006 Workshop on High Performance Switching and Routing (HPSR)*, Póznan, Poland, Jun. 2006, pp. 379–384.

[71] D. Ishii, N. Yamanaka, and I. Sasae, "Self-learning route selection scheme using multipath searching packets in an OBS networks," *OSA Journal of Optical Networking*, vol. 4, no. 7, pp. 432–445, Jul. 2005.

[72] C. Garcia-Aros, O. González, and J. Aracil, "Adaptive multi-path routing for OBS networks," in *2007 9th International Conference on Transparent Optical Networks (ICTON)*, Rome, Italy, Jul. 2007, pp. 299–302.

[73] Z. Shi, Y. TinJin, and Z. Bing, "Ant algorithm in OBS RWA," in *SPIE, Optical Transmission, Switching and Subsystem II*, vol. 5625, Beijing, China, Feb. 2005, pp. 705–713.

[74] G. Pavani and H. Waldam, "Traffic engineering and restoration in optical packet switching networks by means of ant colony optimization," in *Proc. of 2006 2nd International Conference on Broadband Communications (BROADNETS)*, San Jose, CA, USA, Oct. 2006, pp. 1068–1077.

[75] L. Xu, H. Perros, and G. Rouskas, "A simulation study of optical burst switching and access protocols for WDM ring networks," *Computer Networks*, vol. 41, no. 2, pp. 143–160, Feb. 2003.

[76] W. Chen, W. Wang, and W. Hwang, "A novel and simple beforehand bandwidth reservation (BBR) MAC protocol for OBS metro ring networks," *J. of High Speed Networks*, vol. 17, no. 1, pp. 59–72, Jan. 2008.

[77] H. Lin and W. Chang, "CORnet: an OBS metro ring network with QoS support and fairness control," *Computer Networks*, vol. 52, no. 10, pp. 2045–2064, Jul. 2008.

[78] L. Peng and Y. Kim, “Investigation of the design of MAC protocols for TT-TR-Based WDM burst-switched ring networks,” *IEEE/OSA Journal on Optical Communications and Networking*, vol. 1, no. 2, pp. 25–34, Jul. 2009.

[79] G. Zervas, Y. Qin, R. Nejabati, D. Simeonidou, F. Callegati, A. Campi, and W. Cerroni, “SIP-enabled optical burst switching architectures and protocols for application-aware optical networks,” *Computer Networks*, vol. 52, no. 10, pp. 2065–2076, Jul. 2008.

[80] E. Wong and M. Zukerman, “An optical hybrid switch with circuit queueing for burst clearing,” *IEEE/OSA Journal of Lightwave Technology*, vol. 26, no. 21, pp. 3509–3527, Nov. 2008.

[81] B. Mukherjee, “Architecture, control, and management of optical switching networks,” in *Photonics in Switching 2007*, San Francisco, CA, USA, Aug. 2007, pp. 43–44.

[82] A. Zalesky, E. Wong, M. Zukerman, H. L. Vu, and R. Tucker, “Performance analysis of an OBS edge router,” *IEEE Photonics Technology Letters*, vol. 16, no. 2, pp. 695–697, Feb. 2004.

[83] C. Qiao, W. Wei, and X. Liu, “Extending generalized multiprotocol label switching (GMPLS) for polymorphous, agile, and transparent optical networks (PATON),” *IEEE Communications Magazine*, vol. 44, no. 12, pp. 104–114, Dec. 2006.

[84] X. Liu, C. Qiao, W. Wei, and T. Wang, “A universal signaling, switching and reservation framework for future optical networks,” *IEEE/OSA Journal of Lightwave Technology*, vol. 27, no. 12, pp. 1806–1815, Jun. 2009.

[85] J. A. Hernández, P. Reviriego, J. L. García-Dorado, V. López, D. Larrabeiti, and J. Aracil, “Performance evaluation and design of polymorphous OBS networks with guaranteed TDM services,” *IEEE/OSA Journal of Lightwave Technology*, vol. 27, no. 13, pp. 2495–2505, Jul. 2009.

[86] C. Qiao, “Labeled optical burst switching for IP-over-WDM integration,” *IEEE Communications Magazine*, vol. 38, no. 9, pp. 104–114, Sep. 2000.

[87] M. A. González-Ortega, C. Qiao, A. Suárez-González, X. Liu, and J.-C. López-Ardao, “LOBS-H: An enhanced OBS with wavelength sharable home circuits,” in *Proc. of 2010 IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, May 2010, pp. 1–5.

[88] C. Qiao, M. González-Ortega, A. Suárez-González, X. Liu, and J.-C. López-Ardao, “On the benefit of fast switching in optical networks,” in *Proc. of*

Optical Fiber Communication and National Fiber Optic Engineers Conference (OFC/NFOEC) 2010, San Diego, CA, USA, Mar. 2010, pp. 1–3.

[89] P. Pedroso, J. Solé-Pareta, D. Careglio, and M. Klinkowski, "Integrating GMPLS in the OBS Networks Control Plane," in *Proc. of 2007 9th International Conference on Transparent Optical Networks (ICTON)*, vol. 3, Rome, Italy, Jul. 2007, pp. 1–7.

[90] P. Pedroso, D. Careglio, R. Casellas, M. Klinkowski, and J. Solé-Pareta, "An interoperable GMPLS/OBS control plane: RSVP and OSPF extensions proposal," in *Proc. of 6th International Symposium on Communication Systems, Networks and Digital Signal Processing (CNSDSP) 2008*, Graz, Austria, Jul. 2008, pp. 418–422.

[91] J. Perelló, S. Spadaro, J. Comellas, and G. Junyent, "Burst contention avoidance schemes in hybrid GMPLS-enabled OBS/OCS optical networks," in *Proc. of International Conference on Optical Network Design and Modeling (ONDM) 2009*, Braunschweig, Germany, Feb. 2009, pp. 1–6.

[92] Y. Yin, H. Guo, J. Wu, X. Hong, C. Tian, T. Tsuritani, N. Yoshikane, T. Otani, and J. Lin, "Dynamic protection and restoration supporting QoS in OBS/GMPLS interworking network," in *34th European Conference on Optical Communication (ECOC) 2008*, Brussels, Belgium, Sep. 2008, pp. 1–2.

[93] X. Wang, H. Guo, Y. Yin, W. Zhang, X. Hong, Y. Zuo, J. Wu, and J. Lin, "Experimental demonstration of dynamic end-to-end lightpath provisioning mechanism in OBS/GMPLS interworking network," in *15th Asia-Pacific Conference on Communications (APCC) 2009*, Shanghai, China, Oct. 2009, pp. 515–518.

[94] A. Gumaste, N. Ghani, B. Bafna, A. Lodha, A. Agrawal, T. Das, J. Wang, and S.-Q. Zheng, "DynaSPOT: Dynamic services provisioned optical transport test-bed - achieving multirate multiservice dynamic provisioning using strongly connected light-trail (SLiT) technology," *IEEE/OSA Journal of Lightwave Technology*, vol. 26, no. 1, pp. 183–195, Jan. 2008.

[95] S. Koo, G. Sahin, and S. Subramaniam, "Dynamic LSP provisioning in overlay, augmented, and peer architectures for IP-MPLS over WDM networks," in *Proc. of 23th Conference of the IEEE Communications Society (INFOCOM) 2004*, vol. 4, Mar. 2004, p. 2866.

[96] K. Shiomoto, "Framework for MPLS-TE to GMPLS Migration," IETF RFC 5145, Mar. 2008.

[97] A. Ayyangar, K. Kompella, J. Vasseur, and A. Farrel, “Label Switched Path Stitching with Generalized Multiprotocol label Switching Traffic Engineering (GMPLS TE),” IETF RFC 5150, Feb. 2008.

[98] H. Zang, J. Jue, and B. Mukherjee, “Photonic slot routing in all-optical WDM mesh networks,” in *Proc. of 1999 Global Telecommunications Conference (GLOBECOM)*, vol. 2, Rio de Janeiro, Brazil, 1999, pp. 1449–1453.

[99] D. Grieco, A. Pattavina, and Y. Ofek, “Fractional lambda switching for flexible bandwidth provisioning in WDM networks: Principles and performance,” *Photonic Network Communications*, vol. 9, no. 3, pp. 281–296, 2005.

[100] V.-T. Nguyen, , R. L. Cigno, Y. Ofek, and M. Telek, “Time blocking analysis in time-driven switching networks,” in *Proc. of 27th IEEE International Conference on Computer Communications (INFOCOM) 2008*, Apr. 2008, pp. 1804–1812.

[101] N.-F. Huang, G.-H. Liaw, and C.-P. Wang, “A novel all-optical transport network with time-shared wavelength channels,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, pp. 1863–1875, Oct. 2000.

[102] S. Subramaniam, E. J. Harder, and H.-A. Choi, “Scheduling multirate session in time division multiplexed wavelength-routing networks,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, pp. 2105–2110, Oct. 2000.

[103] B. Wen, R. Shenai, and K. Sivalingam, “Routing, wavelength and time-slot assignment algorithms for wavelength-routed optical WDM/TDM networks,” *IEEE/OSA Journal of Lightwave Technology*, vol. 23, no. 9, pp. 2598–2609, Sep. 2005.

[104] H. Zang, J. Jue, and B. Mukherjee, “A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks,” *Optical Networks Magazine*, vol. 1, no. 1, pp. 47–60, Jan. 2000.

[105] J. Yates, J. Lacey, and D. Everitt, “Blocking in multiwavelength TDM networks,” *Telecommunication Systems Journal*, vol. 12, no. 1, pp. 1–19, Aug. 1999.

[106] H. Zeineddine and G. V. Bochmann, “Least constrained slot allocation in optical TDM networks,” in *IFIP International Conference on Wireless and Optical Communications Networks (WOCN) 2007*, Singapore, Jul. 2007, pp. 1–5.

[107] ——, “A distributed algorithm for least constraining slot allocation in MPLS optical TDM networks,” in *IEEE International Conference on Communications (ICC) 2009*, Dresden, Germany, Jun. 2009, pp. 1–6.

[108] V. Eramo, A. Cianfrani, M. Listanti, A. Germoni, P. Cipollone, and F. Matera, "Performance evaluation of OTDM/WDM networks in dynamic traffic scenario," in *Proc. of 2010 12th International Conference on Transparent Optical Networks (ICTON)*, Munich, Germany, Jun. 2010, pp. 1–4.

[109] M. Dorigo and T. Sttzle, *Ant Colony Optimization*. Massachusetts, USA: The MIT Press, 2004.

[110] M. Dorigo and L. Gambardella, "Ant colony system: A cooperative learning approach to the traveling salesman problem," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 53–66, Apr. 1997.

[111] M. Aydin, T. Atmaca, H. Zaim, O. Turna, and V. Nguyen, "Performance study of OBS reservation protocols," in *Proc. of 4th Advance Int. Conf on Telecommunications (AICT'08)*, Athens, Greece, Jun. 2008, pp. 428–433.

[112] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[113] J. Teng and G. Rouskas, "Wavelength selection in OBS networks using traffic engineering and priority-based concepts," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 8, pp. 1658–1669, Aug. 2005.

[114] A. Rostami and A. Wolisz, "Modelling and synthesis of traffic in optical burst-switched networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 25, no. 10, pp. 2492–2952, Oct. 2007.

[115] H. Zeng, C. Huang, and A. Vukovic, "A novel fault detection and localization scheme for mesh all-optical networks based on monitoring-cycles," *Photonic Network Communications*, vol. 11, no. 3, pp. 277–286, May 2006.

[116] B. Waxman, "Routing of multipoint connnections," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 9, pp. 1617–1622, Dec. 1988.

[117] IEEE Standards Board, "Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Network - 802.6," IEEE Standard 802.6, Dec. 1990.

[118] Z. Zhang, L. Liu, and Y. Yang, "Slotted Optical Burst Switching (SOBS) networks," *Computer Communications*, vol. 30, no. 18, pp. 3471–3479, Dec. 2007.

[119] M. Klinkowski, D. Careglio, and J. Solé-Pareta, "Comparison of Conventional and Offset Time-Emulated Optical Burst Switching," in *Proc. of 8th Int. Conf. on Transparent Optical Networks (ICTON)*, Nothingham, United Kingdom, Jun. 2006, pp. 47–50.

[120] N. M. Garcia, P. Lenkiewicz, P. P. Monteiro, and M. M. Freire, "Issues on performance assessment of optical burst switched networks: Burst loss versus packet loss metrics," in *Proc. of NETWORKING 2006 Lecture Notes in Compute Science*, vol. 3976, 2006, pp. 778–786.

[121] S. K. Bose, *An Introduction to Queueing Systems*, 1st ed. Springer, 2002.

[122] M. Izal and J. Aracil, "On the influence of self-similarity on optical burst switching traffic," in *Proc. of 2002 IEEE Global Telecommunications Conference GLOBECOM*, vol. 3, Taipei, Taiwan, China, Nov. 2002, pp. 2308–2312.

[123] M. Conti, E. Gregori, and L. Lenzini, "On the approximation of the slot occupancy pattern in a DQDB network," in *Proc. of 11th IEEE International Conference on Computer Communications (INFOCOM) 1992*, vol. 2, Florence, Italy, May 1992, pp. 518–526.

[124] P. Courtois and J. Georges, "On a single-server finite queuing model with state-dependent arrival and service processes," *Journal of Operations Research*, vol. 19, no. 2, pp. 424–435, Apr. 1971.

[125] T. Takine, H. Takagi, and T. Hasegawa, "Analysis of an M/G/1/K/N queue," *Journal of Applied Probability*, vol. 30, no. 2, pp. 446–454, Jun. 1993.

[126] S. Ramanathan, "Multicast tree generation in networks with asymmetric links," *IEEE/ACM Transactions on Networking*, vol. 4, no. 4, pp. 558–568, Aug. 1996.

[127] N. Singhal, L. Sahasrabuddhe, and B. Mukherjee, "Optimal multicasting of multiple light-trees of different bandwidth granularities in a WDM mesh network with sparse splitting capabilities," *IEEE/ACM Transactions on Networking*, vol. 14, no. 5, pp. 1104–1117, Oct. 2006.

[128] S. Even, A. Itai, and A. Shamir, "On the complexity of timetable and multicommodity flow problems," *SIAM Journal on Computing*, vol. 5, no. 4, pp. 691–703, 1976.

[129] M. Bouklit, D. Coudert, J.-F. Lalande, C. Paul, and H. Rivano, "Approximate multicommodity flow for WDM networks design," in *SIROCCO'03: Colloquium on Structural Information and Communication Complexity*, Umeä (Sweden), Jun. 2003, pp. 43–56.

[130] J. Fang and A. Somani, "IP traffic grooming over WDM optical networks," in *2005 Conference on Optical Network Design and Modeling (ONDM)*, Feb. 2005, pp. 393–402.

[131] A. Gumaste and P. Palacharla, "Heuristic and optimal techniques for light-trail assignment in optical ring WDM networks," *Computer Communications*, vol. 30, no. 5, pp. 990–998, Mar. 2007.

[132] J. Fang, W. He, and A. Somani, "Optimal light trail design in WDM optical networks," in *Proc. of 2004 IEEE International Conference on Communications (ICC)*, vol. 3, Jun. 2004.

[133] R. Lin, W.-D. Zhong, S. Bose, and M. Zukerman, "Design of WDM networks with multicast traffic grooming," *IEEE/OSA Journal of Lightwave Technology*, vol. 29, no. 16, pp. 2337–2349, Aug. 2011.

[134] G.-K. Chanq, A. Chowdhury, Z. Jia, H.-C. Chien, M.-F. Huanq, J. Yu, and G. Ellinas, "Key technologies of WDM-PON for future converged optical broadband access networks," *IEEE/OSA Journal on Optical Communications and Networking*, vol. 1, no. 4, pp. C35–C50, Sep. 2009.

[135] J. Chen, L. Wosinska, C. Machuca, and M. Jaeger, "Cost vs. reliability performance study of fiber access network architectures," *IEEE Communications Magazine*, vol. 48, no. 2, pp. 56–65, Feb. 2010.

[136] M. Carroll, J. Roese, and T. Ohara, "The operator's view of OTN evolution," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 46–52, Sep. 2010.

[137] A. Farrel, J.-P. Vasseur, and J. Ash, "RFC 4655: A Path Computation Element (PCE)-based architecture," IETF, Aug. 2006.

[138] MAINS consortium, "EU FP7 MAINS project," [Online]. Available at: http://www.ist-mains.eu/, 2010.

[139] I. Chlamtac, A. Ganz, and G. Karmi, "Lightpath communications: an approach to high bandwidth optical WAN's," *IEEE Transactions on Communications*, vol. 40, no. 7, pp. 1171–1182, Jul. 1992.

[140] Free Software Foundation, "GLPK: GNU Linear Programming Kit," [Online]. Available at: http://www.gnu.org/software/glpk/.

[141] G. Thodime, V. Vokkarane, and J. Jue, "Dynamic congestion-based load balanced routing in optical burst-switched networks," in *Proc. of the 22nd IEEE Global Telecommunications Conference (GLOBECOM) 2003*, vol. 5, San Francisco, CA, USA, Dec. 2003, pp. 2628–2632.