Research on Optical Core Networks in the e-Photon/ONe Network of Excellence

F. Callegati, J. Aracil, L. Wosinska, N. Andriolli, D. Careglio, A. Giorgetti, J. Fdez-Palacios,
C. Gauger, M. Klinkowski, O. Gonzáles de Dios, G. Hu, E. Karasan, F. Matera, H. Overby,
C. Raffaelli, L. Rea, N. Sengezer, M. Tornatore, K. Vlachos

Abstract— **This papers reports the advances in Optical Core networks research coordinated in the framework of the e-Photon/ONe and e-Photon/ONe+ networks of excellence.**

I. Introduction

e-Photon/ONe (E1) is a network of excellence funded by the EU in 2004 and 2005, with the goal to promote the integration of the several research activities ongoing in Europe on optical networking. The funding period has been recently extended for two more years in the e-Photon/ONe+ (E1+) network, stemming from the previous experience. One of the instruments implemented in E1 and E1+ are the Virtual Departments (VDs), working groups that gather researchers around big topics to exchange results, promote joint research and identify new topics. VD 1 (VD-C in E1+) is devoted to core optical networks and this paper reports an overview of the results achieved, with particular focus on the joint research activities.

It was decided to organize the broad topic of optical core networks in thematic areas, according to a cross-reference approach, based on two perspectives: switching paradigms and related technologies, and high-level network functions. This was motivated by the consideration that different networking technologies will very likely co-exist and co-operate, following the same architectural principle of the Internet.

This classification brings to a matrix of thematic areas. Research has been oriented accordingly and joint research activities were started to cover the topics not addressed by single partners or that could leverage synergies. The main idea behind the cross reference matrix was to make as evident as possible the re-usability of research. For instance a partner with expertise on QoS for OPS networks could be invited to exploit such expertise also for OBS or OCS. This approach did work in general, even though little integration was achieved between research on OCS and OBS/OPS, mainly because of the rather different issues relevant to these scenarios.

II. Optical Circuit Switching

In optical circuit switching traffic is aggregated to create lightpaths, very high capacity optical circuits that require careful optimization of network resources and an efficient fault management. These issues have been extensively studied in the literature. In the following are outlined the main streamlines of research and some related results.

A. Shared path protection in multi-class optical networks

Protection in multi-class optical networks has been investigated assuming high-class traffic exploits shared path protection and low-class traffic resorts to best effort dynamic restoration. The impact on the network performance of the selective reutilization (either before or after failure occurrence) of the idle capacity allocated for high-class traffic protection has been analyzed, proposing and evaluating two Idle capacity Reuse (IR) schemes. In the Provisioning-phase Idle Reuse (P-IR) scheme, the capacity allocated for high-class traffic protection is utilized under normal (i.e., failure free) working conditions to carry low-class traffic. In the Restoration-phase Idle Reuse (R-IR) scheme, enforced after a link failure occurrence, the protection capacity unutilized in the specific failure scenario is used to facilitate the low-class traffic recovery. P-IR allows to significantly increase the amount of accepted traffic. However the better performance achieved in the provisioning phase is counterbalanced by a worse performance upon failure occurrence, because part of the low-class traffic is preempted to recover the high-class traffic, as shown in Figure 1.

B. Availability-Constrained Optical Circuits in the Presence of Optical Node Failures

The impact of optical node failures was investigated considering a WDM network scenario where circuit redundancy was

F. Callegati and C. Raffaelli are with D.E.I.S. Unversity of Bologna, Viale Risorgimento 2, 40123 Bologna (Italy)
J. Aracil is with Escuela Politecnica Superior, Universidad Autonoma de Madrid, Campus de Cantoblanco, 28049 Madrid, (Spain)
L. Wosinska is with is with the Royal Institute of Technology (KTH), School of Information and Communication Technology (ICT), Isafjordsgatan 24, 164 40 Kista, (Sweden)
N. Andriolli and A. Giorgetti are with Scuola Superiore Sant'Anna, Via Moruzzi 1, 56124 Pisa, (Italy)
D. Careglio and M. Klinkowski are with Univeristat Politcnica de Catalunya, Jordi Girona 1-3, 08034 Barcelona, (Spain)
J. Fdez-Palacios and O. González de Dios are with Telefónica Investigación y Desarrollo, C/ Emilio Vargas N 6 28043 Madrid (Spain)
C. Gauger and G. Hu are with University of Stuttgart, IKR, Pfaffenwaldring 47, 70569, Stuttgart (Germany)
E. Karasan and N. Sengezer are with Bilkent Universitesi, Elektrik Elektronik Muhendisligi, PK:06800 Ankara (Turkey)
F. Matera and L. Rea are with Fondazione Ugo Bordoni, via B. Castiglione 59, 00142 Roma (Italy)
H. Overby is with Norwegian University of Science and Technology, O. S. Bragstadsplass 2B (Norway)
M. Tornatore is with Dipartimento di Elettronica e Informazione, Politecnico di Milano, Via Ponzio, 34/5 20133 Milano, (Italy)
K. Vlachos is with Computer Engineering and Informatics Dept. University of Patras, GR26500, Rio, (Greece).

achieved by means of Shared Path Protection (SPP) switching, in combination with Differentiated Reliability (DiR) in multiple failure scenario. Optical Cross-connect (OXC) equipment reliability was estimated using proven component level reliability models. A selection of representative OXC architectures and optical switching technologies was examined to assess the influence of the node equipment choice on the overall network performance. The results show that in many cases assumption of negligible not failures is not acceptable, suggesting that the OXC architecture has to be driven by reliability requirements, in addition to other conventional metrics, e.g., cost, scalability, etc.

C. Routing issues in transparent optical networks

Research is ongoing about the design of a control plane to encompass physical impairments either assuming a centralized or distributed scenario. GMPLS extensions were proposed for the dynamic estimation of the optical signal quality and connection reliability during lightpath set up. The most relevant physical impairments that need to be considered for Routing and Wavelength Assignment (RWA) in transparent WDM networks have been defined. On-line, non-intrusive monitoring strategies for providing physical-layer information to the GMPLS (Generalized Multiprotocol Label Switching) control plane were designed and evaluated experimentally. This topic is considered of significant importance to be taken into account in transparent network domains and is subject of a joint research activity, stemming from a thematic workshop organized at KTH in Sweden.

D. On Temporary Inconsistency of the Link State Database

A fundamental issue in service guaranteed WDM networks is the network state information update. Connection admission decisions based on outdated or inconsistent network state information may lead to severe performance degradation. Most frequently, interior gateway routing protocols are in charge of network state information dissemination, therefore their efficiency determines network performance, as well. A theoretical model was developed to study the probability that wrong decisions are made when the network state is continuously changing due to parallel connection admission and teardown actions and equipment or cable failures. The study focused on the performance of OSPF (Open Shortest Path First).

E. Flow Based Traffic Engineering

Traffic engineering (TE) for WDM was investigated by defining a flow based traffic model to represent the changes in traffic demands over time. The average hourly traffic demand between each node pair was calculated based on the time zone difference between the nodes. A specific traffic engineering strategy, called Dynamic Cost TE, was defined that is based on routing the LSPs along the shortest paths when the network is lightly loaded and using the paths with most available capacity when the traffic load increases. The Dynamic Cost TE strategy was compared with Best Paths TE. In the Best Paths TE strategy, the path for each LSP is fixed during each hourly period, and the paths are updated for all node pairs at the beginning of each hour. The best paths are calculated using an heuristic based on tabu search. In Figure 2, the bandwidth blocking probability for these two TE strategies and the shortest path routing are compared for different values of the relative traffic uncertainty, defined as the ratio of the variance of the Gaussian distributed traffic uncertainty and average offered traffic.

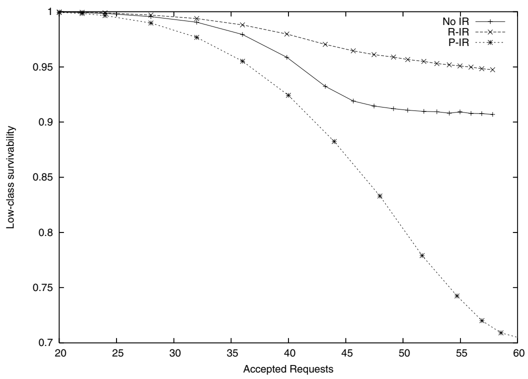

Fig. 1. Low-class survivability in a multi-service wavelength-routed optical network.

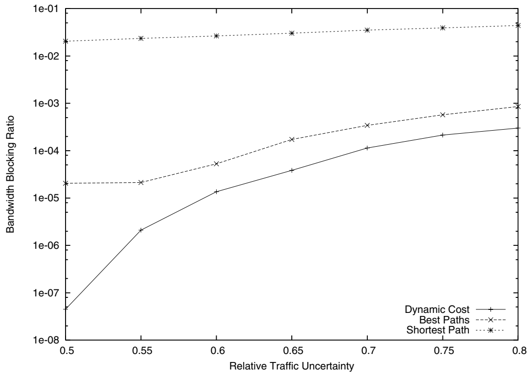

Fig. 2. Bandwidth Blocking Ratio as a function of the Relative Traffic Uncertainty for three routing strategies applied on a 10-node topology

F. Exploiting the knowledge of lightpaths duration

In some cases the holding time of connection requests can be known in advance based on SLAs or contracts with customers. In this case the knowledge of the holding-time can be used to improve backup resource sharing, for instance with shared-path protection (SPP). We propose to minimize both the additional capacity and the cost of additional wavelengths multiplied by the estimated time it has been provisioned. In our approach K minimal-cost paths are computed and then, for each of these K paths, the backup is determined by exploiting a holding-time aware link-cost-assignment. A figure of merit for comparing backup resource efficiency is the resource overbuild (RO): the lower RO the better backup sharing. Our results presented in Figure 3 show the benefit of such approach in terms of level of RO, compared to the traditional one ($K = 1$).

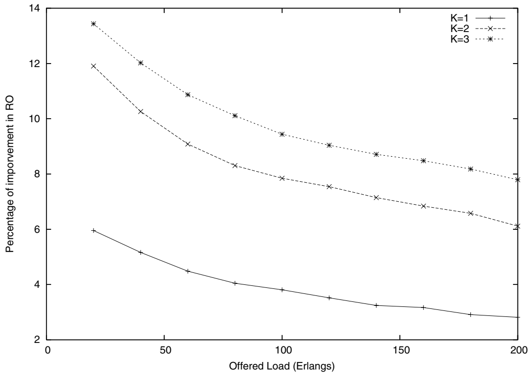

Fig. 3. RO as a function of calls arrival rate.

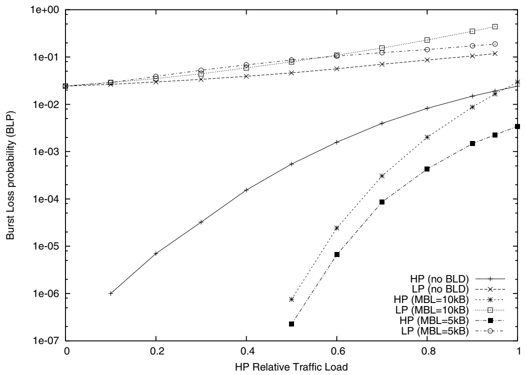

Fig. 4. Burst Loss Probabilities for HP and LP class as a function of HP traffic load.

III. Optical Burst Switching

OBS is the topics to which the participants to VD1 devoted most effort. A specific task force, the OBS task force, was set up with the aim to put together already available results, create a common reference basis and start joint research activities. The research topics addressed cover several aspect of the engineering and design of an OBS network.

A. The Burst Length Differentiation technique for QoS provisioning

In OBS architectures, it can be observed that shorter bursts have much more chances to access wavelengths and to fill gaps between bursts already scheduled than longer bursts. Taking into account these arguments it may be advantageous to assign different burst lengths to different classes in order to give better performance to some of them. In particular, shorter burst units could carry high priority (HP) loss/delay-sensitive traffic, while low priority (LP) traffic would be aggregated into longer bursts. Such idea brings us to the proposal of the Burst Length Differentiation (BLD) technique. It assumes that each traffic class uses mixed time/burst-length assembly algorithm to build bursts in edge nodes; HP bursts are aggregated with lower timer and maximum burst length thresholds than LP bursts.

The BLD technique was used combined with other selected *relative* QoS mechanisms in order to boost the performance characteristics of HP traffic. An example of the burst loss probability is plotted in Figure 4, where both burst inter-arrival times and mean burst length (MBL) are Gaussian distributed. When BLD is applied, we assume that MBL$_{LP}$ is equal to 40 kbytes while MBL$_{HP}$ is either 5 or 10 kbytes. The figure shows that a very good differentiation can be achieved between HP and LP, with a significant improvement when adding BLD.

B. Edge node design

The delay experience by data in a generic edge node is mainly composed by the burst assembly delay and the burst queueing/scheduling delay. While the assembly delay can be bounded by a timer in the time-based burst assembly scheme, the queueing delay in the transmission buffer needs to be carefully inspected.

The probability distribution of the delay experienced by the individual IP packets of a test IP flow was studied, applying FIFO scheduling and estimating the total delay as the time interval between an IP packet arrival at an assembly queue and the time the burst containing this packet is delivered to the optical channel. It was found that the burst queueing delay is about one magnitude smaller than the total delay, therefore the queueing delay is actually minor in comparison to the assembly delay, even at high system loads.

C. Performance of TCP over OBS

The TCP throughput has been analytically expressed as a function of the burst loss probability and of the structure leading to remarkable insights of the influence of OBS network and nodes design on end-to-end performance. The reference scenario evaluated assumes an access bandwidth B_a = 100 Mbit/s, core bandwidth B_o= 2.5 Gbit/s, RTT=120ms (20 ms of delay in access networks, 20 ms of delay in core). The TCP maximum segment size is 512 bytes, and all other protocol related quantities are set to typical values. A 8×8 switch equipped with wc shared-per-node wavelength converters was considered as transit node in the OBS network section. Figure 5 shows the TCP send rate as a function of the offered load per input wavelength, varying the number N of wavelengths per fiber as a parameter. A pure timer-based burst assembly strategy is used with 3 ms of burstification time. The TCP throughput drops from a rather good level to almost zero by increasing the load. The dropping edge shifts to higher loads as N increases up to a maximum value ($N = 32$ in this case), after which the increased blocking due to the lack of wavelength converters in the nodes enters into play. The figure outlines the strong impact of wc on the TCP send rate and that an increase of N not necessarily improves the throughput but, in some cases may reduce it. Therefore, the core node design should be considered to optimize the overall network network.

D. OBS Signaling

Within the framework of VD1, the EBRP protocol was developed. It is a two-way scheme, but, unlike typical Tell-and-Wait schemes, reserves resources only for a given duration

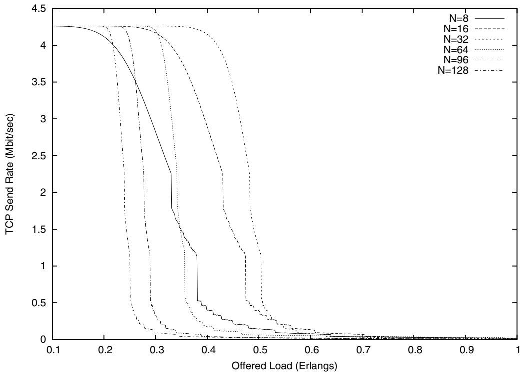

Fig. 5. TCP send rate as a function of the offered load per wavelength, varying the number of wavelengths per link as a parameter, $wc = 64$

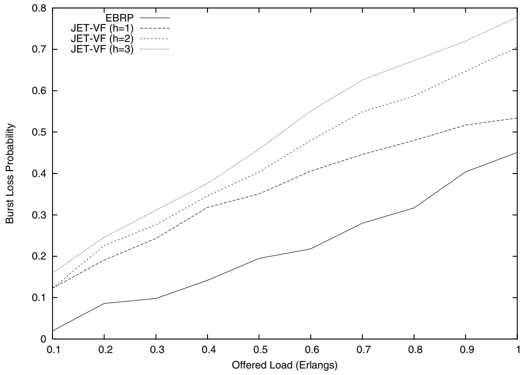

Fig. 6. Burst loss probability of EBRP versus Just-Enough-Time with void filling (JET-VF) for bursts traversing through h=1, 2 or 3 nodes, with $k = 1.5$, $m = 1$ and $n = 0.5$.

(delayed reservation) similarly to the one-way schemes. However, delayed reservations are "relaxed", by the introduction of a generalized reservation function $RD(T_{\text{data}}, h) = kT_{\text{data}}^{m}h^{n}$ where k, m and n are positive reals, T_{data} is the duration of the burst and h is the number of the remaining hops across the network path. Using this reservation function, capacity requested by the bursts may exceed their actual holding time and it can be used to increase the successful forwarding probability and also to provide service differentiation by assigning different reservation duration parameters to different classes of traffic.

The performance of the proposed scheme has been assessed against other typical one-way and two way schemes. Figure 6 compares the performance of the EBRP scheme with the just-enough-time protocol with void filling (JET-VF), for bursts traversing through $h = 1, 2$ and 3 nodes before they reach their destination. For these parameters, the yielding RD reservation function was found to equalize blocking for all bursts, independent of their destination. The simulation results were carried out on the NSFnet backbone topology for burst sizes with an exponential distribution and 5MB mean size.

E. OBS routing

In OBS networks, a shortest path approach is usually adopted, and the network state is not considered at all. Therefore, it is proposed and analyzed a routing strategy that includes load estimation. Specifically, the MRDV (Multipath Routing with dynamic variance), originally designed for IP networks, was proposed to be used in an OBS network. The aim of the strategy is to balance the load of the network, moving traffic from the most loaded links to the spare ones. Hence, it is expected that the overall blocking probability will be reduced. To decide the amount of traffic to be moved , two metrics, load and blocking probability, were studied. It was found that using a blocking probability policy leads to an excessive aggressive behavior and has more difficulties to reach to a stable state. Hence, the load policy shows a better behavior.

IV. Optical Packet Switching

Most of the activities on node design for OPS networks were part of a different project workpackage. The activities on OPS in VD1 are focused on the crucial issue of congestion resolution, that is more a matter of smart and efficient scheduling in optics where queuing can not be implemented. Both the space (deflection routing), the wavelength (wavelength conversion) and the time (delay buffering) domains can be exploited to reduce congestion and to manage QoS.

A. Congestion Resolution Preserving Packet Sequence

Congestion resolution in the wavelength and time domain was addressed by analyzing the so called Wavelength and Delay Selection (WDS) scheduling problem, i.e. using simultaneously both the time domain (with delay buffers) and the wavelength domain (with load balancing on the output wavelengths), assuming full wavelength conversion. A specific joint activity had the aim to design scheduling algorithms that can preserve the packet flows in an MPLS connection oriented scenario. The scheduling algorithm still exploits the time and the wavelength domain, but safeguard the time relationship between packets belonging to the same flow.

The results provided show that the packet sequence can be maintain at the price of a limited degradation of the packet loss probability (PLP). For instance in 4 × 4 switch with 16 wavelengths per fiber, where each input wavelength carries 3 different LSPs for a total of 192 incoming LSPs, a WDS preserving the packet sequence results in a PLP of about 10^{-4} while the best performing WDS algorithm results in a PLP of about 10^{-5} with 5% of out of sequence packets. This work has placed the basis for a better understanding of the influence of an OPS core on the performance of higher layer protocols such as TCP.

B. QoS Differentiation by Wavelength Partitioning

Resource partitioning when applying the WDS scheduling algorithm can be exploited to support QoS management. In Figure 7 an example of application of partitioning of the access to the wavelengths in a fiber is presented. The focus is on QoS differentiation, in terms of packet loss probability, that can be achieved over a wide range by changing the number of wavelength reserved for HP traffic.

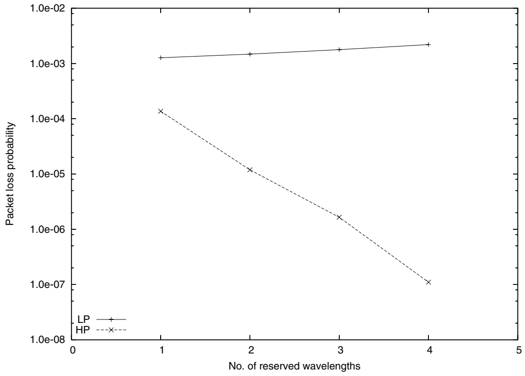

Fig. 7. QoS differentiation between HP and LP traffic in a 5 node OPS network as a function of the number of wavelengths reserved for HP traffic.

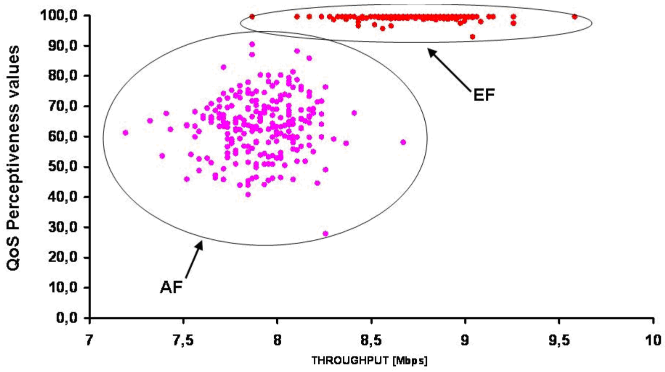

Fig. 8. QoS perceptiveness values vs throughput for EF and AF CoS, in case of a traffic load equal to 900 Mb/s in case of a tennis match video 30 second long.

C. 1+1 Path Protection Combined with Shared Packet Redundancy

Reliability in OPS networks was analyzed in combination with performance in an OPS network employing 1+1 path protection and shared packet redundancy (SPR). In 1+1 path protection, packets are duplicated on two node and link disjoint paths, thus enabling protection against both node and link failure on either paths. With SPR, a number of redundancy packets are added to a set of data packets, which enables a possible reconstruction of data packets lost due to contention.

Several scenarios have been considered, in a simple case with two nodes and link disjoint paths between an ingress and egress node in the OPS network, depending on whether SPR is used or not, and whether one of the paths have failed or not. When SPR is used, redundancy packets are added to data packets for each path separately. Quantitative evaluation of these scenarios is currently ongoing.

V. Experimental test-bed

A high speed core network testbed was implemented in Rome by Fondazione U. Bordoni in cooperation with ISCOM. The goal of the test-bed is to investigate QoS management issue in a multiservice multi-access IP network where QoS is managed by means of DiffServ over MPLS. The tests reported here refer to the interworking between domain and layers (different accesses and core routers) with relevance to the correlation between objective (network) and subjective (perceptiveness) measurements, also with lambda switching approach. The network test bed is based on 3 core routers, one Ethernet traffic generator-analyzer and one multimedia server. The routers are connected to an optical ring-like structure, based on the fibers contained inside a deployed cable between Rome and Pomezia (25 Km). The PCs were connected to the Routers by means of UTP copper cables. The traffic generator-analyzer, with two GbE optical interfaces and eight FastEthernet interfaces, has been used to overload the optical link under test. To obtain a correlation between the perceived quality (from the Human point of view) and the measurement of the network performance, it is necessary to carry out a subjective assessment test, that in our case was implemented according to the ITU-R recommendation BT 500-11.

Movie test sequences were shown to the viewers, by using the three different DiffServ labeling, i.e.: Expedited Forwarding (EF), Assured Forwarding (AF) and Best Effort (BE). To simulate a condition of network overload the traffic generator-analyzer transmits 800 Mbps BE traffic, 80 Mbps of AF traffic, 32 Mbps of EF traffic. As a an example of QoS tests we report the QoS perceptiveness values for a video test (tennis match 30 second long), after an averaged made on the 16 viewers, versus the corresponding network performance in terms of throughput. As shown by figure 8, EF always provided very good network performance and the corresponding perceptiveness was excellent . Conversely in the case of AF the network performance were worse, with a reduction of the throughput (and a small increase of the jitter not reported in the figure) and a consequent QoS degradation.

VI. Conclusion

In this document the main activities on optical core networks developed in the framework of the network of excellence e-Photon/ONe have been summarized. The amount of work and the variety of topics considered lead to the conclusion that research in this field is still very open and lively.

Acknowledgment

This work was funded by the Commission of the European Union, contract no. FP6-001933 e-Photon/ONe and FP6-IST-027497 e-Photon/ONe+. Thanks to all participants to Virtual Department 1.

References

[1] General information, public reports and published paper listing can be found on the e-Photon/ONe and e-Photon/ONe+ web site `http://www.e-photon-one.org`.