SEVENTH FRAMEWORK PROGRAMME

Report on Y2 activities and new integration strategy

FP7-ICT-216863/DEIS-UNIBO/R/PU/D11.2

Project Number: FP7-ICT-2007-1 216863

Project Title: Building the Future Optical Network in Europe (BONE)

Contractual Date of Deliverable: 30/11/2009

Actual Date of Delivery: 29/12/2009

Workpackage contributing to the Deliverable: WP11 : VCE on Network Technologies and Engineering

Nature of the Deliverable Make your choice from: R

Dissemination level of Deliverable Make your choice from: PU

Editors: DEIS-UNIBO/F. Callegati, FUB/F. Matera, TUW/G. Franzl, UPC/M. Klinkowsky, SSSUP/F. Cugini, SSSUP/I. Cerutti, UPV/D. Garcia-Roger, POLIMI/M. Tornatore, CTTC/R. Munoz, UOP/T. Orphanoudakis, FER/N. Skorin-Kapov, UAM/V. Lopez-Alvarez, UPCT/P. Pavon-Marino

Abstract:

This report focuses on the results of the joint activities achieved during the second year of the BONE project in the VCE on Network Technologies and Engineering. The report also presents an overview of the organization of the work and of the results achieved in the executive summary.

Keyword list: optical networks, traffic engineering, OBS, OPS, protection, restoration, Carrier Grade Ethernet, technology survey.

Clarification:

Nature of the Deliverable

R	Report
P	Prototype
D	Demonstrator
O	Other

Dissemination level of Deliverable:

PU	Public
PP	Restricted to other programme participants (including the Commission Services)
RE	Restricted to a group specified by the consortium (including the Commission Services)
CO	Confidential, only for members of the consortium (including the Commission Services)

Disclaimer

*The information, documentation and figures available in this deliverable, is written by the BONE (“*Building the Future Optical Network in Europe*) – project consortium under EC co-financing contract FP7-ICT-216863 and does not necessarily reflect the views of the European Commission*

Table of Contents

Executive Summary

This document reports the activities of the VCE on Network Technologies and Engineering.

Participation

To date 148 researchers from 40 partner institutions registered to WP11 in the BONE directory service. This is a measure of the general interest for the topics addressed by the WP. The involvement of partners and researchers is uneven, ranging from active participation to several joint activities to a more basic auditory participation.

The partners involved are UPC - BILKENT - TUW - SSSUP - UPCT - UAM - AIT - UCL - FPMs - PoliTO - GET - USWAN - UNIBO - CTTC - UNIMORE - Ericsson - KTH - UniRoma1 - UniRoma3 - BME - FUB - IT - AGH - UoP - TID - UC3M - FER - CORITEL - FT - UPVLC - UST-IKR - UDE - TUB - PoliMI - COM – Uessex - IBBT - RACTI - HWDE – TELENOR.

Joint research activities

Joint Research Activities (JAs) are key to “integration”. Therefore the main R&D effort in WP11 was devoted to running the JAs. Some of these JAs just ended at the end of the second year. The WP leader took care to avoid overlapping with JAs in other WPs and within the WP. For this reason some topic, that are indeed relevant for WP11, are not investigated in any JA. These topics are investigated in the framework of the TPs and WP11 will rely on the results produced there. Moreover it is worth reminding that, at least in the understanding of the WP leader according to his previous experience, a VCE does not have as major goal the steering of the direction of the research (more typical for a TP). The VCE should aim at favouring integration of the research by promoting the collaboration of the research groups about research topics that spontaneously stem from the community as a result of the research interests of the participants. In this way the research carried on in the VCE is a sort of “metric” of what the community believes is important in the field.

Activity #	Activity Title	Partners	Responsible	Status
1	Common Architecture/Hierarchy for integration of OCS, OBS, and OPS in a common transport plane	TUW, DTU, AGH, GET/ENST, PoliTO, Uessex	Gerald Franzl	**Completed**
2	Comparative techno-economic network planning in OCS/OBS/OPS networks	UPCT, TID, UNIRM, FT, AGH	Pablo Pavon Marino	**Active**
3	Experimental tests of Carrier Ethernet techniques	FUB, UNIROMA3, UNIROMA1, UNIBO, UNIMORE, KTH, CORITEL	Francesco Matera	**Active**
4	Extension of the Flow-Aware Networking (FAN) architecture to the IP over WDM environment	GET/ENST, UaM, AGH	Victor Lopez Alvarez	**Active**
5	Network planning algorithms considering fault tolerance, security threats and periodic traffic patterns	FER, KTH, UPCT, IT, AIT, GET	Nina Skorin-Kapov	**Active**
6	On exploring Admission Control Mechanisms for OBS networks	UaM, UC3M, UniMORE	Jos□ Alberto Hern□ndez	**Completed**
7	PCE for Multi-domain Traffic Engineering	SSSUP, PoliMI	Filippo Cugini	**Active**
8	Resilience analysis of double rings with dual attachments	UaM, TID	Jos□ Alberto Hern□ndez	**Completed**
9	Traffic engineering and topology design in metro networks	SSSUP, GET/ENST	Isabella Cerruti	**Active**
10	Survey on OBS routing methods for OBS networks	UPC, IT	Mirek Klinkowski	**Completed**
11	Optical Buffering Technologies Survey	FT, USWAN, TUW, UNIBO, UPC, KTH, Ericsson	Hisao Nakajima, Karin Ennser	**Stand By**
12	Effects of outdated control information on routing in optical networks	POLIMI, CTTC	Achille Pattavina	**Active**
13	Scalability of IP networks and routers	Telenor (FT, TID, UoP)	Evi Zouganeli	**Stand By**
14	NETBENCH "Benchmarking of network architectures for guaranteed service provisioning"	UoP, ESSEX, PoliTO, UniBO, GET, UaM, UC3M	Theofanis Orphanoudakis	**Active**
15	Adaptive Admission Control in Dynamic GMPLS Networks	UC3M, UPV	David Larrabeiti	**Active**

Table I. List of WP11 Joint activities

Publications

The various joint research activities achieved several interesting scientific results in Y2. A total of 30 joint papers were published in Y2. The complete list is attached.

The papers were co-authored by 85 researchers belonging to 25 institutions for a total of 164 authorships, with an average of 5.5 authors per paper.

The number of authorships per partner are plotted in .

The joint papers were published in journals (10) and international conferences on optical networking (20) as plotted in . It is worth noting a good participation into ICTON, OFC and Photonic in Switching, that are indeed closely related to topics of interest of WP11.

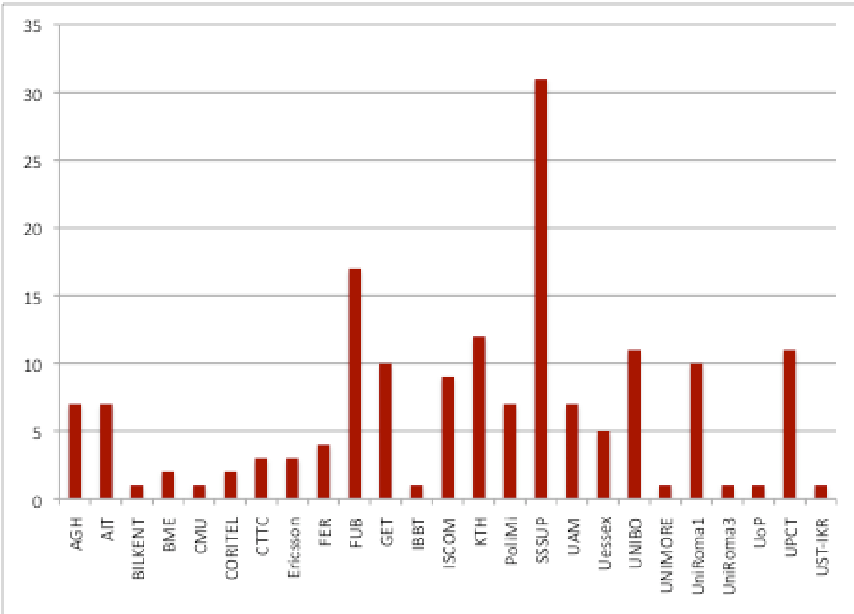


Figure 1. Number of joint paper authorships per partner

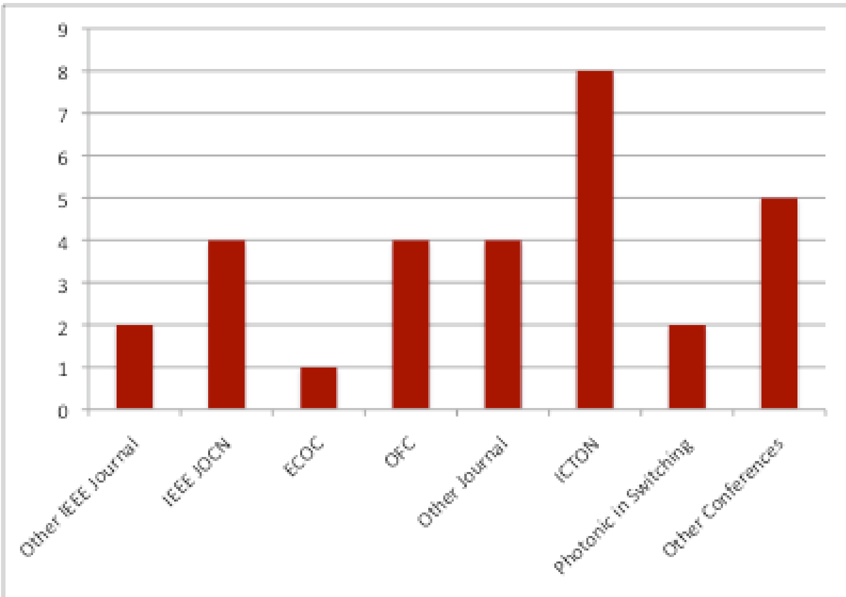


Figure 2. Summary of the place of publication of the joint papers.

Among these papers it is worth mentioning the first in the list, as a significant example of an effort to provide maximum visibility to the results of NoE. This is a paper summarizing the findings of the JAs carried on during the e-Photon/One+ project, that was prepared and completed during the first year of BONE. The paper makes and effort to provide a unified view of the research challenges in the broad area of optical core networks, followed by a summary of the main findings of the joint activities.

- F. Callegati (UNIBO), F. Cugini (SSSUP), P. Ghobril (Orange Labs), S. Gunreben (UST-IKR), V. López (UAM), B. Martini (SSSUP), P. Pavón-Mariño (UPCT), M. Perényi (BME), N. Sengezer (BILKENT), D. Staessens (IBBT), J. Szigeti (BME), M. Tornatore (PoliMI), Optical Core Networks Research in the e-Photon-ONe+ Project, IEEE/OSA Journal of Lightwave Technology, Vol. 27, No. 20, pp. 4415-4423, October 2009. (WP11)
- A. Coiro (UniRoma1), A. Valenti (FUB), S. Pompei (FUB), F. Matera (FUB), P. Testa (CORITEL), M. Settembre (ELSAG Datamat), Network Evolution Toward a Carrier-Grade Ethernet Transport Network, Fiber and Integrated Optics, Vol. 28, No. 6, pp. 393-411, December 2009. (WP11)
- J. Domzal (AGH), R. Wójcik (AGH), K. Wajda (AGH), A. Jajszczyk (AGH), V. Lopez (UAM), J. A. Hernandez (UAM), J. Aracil (UAM), C. Cardenas (GET), M. Gagnaire (GET), A Multi-layer Recovery Strategy in FAN over WDM Architectures, Design of Reliable Communication Networks, Washington D. C. (USA), October 2009. (WP11)
- B. Uscumlic (GET), A. Gravey (GET), P. Gravey (GET), I. Cerutti (SSSUP), Traffic Grooming in WDM Optical Packet Rings, ITC'09, September 2009. (WP11)
- B. Uscumlic (GET), A. Gravey (GET), I. Cerutti (SSSUP), P. Gravey (GET), M. Morvan (GET), The Impact of Network Design on Packet Scheduling in Slotted WDM Packet Rings, Proceedings Photonics in Switching, September 2009. (WP11)
- N. Skorin-Kapov (FER), P. Pavon-Marino (UPCT), B. Garcia-Manrubia (UPCT), R. Aparicio-Pardo (UPCT), Scheduled Virtual Topology Design Under Periodic Traffic in Transparent Optical Networks, Proc. of the Sixth International Conference on Broadband Communications, Networks and Systems (BROADNETS 2009), Madrid, Spain, September 2009. (WP11)
- A. Valenti (FUB), S. Pompei (FUB), F. Matera (FUB), G. M. Tosi Beleffi (ISCOM), D. M. Forin (ISCOM), Quality of Service control in Ethernet Passive Optical Networks based on Virtual Private LAN Service, IET Electronics Letters, Vol. 45, No. 19, pp. 992-993, September 2009. (WP11 WP13)
- A. Jirattigalachote (KTH), L. Wosinska (KTH), P. Monti (KTH), K. Katrinis (AIT), A. Tzanakaki (AIT), Impairment Constraint Based Routing (ICBR) with Service Differentiation in Survivable WDM Networks, ECOC 2009, Vienna, September 2009. (WP11 WP22)
- A. Coiro (UniRoma1), A. Valenti (FUB), S. Pompei (FUB), G. M. Tosi Beleffi (ISCOM), F. Curti (ISCOM), D. Forin (ISCOM), A. Rufini (UniRoma1), Experimental demonstration of the All Optical Network Wavelength Conversion in a Passive Optical Network, Photonic Switching 2009 (PS2009), Pisa, September 2009. (WP11 WP13)
- P. Pavon-Marino (UPCT), R. Aparicio-Pardo (UPCT), B. Garcia-Manrubia (UPCT), N. Skorin-Kapov (FER), Virtual topology design and flow routing in optical networks under multihour traffic demand, Photonic Network Communications, August 2009. (WP11)
- A. Valenti (FUB), P. Bolletta (UniRoma1), S. Pompei (FUB), F. Matera (FUB), Experimental Investigations on Restoration Techniques in a Wide Area Gigabit Ethernet Optical Test Bed based on Virtual Private LAN Service, 11th International Conference on Transparent Optical Networks (ICTON 2009), Vol. 978-1-4244-4826-5/09, Ponta Delgada, July 2009. (WP11)
- A. Silvestri (FUB), A. Valenti (FUB), S. Pompei (FUB), F. Matera (FUB), A. Cianfrani (UniRoma1), Wavelength Path Optimization in Optical Transport Networks for Energy Saving, ICTON 2009, July 2009. (WP11 WP14 WP24)
- V. Eramo (UniRoma1), A. Germoni (CORITEL), A. Cianfrani (UniRoma1), F. Lo Buono (UniRoma1), Performance Evaluation of a QoS Technique for Bufferless Optical Packet Switches, ICTON 2009, July 2009. (WP11 WP14 WP24)
- A. Valenti (FUB), S. Pompei (FUB), L. Rea (FUB), F. Matera (FUB), G. M. Tosi Beleffi (ISCOM), F. Curti (ISCOM), S. Di Bartolo (ISCOM), G. Incerti (UniRoma3), D. Forin (ISCOM), Experimental Investigation of Quality of Service in an IP All-Optical Network Adopting Wavelength Conversion, IEEE/OSA Journal of Optical Communications and Networking, Vol. 1, No. 2, pp. A170-179, July 2009. (WP11 WP13 WP15)

- T. Politi (UoP), C. Raffaelli (UNIBO), A. Tzanakaki (AIT), D. Simeonidou (UEssex), L. Wosinska (KTH), Optical Networks for the Future Internet: Introduction, Editorial for the special issue of IEEE/OSA Journal of Optical Communications and Networking, Vol. 1, No. 2, pp. FI1-FI3, July 2009. (WP11 WP14)
- M. Tornatore (PoliMI), F. De Grandi (PoliMI), R. Muñoz (CTTC), R. V. Martínez Rivera (CTTC), R. Casellas (CTTC), A. Pattavina (PoliMI), Effects of Outdated Control Information in Control-Plane-Enabled Optical Networks with Path Protection, IEEE/OSA Journal of Optical Communications and Networking (JOCN), Vol. 1, No. 2, pp. A180–A193, July 2009. (WP11)
- J. Domzal (AGH), R. Wojcik (AGH), A. Jajszczyk (AGH), V. Lopez (UAM), J. A. Hernandez (UAM), J. Aracil (UAM), Admission Control Policies in Flow-Aware Networks, ICTON'2009, Azores, Portugal, July 2009. (WP11)
- M. Casoni (UNIMORE), C. Raffaelli (UNIBO), TCP Performance over Optical Burst-Switched Networks with Different Access Technologies, OSA/IEEE Journal of Optical Communications and Networking (JOCN), No. 1, pp. 103-112, June 2009. (WP11 WP24 WP13)
- F. Cugini (SSSUP), N. Andriolli (SSSUP), A. Giorgetti (SSSUP), F. Paolucci (SSSUP), L. Valcarenghi (SSSUP), P. Castoldi (SSSUP), P. Iovanna (Ericsson), A. Welin (SSSUP), PCE in multi-layer MPLS over WSON networks, 5 International Conference on IP over Optical Networks, Tokyo (Japan), June 2009. (WP11)
- L. Valcarenghi (SSSUP), P. Monti (KTH), I. Cerutti (SSSUP), P. Castoldi (SSSUP), L. Wosinska (KTH), Issues and solutions in mobile WiMAX and wired backhaul network integration, Proceedings of ICTON 2009, June 2009. (WP11 WP12)
- B. Garcia-Manrubia, (UPCT), R. Aparicio-Pardo (UPCT), P. Pavon-Marino (UPCT), N. Skorin-Kapov (FER), J. Garcia-Haro (UPCT), MILP Formulations for Scheduling Lightpaths under Periodic Traffic, 11th International Conference on Transparent Optical Networks, ICTON 2009, Island of São Miguel, Azores, Portugal, June 2009. (WP11)
- A. Jirattigalachote (KTH), K. Katrinis (AIT), A. Tzanakaki (AIT), L. Wosinska (KTH), P. Monti (KTH), Quantifying the Benefit of BER-based Differentiated Path Provisioning in WDM Optical Networks, ICTON2009, Ponta Delgada, June 2009. (WP11 WP27 WP22)
- V. Eramo (UniRoma1), A. Cianfrani (UniRoma1), A. Germoni (CORITEL), M. Listanti (UniRoma1), F. Matera (FUB), Routing and Wavelength Assignment in OTDM/WDM Networks with Physical Impairments, NON 2009, June 2009. (WP11 WP14 WP24)
- A. Tzanakaki (AIT), K. Georgakilas (AIT), K. Katrinis (AIT), L. Wosinska (KTH), A. Jirattigalachote (KTH), P. Monti (KTH), Network Performance Improvement in Survivable WDM Networks considering Physical Layer Constraints, ICTON 2009, Ponta Delgada, June 2009. (WP11 WP27)
- N. Skorin-Kapov (FER), O. Tonguz (CMU), N. Puech (GET), Towards efficient failure management for reliable transparent optical networks, IEEE Communications Magazine, Vol. 47, No. 5, pp. 72-79, May 2009. (WP11)
- M. Savi (UNIBO), G. Zervas (UEssex), Y. Qin (UEssex), V. Martini (SSSUP), C. Raffaelli (UNIBO), F. Baroncelli (SSSUP), B. Martini (SSSUP), P. Castoldi (SSSUP), R. Nejabati (UEssex), D. Simeonidou (UEssex), Data-Plane Architectures for Multi-Granular OBS Network, OFC 2009, March 2009. (WP11 WP14 WP25)
- F. Cugini (SSSUP), F. Paolucci (SSSUP), L. Valcarenghi (SSSUP), P. Castoldi (SSSUP), A. Welin (Ericsson), PCE Communication Protocol for Resource Advertisement in Multi-domain BGP-based Networks, Proc. of OFC/NFOEC 2009, San Diego - CA (USA), March 2009. (WP11)
- N. Andriolli (SSSUP), F. Cugini (SSSUP), L. Valcarenghi (SSSUP), P. Castoldi (SSSUP), A. Welin (Ericsson), Virtual Network Topology Manager (VNTM) and Path Computation Element (PCE) Cooperation in Multi-Layer GMPLS Networks, Proc. of OFC/NFOEC 2009, San Diego - CA (USA), March 2009. (WP11)
- B. Martini (SSSUP), A. Campi (UNIBO), F. Baroncelli (SSSUP), V. Martini (SSSUP), K. Torkman (SSSUP), F. Zangheri (UNIBO), W. Cerroni (UNIBO), P. Castoldi (SSSUP), F. Callegati (UNIBO), SIP-based Service Platform for On-demand Optical Network Services, OFC 2009, San Diego - CA (USA), March 2009. (WP12 WP11)
- A. Pantaleo (PoliMI), M. Tornatore (PoliMI), A. Pattavina (PoliMI), C. Raffaelli (UNIBO), F. Callegati (UNIBO), Dimensioning for In-Band and Out-of-Band Signalling Protocols in OBS Networks, IET Communications, Vol. 3, No. 3, pp. 418-427, United Kingdom, March 2009. (WP11 WP24)

An important achievement in terms of joint publications was also the book "*Enabling Optical Internet with Advanced Network Technologies*", J. Aracil and F. Callegati editors, published by Springer, UK, July 2009. This was the joint effort of 32 researchers belonging to 15 institutions partner of the BONE consortium. This work was started during the e-Photon/One+ project and finalised during BONE (the book acknowledges both project). The book was organized into 5 chapters plus an introduction, which were edited by a chapter editor with the help of several contributors.

This is a 230 pages book providing an overview of advance optical networking, focusing on the expertise of the research groups involved. It counts 286 references and a number of figures and examples.

The list of chapter, with chapter editors and contributors is as follows.

- Introduction - Franco Callegati ed. (UNIBO), Javier Aracil (UAM), Vıctor Lopez (UAM)
- Introduction to IP over WDM - Davide Careglio ed. (UPC), Javier Aracil (UAM), Juan Fernandez Palacios (TID), Andrzej Jajszczyk (AGH), David Larrabeiti (UC3M), Vıctor Lopez (UAM), Xavier Masip (UPC), Sergio Sanchez (UPC), Salvatore Spadaro (UPC)
- Optical Packet Switching – Carla Raffaelli ed. (UNIBO), Slavisa Aleksic (TUW), Franco Callegati (UNIBO), Walter Cerroni (UNIBO), Guido Maier (POLIMI), Achille Pattavina (POLIMI), Michele Savi (UNIBO)
- Optical Burst Switching – Jose Alberto Hernandez ed. (UAM), Vıctor Lopez ed. (UAM), Jose Luis Garcıa Dorado (UAM), Reza Nejabati (UESSEX), Harald Overby (NTNU), Ahmad Rostami (TUB), Kyriakos Vlachos (UPATRAS), Georgios Zervas (UESSEX)
- Advanced Optical Burst Switched Network Concepts – Reza Nejabati ed. (UESSEX), Javier Aracil (UAM), Piero Castoldi (SSSUP), Marc De Leenheer (UGent), Dimitra Simeonidou (UESSEX), Luca Valcarenghi (SSSUP), Georgios Zervas (UESSEX), Jian Wu (BUPT-China)
- Optical Switch Fabrics and Their Application, Kyriakos Vlachos ed. (UPATRAS), Lambros Raptis (Attica Telecom), Antonio Teixeira (UA), Giorgio Maria Tosi Beleffi (ISCOM), Kostas Yiannopoulos (UPATRAS)

Mobility

11 Mobility actions took place in WP11 to date, 4 in Y1 and 7 in Y2. The number of mobility actions in Y2 almost doubled when compared with year 1. This may be considered an indication of an improving degree of collaboration and integration between the researchers actively participating in WP11.

The complete list of mobility actions follows, emphasized those occurred in Y2.

- "Network planning algorithms considering fault tolerance, security threats and periodic traffic patterns" JiaJia Chen, PhD student at KTH, hosted by FER from 18/04/2008 to 05/05/2008
- "Benchmarking of network architectures for guaranteed service provisioning" Michele Savi, PhD student at UNIBO, hosted by UEssex from 27/05/2008 to 07/12/2008
- "Network planning algorithms considering fault tolerance, security threats and periodic traffic patterns" Nina Skorin-Kapov, Assistant Professor at FER, hosted by KTH from 29/08/2008 to 14/09/2008
- "Network planning algorithms considering fault tolerance, security threats and periodic traffic patterns" Nina Skorin-Kapov, Assistant Professor at FER, hosted by UPCT from 22/11/2008 to 07/12/2008
- *"Algorithm approach for the design of a WDM ring with optical traffic grooming" Isabella Cerutti, Assistant Professor at SSSUP, hosted by GET from 25/01/2009 to 01/02/2009*
- *"An analytical reliability analysis of DRDA architecture" Jerzy Domzal, assistant at AGH, hosted by UAM from 21/02/2009 to 07/03/2009*
- *"Evaluation of Congestion Control Mechanisms for Flow-Aware Networking (FAN) networks under Grid environment" Robert Wojcik, Ph D student at AGH, hosted by UAM from 21/02/2009 to 07/03/2009*
- *"Network planning algorithms and optimization considering periodic traffic patterns" Nina Skorin-Kapov, Assistant Professor at FER, hosted by UPCT from 14/04/2009 to 05/06/2009*
- *"SLA-aware Provisioning in Optical Transport Network" Massimo Tornatore, Assistant researcher at PoliMI, hosted by KTH from 07/05/2009 to 09/05/2009*
- *"Planning of Scheduled Virtual Topologies Under Periodic Traffic in Transparent Optical Networks" Ramon Aparicio-Pardo, PhD Student at UPCT, hosted by FER from 21/09/2009 to 23/11/2009*
- *"Comparative evaluation optical switching technologies (related to JA2, WP11)" Pablo Pavon-Marino, Associate Professor at UPCT, hosted by PoliTO from 23/11/2009 to 26/11/2009*

Meetings

WP11 organized a plenary meeting, jointly with WP12 WP21 WP22 WP24 WP26. This meeting was planned, similar to Y1, in June, shifted of about 6 months with respect to the project plenary meeting. In Y2 the focus of the meeting was on presenting the technical achievements of the JAs.

The meeting was in Bologna, lasted 2 days (June 8 and 9) and was organized as a workshop, with one session per WP, including a general presentation of the WP current results and some technical presentation of the selected JAs. About 60 researchers attended the meeting, as reported in the related minutes.

Several meetings in real life or virtual were held in the various Jas to discusses the workplan and the advance of results.

Completed JAs

Common Architecture/Hierarchy for integration of OCS, OBS, and OPS in a common transport plane (JA1)

This Joint Activity (JA1) studied options to integrate all three optical switching paradigms (OCS, OBS, and OPS) in a single transport plane as outlined in the first year report (figure below). Two approaches could be separated: Horizontal – providing all kinds of switching in the same network-layer without a hierarchy, and vertical – cascading the different switching options in a hierarchy of network-layers, as commonly done today, but integrating control and management of such multiple network-layer (mL3) architectures for better utilization of the available bandwidth. There are a lot of possibilities in between, these were not excluded, however, none were proposed. The prime viewpoints considered are the transport properties respectively characteristics and the thereby defined Quality of Service (QoS) an application can demand, and the feasibly achievable utilization of resources.

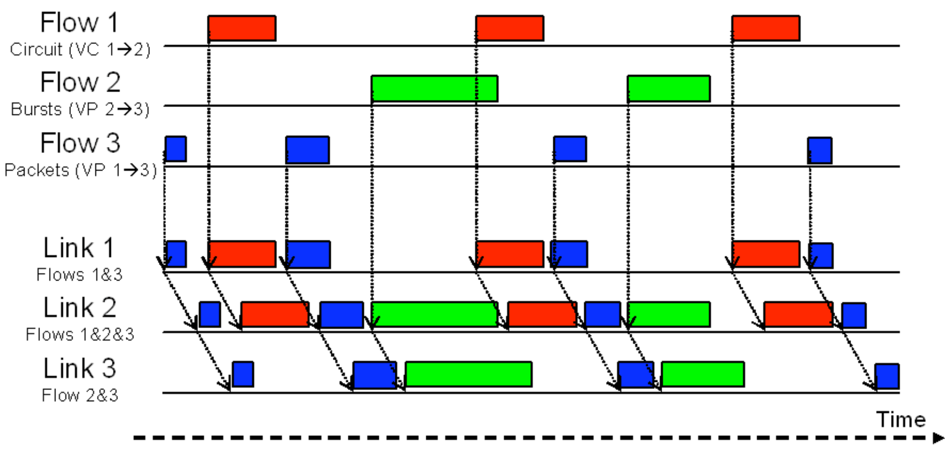


Figure 3. Multiplexing of different types of flows (time-slots, bursts, packets) sharing common transport capacities on connected links to perform transparent end-to-end transport

1.1.1 Objectives

a) Several single partner studies on specific related issues preparing the ground for a sound proposal on unified transport plane providing potential for joint work, short term mobility and publications.

b) A survey on situation and options as currently realistically imaginable based on existing – at least in laboratory - hardware components.

c) A jointly written journal paper presenting one or more options for a transport plane serving all kinds of application demands - including performance analysis/forecast.

Interest in the JA initially was promising; however, as many of the related topics are covered by other more specific JAs and WPs, the intended comparison with legacy packet over circuit technology as well as the all-IP approach using some virtual circuit over packet strategy, could not be achieved. Seemingly the topic is too common or too restrictive to be of interest. We need to admit that publication of the well known in scientific media is not welcome, thus a scientific revenue not achievable for such activities.

1.1.2 Summary of achieved Results

Most JA related effort in 2009 was focused on the publication of the developed proposal. Being based on OBS two attempts were rejected. Finally we rewrote it to be of tutorial type, centred it on OBS, and highlight the potential OBS would have if the control would be extended to support any type of connection. We show that technically this is feasible. However, to be efficient OBS demands to use many wavelengths in parallel; this forbids integration in GMPLS hierarchy - making the proposal currently very unpopular. A complete copy of all results can not be included here as a publication in this report would seriously endanger acceptance of future submissions of related paper(s).

Proposed architecture

The studied scenario (depicted below) compares to radio access: the medium to connect a terminal with the application-server is shared in two domains – the access line and the distribution network. In contrast to radio networks, here the 'critical' resource sharing happens in the distribution network and not at the access lines. While radio bandwidth is a limited resource, the capacity of the distribution network can efficiently be engineered to needs. The approach therefore is scalable to any number of customers. That less capacity needs to be spent on error correction is another goody due to the robustness communication via light grants, as is the transmit power saving due to the guidance the fibre provides.

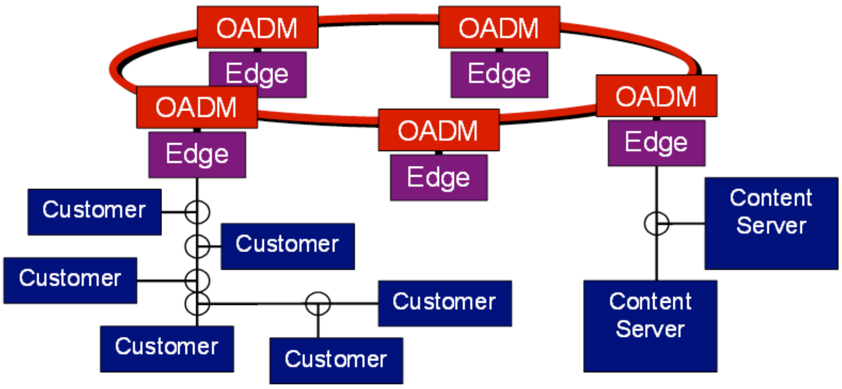


Figure 4. Network example

In order to maximally exploit the advantage of transparency and the power savings from keeping data in the optical domain end-to-end, we do not want to buffer bursts at line terminals respectively edge nodes. If we could assume that a sufficient number of channels is provided, meaning the loss rate due to resource occupancy is acceptable, no additional control mechanism is required. However, therefore we would need many channels, practically achievable only with an equally huge number of any-to-any wavelength converters per node. More practical is an approach foreseeing a mechanism to control the access to resources. Most of the known access protocols could be used; to model any appropriate we relate to the general insertion buffer ring. Arrivals are buffered until capacity on the ring is available. This results in an interrupted service process that can be modelled by an interrupted Poisson process (IPP). Assuming Poisson arrivals we can use the PASTA (Poisson arrivals see time averages) convention to calculate the global behaviour. As there exist only deterministic delay components during transportation, the metric of interest is the waiting-time that bursts stay in the transmit buffers at ONTs. Assuming balanced load (equal load per hop and channel on the ring) this can be derived by the M/G/n queueing model shown.

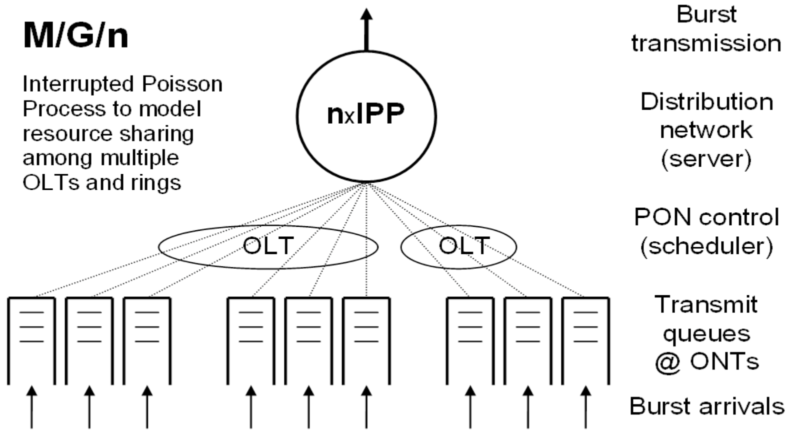


Figure 5. Queueing model for a concentration ring topology

A contradiction to packet insertion rings is the absence of transit buffers within the ring. For packets these are necessary to delay a packet arriving from the ring while a local packet is already in the process of being inserted. With OBS we do not need to delay DBs as we know in advance when the next burst will arrive – this is an offset-time ahead advertised by the according BCP. Still, there is a problem: prior insertion of a local burst in a sufficiently large gap the according BCP needs to be released. The offset-time this is to be done ahead would need to be big enough to inform the next node in time. This circular dependence can be solved by inserting delay lines that only delay the DBs and not the BCPs as shown in the architecture depicted.

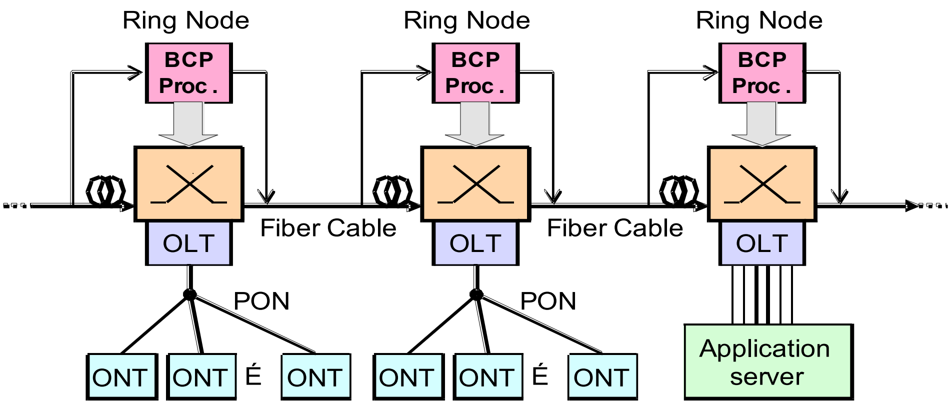


Figure 6. Network architecture

Insertion of these local delay lines solves the BCP timing problem and causes hop-by-hop increasing offset-times. Potentially this could even compensate the fairness problem of insertion buffer rings in case of unbalanced load. Note also that in practice dispersion compensation and fibre amplifiers can be part of the local delay lines.

The mean waiting-time for M/G/n derived using the Allen-Cunneen formula for GI/G/n is given by

$$E[T_W] = \frac{P_n}{\mu\cdot(1-\rho)}\cdot\frac{1+c_B^2}{2\cdot n},\quad P_n = \frac{(n\cdot\rho)^n}{n!\cdot(1-\rho)}\cdot\pi_0,\quad \pi_0 = \frac{1}{\sum_{i=0}^{n-1}\frac{(n\cdot\rho)^i}{i!}+\frac{(n\cdot\rho)^n}{n!}\cdot\frac{1}{1-\rho}} \tag{1}$$

where E[.] is the expectation operator, T_W the waiting time function, P_n the probability for waiting in the corresponding M/M/n system, load $\rho=\Sigma\lambda/(n\cdot\mu)$, π_0 the probability that the M/M/n system is in idle state, and c_B the coefficient of variation introduced by the interrupted Poisson process.

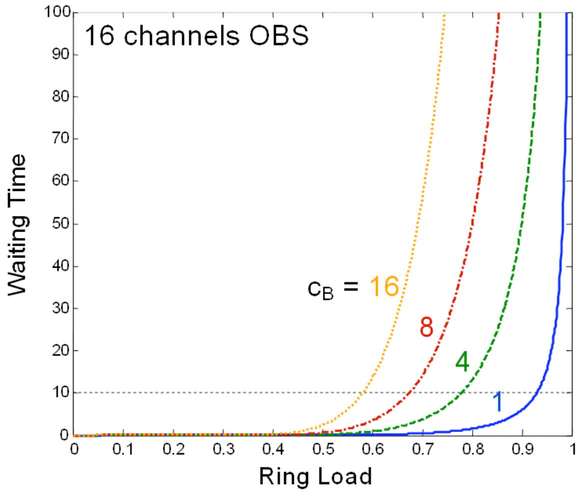


Figure 7. Waiting time for different coefficient of variation (c_B)

The proposed scheme would replace all lower layers, still giving access to all types of connections (transport network services). Switching is performed dynamically in the physical layer; therefore the common layer 2 protocols need to be shifted to the terminals of the transparent optical network domain. In between domains different O/E/O switch architectures (for example OTN/DTM/CGE) could be used to seamlessly interconnect domains for circuit/burst/packet switched connections, and layer 3 switches to route flows, i.e., to add/drop them to/from the different connections.

Application aware transport

IP packet tunnelling toward servers connected to the internet exchange (IP-layer) evidently needs to be supported. In a triple play environment multimedia broadcast is a necessary service: as 50 HD and 200 SD MPEG4 compressed TV-channels sum-up to < 2 Gbps, assembling all streams in parallel enables transport of the entire live program in a single stream of bursts. If the all-optical edge nodes allow duplication, efficient broadcast reaching all customers can be realised. For network storage special file transfer allowing the transmission of entire files by a single burst can be envisaged and also support of applications not known today because the core components are entirely transparent, autonomous, and service independent. There is a drawback: too small traffic volumes can not be efficiently transported. If an application is not extremely latency insensitive, meaning does not allow collecting the content until a DB is filled, there is no way to efficiently transport it. Such shall be groomed with other IP traffic if appropriate. Still, it is not prohibited or impossible to transport partly empty bursts if necessary.

A key procedure of any OBS implementation, and derivation of it, is transport unit assembly. The assembly process is intrinsic and dominates the quality of the service (latency and jitter) the transport channels provided offer to client layers. In the literature many proposals for assembly procedures can be found. Here we concentrate on issues relevant for the core network, namely generated burst-length and burst scheduling (inter-arrival-time) distributions, as these are the traffic generation characteristics relevant for network layer performance. To realize connection types that match flow characteristics we find two basic parameters that can be tuned as needed: a) burst-length and b) inter-burst time interval. Basic scheduling characteristics for different flows are listed in table I.

TABLE I

BURST GENERATION CHARACTERISITICS

Capacity demand	Jitter/Latency sensitivity	Burst-length distribution	Inter-burst time
bursty	no	one can be selected the other follows capacity demand	
bursty	latency bound	similar capacity demand	upper bound
variable (constant mean)	no	one can be selected the other follows capacity demand	
variable (constant mean)	jitter sensitive	equal capacity demand	upper bound
constant	no	one can be selected the other is thereby determined	
constant	both	inter-burst-time determined	upper bound

To support a certain bit-rate of a flow (B_{flow}) the bit-rate provided by consecutively scheduled bursts (flow of bursts) is electrical line-rate (B_{line}) times burst-length (T_{burst}) over mean inter-arrival time ($T_{inter-arrival}$): $B_{flow} = B_{line} \times T_{burst} / T_{inter-arrival}$. Note that electrical line-rate is not necessarily a network-constant: it is the bit-rate with which a terminal modulates the optical carrier (wavelength) - and that can differ from terminal to terminal.

Constraints from the core network shall be considered by assembly in addition to the traffic demands'. These are efficiency related and not as stringent as the service related.

- Minimum burst-length > guard-time
- Mean burst-length deviation shall be small
- Generated burst-rate << BCP processing capacity

Considering the multi-rate variant of the Erlang_B formula we realize that shorter bursts are easier to schedule than extremely long ones. To completely eliminate this unfairness, all burst-lengths would need to be equal. The scheme that supports most service types is to apply individual and adjustable upper and lower bounds on both, burst-length and scheduling-rate. The upper bound on scheduling-rate (inverse inter-arrival-time of bursts caused by one flow) shall be controlled by the network to avoid congestion in the control plane and thereby prevent that a DB overtakes its BCP. An adaptive, window based scheme alike TCP (transport control protocol) might be used as well as more sophisticated access control and flow shaping mechanisms. The lower bound on scheduling-rate is necessary to obey the service's jitter constraint: small scheduling-intervals introduce less jitter. This contradicts the suggestion to use global targets on burst-size and inter-arrival time. Any theoretically optimal solution is in reality out of reach, as the mathematically optimal assembly parameters will entirely fail if individual traffic demands are not met. However, one of the two parameters can, in general, at least to some extend, be tuned to network needs.

End-to-end performance

Since End-to-End performance is a problem of queueing- respectively loss-networks we assume independence among network nodes as commonly applied. Equations (2) and (3) provide performance approximations, where n is the number of hops, i indicates the sequence of nodes, and j the sequence of links along a path.

$$\mathrm{E}[\mathrm{loss}_{\mathrm{path}}] \le \left(1 - \prod_{\mathrm{path}} \left(1 - \mathrm{E}[\mathrm{loss}_{\mathrm{ij}}]\right)\right)$$

$$\mathrm{E}_{\mathrm{M/M/n/n}}[\mathrm{loss}_{\mathrm{ij}}] = \left(1 - Pb_{ij}\right)\rho_{ij}$$

$$\mathrm{E}_{\mathrm{D/D/n/n}}[\mathrm{loss}_{\mathrm{ij}}] = \begin{cases} 0 & , \rho_{ij} \le 1 \\ \dfrac{\rho_{ij} - 1}{\rho_{ij}} & , \rho_{ij} > 1 \end{cases}$$

$$\mathrm{E}_{\mathrm{M/M/1/s}}[\mathrm{loss}_{\mathrm{ij}}] \le \left(k_j \mu_j - \lambda_{ij}\right) \frac{\rho_{ij}^{\,s+1}}{1 - \rho_{ij}^{\,s+1}}$$

$$\rho_{ij} = \frac{\lambda_{ij}}{k_j \mu_j} \qquad (2)$$

In (2) $loss_{ij}$, λ_{ij} and ρ_{ij} indicate loss, arrival rate and load for link j (with k_j channels and service rate μ_j each) at node i. Pb_{ij} is the corresponding loss-rate given by Erlang_B formula. We note that (2) provides a secure upper bound if the M/M/n/n model represents a worst case only, i.e., when streamlining effect and network access control assure traffic aggregates smoother than Poisson. The time needed to transport a burst containing a particular payload unit, is given in (3):

$$\begin{aligned}\mathrm{E}_{\text{one-way}}[\mathrm{T}_{\text{F-path}}] &\le T_{assembly} + T_{guard} + \sum_{path}\left(T_{BCP}(i) + T_{prop}(j)\right) \\ &+ \mathrm{E}[\mathrm{loss}_{\mathrm{path}}]\left(\frac{1}{2}T_{RTT} + T_{guard} + \sum_{path}\left(T_{BCP}(i) + T_{prop}(j)\right)\right)\end{aligned}$$

$$\begin{aligned}\mathrm{E}_{\text{two-way}}[\mathrm{T}_{\text{F-path}}] &\le T_{assembly} + T_{RTT} + T_{guard} + \sum_{path}\left(T_{BCP}(i) + T_{prop}(j)\right) \\ &+ \mathrm{E}[\mathrm{loss}_{\mathrm{path}}]\left(\frac{3}{2}T_{RTT} + T_{guard} + \sum_{path}\left(T_{BCP}(i) + T_{prop}(j)\right)\right)\end{aligned} \qquad (3)$$

In (3) we introduce the round-trip time T_{RTT} to consider delay from signaling. In case of one-way signaling T_{RTT} only applies if a burst needs to be re-scheduled (time required to inform the source about the loss equals half the round-trip time). For two-way reservation the entire round-trip time is required to signal the path, and 1.5 round-trip times in case of loss. For not managed transport (UDP like) the part weighted by E[$loss_{path}$] does not contribute to the 'length' of a connection and the impact of loss is shifted to upper layers.

To depict results we need to assume some load distribution. Equal load distribution is the best case. However, it represents a valid worst case approximation if we apply the maximum load from the most stressed link for all links. To calculate the curves shown in 1.1.2 the following assumptions were made: OBS: $T_{assembly}$ = 10 (average) burst lengths, T_{guard} = 0.1 burst length, T_{BCP}(length of BCP) = 0.01 burst length, JET signaling; OPS: packet size = 0.1 burst length, buffer size = 10x link bandwidth (for loss-rate calculation only), OPS header processing ≤ packet size; both: link length = 0.5 burst length, path lengths are 2, 4, and 8 hops respectively, T_{RTT} = 2 path delays, no re-assembly of bursts in case of loss, every hop identically loaded. The reciprocal load increase due to loss is not considered, i.e. iterative load-point evaluation is assumed.

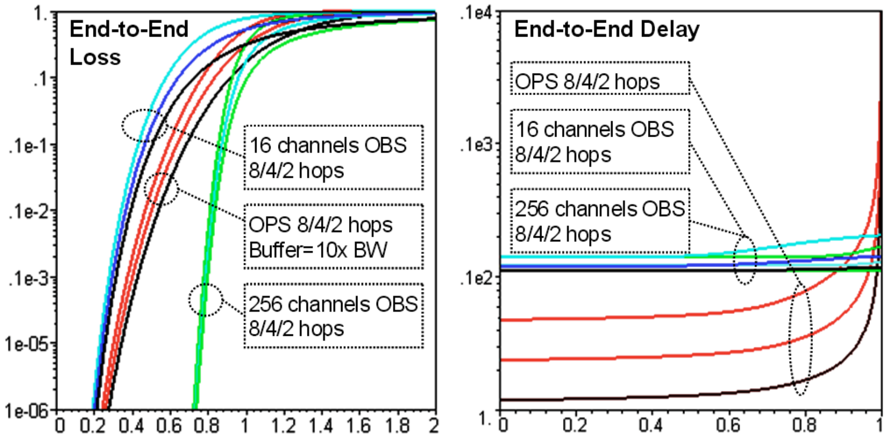


Figure 8. Loss and Latency of OBS and OPS for different path-lengths

The figure shows that OBS with 16 parallel channels causes higher loss than OPS, for 256 parallel channels OBS is better. Considering latency, OPS clearly outperforms OBS. That OBS latency is not very load-dependent is due to being related to the design's load target. Shown are traces for 30% and 80% load target (16 respectively 256 parallel channels) but differ only due to different loss. Due to the small average BCP processing time (required to achieve negligible BCP loss-rates from BCP queueing) latency for OBS clients is clearly dominated by the assembly process.

Cooperation with GMPLS

A hot topic is the implementation of the Control Plane (CP) of the OBS network. Currently, the GMPLS control plane is the dominating solution for integrating circuit and packet based services. OBS was not foreseen at the time of development of GMPLS and different options for the GMPLS/OBS integration can be found in the literature. Assuming a generic transport solution that supports various switching capabilities the integration of an artificial hierarchy would be counterproductive. OBS over a GMPLS controlled infrastructure fails entirely due to the massive demand for parallel resources, which contradicts GMPLS principles. However, an OBS network used as the underlying transport network, connecting real as well as virtual domains of different switching capabilities, is an option for integration of GMPLS. This approach comes closest to the idea of integrating all known services: regular bursts are handled by common OBS control; periodic bursts (i.e., constant bit-rate connectivity) and wavelength channels are controlled by slightly modified GMPLS signalling for resource reservation. In this integration scenario there are two options for the interplay between the GMPLS and the OBS control planes. Depicted below, the OBS network is the underlying transport network and the OBS control plane is kept separate and independent of the controls of the overlay legacy networks.

The potentially most powerful integration option is depicted next. It uses modified GMPLS signalling to perform the requirements for DB signalling, which would facilitate a horizontal integration between GMPLS-controlled legacy networks and the OBS transport network.

Under such an approach the GMPLS signalling is responsible for the actual resource reservation. This implies changes in the GMPLS resource reservation procedure (which needs to support one-way in addition to two-way) and a specific adaptation function at the edge of the OBS network for translation of the traditional GMPLS information into OBS-compliant information.

1.1.3 List of relevant publications

Most JA related effort in 2009 was focused on the publication of the developed proposal. Two attempts were rejected; the first due to poor quality of writing and missing references; the second, now with an extensive list of references, was not considered due to being rated as overview paper. Finally it was rewritten to be of tutorial type and submitted as such.

1. G. Franzl (TUW), S. Aleksic (TUW), B. Statovci-Halimi (TUW), S. Sarwar (TUW), *Quality of Transmission Management in Dynamically Routed All-Optical Networks*, Proceedings of NOC2008, Krems, Austria, July 2008.
2. G. Franzl, and A. Manolova, *On performance and control of burst switched optical networks supporting legacy and future service types*, OSA Journal on Optical Networking, special issue on Optical Networks for the Future Internet, rejected December 2008.

3. G. Franzl, and A. Manolova, *On performance and control of burst switched optical networks supporting legacy and future service types - scalability, flexibility and integrability review*, IEEE Transactions on Networking, returned May 2009.
4. G. Franzl, *An optical burst switched access and distribution architecture*, 11th International Conference on Transparent Optical Networks (ICTON 2009), Vol. WAOR, No. III, pp. Mo.D3.5, Ponta Delgada, Sao Miguel, Azores, Portugal, June 2009.
5. G. Franzl, and A. Manolova, *On performance and control of burst switched optical networks supporting legacy and future service types - a tutorial on scalability, flexibility and integrability*, IEEE Communications Surveys and Tutorials, submitted August 2009, rejected.

As mentioned initially, the intended comparison with packet over circuit and virtual circuits could not be achieved. Please see publications from other JAs listed in this report, publications from WP24 "Edge-to-core adaptation for hybrid networks" and WP26 "Alternatives for multi-layer networking with cross-layer optimization", as well as related publications from outside BONE to compare the proposed integrated approach with other solutions/architectures.

The JA leader thanks all partners for their interest, especially those actually contributing. Eventually the final report shows that we did achieve some nice results and a proposal valid to be investigated in more detail. Hereby the work in this joint activity ends, but not the collaborative work on the topic. For the time being thanks to all, best regards, and good cooperation in ongoing JAs and future tasks.

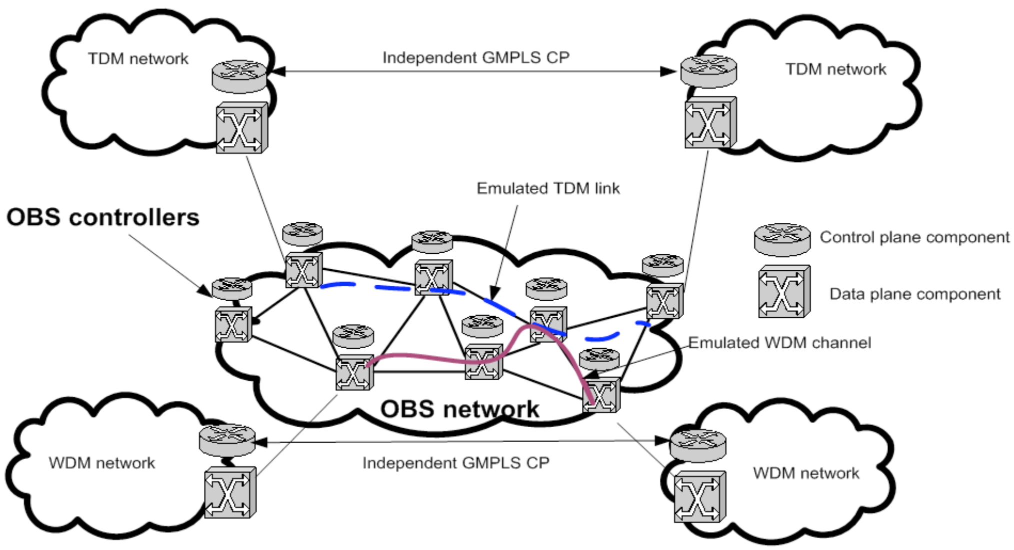


Figure 9. OBS controlled optical core infrastructure providing different services in order to connect different types of GMPLS controlled clients

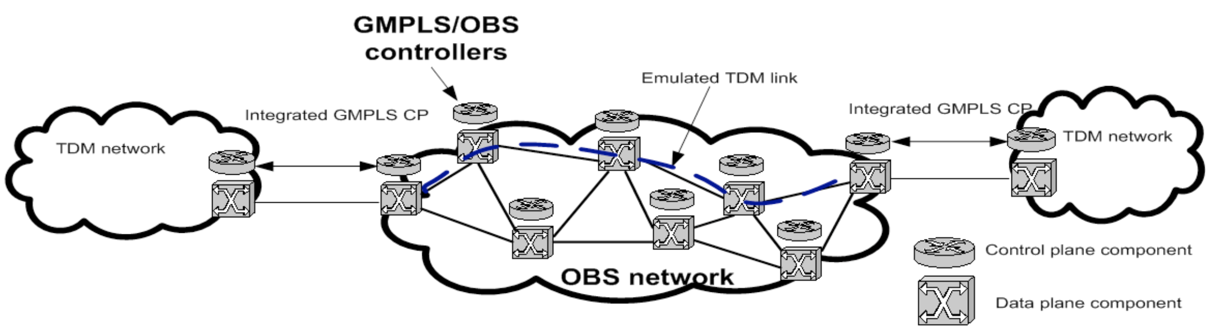


Figure 10. Integrated GMPLS/OBS control merging the merits integrating OBS transport to provision GMPLS controlled services

1.1.4 Objectives

In Optical Burst Switched (OBS) networks employing Just Enough Time (JET) signalling, Burst Control Packets (BCPs) request resources at intermediate nodes following a one-way reservation protocol, that is, requested resources are not confirmed back to the source. Hence, data bursts are transmitted without any guarantees, and it sometimes occurs that these are dropped at a certain hop in the source-destination path, hence wasting resources at previous hops. This effect is specially harmful if some connections are abusing of the global shared resources, violating their respective Service Level Agreements, thus causing: (1) global performance degradation; and, (2) unfair service received by other connections.

This Joint Activity has proposed the Random Packet Assembly Admission Control (RPAAC) algorithm, a new traffic engineering mechanism which monitors the network load and proactively drops incoming packets before they may actually become harmful to the network. Essentially, RPAAC is required to identify which flows are causing network overload and discard some of its data packets at the ingress
nodes, thus before they have actually made any reservation. This is achieved via adjusting the value of the packet selection probability, which regulates the size of bursts and the percentage of proactively dropped traffic, on attempts to either prevent or alleviate network congestion. In addition to this, a timer-based burst-assembly process with random exponentially-distributed timer is employed, which keeps the Poisson nature of the traffic injected to the OBS network, making the analytical process tractable, as shown in the forthcoming.

In addition to congestion prevention, the RPAAC mechanism can also be applied to penalise flows which exceed their Service Level Agreement (SLA) contracted with the network operator. In such case, the RPAAC mechanism adjusts the packet selection probability in the assembly process in a way such that those flows exceeding their SLAs (abbusive or misbehaving flows) are penalised more (lower Admission Control probability) than others which meet their SLAs contracted. Again, this acts as an admission control mechanism that throttles traffic before it causes a serious performance degradation, but also it encourages flows to meet their SLAs, with a clear application in traffic
Engineering.

1.1.5 Summary of achieved Results

This Joint Activity has defined the Random Packet Assembly Admission Control mechanism (or RPAAC), a policy which is designed to drop packets at the border OBS nodes when the network load exceeds a certain threshold. Essentially, incoming packets at the optical network are either accepted or rejected at the burst assembler with a certain probability, which is adjusted by the admission control algorithm. The algorithm includes a monitoring stage whereby all network links are examined whether or not they exceed a load level, in order to calculate such amount of excess traffic per link, and it further detects whichparticular flows arecausingsuchexcess. This information is sent back to the border nodes which adjust the admission control probability for each individual flow. Additionally, the RPAAC follows a proportional penalisation policy, that is, those flows which are abusing of the network resources suffer throughput decrease proportionally to the amount of traffic excess observed. This gives a more fair resource sharing between the connections existing in the network.

Finally, a set of numerical examples over the well-known NSFnet network topology have been carried out and it is shown that RPAAC is capable of reducing and maintaining the load levels within acceptable performance levels.

1.1.6 List of relevant publications

- P. Reviriego, J. A. Hernández, J. Aracil: *Assembly admission control based on random packet selection at border nodes in Optical Burst-Switched networks.* Phot. Network Communications, vol. 18, no. 1, pp. 39-48 (August, 2009)

1.1.7 Objectives

Ring topologies for metropolitan-area network are typically not protected against multiple failures, which unfortunately occur more often than desired. This work proposes an easy evolution towards meshed topologies based on Double Ring topologies with Dual Attachment (DRDA) for the metropolitan-area core network.

The main goals of this joint activity are to study the architecture of DRDA topologies and to evaluate their reliability and availability metrics following the well-known pre-configured protection cycles (p-cycles) as their main recovery mechanism.

We are interested in calculating the probability distribution of the time to disconnection, the mean time to disconnection values, the network availability and the additional capacity required to guarantee a determined availability level, etc. These parameters are very useful to network operators. For instance, the results of this study provide the mean time to repair (MMTR) value that a given ISP provider must guarantee in order to achieve a desired service-time availability. This information is of special interest to providing multicast services which demand full-time connectivity, such as the distribution of IPTV services. In such case, a single isolated node translates to thousands of users without IPTV service, which is unacceptable for most network operators. Furthermore, this work gives a mathematical framework or reference to all those network operators who are willing to deploy highly resilience metropolitan area networks at a moderate cost.

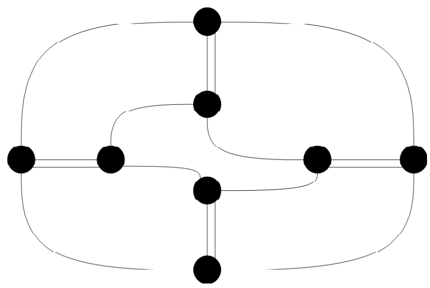


Figure 11. 8-nodes DRDA Network Topology

In DRDA topologies, (see 1.1.7) two former isolated bidirectional rings (the inner and the outer rings) are interconnected such that every node in the outer ring is directly connected with its associated node in the inner ring, via double links (dual attachment). This topology provides a high level of connectivity and reliability. A DRDA network with k nodes consequently has 2k links. For instance, in an 8-node DRDA there are 8 nodes and 16 links. The main advantage of DRDA topology is that it permits to define two dual p-cycles, as shown in 1.1.7.

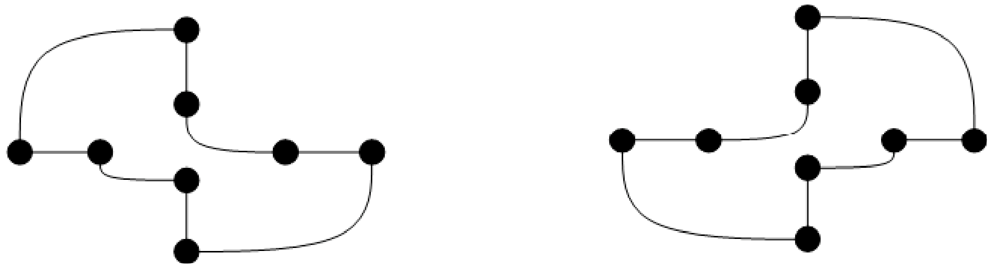


Figure 12. Dual p-cycles structures on DRDA Network Topology

We can define three types of spans whose failure can be recovered by the two dual p-cycles:

- Full-straddling span failure (F-S): In this case, a span failure is recovered by the link's dual p-cycle, over the two different paths defined on it (see 1.1.7 top). To recover the C units of capacity carried by the failing link, the two paths on the dual p-cycle are required to provide C/2 of protection capacity each.
- Semi-straddling span failure (S-S): In this second case, a failure is recovered again by its dual p-cycle, but this time this is done by only the shortest path among the two possible ones defined by the p-cycle (see 1.1.7 middle). This case requires the backup path defined over the p-cycle to have C units of extra capacity.

- On-cycle span failure (O-C): In this final case, a span failure is recovered by its actual p-cycle, that is, by the same p-cycle which contains the link that happened to fail (see 1.1.7 bottom). To recover from this failure, C units of extra capacity are required to transport the data on the opposite direction of the link's actual p-cycle.

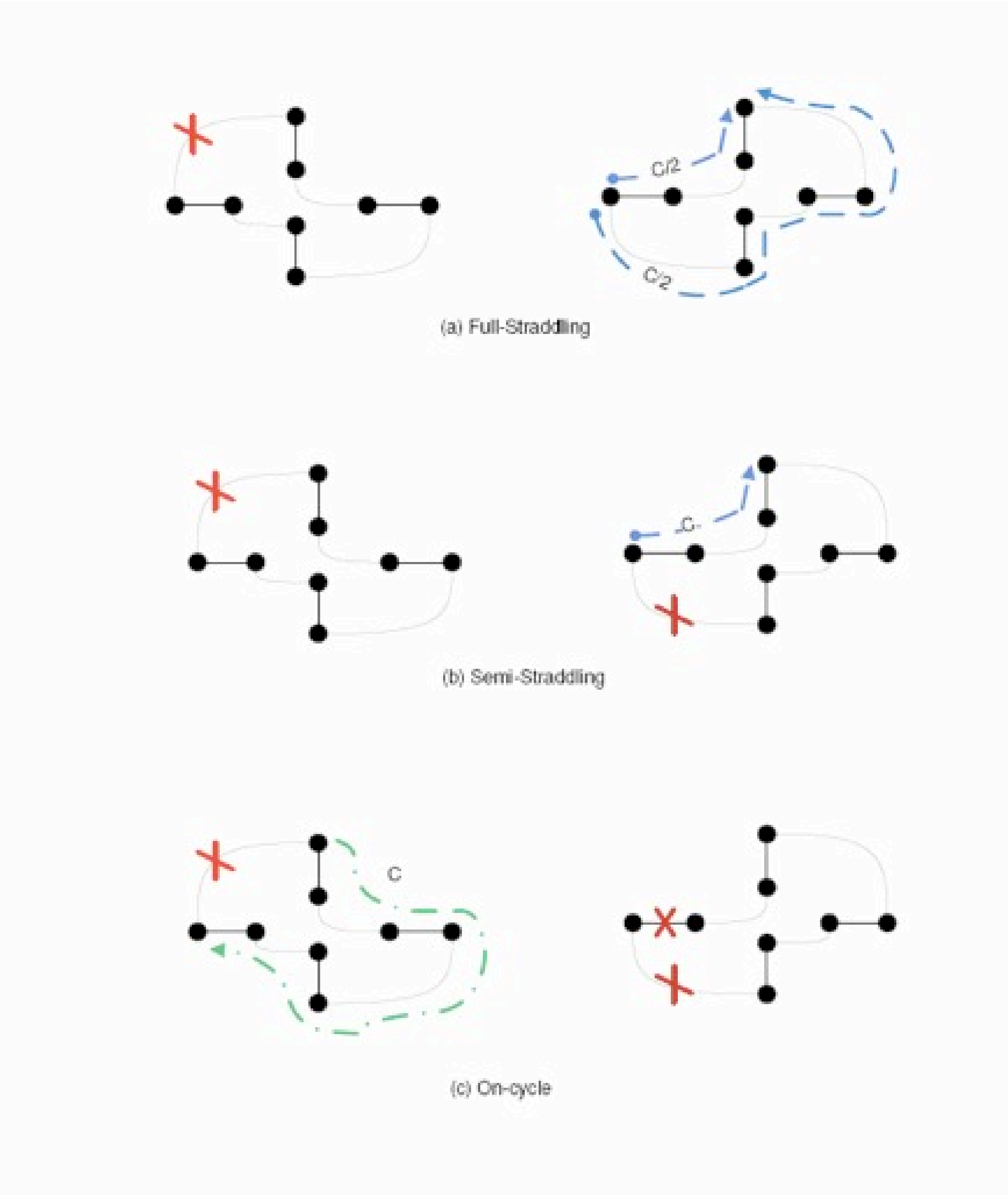


Figure 13. Recovery strategies upon single link failure: Full-Straddling (top); Semi-Straddling (middle); and On-cycle (bottom)

Depending on which p-cycle the failure occurred and the number of failures, we can distinguish several cases, which are shown in 1.1.7.

Failure Case	Failure Type	Required Capacity	Maximum Utilization
n:0 ($1 \leq n \leq 8$)	n F-S	nC/2	2/(n+2)
1:1	2 S-S	C	½
m:1 ($2 \leq m \leq 8$)	m S-S + 1O-C	(m+1)C	1/(m+2)
m:n ($2 \leq m,n \leq 8$)	DISCONNECTION	-	-

Figure 14. Different cases of failure

- (n:0): This state considers n span failures that occurred on one p-cycle, but no failures occurred on its dual p-cycle, where $1 \leq n \leq k$. All such failures on the same p-cycle are thus recovered on its dual p-cycle following a Full-Straddling strategy. This state thus consumes Cb = nC/2 of extra capacity on the dual p-cycle to recover from all failures.
- (1:1): This state considers two link failures, one on each p-cycle. In this case, the two link failures are recovered on their dual p-cycles, following a Semi-Straddling strategy. This state requires at least Cb = C units of extra capacity on each p-cycle.
- (n:1): In this state, one p-cycle has suffered n link failures, with $2 \leq n \leq k$, and its dual p-cycle only one link failure. Any failure on the first p-cycle can be recovered in its dual one by means of only one path, that is, following a Semi-Straddling recovery mechanism. The single failure on the second p-cycle is recovered over it following an on-cycle policy. Therefore, this state may require an amount of backup capacity of, at least, Cb = (m + 1)C to recover from all link failures.
- (m:n): When m, $n \geq 2$, this case leads to disconnection, that is, the all-terminal connectivity between any two pair of nodes is not guaranteed regardless of the amount of backup capacity provisioned.

Only the state (m : n) with m, $n \geq 2$ brings a disconnection situation to the DRDA regardless of the amount of extra capacity provided by the two dual p-cycles. The following analysis assumes that links in the DRDA happen to fail independently from one another. Additionally, link failures occur with a memoryless nature, that is, the inter-failure times are exponentially distributed with rate

λ failures per unit of time. The value of $1/\lambda$ shall be referred to as the Mean Time Between Failures (MTBF). Also, links are assumed to be repaired by the network operator following again an exponential distribution with rate μ repaired links per unit of time. Now, $1/\mu$ is referred to as the Mean Time To Repair (MTTR). The main goal of this study is to find the Time To Disconnection (TTD) probability distribution function of a generic k-DRDA topology and derive its average value of Mean Time To Disconnection (MTTD), given an observed average link failure value of MTBF and provided that the network operator can guarantee a certain average link repair time given by its MTTR value. To simplify the model, no more than four link failures are assumed to occur simultaneously. Indeed, the probability to have more than four link failures simultaneously is less than 10e−6 for MTBF = 60 days and traditional MTTR values of hours and days.

With these assumptions, a given k-node DRDA can be easily characterised and analysed with the nine-state Reliability CTMC shown in 1.1.7, which reads as follows: The generic state (m:n) gives the number of failures on the two p-cycles, together with the required units of backup capacity Cb for that state. For instance, the state labeled (2:1) means that two of the k links of one p-cycle have failed together with one of the k links of its dual p-cycle. In such a case, 3C additional units of backup capacity are required. The diagram in 1.1.7 also gives the transition probabilities between states.

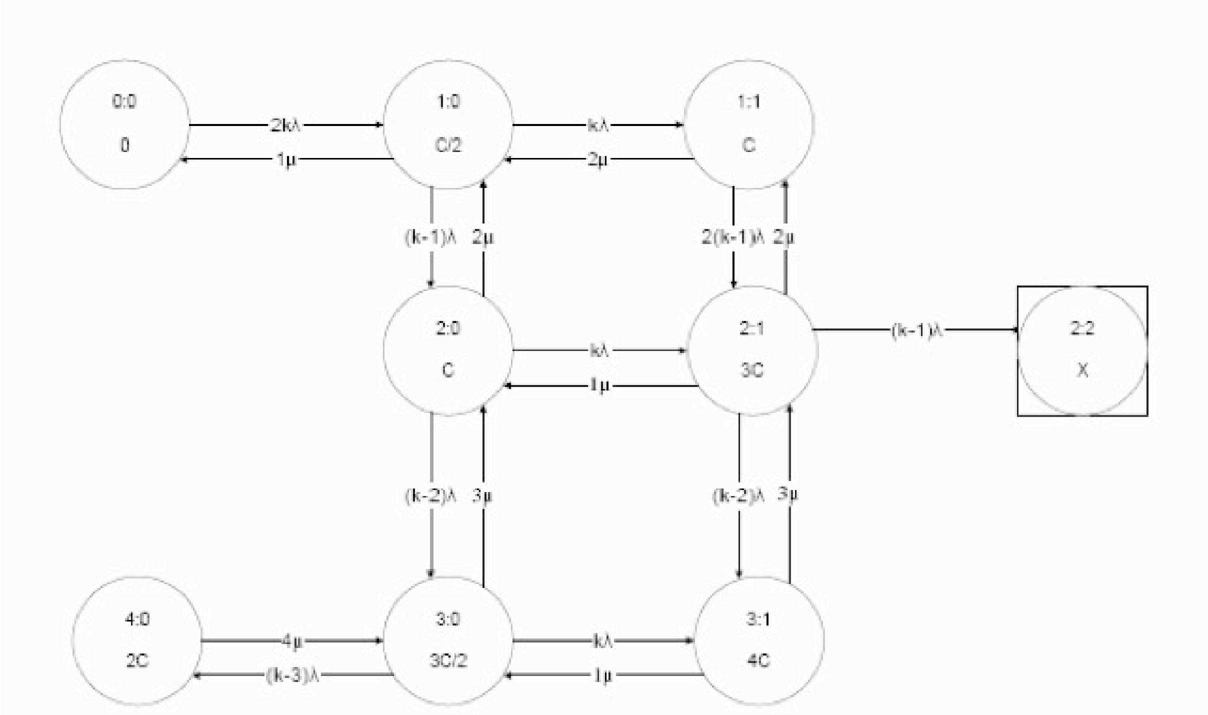


Figure 15. The 9-state Reliability Markov Model for a generic k-DRDA

To compute the Time To Disconnection probability distribution from initial state (0:0), it is just required to choose the entry ((0:0),(2:2)) in the matrix A(t) = e−Gt where A(t) gives the distribution function of the passage time between any two states of the CTMC within the time interval [0, t], and matrix G is the transition probability matrix for this chain.

Finally, it can be defined the Availability Markov model which is the same diagram as the Reliability Markov Model depicted in Fig.4 but with state transition back from (2:2) to (2:1). This model makes it possible to compute the stationary probability distribution of all states, therefore, the amount of backup capacity required by the DRDA on average per unit of time.

Deriving such stationary probabilities of states requires solving the linear equation system:

$$\Pi_i q_i = \sum_{j \in S} q_{ij} \Pi_j \text{ for } i \in S$$

$$\sum_{j \in S} \Pi_j = 1$$

1.1.8 Results

This section provides a set of numerical examples to show the applicability of the equations derived in previous sections. Basically, these include the study of DRDA topologies with (a) different MTBF and MTTR values and (b) different topology sizes. Moreover, it is shown that a network operator must guarantee the appropriate MTTR in order to assure a given service-time availability for different observed MTBF values. Finally, it is studied how much backup capacity it is necessary to provide in order for a network operator to guarantee a given service availability level.

Concerning designing purposes, it is important to find the MTTR that a network operator must be compromised to in order to guarantee a given disconnection probability over a period of time of one year (360 days), assuming the network has been observed to suffer one failure every MTBF. In light of this, 1.1.8 answers this question: it shows the probability to have disconnection over one month (30 days), over one half year (180 days) and over one year (360 days) considering MTBF = 60 days and different values of MTTR.

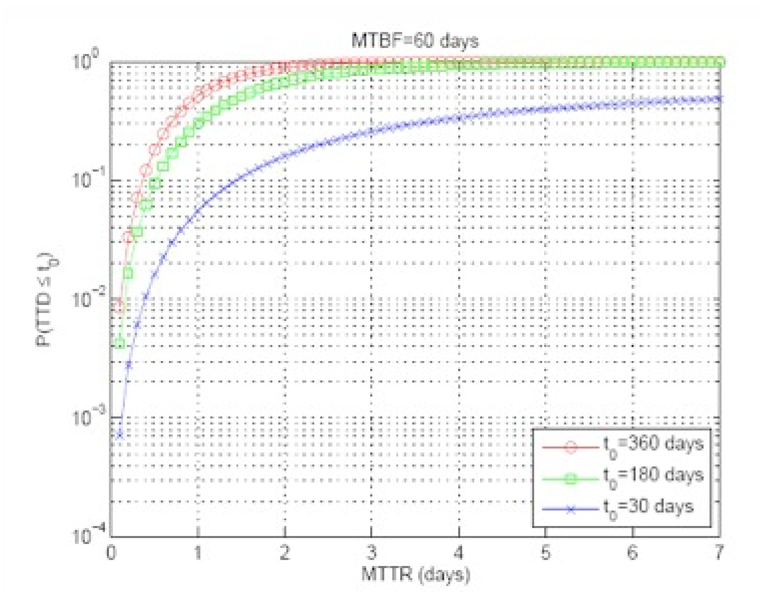


Figure 16. Time To Disconnection probability for different guaranteed MTTRs by the network operator

In terms of service availability, 1.1.8 shows the Service Time Unavailability (STU), computed as:

$$STU = \Pi_{(2:2)}$$

That is, STU represents the average proportion of time in which the network is not available. For instance, if a given DRDA is observed to suffer one failure every MTBF = 180 days and the network operator guarantees an average repair time MTTR in the range of 1 day or below, the target five-nines availability is provided.

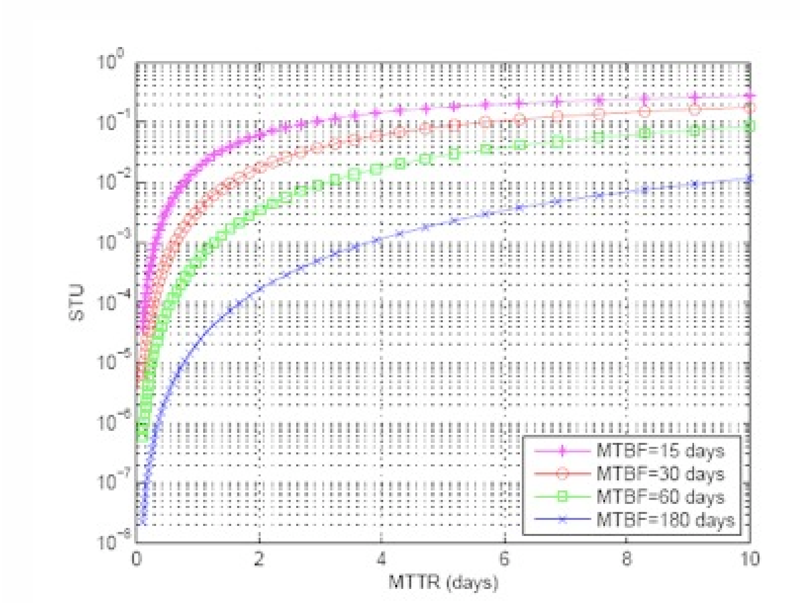


Figure 17. Time Service Unavailability for different combinations of MTTR and MTBF

1.1.8 shows the MTTD values for different k-size DRDAs. There is a decreasing trend of MTTD with respect to k since the more number of links (remark that a k-DRDA contains 2k links) in the topology, the more subject to failure this is. In conclusion, DRDAs are shown to provide high resilience capabilities, but these decrease with its size k (i.e. number of links subject of failure). The network operator must take care of this aspect when designing a given DRDA to cover a certain metropolitan area. Furthermore, when new nodes are included in the inner and outer rings to cover new neighbourhoods the network operator must be aware that the total service availability gets reduced and must provide faster service repair times (reduce MTBF). The p-cycles provide good resilience properties for MANs, but should not be used as a resilience mechanism form WANs, unless an outstanding MTTR value is guaranteed.

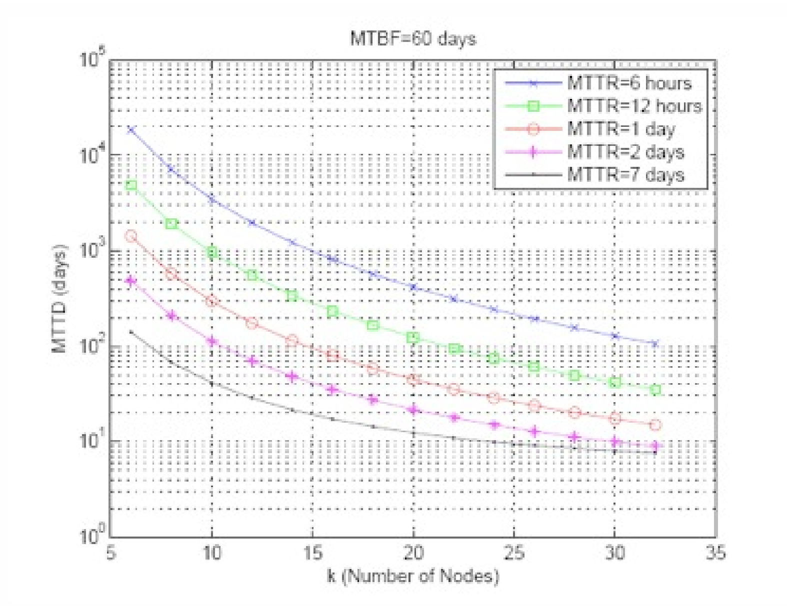


Figure 18. Mean Time To Disconnection for different size of topology with different values of MTTR and MTBF=60 days

1.1.8 gives the stationary probability distribution of the cases that demand backup capacity: Cb = {0,C/2,C, 3C/2, 2C, 3C, 4C}. This gives an idea of the average portion of time over which the network is using such a capacity as backup. Essentially, the provisioning of too much extra backup capacity does not guarantee a much larger service availability level. Only there is significant service availability improvement if Cb = 3C and Cb = 4C. For instance, assuming MTBF = 60 days and MTTR = 1/4 days (6 hours), if the network operator guarantee an extra capacity of Cb = 1C, a saturation probability of 1.28 × 10−3 is obtained, whereas if Cb = 4C is guaranteed, the five-nine service availability is provided.

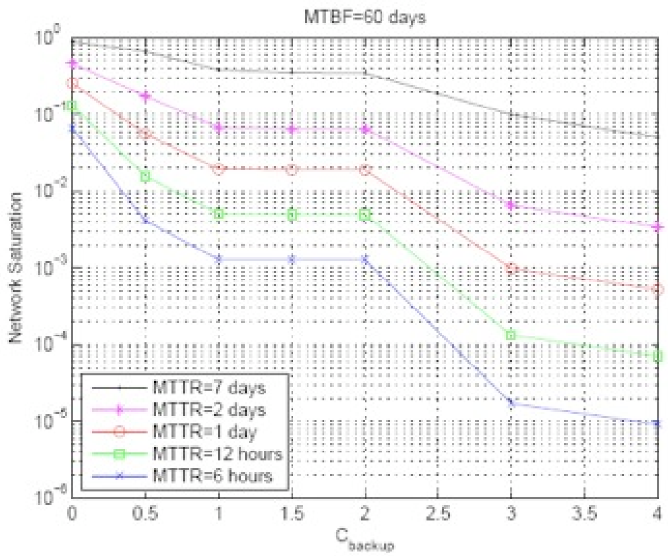


Figure 19. Network Saturation for different additional Capacities with different values of MTTR and MTBF=60 days

1.1.9 Conclusions

This work introduces Double Ring topologies with Dual Attachment (DRDA) and studies their resilience capabilities against link failures. This topology comprises two bi-directional rings of the same size, namely the inner and the outer ring, whereby each node in the inner ring is connected with its associated node in the outer ring via dual-attachment, thus leading to a highly-redundant

topology configuration. Such a solution is particularly useful when each pair of nodes (inner node and its dual attached outer node) are physically close, and the cost of connecting both nodes via dual attachment is small. This is the case for most metropolitan area network of big cities.

Such resilience capabilities are modelled by a Continuous Time Markov Chain which, after solving, provides a useful insight in: (1) The repearing times that a network operator must guarantee to achieve a given service availability; (2) the service availability provided by different size DRDA topologies and their implications in adding new nodes in the inner and outer rings; and, (3) the service availability provided with respect to the amount of backup capacity dedicated to recover from failures.

This information is of special interest to providing multicast services which demand full-time any-to-any connectivity, such as the distribution of IPTV services. In such applications, a root node serves IPTV content to all other nodes in the topology, which further distribute this signal to a number of DSLAMs connected with it. In such a case, the isolation of a single node translates to thousands of users without IPTV service, which is unacceptable for most network operators.

Survey on OBS routing methods for OBS networks (JA10)

1.1.10 Objectives

In last two years, the research groups of Universitat Politècnica de Catalunya (UPC, Spain) and Instituto de Telecomunicações (IT, Portugal) have studied the problem of routing in optical burst switching (OBS) networks. The intention of this study was to put together a survey on routing methods in OBS with the aim of providing a comparison of the performance of selected routing strategies.

There were some motivations behind that work. Firstly, in the literature there is lack of a paper that would gather together and classify a variety of routing strategies that have been proposed for OBS networks. Secondly, to the best of authors' knowledge, there is lack of papers comparing performance of routing methods in OBS. Finally, broad review of the literature would allow recognize still not-deeply addressed or open issues related to the routing problem.

In the first step of this joint activity, the focus was on the review of the literature and two outcomes were achieved: 1) a detailed classification of routing methods, and 2) an investigation of analytical OBS network loss models with the aim of recognizing the applicability of these models for routing problems. The next goal was to perform the evaluation of the performance of selected algorithms using the UPC and IT simulation tools. After the successful validation of the simulators in a unified network scenario, performance evaluation of selected linear and non-linear optimization methods applied in multi-path routing was performed. The final step of this activity was the preparation of a joint UPC-IT journal paper reporting obtained results. The paper was submitted to the Optical Switching and Networking journal in October 2009.

A summary of the results obtained in this activity is presented in the following subsection.

1.1.11 Summary of achieved Results

A1.1.11.1 Routing methods in OBS networks

A great number of routing strategies have been proposed for OBS networks in the literature. In general, these strategies can be classified as alternative (deflection), multi-path (source-based), or single-path routing strategies.

A) Alternative routing

A great part of research on routing problem in OBS networks concerns alternative (or deflection) routing. In alternative routing, when the burst contention occurs, a deflective mechanism reacts to it and re-routes a blocked burst from the primary to an alternative route. Deflection routing can be combined with other burst contention resolution mechanisms.

A great part of research on routing problem in OBS networks concerns alternative (or deflection) routing. In alternative routing, when the burst contention occurs, a deflective mechanism reacts to it and re-routes a blocked burst from the primary to an alternative route. Deflection routing can be combined with other burst contention resolution mechanisms.

Routing strategies considered for alternative routing in OBS networks can be either non-adaptive or adaptive.

In non-adaptive alternative routing both primary and alternative routing paths are fixed (static), and in most cases calculated with the Dijkstra algorithm. A number of alternative paths can be given from a node to the destination. Routing decision is taken in isolation, based only on local node congestion state information.

Adaptive alternative routing strategies apply a proactive calculation of alternative paths as well as their dynamic selection. The calculation of alternative paths can be performed in an optimized way with the assistance of linear programming formulations. These methods need the information about network topology and traffic demands. In the case of dynamic alternative route selection some heuristics methods are used. In particular, either threshold-based, path-rank (priority) or probabilistic route selection techniques are applied. Dynamic route selection methods require the distribution of some link/node state information between respective nodes. Some of the alternative routing strategies support QoS provisioning by routing differentiation with respect to the quality class.

B) Multi-path routing

Multi-path routing strategies in OBS networks aim at distributing adaptively the traffic over a number of routing paths in order to reduce network congestion. Although some proactive optimization techniques can be found, Dijkstra's shortest-path algorithm is still the most explored method for pre-calculation of routing paths. In most cases a small number of disjoint SPs with respect to the number of hops is calculated between each source-destination pair of nodes. In OBS multi-path routing, the selection of routing path is performed by the source node. The path selection can be either according to a given probability, like in the multi-path routing with traffic splitting, or according to the path congestion rank. Some authors propose centralized optimization methods for the calculation of the traffic splitting vector, whilst the others apply distributed heuristic methods. A ranking of the less-congested paths usually is obtained with some distributed heuristic algorithms. In both cases the distributed methods need updates about the network state information, which is usually disseminated from the intermediate/destination nodes to the source nodes. Such signalling messages can be either broadcasted or based on some events, like for instance, the burst dropping event.

C) Single-path routing

Both non-adaptive (static) or adaptive (dynamic) strategies are considered for single-path routing in OBS networks. Static routing is usually based on Dijkstra's shortest path calculation with respect to the number of hops.

Adaptive single-path routing aims in burst congestion avoidance thanks to a proactive path calculation. The path calculation can be performed either in a centralized or in a distributed way. Centralized (or pre-planed) routing in OBS, in most cases, makes use of the optimization theory with (mixed) integer linear programming formulations. In each case it is supposed that a route computation unit has knowledge about network topology and (long-term) traffic demands. On the contrary, distributed routing uses some heuristics. Node state statistics are broadcasted, usually in a periodical manner, and used to calculate link weights (costs) in the respective nodes. Then a Dijkstra-like calculation is applied in order to find the lowest cost route. Some of the adaptive single-path routing strategies also support network resilience by the computation of backup paths.

A1.1.11.2 OBS network loss models

Here, we review basic analytical models proposed for OBS networks. The presented models are based on the information of traffic demands and for a given set of candidate routing paths allow to estimate, in the first instance, the traffic load offered to network links and, in the second instance, the burst loss probability in the entire network. These models, together with appropriate optimization procedures, can be effectively applied in load-balancing routing algorithms, as described later in this Section.

A) Calculation of link loads

A common method of the link load calculation in an OBS network was proposed by Rosberg et al. [Ros03] and it makes use of the so-called reduced load (RL) calculation. This model is an extension of the model proposed by Kelly [Kel86] for circuit-switching (CS) networks. In the OBS network, it is assumed that the traffic offered to a network link is obtained as a sum of the traffic offered to all the paths that cross this link reduced by the traffic lost in the preceding links along these paths.

The main difficulty in the RL model is the calculation of losses in network links. Indeed, there is no closed-form expression to compute link losses if a wavelength-continuity constraint (i.e., a burst can not change the wavelength) is imposed in the network. Therefore, the common simplification in the literature is that the network has a full wavelength conversion capability, i.e., any wavelength can be assigned to a burst in a link if only it is available. In such case the blocking probability on each link is given by the Erlang B-loss formula (see [Ros03]).

The calculation of link loss probabilities together with the calculation of offered burst traffic, given by the reduced load model, leads to a fixed point equation with a solution known as the Erlang fixed-point. The fixed point cannot be solved in a closed form but its approximation can be found through repeated link load and link loss calculations. It is known that the fixed point exists in both CS and OBS networks (see [Kel86] and [Ros03], respectively). Although the fixed point is unique in CS networks, still, its uniqueness has not been proved in OBS networks.

Although, according to a common assumption, the traffic offered to a route is Poisson, still it may be thinned by blocking at the consecutive links and thus no longer follows the Poisson statistics. Since there is no straightforward solution to this problem the common simplification is that the burst arrival process to each link is assumed to be Poisson.

The RL model may bring some computational difficulties, especially, with regard to the calculation of partial derivatives for optimization purposes. Also, it can hardly be used in linear programming formulations due to its nonlinear character. Therefore, it is sometimes convenient to consider a simplified non-reduced load (NRL) model, where the traffic offered to link is calculated as a sum of the traffic offered to all paths that cross this link. The rationale behind this assumption is that under low link losses, observed in a properly dimensioned network, model RL can be approximated by NRL.

B) Network loss models

Having calculated traffic loads offered to network links, given by RL or NRL, and burst loss probabilities on links, given by the Erlang B-loss formula, in the next step we may estimate the network-wide burst loss/blocking probability. Such a calculation for an OBS network has been presented in [Ros03], and it uses the same formulation as the one proposed for circuit-switching networks [Kel86]. Hereafter, we name this model an overall network loss (NL) model. In the NL model, in order to calculate the path loss probability, we assume that burst blocking events occur independently at the network links. The overall burst loss probability is calculated simply as the volume of burst traffic lost in the network normalized to the volume of burst traffic offered to the network.

Another approximate method for calculation of burst losses in the entire network is based on an overall link loss (LL) model [Gir98]. In this method we simply sum up the volumes of traffic lost on individual network links. Note that LL overestimates actual burst losses given by NL, because it counts twice the intersection of blocking events that occur on distinct links. In fact, this estimation may be greater than 1 and thus it cannot be considered as the probability metric. Nevertheless, for link blocking probabilities approaching 0, the blocking events that occur simultaneously vanish rapidly and model LL converges to model NL.

Eventually, taking into account different methods to describe the link load and the network loss calculation, which were presented above, several network loss models can be distinguished: NL-RL, NL-NRL, and LL-NRL. The last possible combination of the link load and the network loss calculation is LL-RL. Because such a model does not bring much gain with respect to the NL-RL one, since it does not avoid the complexity of fixed-point calculation, we do not consider it in this analysis.

In [COST293] we studied the accuracy of both NR-NRL and LL-NRL network loss approximations relative to NL-RL, which is a well-known OBS network loss model. The accuracy of both approximate models is very strict for the blocking probability in the network below 10^{-2}. Therefore, we can assume that any of the presented analytical models can be effectively applied for network optimization purposes as far as the network is properly dimensioned, i.e., it experiences low burst losses.

A1.1.11.3 Optimization methods for load balancing routing in OBS

Below, we focus on load balancing methods that apply multi-path source routing and make use of analytical models presented above. The multi-path routing approach allows improving the network performance by splitting the offered traffic over several paths, which permits to reduce the load on the more congested links. In addition, source routing allows a source to directly control network performance by forcing data to travel over one path to prevent congestion on another. We assume a probabilistic traffic splitting approach, i.e., a fraction of traffic load is routed over each candidate path.

We have studied two routing optimization methods that are based on 1) linear and 2) nonlinear problem formulations. The common idea is to solve an optimization problem in order to find a vector of average fractions (splitting factors) of burst traffic routed through the candidate paths such that minimize the burst contention. In a network scenario, the calculated traffic splitting factors, when uploaded to source nodes, can be used to make routing decisions, such as selection of routing paths, for individual bursts or flows of bursts.

A) A link congestion reduction optimization method

The main objective of this method is to reduce congestion at the bottleneck links by reducing the burst traffic load offered to those links. Clearly, contention will occur more often in the most congested links of the network (bottleneck links), wherein the limited number of wavelengths is being shared by the largest amount of burst traffic. The traffic load offered to a link is estimated with the NRL model. Under this assumption, the problem of traffic load distribution over the set of candidate paths can be solved using Linear Programming (LP).

B) A network-wide BLP optimization method

This method aims at the optimization of overall burst loss probability (BLP). The network-wide BLP is the primary metric of interest in an OBS network since it adequately reflects the congestion state of the entire network. In the optimization problem, this metric is represented by a properly selected objective function. To define the objective function any of the network loss models presented previously may be used. In particular, such objective function can be defined from NL-RL, NL-NRL, or LL-NRL. Note that the usefulness of the approximate loss models is justified by their high accuracy, as discussed in [COST293].

The resulting problem is a non-linear optimization problem (NLP). Taking into account the form of network loss constraints of the RL model and the NRL model, which have to be incorporated into the problem, a particularly convenient optimization method is the Frank-Wolfe reduced gradient method (algorithm 5.10 in [Pio04]); this algorithm was used for a similar problem in circuit-switched (CS) networks [Har76].

In general, gradient methods are iterative methods used in the optimization of nonlinear functions. Gradient methods require the calculation of partial derivatives of the objective function (i.e., the gradient) as a way to find the direction of improvement of this function. In [COST293] we provide appropriate formulae for the gradient calculation as well as we discuss the properties of each of the considered objective functions of NLP.

A1.1.11.4 Numerical results

A) Computational effort of optimization methods

The calculation of the objective function (OF) and the gradient (SOLV) is highly time consuming in the NL-RL model, even in a small network scenario (see 1.1.11.4). On the contrary, such a calculation is not an issue if either NL-NRL or LL-NRL model is used. It is worth to mention that by decreasing the value of the termination tolerance parameter, which determines the termination condition of the solver function, we significantly accelerate the optimization procedure (see SOLV column in the table) without substantial decreasing the routing performance (compare 'BLP' value in both EON28 scenarios). Moreover, we can see that when increasing the number of paths the computation time of the solver function increases considerably in a larger (NSF) network scenario. Also, the application of fast gradient calculation in NL-NRL speeds up significantly computations with respect to the standard method. Finally, the computation times of the LP method (see last column) are significantly faster than of the non-linear optimization method, what is expected in view of the simpler approach. It is worth to mention that the computation times of this latter method might be possibly improved since to solve the problem we used a general-purpose solver function of Matlab, which was not optimized for this method.

Figure 20. Comparison of computation times.

Network	Paths	BLP	NL-RL		NL-NRL			LL-NRL		LP
			OF	SOLV	OFs 1)	OFf 2)	SOLV	OF	SOLV	
SIMPLE	2	$2.4\cdot10^{-3}$	64sec	1.5sec	0.3sec	0.1sec	1.4sec	0.1sec	1.4sec	-
SIMPLE	4	$2.4\cdot10^{-3}$	243sec	3sec	0.46sec	0.1sec	3.4sec	0.1sec	3.1sec	-
NSFNET	2	$4.6\cdot10^{-2}$	>5h		9.9sec	0.38sec	22.3sec	0.37sec	24sec	0.18sec
NSFNET	4	$3.1\cdot10^{-2}$	>5h		139sec	1.3sec	850sec	1.1sec	852sec	0.31sec
EON28	2	$3.07\cdot10^{-2}$	>5h		1483sec	7.4sec	1757sec	7.2sec	1789sec	-
EON28 3)	2	$3.09\cdot10^{-2}$	>5h		356sec	1.8sec	509sec	1.6sec	508sec	-

1) standard gradient calculation method is used

2) fast gradient calculation method (as in [Kli07]) is used

3) the termination tolerance parameter equal to 10^{-3}

B) Uniform and non-uniform traffic

We have evaluated the average (network-wide) Burst Loss Probability (BLP) performance of (proactive) multi-path source routing that operates with optimized traffic splitting vectors, calculated by linear and non-linear optimization methods. In the remainder, the routing algorithm that applies the LP optimization method is referred to as the Load Balancing based on Lossless approximation (LBL) algorithm, whereas the algorithm using the NLP optimization method is referred to as the Network-wide Burst Loss model-based (NBL) algorithm. NBL makes use of the NL-NRL network loss model. As a reference, a Shortest Path Routing (SPR) algorithm is considered. In this study, static traffic conditions are assumed, i.e., long term traffic demands are given and they do not change during the evaluation period.

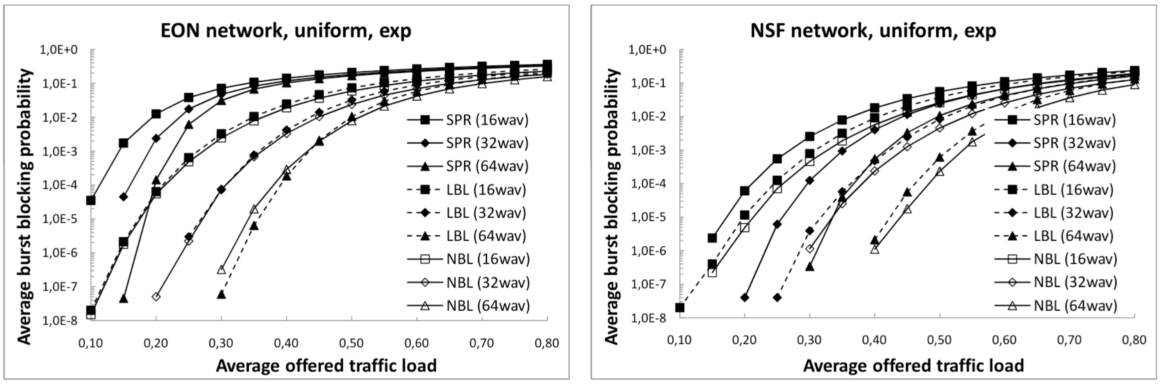


Figure 21. BLP as a function of the average offered traffic load under uniform traffic in EON and NSF.

1.1.11.4 plots the average BLP as a function of offered traffic load under uniform traffic with exponential burst inter-arrival times (IAT) for different number of wavelengths per link for the EON and the NSF network. In both networks we can observe similar results. Both LBL and NBL outperform SPR, whichever link dimension is given. The performance gain of multi-path routing is significant, especially under low and moderate traffic loads. Comparing LBL and NBL, we can see that optimization methods used in both algorithms allow achieving similar results with a slight prominence of NBL that benefits from the application of a more accurate network loss model. It is worth to notice that NBL distributes the load very effectively over the network, even when it explores only k=2 paths per each source-destination pair of nodes. Similarly, although LBL may distribute traffic over k=12 candidate paths, still it makes use of at most two paths per node pair in our simulations.

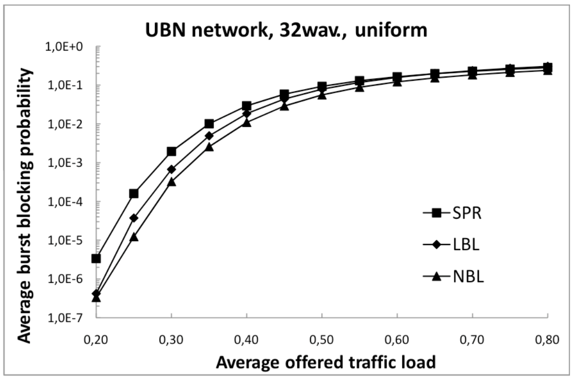


Figure 22. BLP as a function of the average offered traffic load under uniform traffic in UBN.

In 1.1.11.4 we can see that the gain of multi-path routing over shortest path routing may not be so high in a network that is relatively well balanced in terms of the node connectivity, such as the UBN network. Indeed, in such network there are no significant bottlenecks even when routing is just made through the shortest paths. To show that this behaviour is maintained for other traffic patterns, we present in 1.1.11.4 results for the case of non-uniform traffic, considering different traffic matrices (1-10 and 1-20 mean the levels of nonuniformity).

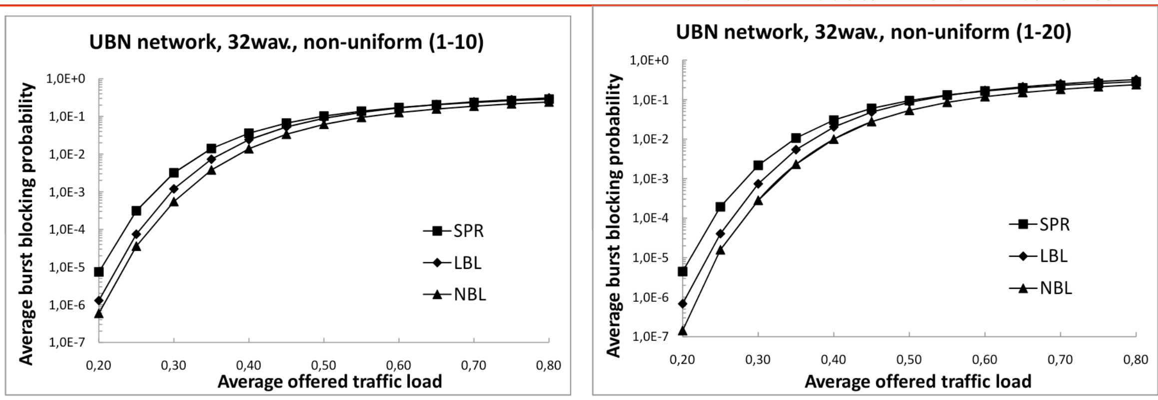


Figure 23. BLP as a function of the average offered traffic load under non-uniform traffic in UBN.

C) Non-Poisson traffic

Apart from the above results obtained for exponentially distributed burst inter-arrival times, in 1.1.11.4 we present performance results obtained for log-normal distributed IATs. First we can observe that in the presence of other traffic characteristics than the Poisson ones, the studied routing optimization models preserve their properties and allow to improve the network performance, in comparison to SPR, even they are based on different traffic assumptions. As an additional conclusion it must be referred that both linear and non-linear optimization methods achieve equally good results in the whole range of traffic loads.

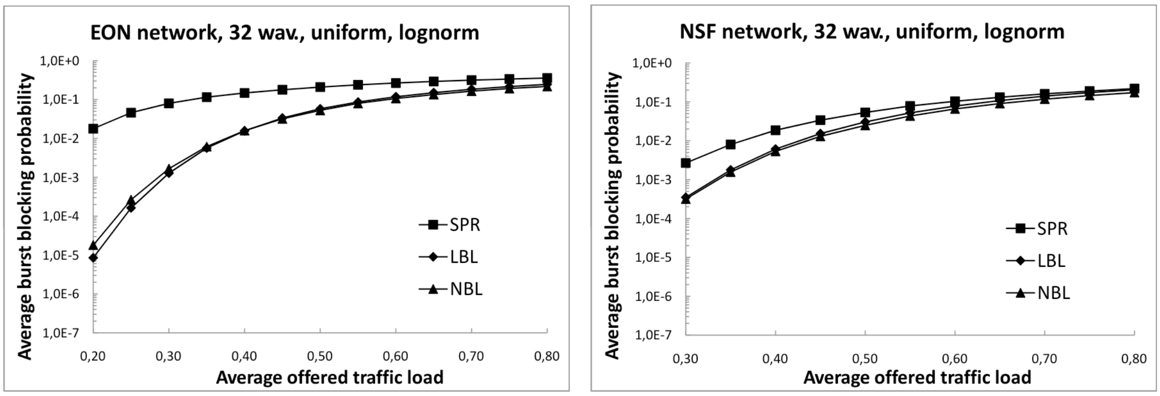


Figure 24. BLP as a function of the average offered traffic load with log-normal burst IAT distribution.

1.1.12 Concluding remarks

The intention of this activity was to review the state-of-the-art literature on routing algorithms, network models, and optimization methods, in order to identify directions for further research, instead of covering completely all the issues related to the routing problem in OBS networks. Our study shows that both linear and non-linear optimization methods deal effectively with the routing problem in OBS networks. Proposed methods are designed in particular for networks with a full wavelength conversion capability. To support networks with wavelength continuity constraints imposed, more advanced network loss models are required. It must be noticed that so far few routing solutions intended for wavelength conversion-less networks have been proposed in the literature. Other observation that may stimulate research is that there can hardly be found optimization-based proposals for the problem of quality of service routing in OBS. Finally, another problem, perhaps not much easier if even feasible, concerns optimized distributed single-path routing.

1.1.13 List of relevant publications

[COST293] A. Koster and X. Munoz, Graphs and Algorithms in Communication Networks - Studies in Broadband, Optical, Wireless, and Ad Hoc Networks. Springer, 2009.

[Gir98] A. Girard and B. Sanso, Multicommodity flow models failure propagation, and reliable loss network design, IEEE/ACM Trans. on Net., vol. 6, no. 1, pp. 82-93, February 1998.

[Har76] R. Harris, The modified reduced gradient method for optimally dimensioning telephone networks, Australian Telecom. Research, vol. 10, no. 1, pp. 30-35, 1976.

[Kel86] F. P. Kelly, Blocking probabilities in large circuit-switched networks, Advanced Applied Probability, vol. 18, pp. 473-505, 1986.

[Kli07] M. Klinkowski, M. Pioro, D. Careglio, M. Marciniak, and J. Sole-Pareta, Non-linear optimization for multi-path source routing in obs networks, IEEE Communications Letters, vol. 11, no. 12, December 2007.

[Pio04] M. Pioro and D. Medhi, Routing, Flow, and Capacity Design in Communication and Computer Networks. Morgan Kaufmann, 2004.

[Ros03] Z. Rosberg, H. L. Vu, M. Zukerman, and J. White, Blocking probabilities of optical burst switching networks based on reduced load fixed point approximations, in Proceedings of IEEE INFOCOM 2003, New York, NY (USA), March-April 2003.

Second year results for running JAs

Experimental tests of Carrier Ethernet techniques

1.1.14 General objectives and summary of results from Y1

Aim of such Joint Activity is the experimental demonstration of techniques for core and metro networks regarding all the routing processes that are migrating from “layer 3” to “layer 1”, with particular interests for protection, restoration, traffic engineering, congestion resolution and resources allocation. In this migration towards layer 1, “Carrier Ethernet” approach will play a key role, and therefore it will be dominant in such JA.

The activities performed in the first year were mainly focused to the management of Quality of Service (QoS) in networks based on Ethernet transmission with introduction of novel optical devices as All Optical Wavelength Conversion (AOWC). These activities were based on experimental investigations carried out by means of the multi-access multiservice IP test bed implemented in the framework of the E-Photon/One project. In particular, in the first year our investigations regarded some developments of Virtual Private LAN Service (VPLS), assumed as a technique close to layer 2 processing, both to control traffic in access networks based on ADSL2+ and to enable the introduction of the AOWC in optical networks.

1.1.15 Advances in Y2

In 2009 we have continued the investigation on techniques for core and metro networks regarding all the routing processes that are migrating from “layer 3” to “layer 2”, with the goal to reach a complete transport based on Ethernet (Carrier Ethernet). Therefore, specific studies have regarded all the techniques that, in different ways, are approaching to Carrier Ethernet: VPLS, MPLS-TP, IEEE 802.1ad (Q in Q), IEEE 802.1ah (MAC in MAC) and PBB-TE, taking into account topics of protection, restoration, traffic engineering, congestion resolution and resource allocation, especially for architectures based on GMPLS control plane. The outcome of such a study is reported in [1].

A1.1.15.1 Towards Carrier Ethernet networks

Ethernet has reached a great success since 1980 as the best technology for LANs; it has become the technology of choice in both home and enterprise local area networks due to its low-cost, flexibility and plug-and-play characteristics. Ethernet transmission rate has increased from 10 Mbit/s, when Ethernet began, to 10 Gbit/s upon the approval of the IEEE 802.3ae standard in 2002, and the work of IEEE 802.3 Higher Speed Study Group (HSSG) will reach 100 Gbit/s in 2010. Moreover, 95% of all data traffic either originates or terminates as Ethernet. All these factors have led many telecom carriers to consider Ethernet as a potential convergence solution for Next Generation Networks. Nevertheless, in order to be adopted in carrier networks, Ethernet must respect requirements that carriers demand in terms of wide area scalability, resiliency, fast fault restoration, Operations, Administration, and Maintenance (OAM) capabilities, and end-to-end Quality of Service (QoS) support.

The term Carrier Ethernet (CE) refers to many technologies that enable the use of Ethernet in Provider networks, but it can be ambiguous. When discussing Ethernet, it is important to specify the context; from a carrier perspective, Carrier Ethernet can refer to:

- Ethernet services;
- Ethernet transport.

Ethernet services refer to packet-based services that ensure the delivery of Ethernet data frames between two or more User Network Interfaces (UNIs), representing points of demarcation between the customer equipment and the network service provider, in respect of a Service Level Agreement (SLA).

Metro Ethernet Forum (MEF) has provided definitions of Carrier Ethernet services, in terms of a set of attributes and parameters, in a number of recommendations for a fastest penetration of Carrier Ethernet services in the market. Ethernet services, however, do not imply an Ethernet transport, but they can be provided using a different transport technology. Most implementations of Ethernet services, in fact, use Ethernet over SONET/SDH or Ethernet over MultiProtocol Label Switching (MPLS).

Ethernet transport, instead, refers to Ethernet as a transport technology, for the transport of different kind of services; the term Carrier-Grade Ethernet (CGE) is referred to a set of technologies that are designed for giving Ethernet the transport feature it is missing. Even if the use of Ethernet has been until now confined to local area and access networks, Provider Backbone Bridging-Traffic Engineering (PBB-TE) has made Ethernet a suitable choice as carrier-grade technology for wide area transport networks.

A detailed investigation on Ethernet transport has been reported in [1], where we have described Ethernet data plane evolution steps, and we have evaluated the technology drawbacks of each step and how such limitations are overcome. We also have identified the main functionalities that the Ethernet control plane shall support in order to guarantee carrier grade reliability and simplified traffic engineering. Finally, we have presented a transport network scenario in which an Ethernet layer is inserted: here it has been described the role that Ethernet can play in next generation transport network, and the advantages deriving from the adoption of this technology, in terms of network resources optimization and convergence toward a packet-based transport network.

The main reason of a transport based on Ethernet can be understood by looking at 1.1.15.1, that shows on the left side the current network layers and on the right side the network evolution based on Ethernet.

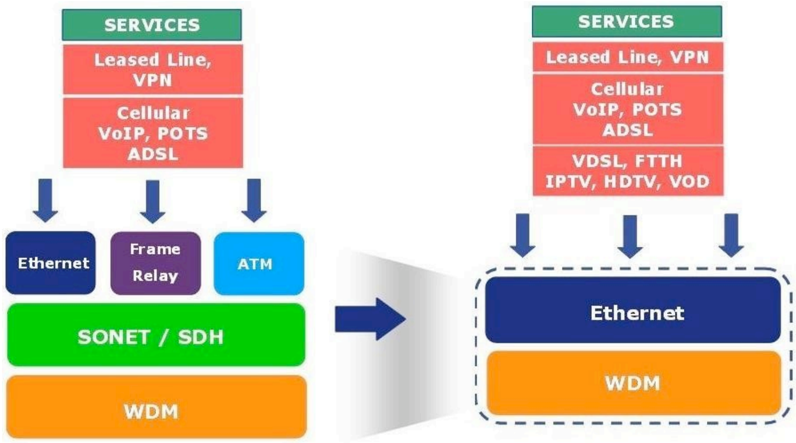


Figure 25. Network evolution

Today services are conveyed by using different techniques as ATM, Frame Relay and obviously Ethernet, to be carried in SONET/SDH channels in a WDM transmission. This step by step process appears quite onerous and, therefore, it appears much simple an evolution that cuts a layer to convey all the services on Ethernet channels in a WDM environment. This evolution is strongly supported by the rising of new generation services that is leading to profound changes in telecommunication networks. In particular, transport networks are undergoing a transition from legacy Time Division Multiplexing (TDM) networks to packet networks for delivering services. In this scenario, carriers are examining two technologies and their evolutions: Ethernet and Provider Backbone Bridging–Traffic Engineering (PBB-TE), and a transport-oriented version of MPLS, MPLS Transport Profile (MPLS-TP).

The PBB-TE is an Ethernet evolution where connectionless nature is disabled, and Spanning Tree Protocols (STP), Media Access Control (MAC) address learning and flooding features are turned off in order to obtain a deterministic behaviour and a connection-oriented approach.

In order to be considered as transport technology, MPLS needs some changes: as a consequence, a new MPLS version, MPLS-TP, has been defined by Internet Engineering Task Force (IETF). MPLS-TP works in the same MPLS framework where just transport-oriented characteristics are implemented.

Both PBB-TE and MPLS-TP are expected to be Layer 2 transport technologies with low costs and with performances, in terms of Quality of Service (QoS) and Operation, Administration and Maintenance (OAM) functionalities, comparable to Synchronous Digital Hierarchy (SDH). However, it has to be pointed out that these technologies are not yet standardised, and carriers are debating about choice of these future transport techniques.

Further themes, with normative proposal, for upgrading Ethernet towards a transport technology are listed below:

- **Synchronous Ethernet:** The ITU-T's G.8262 recommendation tackles the synchronization of packet networks to ensure successful transmission of time-sensitive services.
- **Ethernet protection switching and ring topologies:** The G.8031 and G.8032 standards specify several fault-tolerant architectures for an Ethernet service. The ability to provide 50-msec failover between primary and secondary paths and support a variety of both ring and point-to-point topologies protects an Ethernet-based transport layer against fiber cuts or system failures.
- **SLA assurance:** Standards such as 802.1ag and Y.1731 define ways for network operators to perform in-service performance monitoring on their Ethernet services. In addition, stress-testing tools utilizing the RFC 2544 test suite show a carrier whether its service delivery chain (which encompasses every end-to-end element, including across third-party infrastructure) is primed to meet SLA parameters before the Ethernet service is handed off to the customer.
- **Circuit emulation:** The MEF's Circuit Emulation over Ethernet (CESoE) borrows on techniques developed to transport TDM services across ATM. CESoE provides the means to transparently carry T1/E1 and other established private-line offerings across Ethernet and is key in Ethernet's evolution from being a service to a transport infrastructure capable of carrying other services.

A1.1.15.2 The current Ethernet scenario: the Virtual Private LAN Service (VPLS)

In the context of the evolution towards a Carrier Grade Ethernet, Virtual Private LAN Service (VPLS) represents a transition to Layer 2 transport. Indeed, it is a combination of Ethernet and MPLS and it inherits their characteristics. VPLS has Ethernet simplicity and cheapness and it exploits MPLS features as, for an instance, mechanisms that allow implementing QoS control, as shown in [2-6].

Currently, in the WAN context, VPLS permits to achieve the best network performance in terms of traffic management and QoS. VPLS offers a sort of Layer 2 Virtual Private Network where the customers are connected by Ethernet Line/LANs. It has to be pointed out that VPLS does not perform proper multicasting since it uses a broadcast traffic forwarding. VPLS is a Layer 2 Virtual Private Network (VPN) where the customers are connected by Ethernet Line/LANs; with respect to the conventional Layer 2 VPNs, where the customers are connected in a point to point way, in VPLS customers can be connected by a multipoint Ethernet LAN. The main feature is that users sharing VPLS seem to be in the same LAN, independently by their own geographic position.

However, it has to be pointed out that VPLS is not a full layer 2 technology since it requires a layer 3 processing, even though it results much interesting to implement a geographical network fully based on Ethernet transmission. Wide potentialities of VPLS can be achieved together with VLAN tagging technique performed at the edge of the network.

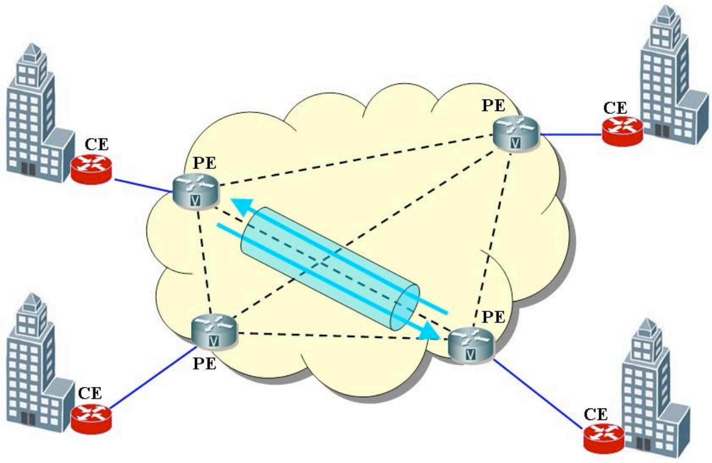


Figure 26. VPLS network

The network elements involved in VPLS are Customer Edge (CE) and Provider Edge (PE) nodes (1.1.15.2) the CE is completely VPLS unaware and it is connected, by means of an Ethernet connection (typically a VLAN), to the PE. The Provider Edge is the main VPLS element: it adds (inner) VPLS label and (outer) MPLS label to customer frames and it forwards them to the remote PE.

In the framework of the BONE project some experimental activities were carried out on a VPLS test bed showing the reliability of such an infrastructure for a transport network based on GbE. Details on some experiments can be found later in the deliverable. The test bed is reported in 1.1.15.2.

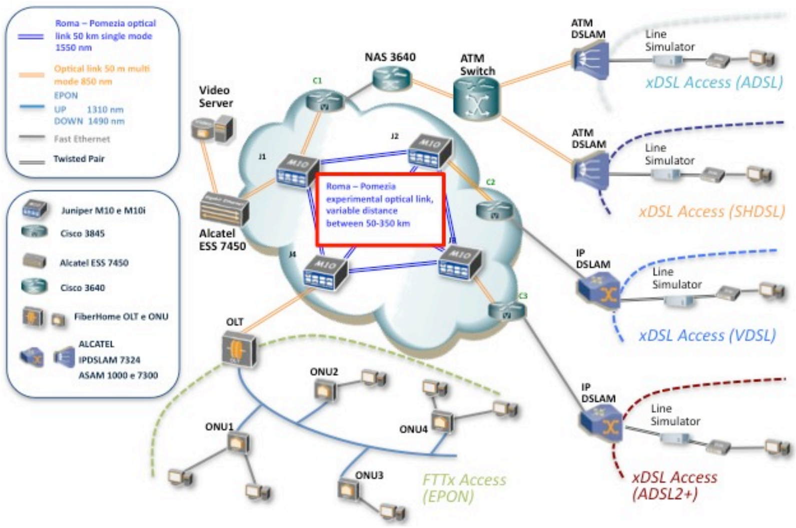


Figure 27. Experimetnal Test Bed

In PEs we set MPLS to carry VPLS traffic in a MPLS network, and Open Shortest Path First (OSPF) as Interior Gateway Protocol (IGP) for MPLS LSPs establishment. Furthermore, according to IETF RFC 4761, we set Border Gateway Protocol (BGP) as protocol for signaling (i.e. label exchange) and auto-discovery mechanisms (for PE identification) in VPLS.

The experiments performed on such a test bed confirmed that VPLS already has many characteristics to implement WAN only based on Ethernet transmission. In particular here we list the main achievements:

- WDM GbE transmissions can be used on long distance and we tested that by using optical amplifiers the distances among the routers can reach 350 km [6];
- VPLS can be used to manage the traffic in the core network in order to control the QoS for an instance in access segments with a limited bandwidth (as for an instance in xDSL device, but also in PON network that have to share the bandwidth among several users). In particular as reported in [4, 5],

VPLS can be used to manage reliable unbundling in PON, guaranteeing bandwidth and QoS to the users.

- VPLS is compatible with novel optical technologies as for an instance the all optical wavelength conversion [2, 3]. In particular we tested the network performance of a WAN GbE with the presence of AOWC based on the Four Wave mixing in DS fiber and we verified that AOWC does not introduce any signal degradation and it can be achieved in a very fast time, compatible with the out of service required by wide bandwidth real time services as HD TV streaming.

Being a combination of Ethernet and MPLS, it is possible to use MPLS restoration and protection techniques in VPLS; in particular, we implemented Fast Reroute and Standby Secondary Path in VPLS. We demonstrated that VPLS allows to have restoration times under 50 ms, with performances equal to SDH/SONET in terms of restoration and reliability capabilities; in particular, VPLS FRR has better performances than VPLS SSP.

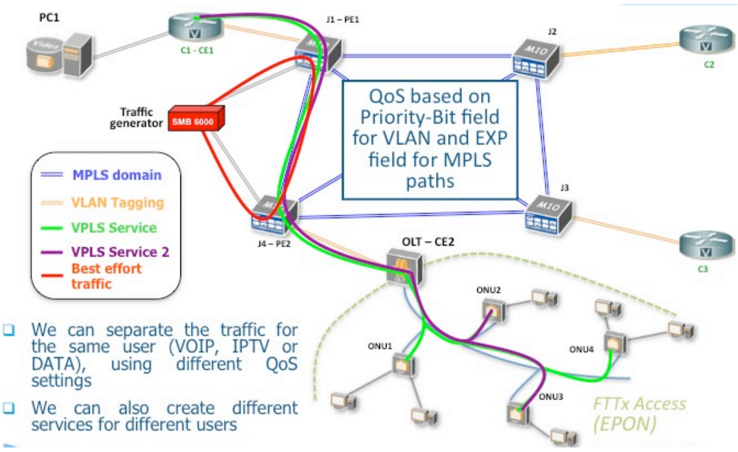


Figure 28. Experimental Setup

As shown in 1.1.15.2 Cisco routers (C1-CE1) and the ONUs behave as CE nodes, and they are connected (by VLAN) to Juniper routers (J1-PE1, J2-PE2), configured as PE nodes.

According to these considerations, the test bed is configured to operate with CoSs based on VPLS and extended to the CE sites by means of VLAN Tagging technique. Therefore, a tunnel based on VPLS&VLAN Tagging between a Server (PC1) (connected to CE1) and a user PC (e. g., the PC connected to ONU2) can be established according to specific service requirements. In this test bed, eight different CoSs can be defined, according to the values of three bits, in agreement with IEEE 802.1p (within IEEE 802.1Q) and DiffServ over MPLS standards. In particular, the VLAN TAG CoS field (3 bits) is employed to guarantee QoS among CEs and PEs; since PE adds an outer MPLS label, the QoS among PEs is guaranteed by using the MPLS LABEL EXP field (3 bits). It has to be pointed out that recently RFC 5462 changed EXP field in Traffic Class Field.

The eight CoSs were defined in such a way that CE and PE treat the corresponding packets with different priority. In particular, in the Gold Class (User Priority Bits: 101) the packets have the highest priority and they are always immediately forwarded, even in condition of congestion; therefore, such a CoS permits to guarantee the necessary performance (throughput, jitter, data loss) independently of the network congestion. Less priority was assigned to a Silver Class (User Priority Bits: 011) and so on, up to the Best Effort (BE) Class (000) that has no priority.

A1.1.15.3 Restoration in VPLS

The birth of bandwidth-greedy services needs some change in telecommunication networks where packet networks seem to be the future way. PBB-TE and MPLS-TP are interesting future solutions but they are not standardised, and some issues have to be addressed, as for example OAM features. In this scenario, VPLS represents a transition solution to deliver next generation packet services.

In order to fully satisfy service requirements, VPLS has to be able to guarantee QoS and to implement OAM functions. In other works, we proposed methods for QoS control in VPLS [6], and we made some investigations on restoration procedures based on Wavelength Conversion that we called Alternative Wavelength Path (AWP) [3].

Since VPLS realizes Ethernet connections by means of MPLS LSPs, we proposed two restoration procedures in VPLS exploiting MPLS capabilities. In the first one, we implemented Fast Reroute technique in VPLS (that we called VPLS FRR), in the second one we proposed a restoration procedure based on Standby Secondary Path technique in VPLS (called VPLS SSP). The implementation of the two techniques are shown in 1.1.15.3 a) (on the left side) and b) (on the right side).

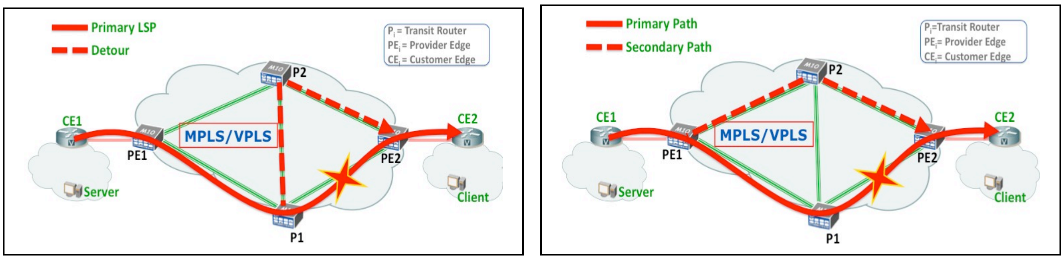


Figure 29. Restoration Techniques in VPLS

There are two Customer Edges (CE1 and CE2) interconnected by means of a VPLS instance that is composed of two Provider Edges (PE1 and PE2) and a transit router (P1). In the first case, we set also a protection mechanism based on Fast Reroute. This technique allows to automatically reroute traffic on a LSP if a link or a node fails. The ingress router (i.e. PE1) signals all the downstream routers that Fast Reroute is enabled, and then each router tries to set a detour path to immediate downstream router. In this way, it is possible to achieve fast rerouting by pre-establishing a set of detours along the LSP. In VPLS, PEs are connected by means of LSPs, and thus it is possible to use Fast Reroute for traffic protection. In particular, as shown in 1.1.15.3(a), we set Fast Reroute in VPLS: when link between P1 and PE2 fails, traffic is routed toward PE2 on a pre-established detour path crossing P2.

In the second case, we set a protection based on Standby Secondary Path. With this technique it is possible to set a primary path and a secondary path for an LSP providing an end to end protection. When the primary path fails, the ingress router reroutes traffic on the pre-established secondary path: in this way, it does not need to compute a new route and to signal new path for the LSP. As shown in 1.1.15.3(b), we set Standby Secondary Path protection in VPLS: when P1-PE2 link fails, PE1 reroutes traffic on the secondary path crossing P2.

Several tests were carried out both for VPLS FFR and VPLS SSP. In particular, we measured restoration time, throughput degradation and perceived quality. For such an aim, we used a traffic analyzer, Anritsu MD1230B, and a sniffing software, Wireshark.

For the restoration time measurement, we sent 1 Gbit/s test traffic from CE1 to CE2; the failure event occurs (by means of an optical switch controlled by a software running on a Linux PC) every 2 minutes for 1000 times, it lasts for 50 ms, and we can assume that each failure event is independent from each other. In 1.1.15.3(a, left side) and (b, right side) results are shown; in particular, we can observe an Average Restoration Time about 26.3 ms and 36.7 ms for VPLS FRR and VPLS SSP respectively.

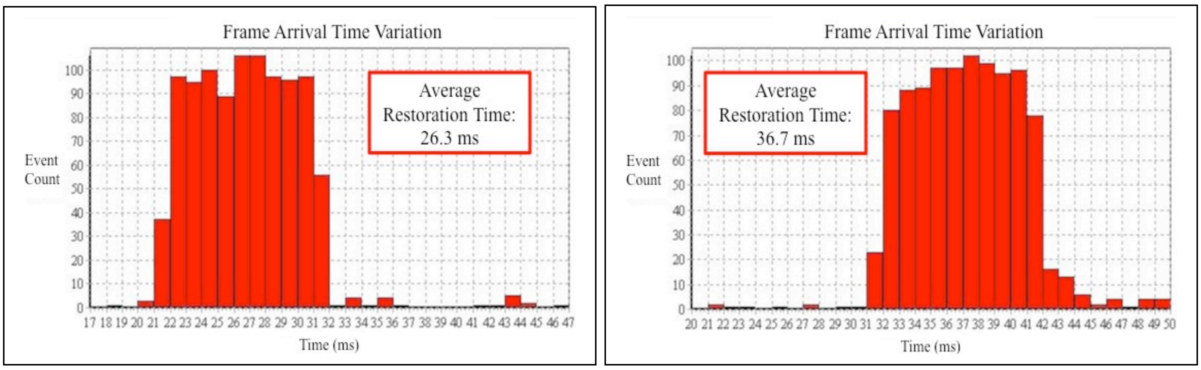


Figure 30. Restoration Time

These measurements represent a quantitative system reliability estimation, which finally points out that for 99.9% of link failure events restoration time is sub-50 ms, making VPLS restoration times and reliability equal to SDH/SONET performances. In addiction, these tests outline that VPLS FRR and VPLS SSP techniques are faster than 30 ms and 40 ms respectively, for nearly 90% of link failure events.

To perform network and subjective measurements, an MPEG2 High Definition (HD) video streaming (about 20 Mbit/s) was sent, over the VPLS instance, between a Server (connected to CE1) and a Client (connected to CE2), as shown in 1.1.15.3. In the 1.1.15.3(a, left) and (b, right), we report a sample for the throughput when failure occurs; we can observe a service interruption that confirms results reported above for VPLS FFR and VPLS SSP.

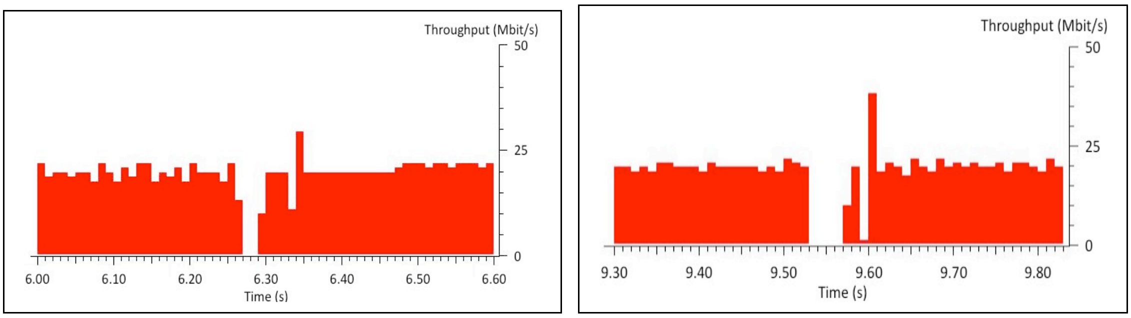


Figure 31. Throughput of MPEG2 HD Video Streaming

In 1.1.15.3(a, left) and (b, right), screenshots of Client's video quality are shown.

Figure 32. Perceived quality of MPEG2 HD Video Streaming

We can underline that VPLS SSP has a stronger degradation than VPLS FRR. Both Restoration Techniques ensure a continuous service also for delay sensitive applications as HD video streaming, but it is possible to point out that VPLS FRR faster restoration times could be appreciate in a subjective evaluation of a Streaming Service.

In this work, VPLS has been implemented in a real wide area optical test-bed and some restoration techniques have been proposed. Being a combination of Ethernet and MPLS, it is possible to use MPLS restoration and protection techniques in VPLS; in particular, we implemented Fast Reroute and Standby Secondary Path in VPLS. We demonstrated that VPLS allows to have restoration times under 50 ms, with performances equal to SDH/SONET in terms of restoration and reliability capabilities; in particular, VPLS FRR has better performances than VPLS SSP.

Therefore, VPLS can be assumed as a Layer 2 technology with low costs for allowing transition toward packet transport techniques as PBB-TE and MPLS-TP.

We can conclude that VPLS can be assumed as a Layer 2 technology with low costs for allowing transition toward packet transport techniques as PBB-TE and MPLS-TP.

We also have shown that satisfying Ethernet solutions have already available using for an instance VPLS architectures.

A1.1.15.4 TCP Performance in Hybrid Multigranular OBS Networks

During the first year of the project UNIMORE has studied the TCP performance when high speed Ethernet over Passive Optical Networks (EPON) are interconnected by means of a core optical network based on the OBS paradigm focusing, in particular, on the edge node and its functions.

In the second year UNIMORE has shifted the focus to the core router architecture, investigating the impact on TCP of different optical technologies.

OBS networks are designed according to optical core switches' configuration times, by properly choosing the offset times to perform switch resource reservations in advance, prior to OBS payload transmission.

OBS payloads are prepared at the ingress edge nodes by aggregating information coming from access networks, such as EPONs.
To further enhance OBS flexibility, hybrid multi-granular core nodes can be employed. The main idea behind multi-granular hybrid switching is to support multi-granular application requirements with different technologies which are characterized by different configuration times. The main target of this design concepts is to achieve short term feasibility while maintaining low switch costs. These switches are said hybrid because they employ different technologies like MEMS and SOAs to implement slow and fast switching paths, respectively.
As a consequence of hybrid switch solutions, paths with different set up delays are available in the OBS network.

As far as the OBS network design, hybrid solutions impact on the values of the base offset that takes into account the time needed to set up a path through the core switches. Longer offset times are needed in the slow path case than for fast paths.

Different offset times differently influence upper layer performance and these effects are worth to be evaluated.
The employment of different optical technologies in core nodes implies different set-up times for the corresponding switching matrices. This means that different offset times have also to be considered, depending on the kind of connection IP datagrams belong to. We assume as fast connections all connections whose IP datagrams are put in bursts which are switched in core nodes employing the fastest optical technologies, which means shortest set-up times, and as slow connections when the related bursts are switched employing slower optical technologies, which leads to longer set-up times.
Numerical investigations [7] have shown that for low burst losses, fast connections can provide remarkable higher TCP throughputs than slow connections, whereas for burst losses greater or equal to 0.001 the average throughput is roughly the same, despite of different transmission rates and technologies.

1.1.16 List of relevant publications

1. Coiro et al, “Network evolution towards a Carrier Grade Ethernet transport network”, in press on Fiber Integrated Optics.
2. Matera, F. Rea, L. Valenti, A. Pompei, S. Beleffi, G.M.T. Curti, F. Forin, D. Incerti, G. Di Bartolo, S. Settembre, M. “Network Performance Investigation in a Wide Area Gigabit Ethernet Test Bed Adopting All-Optical Wavelength Conversion”, IEEE Photonics Technology Letters, Vol. 20, Issue 24, pp. 2144-2146, December 2008
3. A. Valenti, S. Pompei, L. Rea, F. Matera, G. Tosi-Beleffi, F. Curti, S. Di Bartolo, G. Incerti, and D. Forin, “Experimental Investigation of Quality of Service in an IP All-Optical Network Adopting Wavelength Conversion”, IEEE/OSA J. of Optical Communications and Networking, Vol. 1, Issue 2, pp. A170-A179, July 2009
4. F. Matera, A. Valenti, S. Pompei, G. M. Tosi Beleffi, and D. Forin, "Unbundling and Quality of Service in Ethernet Passive Optical Networks based on Virtual Private LAN Service technique” in Proc. Of CONTEL 2009, Zagreb, June 8-10.
5. L. Rea, A. Valenti, S. Pompei, L. Pulcini, M. Celidonio, D. Del Buono, G.M. Tosi Beleffi “Quality of Service control in a multi-access integrated network based on Virtual Private LAN Service”, in Proc. Of IPHOBAC 09, Duisburg, May 18-20.
6. A.Valenti, P. Bolletta, S. Pompei, F. Matera, “experimental investigations on restoration techniques in a wide area Gigabit eterne optical test-bed based on Virtual Private LAN Service”, in proc. Of ICTON 09, Ponta Delgada, June 29 – July 2
7. M. Casoni C. Raffaelli, "TCP Performance in Hybrid Multigranular OBS Networks", Proc. of IEEE Workshop on Optical/Burst Switching 2009, Madrid (Sp), September 14, 2009.

Extension of the Flow-Aware Networking (FAN) architecture to the IP over WDM environment

1.1.17 General objectives and summary of results from Y1

The research in Quality of Services (QoS) technologies started many years ago, but now the networking engineers start to understand the keys to QoS provisioning. The traditional traffic engineering is based on demand predictions, while the modern Traffic Engineering request for measurement-based approaches or state based approaches (automatic approaches).

Flow-Aware Networking (FAN) appears as a new approach to offer QoS. Flow-Aware Networking (FAN) observes and makes decisions at flow level instead of packets level. Flow-based traffic engineering uses more tractable mathematical models than packet-based traffic engineering (self-similar models). Flow-based traffic engineering takes more into account the probabilistic relation between demand-performance-capacity (represents almost human behaviour and is near to telephone systems).

Operator networks are migrating to an IP over WDM scenario where IP layer is directly connected to the optical layer. Moreover, the optical layer is becoming more and more intelligent and it is able to establish and tear down lightpaths automatically. This new topology creates not only new problems but also solutions, such as the utilization of the optical layer to solve congestion problems at IP layer. We call this solution Multilayer Flow-Aware Networking (MFAN).

This joint activity is focus on the enhancement of the congestion avoidance in Flow-Aware Networking technology. The main results of this activity are:

- Define the node architecture for multi-layer networks as well as the policies to deal with the incoming flows to the system.
- Define new congestion control mechanisms.

The solution proposed is shown in the next Figure. A MFAN node is composed by a FAN queue at the IP level and a module that is able to ask for extra resources to the optical layer when congestion is detected at the IP level. When FAN queue can not deal with the incoming flows, extra resources are applied to the optical layer.

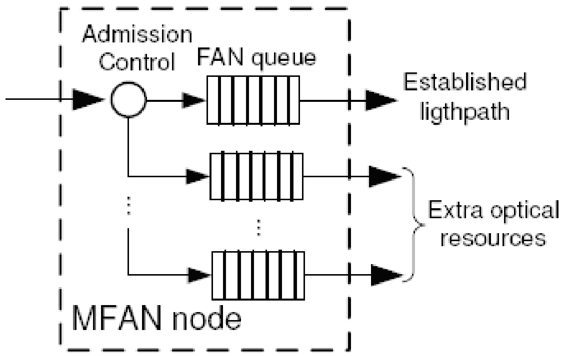


Figure 33. Multi-layer Flow-Aware Networking (MFAN) node architecture

In the first year of this joint activity, we evaluated the performance of Multilayer Flow-Aware Networking (MFAN) in a UDP dominated scenario, where multimedia applications such as VoIP were the main traffic source. In previous work we assumed that the congestion was due to TCP traffic, but UDP traffic can congest the system. It is necessary to know the performance of the policies when this situation occurs (see next figure).

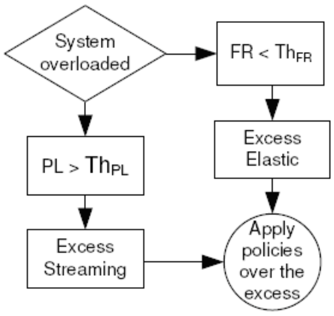
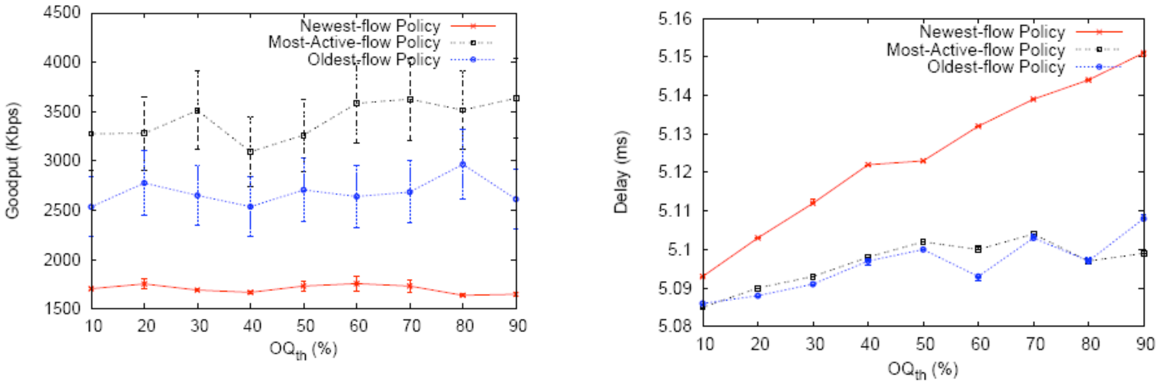


Figure 34. Optimal decision of the flows.

The performance of MFAN in a UDP dominated scenario is different than in the TCP scenario. Due to the nature of the UDP flows, Most-Active policy is able to extract the more suitable flows from the IP layer allowing a better performance in the optical queue. This can be seen in the following figures. Delay is the similar to Oldest policy but the goodput achieved is higher.



TCP flows goodput in the optical queue — Mean delay of the UDP packets in the optical queue

Figure 35. Performance in a UDP dominated scenario in a dumb-bell topology

To check this behavior, the amount of TCP traffic was varied from an 80% to a 20%. This experiment concluded that the performance of Oldest policy is better when the amount of TCP traffic is more than a 50%, while Most-Active policy is more suitable when UDP traffic is congesting the IP level.

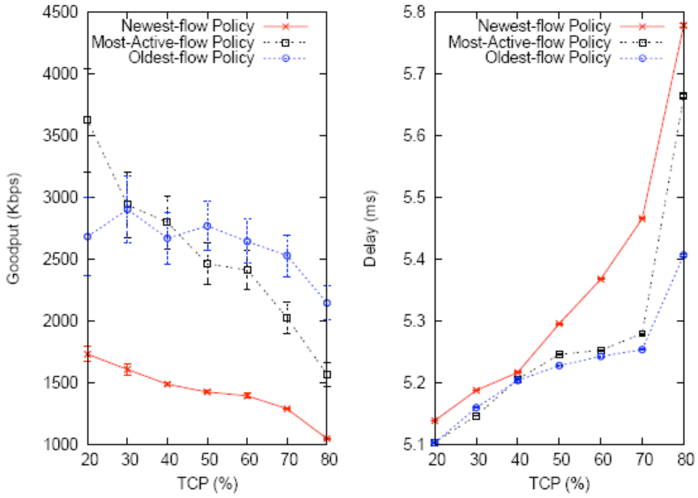


Figure 36. Policies performance evaluation when the percentage of TCP varies

This experiment shows that a solution based on a "Hybrid policy", which combines the benefits of Oldest and Most-Active policies, it is the solution for all kind of traffic profiles.

1.1.18 Advances in Y2

In the second year of this JA, we have focus on a *recovery mechanism for FAN based on IP over WDM architectures* and on admission control policies for Flow-Aware Networking solutions. Thanks to these research lines, we have achieved two conference articles.

The motivation to study *the recovery mechanism for FAN based on IP over WDM architectures*, it is because FAN does not ensure the reliable transmission. There are no protection and restoration mechanisms in FAN which means that in case of link or node failure the traffic has to be redirected. The disadvantages of such a situation are summarized in the following table:

TABLE II
THE ADVANTAGES AND DISADVANTAGES OF FAN ARCHITECTURES, BASIC, FAN/WDM AND FAN/WDM WITH EHOT

Link failure	Advantages	Disadvantages
Basic FAN	simple (only rerouting)	relatively slow rerouting: - in congestion rerouted flows are not accepted; - in congestionless the number of accepted flows after rerouting may be too large
MFAN without EHOT	chance for fast recovery in optical layer	rerouting in IP layer and recovery in optical layer at the same time
MFAN with EHOT	chance for recovery without rerouting in the IP layer (fast and short outage in transmission and without changes in flows assignment)	complexity of the algorithm

We evaluated the performance of FAN over WDm and we propose to implement the MFAN architecture to improve the chances for making the link restoration process quicker than 50 ms. We proposed a MFAN solution with the EHOT algorithm that improves the performance of single-layer FAN and FAN over WDM.

The second study focus on *admission control policies for Flow-Aware Networking* is a first study to evaluate the performance of Enhanced Flushing Mechanism (EFM). The EFM algorithm identifies all elastic flows to be removed from the PFL, when there is congestion. This congestion control mechanism has one important drawback. The number of accepted flows in the PFL right after a flushing procedure can be too high (see next table). Consequently, we integrate EFM mechanism with the flow classification of MFAN (Oldest and Most Active Flows). If this flow information is included into the mechanism the number of flows in reduced, but there are some flows that are not accepted again. The final solution we proposed is to use the EFMP (EFM with Priority) to ensure quick acceptance of removed flows from the PFL while ensuring all advantages of the mechanism described previously.

Table 1. The properties of admission control policies

Architecture or mechanism	*waiting time* [s]	No. of admitted flows	Advantages	Drawbacks
EFM (5s)	*1.43±0.91*	*160.43±10.91*	• quick acceptance of streaming flows • low cost	• instability of fair rate • high number of accepted flows
EFM (1s, oldest flow)	*3.55±0.53*	*23.55±0.62*	• quick acceptance of streaming flows • low cost • fair rate assurance • low number of accepted flows	• removed flow may not be accepted again
EFM (1s, most active flow)	*3.68±0.45*	*23.13±0.45*		
EFMP (1s, oldest flow)	*4.73±1.06*	*23.48±0.56*	• quick acceptance of streaming flows • low cost • fair rate assurance • low number of accepted flows • quick acceptance of removed flows	• none
EFMP (1s, most active flow)	***3.86±0.80***	***24.26±0.64***		

1.1.19 List of relevant publications

1. J. Domzal, R. Wójcik, K. Wajda, A. Jajszczyk, V. López, J.A. Hernández, J. Aracil, C. Cárdenas and M. Gagnaire: *A Multi-layer Recovery Strategy in FAN over WDM Architectures* in Design of Reliable Communication Networks (DRCN), October 2009.
2. J. Domzal, R. Wójcik, A. Jajszczyk, V. López, J.A. Hernández and J. Aracil: *Admission control policies in multi-layer flow-aware networks* in International Conference on Transparent Optical Networks (ICTON), June 2009.

Network planning algorithms considering fault tolerance, security threats and periodic traffic patterns

1.1.20 General objectives and summary of results from Y1

In WDM (wavelength division multiplexing) optical networks, all-optical channels, called lightpaths, are established between pairs of nodes forming a virtual topology over the physical interconnection of optical fibers. Lightpaths can span multiple physical links without undergoing opto-electronic conversion at intermediate switches. Network planning in such transparent WDM optical networks is based on solving the problems of Virtual Topology Design (VTD), Routing and Wavelength Assignment (RWA), and Traffic grooming. Virtual topology design determines the set of lightpaths to be established based on a given static, scheduled or dynamic traffic pattern. The process of finding a physical route and assigning a wavelength to each of the users' lightpath demands, is called routing and wavelength assignment. It is subject to two main constraints. The wavelength clash constraint prohibits assigning the same wavelength to lightpaths traversing a common directed physical link. The wavelength continuity constraint dictates assigning the same wavelength along a lightpath's entire physical path. Finally, the process of routing packet switched traffic over the established virtual topology is referred to as traffic grooming. In some literature, the combination of all these subproblems together is referred as Virtual topology design but here we consider this term to refer to just the process of determining the lightpath demand set.

Since optical networks employ extremely high data rates, any malfunction puts large amounts of data at risk of getting lost or corrupted. Deliberate physical-layer attacks aimed to deteriorate the QoS and/or deny service, analyze traffic or eavesdrop, can cause severe damage to proper network functioning. Furthermore, due to network transparency, malicious signals can propagate through the network. All of this makes fault and attack management a key element of network security. A significant number of network safety approaches, focus on monitoring and localizing faults and attacks, i.e. deal with their consequences through different network restoration techniques after they have already occurred. However, attack prevention approaches using wavelength assignment, aimed at reducing the potential damage and propagation of jamming attacks, to the best of our knowledge, have not yet been considered in the literature beyond our ongoing work.

In the framework of this joint activity, we proposed a novel approach to optical networks security. While most existing prevention measures are hardware-oriented, our goal is to achieve significant prevention measures without the need for specialized equipment, i.e. at minimal extra cost, through careful network planning. The idea is to consider the potential consequences of various physical-layer attacks, faults and security threats (e.g., power jamming, fiber cuts, eavesdropping) and arrange the set of lightpaths in such a way as to minimize the possible service disruption which can be caused in case of such scenarios. By limiting the number of lightpaths that can be affected by an attack, not only is network service disruption minimized, but failure detection and localization algorithms can be more efficient.

The main goal of this joint activity is to design a cost-effective security planning framework for optical networks. In the first phase of the activity in Y1, we investigated various physical-layer attack and fault scenarios and designed new planning algorithms, specifically algorithms to route and assign wavelengths to lightpaths, in a way which can create a more robust network with respect to out-of-band crosstalk attacks. Namely, long distances and high-power signals can introduce nonlinearities in fibers causing crosstalk effects between channels on different wavelengths along fibers, called out-of-band crosstalk. We proposed a new objective criterion for the Routing and Wavelength Assignment problem, developed a tabu search algorithm in [1] and formulated the problem as a MILP (mixed integer linear program) in [2]. The new objective, called the Lightpath Attack Radius (LAR) is aimed to arrange a given set of lightpaths in such a way as to minimize potential disruption, i.e. minimize the maximum number of lightpaths which can be disrupted by out-of-band crosstalk attack scenarios. Consequently, if fewer lightpaths are attacked, not only is network service disruption minimized, but detection and localization algorithms can be faster since they search for the source among fewer potential lightpaths.

We also took a second approach to dealing with optical networks security in [3] by investigating the topological structures of the established lightpaths, not just considering their individual attacking relations, but their global interconnectivity. Namely, fast and successful reaction and restoration mechanisms can prevent from loosing large amounts of data which can cause severe service disruption. These mechanisms rely on monitoring information exchanged via optical supervisory channels along each physical link. We proposed applying common structural properties of self-organizing systems (specifically, the small-world and scale-free models) to help achieve faster and more robust system-wide communication in the context of failure management.

The third avenue of research in optical networks planning conducted within the framework of this JA, focuses on multi-hour, or scheduled, virtual topology design. This implies searching for the temporal evolution of virtual topologies, and its corresponding traffic grooming, which efficiently adapt to known traffic variations. We assume known multihour, i.e. scheduled or periodic, traffic patterns. Since this problem is NP-complete, heuristic algorithms are needed to help solve it sub-optimally. Besides developing efficient algorithms for multi-hour (scheduled) VTD, the goal of this investigation is to assess the potential advantage of reconfigurable equipment, i.e. its benefit over the static case.

In Y1, we formulated the Scheduled Virtual Topology Design problem as a Mixed Integer Linear Program (MILP) with the objective to minimize the number of transmitter and receivers necessary to establish a scheduled set of lightpaths which can satisfy given estimated traffic demands [4]. The MILP formulation was implemented and tested in the MatPlanWDM, a MATLAB-based software developed at UPCT and publicly available at the MATLAB central web site. It is composed of an application kernel, a set of libraries of related algorithms and a graphical user interface whose general goal is to aid the implementation and evaluation of optimization algorithms for lightpath-based optical networks. Due to the high complexity of the problem, heuristic approaches necessary to find good solution for large problems were developed in Y2.

1.1.21 Advances in Y2

A1.1.21.1 Attack aware RWA

As a continuation of the research efforts started in Y1 on the topic of attack-aware routing and wavelength assignment, we extended our work from [1] and [2] in [5] by further developing and testing the proposed tabu search algorithm for routing lightpaths to minimize the attack radius in the network with respect to out-of-band (inter-channel) crosstalk. Furthermore, we also investigated the consequences of in-band (i.e. intra-channel) crosstalk attacks and developed wavelength assignment schemes to minimize its propagation in [6] and [7]. Namely, in optical switches, channels on the same wavelength can interfere with each other causing in-band or intra-channel crosstalk. Intra-channel crosstalk is much more influential than inter-channel, since optical filters cannot remove the acquired undesirable leakage. A deliberate intra-channel crosstalk attack is achieved by injecting a high-powered signal (e.g. 20 dB higher than the other channels) on a legitimate lightpath causing such significant leakage that attacked signals can acquire attacking capabilities themselves.

Consequently, such an attack can propagate though the network, affecting links and nodes not even traversed by the original attacking signal. As shown in 1.1.21.1(a), this allows an attack to propagate and affect lightpaths that do not share any common physical components with the original attacker. Large amounts of data can get lost or corrupted in case of a single failure, making fault and attack management a crucial factor of proper network functioning. To the best of our knowledge, none of the RWA approaches developed so far aim at improving network safety by minimizing intra-channel crosstalk propagation.

In [6] and [7], we introduced a new objective for WA, called the Propagating Crosstalk Attack Radius, or P-CAR. We define the P-CAR (LPi) of lightpath LPi as the total number of lightpaths that a jamming signal propagating on LPi can affect via intra-channel crosstalk occurring in switches between lightpaths on the same wavelength. The maximum P-CAR (LPi) over all lightpaths in a RWA scheme determines the P-CAR value of that scheme. The objective of our WA approach is to minimize the P-CAR, thus minimizing the maximum damage that potential intra-channel crosstalk attacks can cause. As our secondary objective, we try to minimize the number of wavelengths used, preserving space for future network expansion. We consider the worst case scenario in which we assume indefinite attack propagation. In reality, the attacking potential of a signal decreases along its path due to several factors, such as the traveled distance and the number of switches traversed. However, since we are offering a strictly preventive WA technique, our model aims to minimize the worst-case crosstalk attack propagation with minimum loss of power. In our future work, we plan to extend our model by including scenarios with stricter assumptions on crosstalk level deterioration.

In 1.1.21.1(b), a simple RWA scheme is given, with 4 lightpath requests, LP 1 - LP 4, routed on their shortest paths. They do not share any common directed physical links and can thus be assigned the same wavelength. (c) shows their common switches. An attack originating from, e.g., LP 1 can affect LP 2 via their common switch A, which then, in turn, after the point where it has been attacked itself, gains attacking potential and attacks LP 3 in switch B and LP 4 in switch D. Therefore, P-CAR (LP 1) = 3 + 1 = 4. Similarly, P-CAR (LP 2) = 4, P-CAR (LP 3) = 3 and P-CAR (LP 4) = 3. We model attacking relations among lightpaths with a simple structure we call the attack graph, shown in 1.1.21.1(d). In it, nodes represent lightpaths routed on the same wavelength, connected to others if an attack can spread from one to the other, directly or indirectly, through intermediate lightpaths. The P-CAR (LPi) equals the out-degree of its corresponding node in the attack graph, incremented by 1.

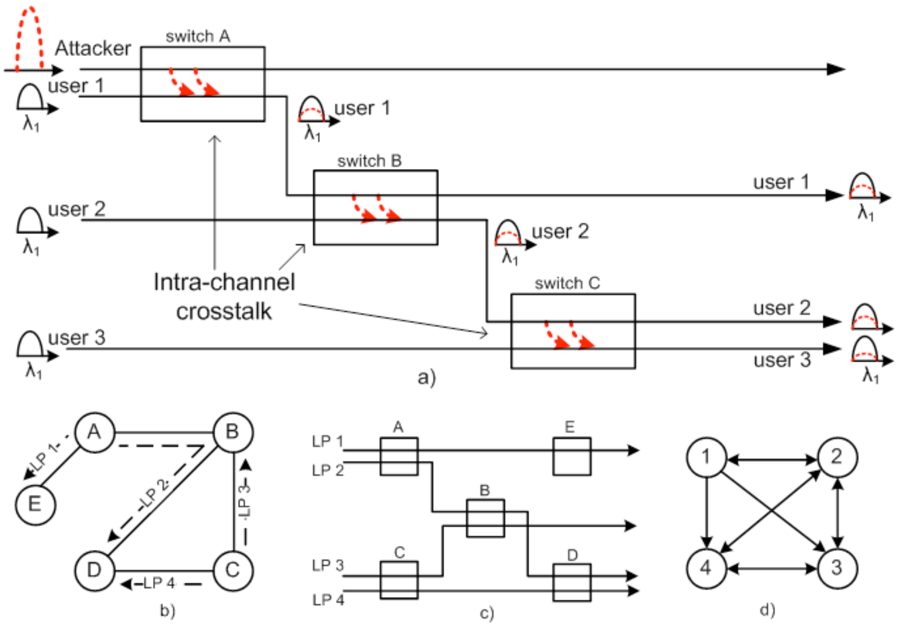


Figure 37. (a) Intra-channel crosstalk attack propagation. The attack can propagate from the Attacker, via user 1, to users 2 and 3, even though they do not share any common physical components. (b) A simple RWA scheme with four lightpaths routed on the same wavelength. (c) Common switches, i.e. potential points of intra-channel crosstalk attacks, and the corresponding attack graph for the given lightpaths (d).

To solve RWA with respect to our new objective, minimizing the P-CAR, we designed heuristics based on bin-packing. We assume fixed shortest path routing and, thus, only concentrate on wavelength assignment. The algorithms run as follows. Given is a set of lightpaths, a physical topology and a fixed number of available wavelengths. We use a layered graph approach, where each layer of the physical topology represents a different wavelength. To solve wavelength assignment, we must route each lightpath on one of the available layers, such that lightpaths routed on the same layer are link disjoint. We assume there are enough wavelengths to cover the entire set of given lightpath demands and our objective is to minimize the maximum P-CAR over all layers. The BF_PCAR_WA (Best Fit P-CAR Wavelength Assignment) algorithm takes lightpaths in random order and routes them on the layer which yields the lowest P-CAR after accommodating the lightpath, i.e. on the layer which yields the 'best fit'. The P-CAR on each layer is calculated by finding the attack graph corresponding to that wavelength using a recursive algorithm and then finding its maximum degree incremented by one. The BFD_PCAR_WA (Best Fit Decreasing P-CAR Wavelength Assignment) algorithm first sorts the lightpaths in decreasing order of their shortest paths and then proceeds as BF_PCAR_WA. For comparison purposes, we also consider two straightforward approaches we call the FF_WA (First Fit Wavelength Assignment) and FFD_WA (First Fit Decreasing Wavelength Assignment) algorithms which place lightpaths on the first layers onto which they fit in random order or in decreasing order of their shortest paths, respectively.

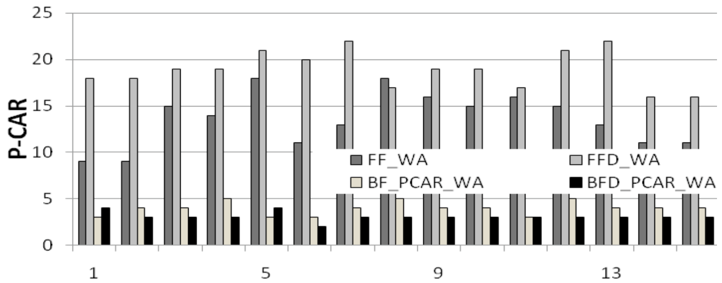


Figure 38. The P-CAR values obtained in each of the 15 test cases.

We implemented and ran the described algorithms on a 30-node hypothetical European basic network for sparse and dense traffic patterns with 5 and 15 transceivers per node, respectively. In both cases, we generated 15 different traffic matrices, where a fraction F of the traffic is uniformly distributed over [0, C/a] while the remaining traffic is uniformly distributed over [0, C ·ε /a]. The values were set to C=1250, a=20, ε=10, and F=0.7 as in previous approaches in literature. Figure 2 shows comparison of the P-CAR values obtained by all four algorithms for the dense traffic set of the 30-node network. Due to lack of space, the results for the sparse data set, analogous to those for the dense, are omitted. It is clear to see that our newly developed heuristics obtain much smaller P-CAR values than the classical approaches of FF_WA and FFD_WA, significantly decreasing the number of lightpaths a jamming signal propagating on any lightpath can affect.

In summary, we proposed a novel approach to dealing with the problem of propagating intra-channel crosstalk attacks in transparent optical networks. Instead of developing monitoring and reaction techniques which require extra equipment and resources to deal with attacks after they occur, we propose to limit the maximal damage caused by such attacks in the planning process through careful wavelength assignment. We introduce a new objective, called the Propagating Crosstalk Attack Radius (P-CAR), and propose heuristic algorithms to obtain suboptimal solutions. Future work in Y3, we will include combining our wavelength assignment algorithms with attack-aware routing to minimize the potential damage caused by a set of attacks, including inter-channel crosstalk, power jamming exploiting gain competition in optical amplifiers, and tapping attacks

A1.1.21.2 Applying self-organizing principles to failure management in transparent optical networks

Besides attack-aware optical networks planning, we also investigated self-organized structures and their application to the topology of the failure management or supervisory plane, extending our work from [3]. As already mentioned, a failure management system is used to deal with failures in transparent optical networks (TONs), which could be due to both component faults or deliberate attacks which aim to disrupt the proper functioning of the network. Due to the transparency inherent in TONs, nodes do not have access to service-bearing wavelengths except where data lightpaths terminate. Thus, management and control information is carried over a separate supervisory wavelength which is opto-electronically processed at each node. We refer to this interconnection of supervisory channels as the supervisory plane. In case of failure, failure management receives alarms from the monitoring equipment available (via the supervisory plane) and then attempts to locate and isolate the source. Meanwhile, the source and destination nodes of failed lightpaths are notified of the failure, after which they launch their restoration mechanisms.

The small-world and scale-free properties are structural properties that have been observed in many self-organizing complex systems. Self-organizing systems are those in which local low-level interactions and processes between individual entities spontaneously achieve global properties with certain functionality. Since structure affects function, these systems often self-organize into structures which enable efficient and successful operation. We propose applying these concepts to TONs in order improve the efficiency of failure management in transparent optical networks. This is a first step in applying self-organization to optical networks. Our ultimate goal, and vision for the future, is to develop a self-healing and self-management approach which will be able to supervise the functioning of TONs in the presence of increasing complexity and unforeseen attacks.

In [8] we proposed creating a hybrid supervisory plane whose structure is such that it can speed up and improve critical security information exchange and, thus, improve the network's ability to reconfigure and reestablish communication in the presence of failures. The optical supervisory plane is maintained on a separate wavelength than data connections on links between OXCs, consequently yielding the same topology as the physical interconnection of optical fibers (or a subset of them) in the network. Since this topology is dependant on geographical distances, it is not a random network but more lattice-like in nature. Such networks exhibit fairly high clustering, but have rather long average path lengths, while the degree distribution is centered around an average degree decaying very rapidly. Recall that supervisory messages are opto-electronically processed at each node, making communication between distant parts of the network fairly slow.

In order to achieve faster communication for more efficient failure management, we aim to reduce the average path length of the supervisory plane, while maintaining high clustering to aid failure detection mechanisms and improve resilience to false alarms. We propose creating a hybrid supervisory plane by maintaining the bidirectional point-to-point supervisory channels along each physical link, but also introducing a few long-range 'supervisory lightpaths' between distant nodes. Thus we could create a small-world supervisory plane, clustered as a result of the physical topology, but with a low average path length due to the small number of transparent shortcuts. Communication via these lightpaths would be somewhat slower than between physically neighboring nodes due to longer propagation delays, but would still be very fast as a result of their transparency. In addition to the small-world property, these shortcuts could be arranged to yield a scale-free topology which could possibly help create a more robust supervisory plane.

We developed algorithms in [3] and [8] to generate such topologies using combined approaches of the 'rewiring' and BA growth algorithms. Namely, we establish supervisory lightpaths between nodes with a small rewiring probability, but via preferential attachment. In the context of preferential attachment, we define the attractiveness of a node to be dependent on more than just its connectivity, but a combination of factors. The attractiveness A of a node i is as follows:

$$A(i) = Mon_i DP_i^{tr} + DP_i^{in} + DP_i^{out} + SP_i^{in} + \alpha SP_i^{out} \qquad (1)$$

Here, Moni represents the optical monitoring capabilities of node i (we assume 0/1 values where nodes either have full or no monitoring capabilities). This value is multiplied by the number of data lightpaths traversing the node, DPitr, which it can monitor. The next element in the attractiveness function, DPiin, represents the in-degree of the data plane, i.e. the number of data connections which terminate at the node and are opto-electronically processed. These connections can be subject to extensive monitoring in the electrical domain. The third element, DPiout, represents the number of data lightpaths originating at node i, where the node can obtain detailed information regarding the traffic being sent. SPiin represents the in-degree in the supervisory plane, giving a measure of how well-informed the node is, while SPiout represents its out-degree giving a measure of the nodes reputation among other nodes. This last element is multiplied by a factor α, letting us tune the influence of this element to a desired value.

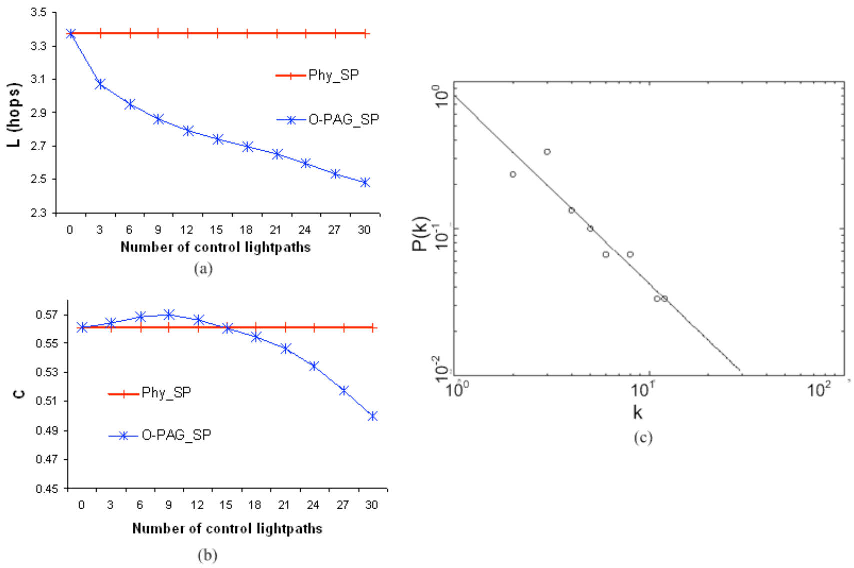


Figure 39. (a) The average path length, L, and (b) the clustering coefficient, C, of topologies obtained by the O-PAG algorithm on the basic European topology from [6] and (c) the degree distribution of a typical O-PAG supervisory plane topology with 30 supervisory lightpaths

The algorithm, called the Ordered Preferential Attachment via Growth (O − PAG) algorithm, begins by interconnecting m0 of the most attractive nodes with lightpaths, and then iteratively selecting the most attractive node not yet connected and assigning to it m< m0 informants from among the already interconnected nodes (provided they are not physical neighbors) with a probability proportional to their attractiveness. The algorithm is terminated after a desired number of informants are chosen. This interconnection of supervisory lightpaths is then super-positioned onto the physical topology representing the point-to-point supervisory channels.

To perform computational tests, traffic matrices and lightpaths were generated in the same way as described in the previous subsection for attack-aware RWA. The monitoring capabilities of nodes were assigned following a monitoring placement policy from literature where all neighbors of non-monitoring nodes and all nodes of degree one have monitoring capabilities. We tested this algorithm on a 29-node European basic network with 96 directed (48 bidirectional) edges from the COST Action 266 project and compared with the standard supervisory plane topology with respect to average path length and clustering. We can see from 1.1.21.2(a) that even a very small number of supervisory lightpaths generated by O-PAG can significantly reduce the average path length with respect to the point-to-point supervisory plane (denoted as Phy_SP), while still maintaining high clustering (1.1.21.2(b)).

Furthermore, as we increase the number of supervisory lightpaths, the algorithm begins to yield supervisory plane hub nodes. The degree distribution of a typical O-PAG supervisory plane topology with 30 supervisory lightpaths is shown in Fig. 3.c. While the advantages of a small world are straightforward, we are currently investigating whether a scale-free degree distribution of the supervisory plane is desirable in the context of optical networks failure management. A power law degree distribution could create a more robust and scalable supervisory plane, however, it would be vulnerable to attacks and generating such topologies would require more supervisory lightpaths. Through this analysis, we see that the topological structure of the supervisory plane model reflects structures observed in self-organizing systems, specifically small-world and scale-free structures. Further work is based on designing such supervisory topologies in a self-organizing manner and investigating the trade-offs between robustness and vulnerability in a scale-free supervisory plane.

A1.1.21.3 Multihour planning

Our third avenue of research in network planning was based on multi-hour, or scheduled, virtual topology design and traffic grooming. We consider the traffic demand to be modelled as a periodic temporal series of traffic matrices known in advance. The objective of our planning problem is to find the most cost-effective set(s) of lightpaths which are capable of fully carrying the given traffic as it changes over time intervals using (non)-reconfigurable equipment. A set of lightpaths established between node pairs in the network at any point in time is called a virtual topology. The virtual topology is the central element of a network multilayer problem. In the upper layer, electronic flows, i.e. traffic demands in Gbps, are routed on top of the virtual topology. Note that lightpaths are of finite capacity, and thus, determining a virtual topology capable of carrying a given traffic demand must be solved in conjunction with flow routing. Thus, the objective of our planning problem is to find the most cost-effective (i) scheduled virtual topology design, and (ii) routing of the electronic flows on top of the virtual topology.

In [9] and [10], we propose a set of algorithms for solving the aforementioned problem, which we denote as SVTD (Scheduled Virtual Topology Design). More formally, a SVTD instance receives a series of traffic matrices, and returns one or more virtual topology designs, together with associated flow routings. We consider two variants of the problem. In the first variant, the switching equipment is assumed to be non-reconfigurable or hard-wired. Therefore, the virtual topology is constrained to be constant along time. We denote this variant as SVTD-NR (SVTD–Non-Reconfigurable). In the second variation of the problem, denoted as SVDT-R (SVTD-Reconfigurable), reconfigurable switching nodes are assumed. This means that an electro-optic transmitter in a node N, can at a moment t1 be used in a lightpath between nodes (N, N1), and at a moment t2 be used in a lightpath between nodes (N, N2). The same concept applies for opto-electronic receivers. Intuitively, fewer transceivers should be necessary to carry a given periodic traffic demand with reconfigurable equipment if peaks in traffic from node N to node N1, coincide with drops in traffic from node N to other nodes. In such a case, otherwise idle transmitters can be reutilized. However, although reconfigurable equipment can yield solutions using fewer transceivers, it must be noted that such equipment is more expensive than its non-reconfigurable equivalent, creating a certain trade-off.

In summary, while in the non-reconfigurable case only the flow routing can change along time, in the reconfigurable-case both the virtual topology and the flow routing can change along time. In [9], we proposed two approaches to address the SVTD-R and SVTD-NR problems. The first approach is based on exact MILP (Mixed Integer Linear Programming) formulations (further extended in [10]), with the objective to minimize the number of transceivers required in the network, yielding optimal solutions to the problem. Clearly, both the SVTD-R and SVTD-NR problems are NP-hard, as the reduced problem with constant traffic is known to be NP-hard (integer capacity planning). Our tests indicate that solving the problem optimally using the MILP formulations can be done in reasonable time only for small network sizes and for a moderate number of time intervals. Consequently, we propose a heuristic approach for the SVTD-R/NR problems, using tabu search in conjunction with smaller MILP formulations for individual time slots within the scheduled planning problem. Tabu search is an iterative metaheuristic which guides the search procedure through the solution space using a memory structure called a tabu list. Although the MILP formulations used within the heuristics also consider NP-hard sub-problems, testing has shown that the proposed approaches are much more scalable than the exact formulation of the full problem, making them suitable for moderately-sized topologies. The effectiveness of the proposed heuristic algorithms is assessed through comparison with optimal solutions obtained by solving the MILP formulations for small topologies. For larger problem instances they are compared with proposed lower bounds on the optimal solution.

We studied the performance of the proposed algorithms for two network sizes: 5 and 18 nodes. For the smaller network, the exact MILP formulations (SVTD-NR/R) were run and compared with the results obtained by the TS-SVTD algorithm to asses efficiency of the heuristic approach. For the medium-sized 18-node scenario, only variations of TS-SVTD were tested, given the intractability of the pure MILP approach. The algorithms were implemented using the MatPlanWDM tool, which interfaces to a TOMLAB/CPLEX library used to solve MILP problems. For both scenarios, two cases were considered: (i) a network with non-reconfigurable equipment, (ii) a network with reconfigurable equipment. A series of traffic matrices were synthesized to feed the planning algorithms. Each traffic matrix series was composed of 24 traffic matrices, one for each hour of the day (i.e., t=1,...,24), obtained from a single base traffic matrix and an activity function. The activity function describes fluctuations in traffic over the course of a typical day. Further details on the traffic generation method can be found in [9].

Table I shows the results obtained for the 5-node network. We evaluated the solutions obtained by solving all proposed variants of the MILP formulation and tabu search algorithm in terms of the total number of transceivers needed. The lower bounds on the number of transceivers are calculated as the maximum total traffic originating or terminating at any node over time, divided by the lightpath capacity and rounded to the nearest integer. Note that the same lower bounds apply to the reconfigurable and non-reconfigurable cases. Results show that the resources planned by the optimal MILP formulation are equal to the computed lower bound in almost all cases, validating their accuracy in this scenario. Slight variations are found in the non-reconfigurable case. As expected, the number of transceivers planned is higher for series of traffic matrices with a higher traffic variability factor R. This increase is more significant at higher loads. The tabu search scheme allocates only a slightly higher amount of resources when compared to the optimal MILP solution indicating its efficiency. The extra-resources planned are in the order of 10% in the non-reconfigurable case, and 5% in the reconfigurable case.

TABLE I

Total number of transceivers needed for the 5-node network

nf	R	Lower Bound	Reconfigurable		Non-Reconfigurable	
			SVTD MILP-R	TS -SVTD	SVTD MILP-NR	TS -SVTD
500 Gbps	10%	109	110	120	110	122
	20%	116	116	125	118	132
	50%	123	123	131	126	154
1000 Gbps	10%	214	214	222	216	228
	20%	223	223	228	226	248
	50%	259	259	266	262	304
2000 Gbps	10%	419	419	423	422	446
	20%	444	444	453	446	482
	50%	518	518	528	524	572

TABLE II

Total number of transceivers needed for the 18-node network

nf	R	Lower Bound	Reconfig urable TS-SVTD	Non-Reconfigurable TS-SVTD
1500 Gbps	10 %	322	720	724
	20%	331	720	728
	50%	359	733	758
3000 Gbps	10 %	631	912	946
	20%	650	918	962
	50%	689	947	1042
6000 Gbps	10 %	1248	1522	1786
	20%	1283	1517	1824
	50%	1371	1566	1984

The solutions obtained for the 18-node network are shown in Table II. At lower loads there is a large gap in the number of transceivers between the lower bounds and the solutions obtained by TS-VTD. Namely, the solutions found double the resources predicted by the lower bounds. At higher loads this gap is much smaller, i.e., in the order of the 25%. The planned number of transceivers does not seem to be significantly affected by the variability factor R, although in the non-reconfigurable case we can see slight variations. If we compare the number of transceivers needed for the non-reconfigurable and reconfigurable cases, we observe that the difference is not significant at lower loads, increasing to some extent at higher loads. However, the maximal reduction in the number of transceivers obtained (i.e., the case for the 18- node network with maximal traffic load and variability) was only 21.07%. The average reduction over all cases was 5.51 %. Considering the significant higher price of reconfigurable equipment, our results indicate that reconfiguration may not be cost-effective, even for cases when traffic is highly variable. Our results indicate that the reduction with respect to the transceivers obtained with reconfigurable equipment was not significant, except in the case of larger network loads associated with high traffic variability. However, even in these cases, the reduction is not dramatic and considering the higher cost of such equipment, does not seem to be cost-effective. For future and ongoing work we will develop a pure heuristic approach for large problems and perform a detailed cost-benefit analysis of using reconfigurable components.

In addition to comparing of the cases with reconfigurable and non-reconfigurable equipment, in [11] we performed a more detailed analysis of just the non-reconfigurable case, but for various types of flow routing schemes. As before, we consider offline virtual topology planning in transparent optical networks given a multi-hour traffic demand, i.e. for known or estimated traffic variations. However, here we assume that the virtual topology can not be changed along time, i.e. we determine a static set of lightpaths capable of routing all traffic as it varies over all time slots. Thus, the network can be based on non-reconfigurable optical switching equipment. This assumption is in line with the current capabilities of the control plane in transparent optical networks and avoids dynamic 'hot' reconfigurations of the virtual topology which can imply short but relevant traffic disruptions. Our target is finding an optimal fixed virtual topology, and the associated traffic flow routing on top of it, which is capable of fully carrying the given traffic in all time slots. The cost to minimize is measured in terms of the number of lightpaths in the network, which define the number of transceivers (transmitters and receivers) needed as in the above problem.

Our main objective was to investigate the trade-offs that arise in the associated planning problem with respect to two different criteria:

- Considering fixed or variable routing of higher layer flows on top of the static virtual topology over time. A fixed flow routing indicates that the traffic from a source node to a destination node is always transmitted via the same set of lightpaths. On the contrary, variable routing imposes no such constraints, but at a cost of higher signaling and network management complexity.
- Considering splittable or unsplittable flow routing. Unsplittable flow routing means that all the traffic between a given input-output pair at any given time is constrained to traverse the same sequence of

nodes. If the traffic between a node pair exceeds the capacity of a single lightpath, the unsplittable traffic must be routed over a set of multiple lightpaths which are all-optically routed across the same sequence of nodes. However, if splittable flow routing is assumed, the traffic between two nodes can be split, where different fractions of the traffic can be routed over a different sequence of lightpaths not necessarily traversing the same nodes. Naturally, splittable routing allows for improved traffic balancing, but at a cost of increased signaling overhead.

We proposed two approaches to address this multihour planning problem and its variants. The first approach was based on exact ILP (Integer Linear Programming) and MILP (Mixed ILP) formulations, yielding optimal solutions to the problem. Clearly, all variants of the problem are NP-hard, as the static virtual topology design problem is. For larger problem instances we proposed a family of heuristic methods based on the concept of traffic domination, in conjunction with a heuristic for the static virtual topology design problem. Also, we proposed a lower bound on the number of transceivers in the network for the case in which the virtual topology can be changed along time (and, thus, reconfigurable switching nodes must be used). This lower bound is valid to evaluate the maximum cost saving in number of transceivers that could be achieved if the virtual topology could be reconfigured along time. We present extensive results in [11], omitted here for the sake of brevity, using synthesized traffic variations with various traffic loads and randomness factors, for networks of up to 8 nodes. In addition, we conduct tests for larger problem instances considering a time series of 672 traffic matrices obtained from a real traffic trace in the Abilene network.

Results support interesting conclusions regarding the cost reduction (in terms of the number of transceivers used) that network operators can achieve in different circumstances. For instance, in the scenarios studied the migration from fixed to variable routing is justified only if splittable routing is also allowed in the network, and only for medium or high loads. However, the maximum reduction obtained is still moderate (~15%) and is a trade-off with increased overhead. Analogously, for network operators evaluating the advantages of splittable routing in constrast to unsplittable routing, our results indicate that the migration is cost-effective for medium load scenarios, with a larger cost reduction if variable routing is allowed. Still, the reductions are moderate (with a maximum of ~20%). In Y3, we will further investigate the trade-offs associated with replacing non-reconfigurable equipment with their reconfigurable counterparts, as well as continue our studies concerning the application of traffic domination to optical networks planning.

1.1.22 List of relevant publications

[1] N. Skorin-Kapov, J. Chen, L.Wosinska, “A tabu search algorithm for attack-aware lightpath routing”, *The Proc. of the10th International Conference on Transparent Optical Networks (ICTON 2008),* Athens, Greece, June 22-26, 2008, pp.42-45.

[2] N. Skorin-Kapov, “MILP formulation for routing lightpaths for Attack-Protection in TONs”, *Proceedings of NAEC'08*, Riva del Garda, Italy, Sept. 2008, pp. 55-62.

[3] N. Skorin-Kapov, O.Tonguz, N. Puech, ”A novel optical supervisory plane model: The application of self-organizing structures”, *The Proc. of the10th International Conference on Transparent Optical Networks (ICTON 2008),* invited paper, Athens, Greece, June 22-26, 2008, pp.10-13.

[4] N. Skorin-Kapov, P. Pavon-Mariño, "Scheduling, Routing and Assigning Wavelengths to Lightpaths in Optical Networks", European Chapter on Combinatorial Optimization, ECCO-XXI, Dubrovnik (Croatia), May 2008.

[5] N. Skorin-Kapov, J. Chen, L. Wosinska, “A New Approach to Optical Networks Security: Attack Aware Routing and Wavelength Assignment“, to appear in *IEEE/ACM Transactions on Networking.*

[6] N. Skorin-Kapov, M. Furdek, “Limiting the Propagation of Intra-Channel Crosstalk Attacks in Optical Networks through Wavelength Assignment”, in *Optical Fiber Communication Conference and Exposition (OFC)* and the *National Fiber Optic Engineers Conference (NFOEC)* (Optical Society of America, Washington, DC, 2009), JWA65, March 2009.

[7] M. Furdek, N. Skorin-Kapov, “ A Scalable Wavelength Assignment Approach for Preventive Crosstalk Attack Localization in Optical Networks”, in *Proc. of the 10th International Conference on Telecommunications (ConTEL 2009),* Zagreb, Croatia, June 8-10, 2009.

[8] N. Skorin-Kapov, O.Tonguz, N. Puech, “Towards efficient failure management for reliable transparent optical networks“, *IEEE Communications Magazine***,** Vol. 47, No 5, pp. 72-79, May 2009.

[9] N. Skorin-Kapov, P. Pavon-Marino, B. Garcia-Manrubia, R. Aparicio-Pardo, “Scheduled Virtual Topology Design Under Periodic Traffic in Transparent Optical Networks|, *In Proc. of the Sixth International Conference on Broadband Communications, Networks and Systems (BROADNETS 2009),* Madrid, Spain, September 14-16, 2009.

[10] B. Garcia-Manrubia, R. Aparicio-Pardo, P. Pavon-Marino, N. Skorin-Kapov, J. Garcia-Haro, “MILP Formulations for Scheduling Lightpaths under Periodic Traffic”, *In Proc of 11th International*

Conference on Transparent Optical Networks (Icton 2009), invited paper, São Miguel, Azores, Portugal, June 2009.
[11] P. Pavon-Marino, R. Aparicio-Pardo,B. Garcia-Manrubia and N. Skorin-Kapov, "Virtual topology design and flow routing in optical networks under multi-hour traffic demand", to appear in *Photonic Network Communications.*

Comparative techno-economic network planning in OCS/OBS/OPS networks

1.1.23 General objectives and summary of results from Y1

The aim of this JA is carrying out network planning studies in realistic scenarios, for optical technologies to be integrated in the short/medium or long term. During year 1, this JA has evolved in two lines:

- A cost comparison of OPS/OBS/OCS networks in a metro ring network scenario
- Techno-economic evaluation of alternatives for the deployment of a photonic mesh

For the former objective, UPCT has developed along the first year of the JA, the network-wide version of the oPASS simulator, suitable for evaluating OBS and OPS networks of heterogeneous topologies. TUW partner is working on estimating the power consumption for the configurations dimensioned, which has a direct influence on OPEX. TUW is estimating the power consumption of the networking equipment but also to take into account the power consumed by the room cooling equipment. Additionally, an UPS (uninterruptible power supply) unit with VRLA batteries for backup can be assumed, which ensures, as an example, one hour of continued operation in case of an interruption of supply from power grid.

For the later objective, the TID group has compared from a techno-economic point of view two types of transponders: DP-QPSK and DQPSK. The study in [Azc08] concludes that, in green-field deployments, the required investments for each option would strongly depend on the network size, so that while DQPSK would require lower investments in medium size scenarios, DP-QPSK would be a more cost effective option in the biggest networks. On the other hand, DP-QPSK channels can be transported over existing 50 GHz optical grids, therefore it would be a better option than DQPSK when 100Gbps channels are introduced over existing 10/40 Gbps photonic networks.

Next table summarizes the above conclusions.

	Mean Link Reach < 3000 Km	Mean Link Reach > 3000 Km
Greenfield photonic mesh at 100 Gbps	DQPSK	DP-QPSK
Migration from an existing 10/40 Gbps photonic mesh	DP-QPSK	DP-QPSK

1.1.24 Advances in Y2

The JA has focused during Y2 on the performance-cost comparison between OPS and OBS approaches, for mesh backbone networks. A testing scenario has been designed to complete an exhaustive battery of tests using the simulation tool developed along year 1. What this JA intends, is to put together a reasonable testing scenario, where the major parameters that can be engineered in an OBS and OPS network are considered, searching for the combination which optimizes the network throughput, guaranteeing a reasonable loss performance target. And then try to answer the question: is a bufferless OBS backbone network a feasible option, from the contention resolution point of view? Are there any significant benefits in the network coming from advancing the header in the one-way reservation paradigm?

The motivation of this work is that the ability to solve contention without optical buffers has been claimed to be a major advantage of OBS paradigm before other approaches like Optical Packet Switching (OPS) networks. Nevertheless, existing evaluation studies supporting this assertion are not exhaustive, and thus not conclusive.

Probably, this is because building a testing scenario is itself a task where relevant non-trivial decisions must be made. Of course, it is not possible to test the interplay of all the combinations of assemblers, with all the schedulers and switching architectures, network topologies, traffic engineering decisions, etc. Identifying the set of parameters and techniques which are going to take part of an evaluation study, should be done with extreme care, if a useful answer is sought. In this JA, an attempt is made in this line. The testing scenario designed has the target of finding the upper limits in burst loss performance that can be achieved with reasonable assumptions by the bufferless OBS paradigm. In other words, relevant design choices in the network have been thought to *favor the contention resolution performance* of the optical nodes.

Some details about the testing scenario:

- 1 edge node connected to each interconnection node.
- General topology for the interconnection nodes.
- Non-blocking full-conversion switch fabrics at the interconnection nodes.
- Fractional routing between 2 traffic connections between every input-output edge nodes.
- Payload distributed using a normal distribution, with constant average and CV={0,0.5,1,1.5}, the coefficient of variation in the payload size distribution.
- W={20,40,80}, number of wavelengths per link between interconnection nodes.
- u={10%,20%,30%,40%,50%,60%,70%,80%,90%}, the maximum utilization allowed in a link between interconnection nodes.
- Δoff ={10,8,5,2,0,-2,-4}, the hop-to-hop reduction in the burst offset. This represents different reservation protocols proposed for OBS networks. Negative values of Δoff imply that the offset actually increases in each hop. This technique has been proposed for compensating the increasing loss probability in longer paths.

Results have been surprising in some aspects. First, the maximum utilization *u* for an end-to-end loss probability of 10^{-4} has shown to be the same for every Δ*off* and *CV* value, and only dependent on the number of wavelengths per fiber *W*. The utilization values obtained are of *u*=0.3 and *u*=0.4 with *W*=20 and *W*=40 wavelengths per fiber respectively. A more satisfactory utilization *u*=0.6 was obtained for *W*=80. These values have to be considered upper bounds, in the sense that are obtained for scenarios designed to favour the contention resolution.

Second, the electronic buffering requirements at the edge nodes is relatively low, and fall very well within the current capabilities of electronic buffers (3.27 MB of buffer, or 2.6 ms of the delay in the worst case tested). It is interesting to see that the ingress delay/buffering requirements are in practice quite similar for the JET/ODD[Cao08]/HPI[Kim04] reservation protocols (JET delays are between 5% and 35% worse than the ones in ODD or HPI).

Third. Several works have studied the dependence between the burst loss probability and the burst length (e.g. see [Cao08]). Results have revealed that losses are distributed quite uniformly. The Jain fairness measure for the conditional distribution of the losses is extremely high (> 0.99). That is quite close to the maximum value 1 which means perfect uniformity.

[Cao08] X Cao, J Wu, X Hong and J Lin K. Elissa, "An Adaptive Offset Time Scheme In OBS Network," Optical Internet, 2008. COIN 2008. 7th International Conference, Oct. 2008.

[Kim04] B. C. Kim, J. H. Lee, Y. Z. Cho and D. Montgomery, "An Efficient Optical Burst Switching Technique for Multi-Hop Networks", IEICE Trans. Commun., vol.E87-B,NO.6 June 2004.

1.1.25 List of relevant publications

[Azc08] Santiago Andrés Azcoitia, Francisco Laura Mateo Quero, Óscar González de Dios, Juan Pedro Fernández-Palacios, "Techno-economic evaluation of alternatives for the deployment of a photonic mesh". NOC 2008 Conference.

A publication is under preparation showing the results associated to the OBS/OPS comparison.

1.1.26 General objectives and summary of results from Y1

The objective of this Joint Activity is to investigate the Path Computation Element (PCE) architecture capability to provide effective Traffic Engineering (TE) solutions in multi-domain BGP-based networks.

PCE architectures have been proposed to perform constraint-based path computations both in single and multi-domain networks. However, in multi-domain networks, the effectiveness of PCE-based computation of inter-domain paths is affected by the limited visibility of TE information which is usually restricted to a single domain. To overcome this limitation, two procedures called Per-Domain and Backward Recursive PCE-based Computation (BRPC) have been proposed. In particular, the BRPC procedure has been proposed for computing, given a specific metric, optimal multi-domain paths. It resorts to the PCE communication Protocol (PCEP) to allow the PCE controlling the destination domain to initiate, in a reverse fashion, the recursive path computation towards the PCE controlling the source domain. However, in BRPC, the sequence of domains to be traversed is given. The PCEP protocol allows the Path Computation Client (PCC), e.g. the Network Management System (NMS) or the source PCE, to specify the sequence of domains to be traversed. Such sequence is included within the PCEP Include Route Object (IRO) carried in the PCEP PCReq message. The problem of defining the optimal sequence of domains to be traversed for the multi-domain path computation (i.e., to be included in the IRO object) has been addressed mainly theoretically or experimentally in management-based networks. Few solutions have been proposed for Border Gateway Protocol (BGP)-based networks. This is motivated by the limited amount of resource information typically exchanged by BGP among domains. Indeed, to preserve network scalability, BGP does not advertise bandwidth information and multiple alternative routes, thus providing the same route choices and causing link congestion. As a consequence, the acquisition of multi-domain resource information from BGP databases drives the PCE to typically consider just one sequence of domains per network prefix. The Per-Domain and BRPC procedures may then be applied along non-optimal sequence of domains, thus potentially affecting the overall network performance.

The objective of this JA is to address the aforementioned issues by introducing the dissemination of additional routing information enabling the effective utilization of PCE-based procedures. The JA started at the end of Y1 with the definition of the objective and the necessary requirements and responsibilities to complete the joint research study.

1.1.27 Advances in Y2

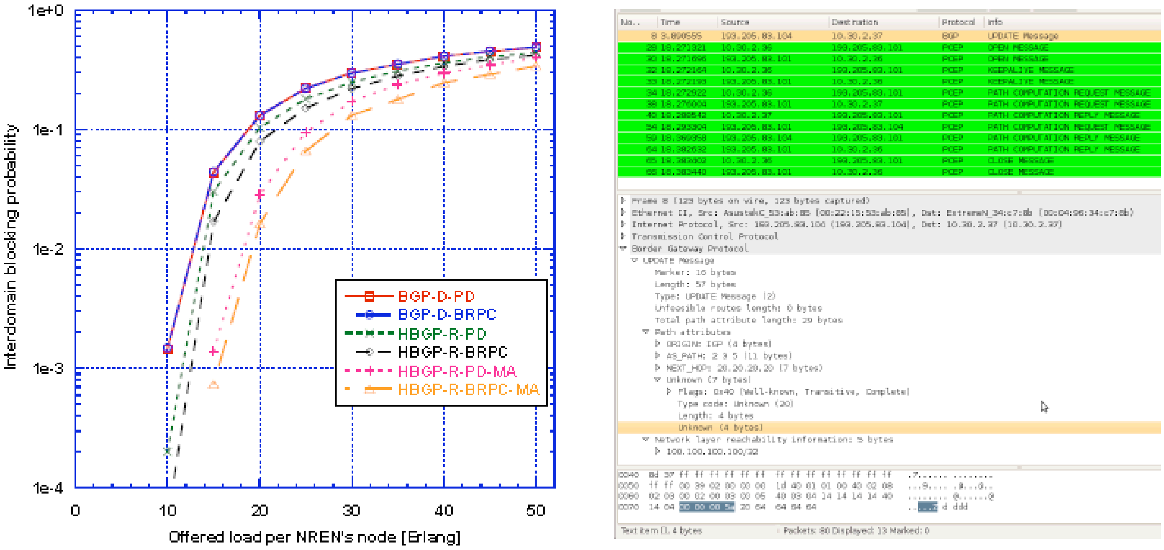
In this study, we focus on the PCE architecture capability to provide effective multi-domain TE solutions. We propose to pre-compute the sequence of domains on the basis of the information provided by a hierarchical path-vector protocol dedicated to TE information. In particular, a hierarchical instance of BGP (HBGP) is proposed to operate within a restricted set of authorized domains. This is in-line with the PCE architecture applicability, which is defined to operate within a limited set of domains with known relationships, like the peering relationships (e.g., up to 20 domains). In terms of confidentiality, it restricts the multi-domain information received by a node to the network view provided by the adjacent domains. In addition, commercial constraints could be adequately supported since transit domains have the capability to potentially treat each customer/prefix in a different way. Moreover, a path-vector view of the network resources is compatible with the view provided by the PCE architecture, which is based on path information exchanged between adjacent domains.

The integrated HBGP-PCE architecture enables the implementation of multiple schemes that differently exploit the various configuration features. They include: the announcement through HBGP of single or multiple routes per each domain or network prefix, the possible announcement of inter-domain bandwidth availability information, the use of Per-domain or BRPC path computation procedures, and the possibility to perform multiple attempts along different sequences of domains.

The performance of the proposed HBGP-PCE architecture has been evaluated through simulations. The reference multi-domain network is a PAN-EUROPEAN topology composed of six ASes: GEANT2 and five National Research and Education Networks (NRENs). Simulations evaluate connection setup time and blocking probability of intra- and inter-domain connections as a function of the offered load. 1.1.27(left) shows an extract of the simulation results. Results show that the worst blocking probability is achieved by a scheme exploiting just the information typically provided by current BGP protocol implementations. On the other hand, the best performance is achieved by a scheme which exploits additional information provided by HBGP as well as the BRPC procedure.

The proposed HBGP-PCE architecture has been implemented in a network testbed composed of commercially available Label Switched Routers (LSRs) configured in N=5 ASes. The implementation successfully includes

PCE prototypes, PCEP Per-domain and BRPC path computation procedures and the introduced HBGP modules. 1.1.27(right) shows an extract of the HBGP and PCEP messages exchanged among the considered network nodes.



[redacted]

Figure 40. HBGP-PCE Schemes: Blocking probability (left). Packet

1.1.28 List of relevant publications

1. L. Buzzi, M. Conforto Cardellini, D. Siracusa, G. Maier, F. Paolucci, F. Cugini, L. Valcarenghi, and P. Castoldi, "Hierarchical Border Gateway Protocol (HBGP) for PCE-based Multi-domain Traffic Engineering", Submitted to ICC 2010 Conference

Traffic engineering and topology design in metro networks

1.1.29 General objectives and summary of results from Y1

During the first year, the JA has focused on the definition of the MAN architecture. In particular, an optical MAN based on optical packet switching and WDM technology was identified as the most suitable candidate to offer high capacity and flexibility. A unidirectional ring topology was considered (the second direction could possibly be used for ring protection but this issue is not considered in this phase of this study). Transmissions are time-slotted and synchronized. On each wavelength, each time slot can support a single fixed-size optical data packet. Each node presents a single fully-tunable transmitter and can transmit at most one packet on any wavelength during a time slot. Each node can be equipped with one or multiple fixed single-wavelength receivers. The main advantage of the proposed architecture is that it permits to blend packet switching techniques with optical transparency in order to offer a flexible MAN architecture able to accommodate large amounts of traffic demands.

After defining the ring architecture, the problem of designing the ring at minimum cost was tackled. The design problem of the WDM ring was formulated with a mixed integer linear programming problem that aims at minimizing the overall cost given by receivers and wavelengths. The peculiar architecture of the proposed ring with respect to previous projects is the possibility to install multiple receivers operating on distinct wavelengths at each node. By exploiting the WDM domain at the receiving side, traffic can, thus, be better aggregated on fewer wavelengths.

1.1.30 Advances in Y2

During the second year, the proposed all-optical packet WDM ring was studied to evaluate the advantages in terms of cost, performance, and scheduling complexity, with respect to WDM optical packet rings with a single dedicated wavelength per node (i.e., single receiver at each node).

A1.1.30.1 Cost Evaluation

The possibility of installing multiple receivers at the nodes allows the sharing of the wavelengths, i.e., packet flows destined to different destination nodes can be supported by the same wavelength(s). From the design point of view, wavelength sharing permits to achieve significant cost reduction in terms of the number of required wavelengths. However, wavelength continuity must be ensured in the intermediate nodes, i.e., packets must use the same wavelength from the source to the destination.

An important result was achieved when assuming that each node is equipped with receivers operating on each wavelength. It is proved that a simple strategy of load balancing of the flows on all the wavelengths allows dimensioning the ring with a smaller number of wavelengths, even if wavelength continuity must be ensured. This shows that, in terms of traffic grooming, the system under study behaves similarly to non-transparent optical packet ring architectures that rely on Optical-Electrical-Optical conversion in transit nodes.

The advantage in terms of number of wavelengths versus ring size, *N*, for a complete and uniform matrix of flows with a rate equal to 10% of the wavelength capacity is shown in 1.1.30.3. The figure compares the number of wavelengths for the case in which each wavelength is shared among all the destinations (WDM case) and for the case in which each node receives on a single dedicated wavelength (nonWDM case).

A1.1.30.2 Dynamic Performance Evaluation

Another benefit of sharing each wavelength among all destinations is given by the possibility of exploiting statistical multiplexing. Indeed, when a destination node receives on multiple (or all) wavelengths, the source node may select any of those wavelengths (if they are not carrying any packet in the given time slot) on a per-slot basis. This statistical multiplexing advantage becomes more evident when stations have little traffic to transmit, as it could happen in access networks that predominantly carry downstream traffic

The statistical multiplexing advantage of the proposed architecture was analytically evaluated, using a queuing theory approach. By assuming that the arrival rate of packets at each node follows a Bernoulli process with rate λ and that the occupancy on the different wavelengths is independent and follows a Bernoulli process with parameter μ, it is possible to model the waiting packets at each node as a Geo/Geo/1 queue, with “Arrival First” policy.

1.1.30.3 identifies working condition where the expected waiting time of the packets is larger than 2 time slots for increasing values of the number of wavelengths, *n*. For a given value of *n*, the area on the right hand side of the curve corresponds to systems where the expected waiting time is smaller than 2 time slots. This curve shows that if μ is at least 0.5, the expected delay is always smaller than 2 as long as $n \geq 2$. Also, if μ is at least 0.2 (that is if each wavelength is occupied 80% of the time), the expected delay is less than 2 if $n \geq 5$. In summary, a higher load that ensures bounded delay can be achieved by exploiting the statistical multiplexing (i.e., at higher value of *n*), even if the source node can transmit at most one packet per time slot.

A1.1.30.3 Scheduling Evaluation

Sharing of wavelengths and statistical multiplexing introduce, however, an additional burden: the packet scheduling issue. Indeed, nodes have to schedule packet transmissions in each time slot, i.e., they need to decide which packet to transmit and on which wavelengths. The trade-off between network design cost and scheduler complexity were considered for the following ring architectures:

1) Multi shared-wavelengths per destination without flow splitting (MW-N) architecture. Each node can receive on one or several wavelengths. Each flow is assigned to a single wavelength (i.e., no splitting) and can be aggregated with other flows having different destinations, on the same wavelength.

2) Multi shared-wavelengths per destination with flow splitting (MW-S) architecture. As MW-N, but each flow can be split and assigned to multiple wavelengths.

In MW-S architecture, a more complex scheduling is required in order to avoid congestion. Two schedulers can be used: “Design Oblivious” (DO) scheduler that randomly selects an available slot with uniform probability, and a “Design Enforcing” (DE) scheduler that enforces, through a Leaky Bucket like mechanism, the proportion of slots used for each traffic flow on each wavelength. The DO scheduler independently selects the destination to serve and the slot to use, which is not the case for a DE scheduler. Performance of DO and DE schedulers are compared on a 6-node ring with nodes labelled A-B-C-D-E-F, a wavelength-to-receiver cost ratio of 1, and a traffic matrix for which the optimal design is summarized in Table 1. Simulations are carried out in ns-2 enhanced with a new MAC layer package.

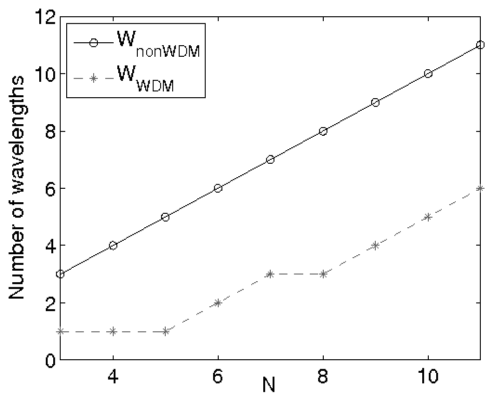


Figure 41. Number of wavelengths versus node size, N, for WDM and non-WDM case

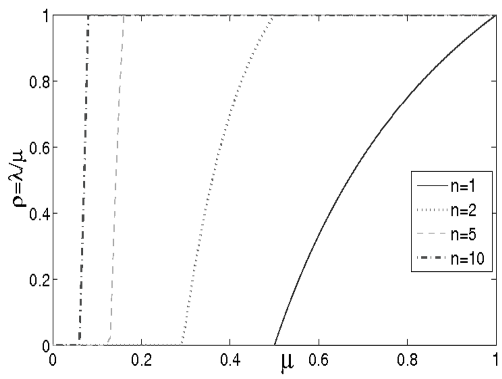


Figure 42. Achievable capacity utilization versus wavelength occupancy, μ, that ensures an expected waiting time of 2 time slots, for different number of wavelengths, n

Wavelength	*A to D*	*B to E*	*E to C*	*F to D*
L1			**0.6**	**0.4**
L2	**0.5**	**0.3**		**0.2**

Figure 43. Optimal design for MW-S. 2 wavelengths are used, C and D receive on L1, D and E receive on L2 and the flow between F and D is split between L1 and L2

Results show that the DE scheduler supports the offered traffic matrix, and enforces the traffic mapping given in 1.1.30.3. On the other hand, the traffic mapping yielded by the DO scheduler is given in 1.1.30.3, which shows that the 0.3 flow from B to E is not supported. This means that node B is congested and thus indicates that DE scheduler should be used in MW-S to enforce designed traffic mapping.

The main drawback of DE scheduler is that DE scheduler has to be reconfigured every time the traffic mapping is modified, even when the numbers of used wavelengths and receivers are not modified. This is a heavy configuration burden, unrealistic for operational networks. DE schedulers also present the drawback of not being work-conserving (some empty time slots may be left unused in presence of queued packets), which induces quality of service (QoS) degradations in terms of transfer delay.

Wavelength	*A to D*	*B to E*	*E to C*	*F to D*
L1	**0.16**		**0.6**	**0.12**
L2	**0.34**	0.18		**0.48**

Figure 44. Traffic mapping with a DO scheduler

To avoid the drawbacks and the use of DE scheduler, flow splitting should be avoided. The use of MW-N architecture in place of MW-S architecture is an attractive alternative from the point of view of scheduling. However, for selected traffic rates, MW-N has higher design costs, .e.g, as displayed 1.1.30.3, where a third wavelength is required.

The main result is that that allowing splitting of packet flows decreases the design cost while imposing a "Design Enforcing" scheduling policy which presents unrealistic configuration burden. Forbidding flow splitting avoids this burden but may lead to a higher design cost.

Wavelength	A to D	B to E	E to C	F to D
L1		**0.3**	**0.6**	
L2	**0.5**			
L3				**0.6**

Figure 45. Table 3: Optimal design for MW-N. 3 wavelengths are used, C and E receive on L1, D receives on L2 and L3

1.1.31 List of relevant publications

The results reported in the previous subsection are the fruit of the joint collaboration of the partners and were published in the following papers:

- B. Uscumlic, A. Gravey, P. Gravey, I. Cerutti, "Traffic Grooming in WDM Optical Packet Rings," Proceedings of 21st International Teletraffic Congress (ITC 21), Paris, France, September 2009.
- B. Uscumlic, A. Gravey, I. Cerutti, P. Gravey, M. Morvan "The Impact of Network Design on Packet Scheduling in Slotted WDM Packet Rings," Proceedings of Photonics in Switching (PS), Pisa, September 2009.

Effects of outdated control information on routing in optical networks

1.1.32 General objectives and summary of results from Y1

In the first year of this joint activity, the research group of Politecnico di Milano (POLIMI) has investigated the effects of outdated control information in control-plane-enabled optical networks. The dissemination of control information (usually provided by routing protocols such as OSPF-TE) is essential in this kind of control-plane enabled networks: in particular, this information has to be continuously updated to allow routing algorithms to efficiently carry on the path computation [1].

The control information depends on various factors, mainly on which protection technique is applied and if wavelength conversion is enabled. E.g., while the standard distributed information (e.g., link state advertisement in OSPF-TE) is the link-state and the free capacity available on each link, in a not-wavelength-convertible case, the single channel status has also to be distributed if the wavelength continuity constraint needs to be satisfied.

Our investigation introduces innovative elements with respect to the previous body of research, since this was the first work covering the peculiar effect on protection (especially, backup sharing in shared protection) and comparing the cases with and without wavelength conversion. We explored a wide range of values of control delays, analyzing the deriving performance loss over three different dimensions: (i) the value of control delay, (ii) the routing strategy adopted (namely, unprotected, dedicated path protected and shared path protected) and (iii) the conversion capabilities (the wavelength convertible case will be referred as Virtual Wavelength Path ,VWP, as opposite to the not wavelength convertible case referred to as Wavelength Path, WP)

Path computation elements (PCEs) in nodes are not always updated and routing may be carried on according to an outdated image of the network. When introducing relevant control delays, new contributions to blocking can be identified:

• RC - Resource Conflict: the PCE routes s connection along a path that is no more available, because in the meanwhile those resources have been taken and the connection is blocked even if there are chances that alternative free paths were available.

• FS - False Saturation Conflict: PCE is unable to route a connection because no routes seem to be available between source and destination nodes. The connection is blocked even if, in the meanwhile, some resources have been released .

• SR- Sub-optimal Routing: The PCE routes a connection along a suboptimal route.

Applying a very general control-delay representation, we were able to provide wide-range simulative study to quantify the effect of signalling on routing performance, mainly by using the blocking probability (P_b) metric: since blocking stems from different causes, the influence of the various P_b components have been also investigated.

P_b curves, independently of the routing strategy adopted, can be described by three well distinct phases:

• I phase - Not influential delay: in this first phase P_b is stationary and it is not influenced by the delay. Essentially performance are equivalent to the ideal (zero-delay) case.

• II phase - Linear increase of P_b: delays in network-state reception start affecting the quality of source-routing, causing a significant and linear increase on P_b

• III phase - P_b saturation: P_b is no more affected by a further increase of the network-state-information delay, since the network image in the source-routing node is totally uncorrelated to the actual network state. The situation is equivalent to route connections with no information about network state.

In the WP cases, P_b values are higher than in VWP case: while in the first phase this is a well-known effect of wavelength continuity constraint, in the saturation phase we should also consider that an higher P_b is achieved because the routing algorithms for VWP cases are more tolerant to outdated information in the VWP case than the WP case.

The contributions of the various components have trends that are strictly interdependent; in particular:

- for very small delays, FS and PS curves grow linearly, while the CB component is almost constant (actually it slightly decreases of a value equal to the sum of PS and FS that are very small in this phase) and largely constitutes the overall probability.
- when the PC contribution equals the CB contribution (at about D = 10^{-3}), the CB curve start rapidly decreasing, while the PC curve keeps growing linearly until it stabilizes for value of D comparable with the average holding time (i.e., about 1).
- also the FS contribution keeps growing until it reaches comparable values with CB. Then it decreases consistently to stability on very small probabilities below 10^{-5}.

The overall P_b is dominated initially by CB, then for higher delay values PC is the dominant term; FS plays a significant role only on a limited range of delays values after PS has intersected CB.

Advances in Y2

The main contribution of this simulative activity has been achieved during this first year, but some further investigation has been conducted also in the second year. In particular in this deliverable we report some discussion on how the effects of outdated information vary for different traffic dynamicity

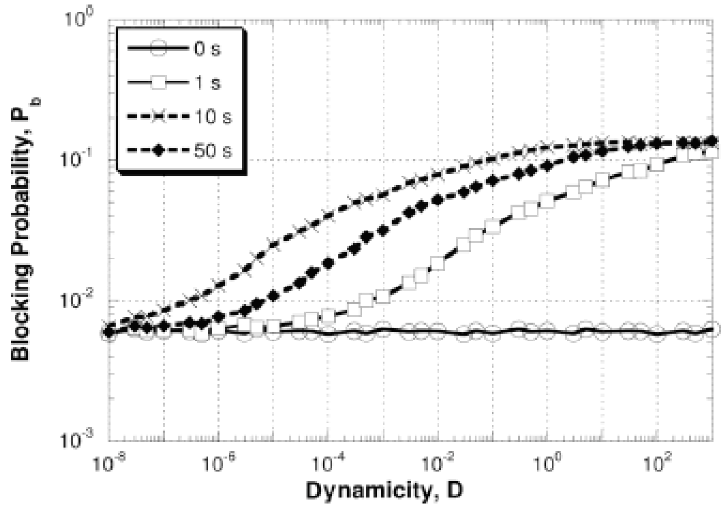


Figure 46. Effects of outdated information for different values of traffic dynamicity

For a fixed value of traffic load in a network, the outdated information may induce different effects according to the characteristics of the traffic. It is intuitive that the same control delay has a weaker effect on traffic composed by long and rare connections than on traffic composed by short and numerous connections. The following parameter *D* expresses the traffic *dynamicity:D* = AR/HT. Note that for a fixed load analysis, *D* must be varied, keeping the product AR_HT constant (in the following simulations AR_HT=150). 1.1.32 compares the *Pb* curves for different values of control delay (0, 1, 10, 50 s), increasing the traffic dynamicity. The effect of control delay is exacerbated by higher dynamicity: for increasing values of control delay, even for very low values of *D* (e.g., 10−6–10−4), the blocking probability starts to increase significantly. Note that in the ideal case without control delay, the dynamicity does not affect performance.

Final results of this investigation have been published in [2].

Within the second year, the CTTC has started the experimental evaluation in ADRENALINE testbed® of the effect of the outdated information in the network performance of distributed Shared Path Protection (SPP) recovery schemes for wavelength-routed optical networks (WSON) with wavelength continuity constraint. SPP is an efficient recovery mechanism that besides achieving an acceptable recovery time, optimizes network resource usage. These benefits are accomplished through so-called backup sharing among existing backup lightpaths. Specifically, backup resources along backup LSPs are pre-reserved, but not cross-connected. The same pre-reserved resource can then be shared by multiple backup lightpath as long as their respective working lightpath may not be affected by a single failure. This condition is termed as sharing violation.

In WSON with a distributed control plane, upon reception of a connection request, the source node executes a constraint shortest path first (CSPF) algorithm to find a feasible end-to-end path, considering as input the topology and network resource state collected in the traffic engineering database (TED) repository. The routing protocol (e.g., OSPF-TE) is responsible for flooding any change occurring in the network state, which permits to update the local TEDs. The node originating a change generates a Traffic Engineering – Link State Advertisement (TE-LSA) to all its neighboring nodes. Then, the neighboring nodes receiving the new TE-LSA, process it (updating of local TED repository, and optionally, pre-caching/pre-computation of routes) and forward it to all its linked neighboring nodes except the node by which it has received the TE-LSA. Finally, this process is repeated by all the nodes.

For this study we consider three information dissemination strategies. The difference between the three approaches relies on the level or granularity of the required network state information disseminated. In particular, for the first two strategies, the information is disseminated only at aggregated bandwidth granularity, but in the first approach only unreserved aggregated bandwidth-based link information is used while in the second approach both unreserved and sharable bandwidth-based link are disseminated. As for the third routing strategy, it considers per-wavelength channel granularity dissemination, based on a wavelength channel bitmap encoding. This study relies on three well-defined SPP routing algorithms [3][4][5] that make use of the information disseminated by the three proposed strategies.

The novelty and goal of this work is to experimentally evaluate the impact of the outdated information in the network performance. We consider four main performance indicators, namely, the blocking probability (BP), setup delay (SD), resource overbuild (RO), and the routing convergence time, that are experimentally measured in ADRENALINE testbed. The restoration overbuild is a figure of merit specific for recovery schemes, and is defined as the amount of channels consumed by backup lightpaths over the amount of channels utilized by working paths. A new key performance indicator introduced in this work is the routing convergence time for an update (CTU), defined as the time period required for an LSA update to be received and processed by all the nodes within the network i.e., the time span starting from the time the LSA is generated or refreshed until the LSA is integrated by the routing controllers into their respective databases. Such processing results in a common view of the network topology in terms of TE attributes of Links and Nodes.

More specifically the CTU focuses on the flooding / dissemination of a LSA and, for our work, we only consider TE Intra Area opaque LSAs, both for Router Address TLVs (type 1), TE Link TLVs (type 2) and experimental TE Node TLVs (type 5). In consequence, the CTU does not consider or include the initial or periodic database synchronization or the creation of OSPF-TE adjacencies which also occur. For the scope of the work, the CTU assumes a stable control plane network (DCN). As expected, there are several network parameters that will impact the CTU value; notably, the network size in terms of nodes and links, the link propagation and transmission delays, and the size and internal processing of the LSA that takes place in each node that receives the LSA, parses its contents and forwards the LSA to adjacent nodes. It is worth to note that the processing that takes place upon the reception of an LSA may vary significantly and it is implementation-defined, since it may range from basic parsing and storing of the LSA to complex management of the TE database (TED) and caching of pre-computed paths.

The selected methodology for the actual sampling and measurement of the CTU in the network is intrusive and requires specific modifications in the implementation of OSPF-TE routing controllers. The core mechanism involves the time-stamping, locally at each node (e.g. in a local log file), of selected events such as the generation, reception, processing and forwarding of the aforementioned LSAs. This uses the fact that the actual number of different LSAs is known in advance (the total number of links and nodes) and the fact that LSAs are uniquely identified. After the network has been operating for a given time, and thus it is considered to be in "permanent" regime, selected logs are retrieved and post-processed, using ad-hoc applications that perform the actual data mining and generation of statistics. Given that the measurement is the result of finding the maximum of a set of time intervals [a, b] for all the pairs (LSA originating node, LSA receiving node) and that time stamps are obtained in different controllers, it requires an accurate synchronization of node clocks. This is a notable challenge that needs to be addressed, since propagation delays, CTU values and clock drifts are in the same order of magnitude of a few milliseconds. Our initial work for this purpose is based on fine-grain tunning of NTP protocol

Experimental performance evaluation will be carried out in the ADRENALINE Control Plane Emulator is an extension of the ADRENALINE testbed used for the experimental performance evaluation of dynamic GMPLS and PCE-based traffic engineering algorithms and enhanced protocols (RSVP-TE, OSPF-TE, LMP and PCEP) in complex and diverse network topologies. The ADRENALINE Control Plane Emulator (Fig. 5) is composed of 42 opctical controllers without associated optical hardware (i.e., emulated hardware). This set of optical controllers introduces a new degree of flexibility in topology configuration, without restrictions regarding either the targeted optical network topology or the link resources (e.g., number of available wavelengths, fibers, etc.). Thus, the optical controllers can be inter-connected following any devised topology, by means of Ethernet point-to-point links. The proposed solution allows the specification of control link parameters for realistic QoS constraints (fixed and variable packet delays, packet losses, bandwidth limitations, etc.) emulating optical links. To do this, it uses virtual local area networks (IEEE 802.1q VLANs), configured both in the layer 2 Ethernet switches and in the GMPLS-enabled controllers within the testbed, with optional GRE or IPIP tunneling.

For the experimentation, the considered network scenario will be based on the NSFNET topology, formed by 14 nodes and 21 links with eight wavelengths per link, and a node switching delay of 20ms. The lightpath-arrival process is Poisson, and the holding time follows a negative exponential distribution, with source-destination uniformly distributed among all (distinct) node pairs. The total offered traffic load (AR*HT) will be set to a fix value, and the dynamicity (AR/HT) and average link propagation delay (ms) of all links will be will be varied during all different experimentations. It is worth noting that in contrast to the simulative activity performed by the Polimi, the variable parameters of the control delay are restricted to the link propagation delay. Each data point will be obtained by requesting 10^4 lightpaths.

1.1.33 List of relevant publications

[1] IETF Network Working Group, "OSPF extensions in support of GMPLS," (2002). Internet Draft.

[2] M. Tornatore, F. De Grandi, R. Munoz, R. Martinez, R. Casellas, A. Pattavina, "Effects of Outdated Control Information in Control Plane-Enabled Optical Networks", IEEE/OSA Journal on Optical Communication and Networking, Vol. 1, no.. 2, pp. 194-204, July 2009

NETBENCH "Benchmarking of network architectures for guaranteed service provisioning"

1.1.34 General objectives and summary of results from Y1

The objectives of this JA are summarized in a number of activities that overall aim to provide a complete (to the extent possible) study and performance evaluation of different optical networking technologies of interest. Optical networking and switching technologies of interest range from OBS and OPS to circuit/wavelength switching, using two-way reservations, etc offering a certain degree of dynamic resource allocation and/or guaranteed end-to-end performance. The methodology followed in this JA is planned to go through the following tasks:

- Collection and documentation of requirements and assumptions to be used in the evaluation phase
- Definition of performance evaluation scenarios
- Modeling and performance evaluation
- Analysis, evaluation and presentation of the results

In Y1 the basic assumptions towards capturing and analyzing requirements have been concluded. The objective was to clarify (especially among the collaborating partners) the basic operational parameters that have been considered in each study and the potential assumptions that have been made.

Starting from the requirements imposed by the network topology, one of the most significant parameter affecting the performance evaluation methodology, details about the reference network topology should be described in terms of: topology (connectivity), path distances and capacity. Depending on the case to be demonstrated either existing backbone network topologies (NSF network, Pan-European network) for which details are available in the literature or "ideal" reference topologies serving specific cases to be demonstrated, were selected. Novel network architectures such as CANON, an approach where the nodes of a wide area network (typically 20-30 nodes) are organized in clusters mainly based on vicinity, traffic, legacy infrastructure and administrative criteria, were also assumed.

The requirements analysis included the functionality of the nodes which should also be specified for each examined scenario. Most optical switching techniques require an adaptation unit employed at the network edge and potentially transparently switch traffic at the inner core network. For example in the CANON approach the nodes are separated as core and edge node depending on being an intra-cluster node (a node inside the cluster) or an inter-cluster node (Master Node, a node to connect with other clusters). This differing functionality of the nodes impose that electrical buffers exist only in the periphery thus once a packet enters the optical domain it will travel until the destination without any further queuing while two granularity levels are used: the transceivers of the cluster node are operating on a slot-by-slot basis while data are exchanged between clusters in much longer fixed-size frames. Thus the switching node architecture and the wavelength conversion capability was included in the list of parameters for evaluation.

In all cases, the traffic load used as an input has to be dynamic fluctuations of traffic approximating realistic conditions over a wide-area backbone network. As an example exponential distributions for interarrival times were selected or a characterization of traffic based on the RedIris network. Also different QoS classes should also be examined.

Since the benchmarking methodology extends to an overall network solution, the control plane architecture cannot be ignored. The usage of OBS alternatives (TAW, JIT, JET) or GMPLS based protocols, create a major effect on the performance evaluation of the system.

Finally the performance metrics have to be pointed out. Such metrics were addressed divided in two major sections

- CAPEX: the switching node and the overall network cost. The cost can be extended in number of components used, cost for implementing the network system, or even in power consumption and switching or node dimensions.
- OPEX: traffic performance. Performance evaluation of the system that can be showed by metrics of delay, packet loss probability, jitter, utilization, etc .

In the results presented in Y1 a study of the alternative CANON hierarchical network architecture in the NSF network was evaluated where the 14 nodes of the NSF network were considered as MNs collecting regional traffic from clusters of 5 nodes. The performance of the proposed scheme was compared to that of just-enough-time (JET) OBS and OBS-INI schemes. As observed in the results CANON and OBS-INI introducing 2-way reservations exhibited higher average delay than plain OBS-JET, remaining though within acceptable limits for any type of service, while at the same time they achieved significantly lower losses than when using OBS-JET. It is worth noting that CANON loss probabilty outperformed the other two solutions due to the exclusive use of two-way reservations in CANON, where losses occur only due to buffer overflows.

1.1.35 Advances in Y2

Work for the second year proceeded along the same lines going into more technical details and producing first results. The benchmarking methodology including results obtained by using network simulators was mainly followed by UoP and UniBo, while UAM is working on traffic load characterization and is planning to provide a characterization of traffic for the RedIris network, specifically regarding busy hour volumes, that can be used at a final stage of evaluation with realistic traffic load models. Evaluation of selected scenarios progressed in two directions investigating the impact of network topology, optical transport and control plane architecture on performance as well as focusing on the optical node architectures and investigating the performance of node architectures with multi-service traffic.

A1.1.35.1 Evaluation of end-to-end performance

The first direction aims to evaluate end-to-end performance in wide are network topologies including core optical networks and integrated access-core optical networks.

Evaluation of large-scale network topologies was performed through modelling and simulation of hybrid topologies and mixed end-to-end resource reservation schemes including:

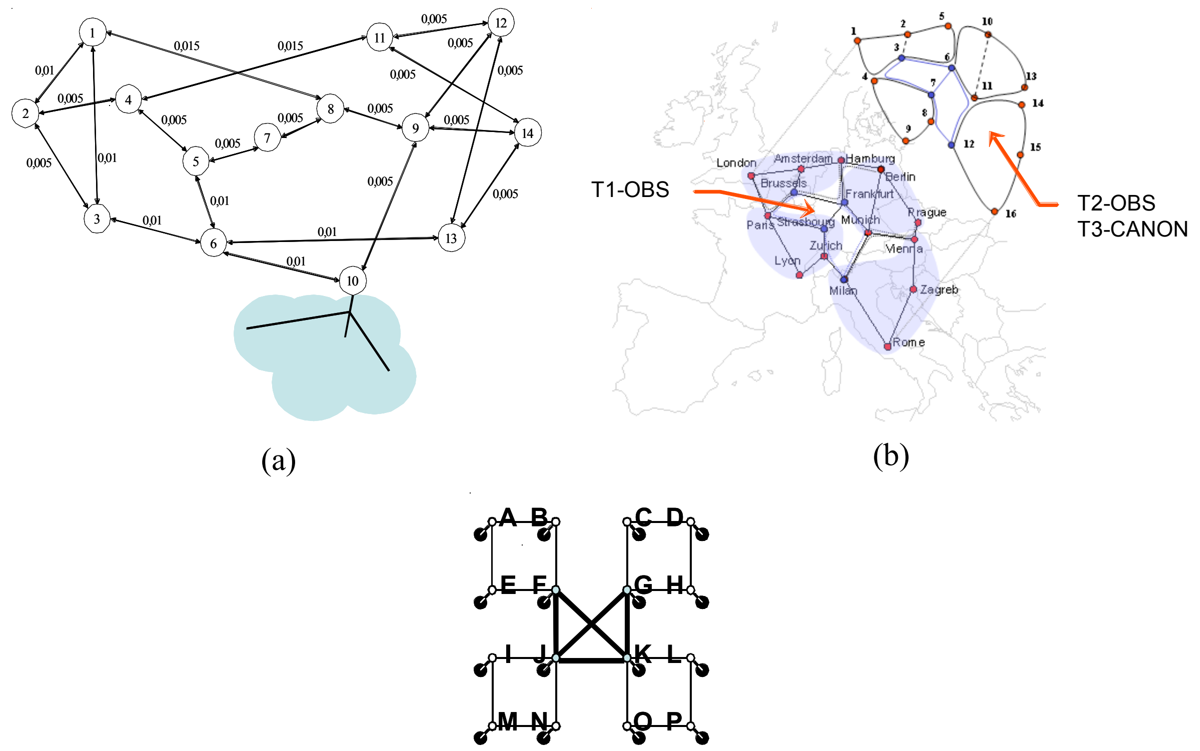
- Evaluation of the impact of network dimension on performance
- Evaluation of traffic grooming techniques for transparent access-core network integration and impact of optical access-core technologies on transport layer protocols

The objective of the above study is twofold; first we evaluate efficient aggregation and switching with appropriated reservation schemes in ultra-high speed optical transport networks and second we investigate the efficient integration of optical access and core networks.

Indicative results

One of the most significant parameters affecting the performance evaluation methodology is the set of requirements imposed by the network topology in terms of: topology (connectivity), path distances and capacity. One applicable well studied case for a backbone network topology is the NSF network shown in

1.1.35.1(a). An equivalent model for Europe is the Pan-European COST 266 backbone optical network, which can be also used as a real case scenario similar to the NSF as showed in 1.1.35.1(b). Since this topology consists from many nodes, different algorithms for clustering have been studied. The advantages of the European network is that it can be used to show the gain from implementing the CANON architecture, from all points of interest, in an existing topology. One other solution is that of the creation of an ideal symmetric network as showed in 1.1.35.1(c).



(a) (b)

(c)

Figure 47. Network topologies used for benchmarking (a)NSF network, (b) the European network (c) ideal/symmetrical network

In a relevant study we demonstrated that using the existing infrastructure of the Pan-European network (1.1.35.1), the fully dynamic CANON can efficiently combine statistical multiplexing gains and improved blocking performance, compared to meshed S-OBS (Slotted OBS), due to a reduction of the contention points, improved network design and resource allocation. OBS suffers excessive burst loss since plain wavelength conversion cannot offer substantial contention resolution at high loads. We showed that by efficiently clustering the optical nodes, loss performance has been improved. By applying the CANON scheme, a reduction in the number of contention points can be achieved, justifying the improved loss performance. In addition, we have shown that CANON is a considerably more cost-effective solution due to the lower overall component count, lower overall power consumption and better utilization of the installed capacity. As shown, OBS needs more wavelengths just to avoid collisions compared to CANON, which keeps burst loss quite low while packing bursts into much fewer wavelengths, increasing utilization. Indicative results are shown below:

As a first scenario we examined the case where sufficient capacity is provided by the network to serve the offered load and transport the generated data bursts/slots with W=16 wavelength channels per link and measured the burst loss probability across all nodes (1.1.35.1). To investigate further the relation between the deployed network resources, the achieved performance and network utilization we compared CANON against other solutions in terms of resources needed to achieve the above loss performance. Since the loss probability factor - albeit its major impact on the system performance - is not the only indicative performance metric, we introduced the utilization factor as a metric for comparison. The utilization factor was expressed as the ratio of the sum of the successfully served traffic for all network nodes over the installed network resources:

$$U = \frac{\sum throughput_i}{\sum capacity_i}$$

Equation 1: utilization factor

At each network node i in Equation 1 above $throughput_i$ denotes the successfully received traffic destined to this specific node and $capacity_i$ denotes its capacity, which in turn is determined by the number of available transceivers per port. The results are shown in 1.1.35.1. Finally of interest is to investigate the power consumption when scaling the size of the network, hence we increase the number of links (fibers) used in the case of the Pan-European clustered network by a factor of K (i.e. the previous analysis for the network performance corresponds to K=1) and the number of wavelength channels per link to W=16, 32 and 64. The results are shown in 1.1.35.1 for the topologies T1-T3, respectively.

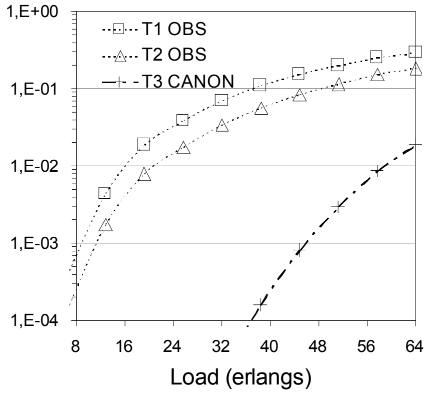


Figure 48. Frame or burst loss probability for W=16

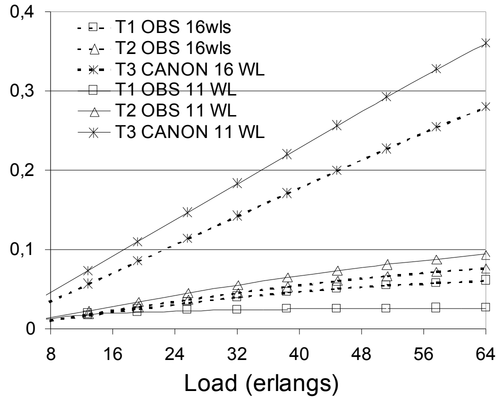
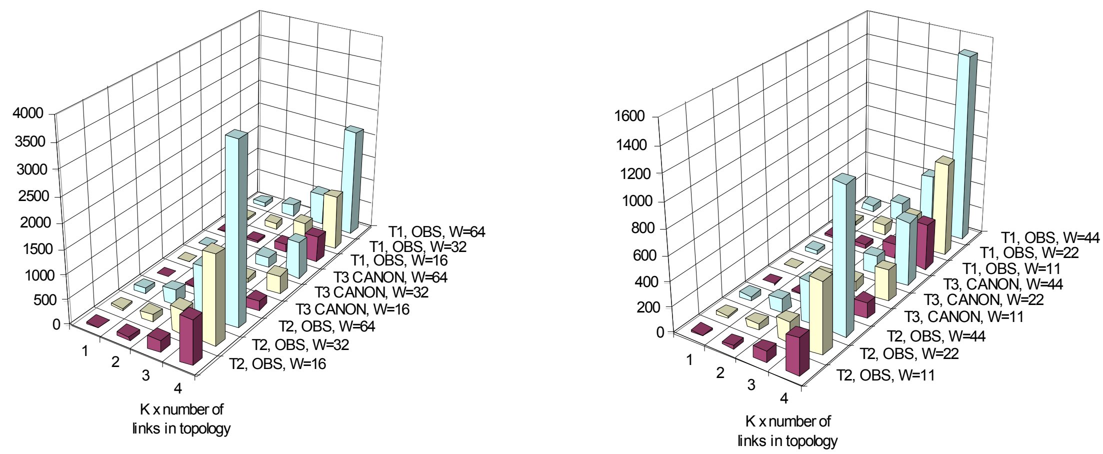


Figure 49. network utilization



(a)

(b)

Figure 50. Estimated power consumption for T1, T2 and T3 topologies and projection of the consumption when the network scales with respect to the number of links in a network K and the number of wavelengths per fiber W a) for W=16, 32, 64 and b) for W=11, 22, 44.

<u>*Evaluation of optical Access and Core networking technologies*</u>

- Performance evaluation of a WDM-PON based integrated access-core optical infrastructure:

To successfully realize the transformation of broadband networks and the Internet to a multi-service platform the structural limitations of current networking architectures must be raised so that information transport infrastructure gracefully evolves to address transparent core-access integration, optical flow/packet transport and end-to-end service delivery capability, overcoming the limitations of segmentation between access, metro and core networks and domains. Towards this end we evaluated an integrated control plane for optical access and core networks, which addresses the above considerations. The proposed control plane can lead to a unified transport infrastructure integrating state-of-the-art components and technologies including Wavelength Division Multiplexing, Passive Optical Networking and Optical Packet Routers with inherent traffic grooming capabilities. The performance of the proposed architecture is assessed my means of simulation in terms of cost, resource utilization and delay. The main innovation is the extension of the control plane and unified resource allocation, so as to minimize the cost of per packet processing, which cannot be tolerated in photonic backhaul networks. The proposed architecture addresses the above by optical burst aggregation over passively split optical access networks employing hybrid WDM-TDM multiplexing into fixed-size containers called frames. This frame aggregation technique can exploit efficient synchronous switching technologies and achieve high utilization of optical core networks. We evaluated this architecture in comparison to typical approaches that have been proposed for dynamic resource allocation utilizing burst switching techniques and have shown that it can better utilize network resources at the cost of a limited increase in delay. More significant performance gains are expected when dynamic resource allocation over the core network is performed, which is currently work in progress.

- Performance evaluation of the TCP protocol over multi-granular networks:
  - The effect of offset time in OBS has been evaluated in the presence of hybrid switching technology with different switching time

The objective of the above study has been to evaluate the performance of TCP with hybrid technologies providing fast and slow paths for supporting multi-granular connections (i.e. multi-granular switching nodes implementing mixed technologies with different properties and speeds e.g. SOA, MEMs etc.). The switch properties result in two alternatives for optical switching of data transport units (bursts): i) fast connections (paths), in which case a sequence of bursts is switched in core nodes with fast optical technologies (short setup times) and ii) slow connections (paths), where a sequence of bursts is switched in core nodes with slower optical technologies (long setup times).Since different switching set up times imply different offsets times in burst switching modes the impact on performance has been measured and was shown that fast and slow paths get best throughput for different values of the TCP window size, while for low burst losses (e.g. lower than 10-) fast paths give remarkably better throughput than slow paths, otherwise the throughput is almost the same, despite of different transmission rates and technologies employed.

- Performance evaluation of TCP protocol with different access techniques to OBS networks
  - Wi-Fi, Ethernet and EPON are considered.
  - Key parameters are investigated; burst assembly time is thoroughly studied as a key parameter to tune performance

The interworking between different access networks and an optical burst switched network is considered here. The end-to-end performance of the TCP (Transmission Control Protocol) is evaluated by jointly accounting for access network protocols and the burst assembly procedure at ingress edge nodes, both with wired and wireless access solutions, based on electrical and advanced optical technologies. The influence of the assembly timeout in different access contexts is presented, and numerical investigations are performed by means of ns-2 simulations. Results show that end-to-end throughput is mostly influenced by the delay introduced by access protocols, which arise in the analyzed different scenarios, and by assembly timeout. These results provide meaningful insights about interconnected systems to the task of overall network design and, in particular, to the setup of the interworking unit parameters.

A1.1.35.2 Performance of node architectures with multi-service traffic

In the expected multiservice scenario of future networks, the core resource usage must be subjected to rules, which in this study are referred to as QoS differentiation mechanisms and are defined in relation to the characteristics of available network elements. The main problem to face in packet-oriented optical networks the lack of an optical equivalent of the electronic random access memory (RAM) that does not permit the use of large buffers and the direct migration of the typical approaches used in electronic IP routers, such as, e.g., the use of active queue management. Thus in this study a bufferless optical packet switches with QoS differentation mechanisms for OPS networks has been evaluated.

The proposed QoS mechanism must be kept very simple to allow fast processing according to the very fast packet–burst forwarding requirements. Several QoS mechanisms have been proposed in the past for this kind of networ witch provide QoS based on a perflow or a perclass classification of the traffic. Furthermore, QoS parameters can be expressed in relation to either relative or absolute bounds. With absolute guarantees, QoS parameters are given as upper bound while with relative guarantees, QoS parameters of a given class are given relative to another class. Within each QoS model different algorithms can be defined depending on the specific networking context to which they are applied. For example QoS differentiation schemes for asynchronous bufferless OPS, such as the wavelength allocation algorithm (WA), the preemptive drop policy (PDP), and intentional packet dropping (IPD).

With the aim to reduce switch cost while keeping contention resolution in the wavelength domain based on wavelength conversion, switch architectures that employ shared wavelength converters (WCs) have been considered. In this kind of architecture, packet loss is caused not only by wavelength contention on the output fibers (OFs), but also by lack of WCs. The most important example of such switches is represented by the shared-per-node (SPN), where a pool of r WCs is shared among all the input wavelength channels (IWCs). This sharing scheme represents a sort of reference, as it allows one to obtain the best savings in terms of WCs, achieving the same loss probability as other bufferless schemes. The wavelength converter sharing scheme has a relevant impact on switch architecture and WC characteristics.

The QoS models mentioned above can be applied to this kind of sharing scheme. In particular a differentiation scheme based on the allocation of the output wavelength channels (OWCs) can be extended to the allocation of the shared WCs. This means that when the OWCs and WCs are assigned to the incoming packets, the QoS mechanism can be applied to arbitrate among classes of service.

A switch architectures that employs fixed-input, tunable-output wavelength converters (FI- TOWCs) is supposed to be easier and less costly to realize, and their employment naturally leads to a new sharing scheme named “shared per wavelength” (SPW). Such switch architectures can be designed in a scalable and feasible way and can be implemented by using multistage organization of modular and scalable switching matrices.

The application of the QoS differentiation scheme to SPW switches in a slotted context with multiple classes of service has been evaluated. Starting from the switch architecture for a slotted best-effort undifferentiated context, a migration to class-based context was presented with the aim to provide a QoS mechanism over a simple and cost-effective architecture. Such a switch, equipped with this QoS differentiation mechanism, can represent a network element to support the functionalities and needs required by applications in optical networks for the future Internet. A new mathematical model for the evaluation of the packet loss performance for the different classes of service and an extended version of the control algorithm that manages packet forwarding according to the classes of service is presented have been developed. Analytical results have been validated against simulation, and the developed numerical analysis showed the effectiveness of the proposed QoS mechanism to provide basic service class isolation.

Overall, this study showed how a simple QoS scheme can be effectively applied to an OPS to make this network element suitable to face basic service differentiation requirements. Within this approach the only aspect of service differentiation that is taken into account is packet loss, because the switch architecture is bufferless. The application of the model allows the maximum average load of each class to be determined to ensure the required maximum loss rate.

1.1.36 List of relevant publications

2. T. Orphanoudakis, A. Drakos, H.-C. Leligou, A. Stavdas, A. Boucouvalas, “Dynamic Resource Allocation with Service Guarantees over Large Scale Optical Networks”, IEEE Communications Letters, Vol. 13, No. 11, November 2009
3. T. Orphanoudakis, H.-C. Leligou, E. Kosmatos, A. Stavdas, “Future Internet infrastructure based on the transparent integration of access and core optical transport networks”, IEEE/OSA Journal of Optical Communications and Networking (special issue on Optical Networks for the Future Internet), Vol. 1, Iss. 2, pp. A205–A218 (2009)
4. M. Casoni, C. Raffaelli, ‘TCP performance over optical burst-switched networks with different access technologies’, accepted for publication in OSA/IEEE Journal of Optical Communications and Networking (JOCN)
5. A. Cianfrani, V. Eramo, A. Germoni, C. Raffaelli, M. Savi, “Loss Analysis of Multiple Service Classes in Shared-Per-Wavelength Optical Packet Switches”, accepted for publication in OSA/IEEE Journal of Optical Communications and Networking (JOCN)
6. T. Orphanoudakis, H.-C. Leligou, E. Kosmatos, A. Stavdas, “Optical metro network architecture based on traffic grooming over hybrid TDM/WDM PONs”, 4th European Conference on Networks and Optical Communications (NOC), June 2009, Valladolid, Spain.
7. M. Casoni, C. Raffaelli, “TCP Performance in Hybrid Multi-granular OBS Networks”, International Workshop on Optical Burst/Packet Switching (WOBS) 2009, Sep. 14, 2009, Madrid, Spain.

8. Carla Raffaelli, Michele Savi, "Multi-granular traffic scheduling in hybrid optical packet switch with electronic buffers", submitted to Globecom 2009

Adaptive Admission Control In Dynamic GMPLS Networks (JA 15)

1.1.37 General objectives and summary of results from Y1

Future optical networks will support data transfer rates of several terabits per second. However, even at such bit rate magnitudes, network resources could be scarce when compared with the expected aggregated bandwidth demand. To provide an acceptable level of Quality of Service (QoS) to the applications an efficient management of the resources is required. The introduction of the Generalized Multiprotocol Label Switching (GMPLS) will standardize the control plane of optical networks, therefore allowing the dynamical allocation of resources, and thus reducing the operational cost of providing QoS levels similar to the experienced with circuit switched, connection-oriented technologies. An optimal session Admission Control (AC) strategy is an efficient way to manage the resources that need to be held in reserve at each intermediate node for sessions belonging to classes of service of superior priority. The design of AC policies must consider session related parameters ---like blocking probabilities for new sessions or forced termination probabilities for ongoing sessions. In fact, blocking probability is an essential QoS parameter in fixed networks. Furthermore, it is now accepted that, compared to scheduling, AC might be a more suitable mechanism of traffic management to provide differentiation among services.

We propose a novel adaptive AC scheme for GMPLS networks that handles at the ingress node the set up of an all-optical packet-based label switched path (LSP) between the edges of the network through one or several label switching routers (LSRs), and tries to maximize the carried traffic while meeting a specific QoS objective. The resources of the system may be identified with the available capacity to accommodate LSPs on a dedicated wavelength channel reserved on each link along the path for as long as the LSP lasts. We suppose that: i) the LSPs belong to two different classes: low-priority (LP-) and high-priority (HP-) LSPs; ii) HP-LSPs enjoy a higher precedence over LP-LSPs, hence if a HP-LSP needs to be established and there are not enough resources, some LP-LSPs may be terminated to attend the demands of HP-LSPs. The proposed AC scheme is an adaptive extension to multiple LSRs of a well-known policy called Fractional Guard Channel (FGC). It enforces a strict upper bound to the QoS objective, which is defined in terms of the termination probability (i.e. the probability that an already established LP-LSP experiences a termination when a HP-LSP set up request arrives to the system). The new scheme is adaptive in the sense that if the offered load (HP-LSP plus LP-LSP set up requests) or the number of resource units changes, or both simultaneously, the AC system will act in response trying to meet the QoS objective for as long as possible.

1.1.38 Advances in Y2

In Y2 we conducted different simulations of the scenario described above. The performance evaluation of the adaptive scheme shows that: i) it is self-adaptive and does not require any configuration parameters beyond the termination probability objectives; ii) It can operate with any arrival process and any distribution of the interarrival times between LSP set up request or LSP duration; iii) it allows the configuration of the QoS objectives in terms of bounds for the termination probabilities. Besides, the operator enforces a prioritization order that guarantees that during overload episodes HP-LSPs will be able to meet their QoS objective, sometimes at the expense of the QoS experienced by LP-LSPs; iv) the resource utilization is close to the one obtained by an optimum FGC policy, which in turn has a performance very close to the optimal policy. By an optimum FGC policy we mean a static policy that is designed for each arrival rate and knowing all system parameters (e.g. the distribution of the LSPs duration, etc.). In practice, real-time estimation of system parameters is a challenging task. Besides, the precision with which system parameters are determined have a major impact on the performance of the FGC policy.3 The proposed scheme does not require any configuration parameter nor does it require knowing any system parameter.

1.1.39 List of relevant publications

1. Adaptive Admission Control of LSPs in GMPLS Networks. David Garcia-Roger (UPV), Ricardo Romeral (UC3M) and David Larrabeiti (UC3M). Working paper.