

Research in Optical Burst Switching within the e-Photon/ONe Network of Excellence¹

J. Aracil, N. Akar, S. Bjornstad, M. Casoni
K. Christodoulopoulos, D. Careglio, J. Fdez-Palacios, C. Gauger,
O. González de Dios, G. Hu, E. Karasan, M. Klinkowski, D. Morató,
R. Nejabati, H. Overby, C. Raffaelli, D. Simeonidou,
N. Stol, G. Tosi-Beleffi and K. Vlachos

The e-Photon/ONe Network of Excellence

Abstract

This paper presents a summary of Optical Burst Switching (OBS) research within the VI framework program *e-Photon/ONe* network of excellence. The paper includes network aspects such as routing techniques, resilience and contention resolution, together with burst switch architectures. On the other hand, we also discuss traffic analysis issues, Quality of Service (QoS) schemes, TCP/IP over OBS and physical layer aspects for OBS.

1 Introduction

This paper presents the current research agenda in Optical Burst Switching (OBS) from the e-Photon/ONe partners within Virtual Department one. E-Photon/ONe is a European Union VI framework program network of excellence that aims at gathering the European expertise in optical networks. The network of excellence comprises 38 partners and it is organized around 5 Virtual Departments (VDs). Such VDs resemble academic departments, but they are "virtual" because members are based in different locations throughout Europe.

In what follows, we will assume that the reader is familiar with OBS [1] and purposely skip introductory material. The paper is structured as follows, section 2 is devoted to signaling and scheduling and section 3 is devoted to

¹ This work was partly funded by the *e-Photon/ONe* Network of Excellence

routing, followed by contention resolution in section 4. Traffic analysis issues are discussed in section 5 and TCP over OBS aspects are presented in section 6. Section 7 is devoted to quality of service and section 8 presents resilience issues. Finally, sections 9 and 10 discuss burst switch architectures and the physical layer for OBS.

Before we present the above mentioned issues let us briefly discuss the strengths and drawbacks of OBS.

There is considerable debate on when and where optical technology will be incorporated to the current networks. Particularly, network operators in *e-Photon/ONe* are devoting a great deal of effort to analyze migration scenarios. Actually, European networks are mainly based on traditional SDH topologies, which were originally built to carry voice traffic. While metropolitan networks are based on SDH rings interconnected with digital-cross-connect systems (DXCs), transport in core networks is provided by point-to-point and ring DWDM systems, usually with a more meshed topology.

In the short term, SDH technology is expected to be gradually migrated to Wavelength Switching (WS). As main drivers for this evolution, we can mention the following: technological availability (appearance of the first ROADMs -Reconfigurable Optical Add-Drop Multiplexers- and OXCs -Optical Cross Connects-), CAPEX -Capital Expenditure- and OPEX -Operations Expenditure- reduction, mainly due to automation and transparency, and increase of revenue coming from new services such as Optical VPNs (Virtual Private Networks). According to this, a feasible trend could be the evolution towards metro aggregation rings based on ROADMs which are connected through a core mesh composed of OXCs with full Generalized Multiprotocol Label Switching (GMPLS) support.

As mentioned above, technological availability makes Wavelength-Switched (WS) networks an upcoming solution for the short term. However, there are a number of factors in the WS scenario which need to be further studied and improved when considering long term scenarios with dramatically increased traffic demands, higher flexibility and granularity requirements and end-to-end QoS demands, in particular for packet-based network services.

In this respect, OBS provides bandwidth granularity between WS and Optical Packet Switching (OPS), and relatively constrained technological requirements (especially for the switching elements). A natural, simple and low cost evolution from WS to OBS scenarios may be achieved by gradually updating the ROADMs and OXCs previously used in the WS scenario in order to support optical burst transmission. Therefore, in a first step, OBS networks may have similar topologies than WS (i.e metro access rings and core meshes). Figure 1 presents the envisioned evolution scenarios for current SDH rings. In

transparent nodes the signal is optically switched while in opaque nodes it is electronically converted.

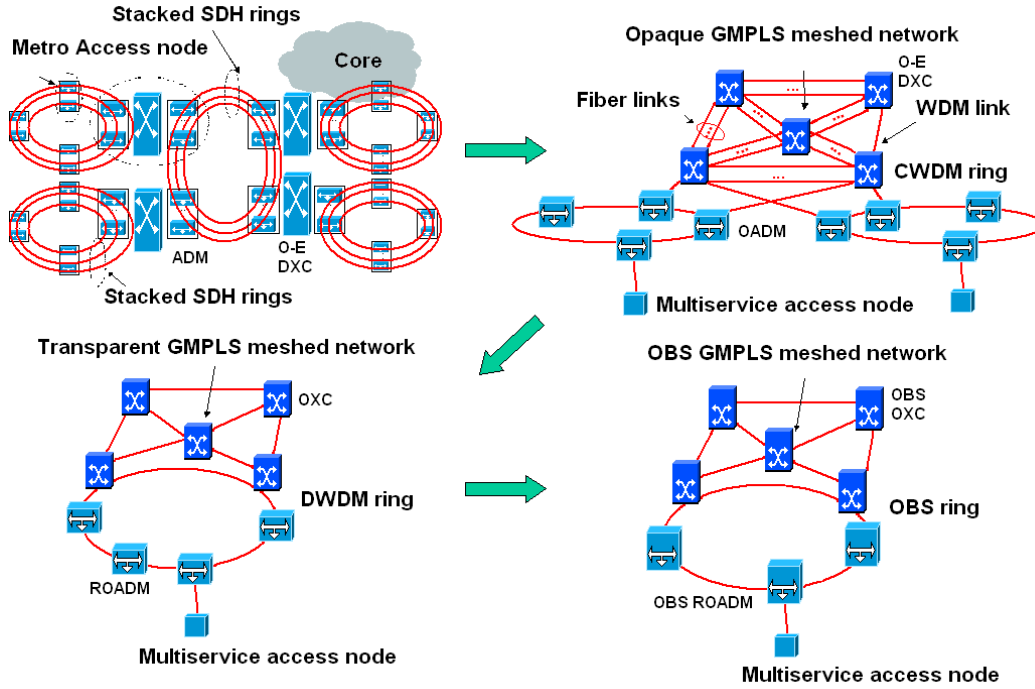


Fig. 1. Network evolution

2 Signaling and scheduling

In OBS networks, the scheduling mechanism is in charge of accommodating resources for the incoming burst. The signaling schemes associated to the scheduling mechanism can be broadly grouped into tell-and-go and tell-and-wait mechanisms. To evaluate the performance of the various scheduling mechanisms found in OBS literature for bufferless networks we developed a discrete-event simulation based on the ns-2 platform. We have simulated a typical tell-and-go mechanism (i) just-enough-time (JET) with LAUC-VF (*Latest Available Unscheduled Channel with Void Filling*) [2] void filling algorithm (ii) tell-and-wait (TAW) and (iii) Efficient Burst Reservation Protocol (EBRP) [3]. EBRP protocol is a novel two-way signaling protocol for efficient burst-level reservations and QoS differentiation in OBS networks. EBRP employs timed and in-advance reservation of resources. However, a key feature in the EBRP protocol is that timed reservations are relaxed, introducing a reservation time duration parameter that is negotiated during call setup. This feature allows bursts to reserve resources beyond their actual size and thus increase their successful forwarding probability. This feature can be used for QoS differentiation.

The simulation was conducted on the NSFNET backbone network topology, where all links are assumed to be bi-directional, propagation delay is proportional to fiber length, processing delay of each SETUP/ACK messages is set to $10\mu\text{sec}$ and the link bandwidth C_{core} is equal to 40 Gb/s. For the simulations conducted, a dynamic traffic model was assumed where burst transmission requests arrive at each edge node following a Poisson distribution with mean λ (arrivals per sec). Burst destinations are evenly distributed, while it is assumed that burst sizes follow a negative exponential distribution with mean value BS (in Mbytes) that corresponds to transmission time $\frac{1}{\mu} = \frac{8*BS}{C_{core}}$. Typical mean burst transmission durations were considered 2-4 ms. (10MB-20MB corresponding BS) that are one order of magnitude less than the mean round trip time of the NSFNET ($T_{RTT-mean} = 26$ msec).

We used burst blocking probability as the main metric for assessing protocol performance. In the case of JET protocol, blocking probability refers to common output contention in the network core nodes. On the other hand, for the two-way reservation schemes, if the connection establishment process is blocked and burst delay time (BD) allows, the corresponding burst manager re-transmits the SETUP until a successful reservation is made or until expiration of the burst delay time. The bursts are kept at the Edge Router (ER) until a successful reservation, thus blocking probability refers to those bursts lost due to ER buffer overflow or due to burst time delay expiration and consequently the core network is free of blocking. An additional performance metric that was used for the 2-way reservation schemes is the average holding time of the bursts at the network ER.

For the experiments of figures 2(a) and 2(b) we have assumed that the network is a single wavelength network, therefore a burst requests and reserves the full bandwidth of the fiber link C_{core} . From these graphs we can observe that although JET can provide a minimum holding time, it is not efficient for single wavelength channel networks due to its high blocking probability. EBRP was proved to outperform the other protocols as its blocking probability is negligible and its holding time is close to $T_{RTT-mean}$ which is the minimum penalty for 2-way schemes ($\lambda=70$ and BS=10Mbytes).

Figure 2(c) shows JET protocol performance for identical traffic characteristics, assuming that the network has full wavelength conversion capability, i.e. each node is able to convert any wavelength on an input port to any other wavelength on an output port. Provided that a fiber has w wavelengths, a burst requests and reserves the full bandwidth of one wavelength. As expected, JET protocol performance improves as w increases. By comparing Figure 2(a) and 2(c) we can note that JET blocking performance reaches the levels of EBRP for single wavelength channel when $w=4$. As stated above, in JET protocol, bursts are transmitted almost immediately (after the related T_{offset}) and thus no measurements for holding time at network edge were obtained.

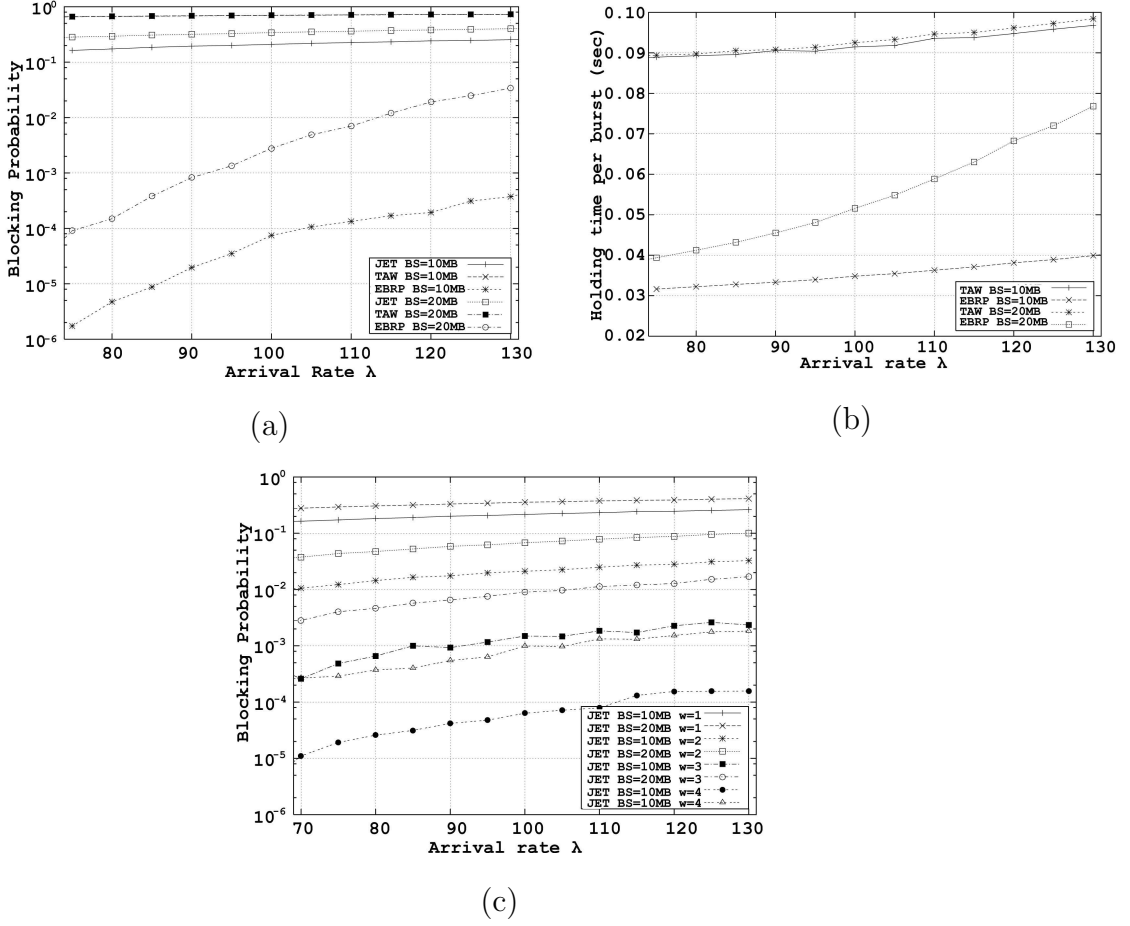


Fig. 2. (a) Blocking probability and (b) holding time of burst at network edge for JET, TAW, EBRP in a single channel network. (c) JET blocking performance for different number of wavelengths per fiber, assuming full wavelength conversion switches

These results show that OBS signaling schemes have evolved to provide solutions to a variety of OBS network scenarios. The network topology and the supported services determine the reservation strategy that should be employed, i.e. whether the network relies on high speed Optical TDM (OTDM) with increased bit rate single channel, or uses WDM with enhanced wavelength conversion capabilities.

3 Routing

In OBS networks, both the offset time between burst and Burst Control Packet (BCP) and the burst dropping probability due to contention are affected by the routing protocol. There are two routing paradigms to be considered: source routing and hop-by-hop routing. In source routing, the route is calculated at

the edge node, and has the advantage of knowing the path length in advance, so the optimal value of the offset time can be calculated. On the other hand, in hop-by-hop routing the output link is chosen at every intermediate node and has the disadvantage of not knowing the length of the path a priori, so the offset must be estimated. If the control packet is routed through a path longer than expected, then the burst will be lost. The main advantage is the possibility to employ routing protocols already deployed in traditional IP networks, like OSPF or IS-IS, whose metrics must be adapted to be used in OBS.

Contention in the network can be reduced by adequately balancing the traffic load. For example, two novel routing strategies to reduce contention are proposed in [4]: a congestion-based static route calculation technique, and a least-congested dynamic route calculation technique. Within e-Photon/ONE, we have developed an OBS routing strategy that is based on MRDV (Multipath Routing with Dynamic Variance) [5], which was initially designed for IP networks. 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 [6].

4 Contention resolution

Since OBS networks with tell-and-go signaling scheme (one-pass reservation) do not assure a successful channel reservation in each switching node, contentions have to be effectively resolved to achieve low burst loss probabilities.

4.1 Principal contention resolution mechanisms

Almost all work on OBS assumes contention resolution by *full* wavelength conversion, i.e., a dedicated wavelength converter is provided for each input or output wavelength. For a low to medium load, this provides a low burst loss probability because all wavelength channels of an output fiber can be shared among all bursts directed towards this output fiber [7–12]. For a high load, the number of wavelength channels has to be very large to reach burst loss probabilities in the order of 10^{-6} or less, e.g., 350 wavelength channels are needed to carry a load of 0.8 Erlang per wavelength channel at this loss rate.

Wavelength conversion has also been complemented by providing a number of Fiber Delay Lines (FDLs) in a FDL buffer. It has been shown that even FDL buffers with rather simple functionality and low technological requirements can improve OBS performance significantly [8,13,2,14].

The feasible length of the FDLs in OBS node is limited by several physical constraints like attenuation, chromatic dispersion and non-linear effects etc. Even if the FDLs are dispersion compensated, the maximum length of a FDL is limited by the power budget. In order to limit the implementation cost as well as the additional amplifier noise, it can be assumed that in maximum a single erbium doped fiber amplifier (EDFA) is used per FDL. Consequently, all FDLs used for contention resolution have to be shorter than a typical EDFA span of 80 km which limits the maximum FDL delay to $266\mu s$ ($c = 2 \cdot 10^8 m/s$). Previous work showed that FDL delays should be in the order of a few mean burst durations in order to be efficient [13,8]. In a FDL buffer with 4 FDLs, e.g., a delay of 8 mean burst durations for the longest FDL is a good choice. From the $266\mu s$ it can be derived that in order for this FDL buffer to be feasible the mean burst length has to be in the order of 10 kbyte, 40 kbyte and 160 kbyte for 2.5, 10 and 40 Gbps respectively. Such performance and technology arguments should be taken into account in the determination of the burst size (e.g. in the burst assembly control in the edge node).

Different architectures for FDL buffers were proposed, which can be categorized into feed-forward (FF) and feedback (FB) architectures (see [15]) as well as into single-stage and multi-stage structures [15]. Reservation strategy is another degree of freedom in the design. The FDL buffer and output channel can be reserved immediately when the reservation request of a burst is blocked for the first time [8,2]. Correspondingly, the burst offset time is increased by the FDL delay period. This is called *PreRes* scheme as the output is reserved prior to the burst entering the buffer. An alternative strategy is *PostRes*, in which case an output reservation is requested after the burst has entered the FDL but before it leaves the buffer [13]. In *PostRes*, all blocked bursts are buffered if buffer space is available. It is shown in [13] that in general *PreRes* leads to a lower burst loss probability than *PostRes* at medium and high loads. However, due to the change of the offset time, *PreRes* can cause a disturbance in the offset-based QoS in certain circumstance, which needs special cares in the parameter setting of the offset time and the FDL length.

Deflection routing has been analyzed in the context of OBS for irregular mesh networks [16–18]. In general, the path a deflected burst takes through the network should be kept as short as possible in order to minimize resource consumption. In OBS schemes which apply offset times, the problem of *insufficient offset times* has to be avoided, i.e., it has to be ensured that there is always a large enough offset between control packet and data burst even if extra nodes are traversed. Thus, it is proposed to use FDL buffers to in-

crease offset times in intermediate nodes prior to deflection [16]. Investigations showed that deflection routing results in only limited improvement in contrast to significant improvement for wavelength conversion or FDL buffering. Actually, deflection routing does not resolve contention locally in a single node but reroutes over-load traffic to neighboring nodes. Therefore, deflection routing is generally regarded as a supplementary solution to the others.

When the combinations of these basic strategies are to be used, the order in which these schemes are applied is essential. They are named by a concatenation of acronyms of the basic schemes. E. g., *ConvFDLDefl* refers to a scheme which tries conversion first, only if this fails it tries to buffer in a FDL and only if this also fails it tries deflection routing. A performance study in [19] shows that when combining full wavelength conversion and FDL buffers conversion should always be used first. Therefore, the schemes which apply wavelength conversion first amount for the most feasible solutions for contention resolution.

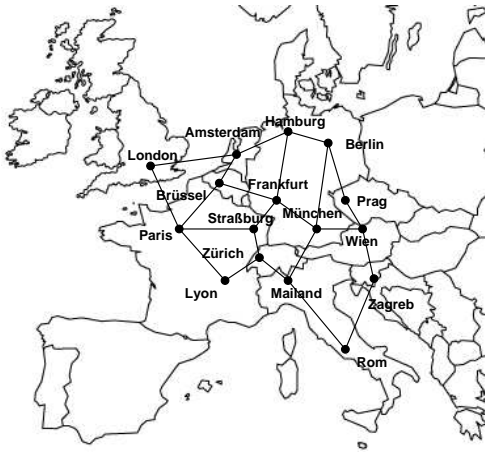


Fig. 3. Pan-European Network Scenario, total traffic=1 Tbps, mean number of λ s/link=5.43

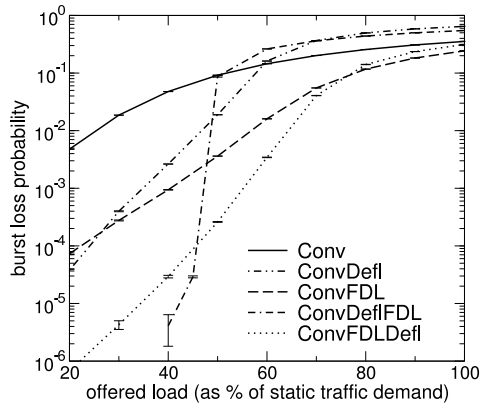


Fig. 4. Results for Pan-European Network

4.2 Comparison of contention resolution mechanisms

In e-Photon/ONE, we have compared different contention resolution schemes with respect to a Pan-European reference network (figure 3). Figure 4 depicts burst loss probability versus relative offered load (c.f.[20] for detailed description of the scenario). It can be seen that for high loads the schemes employing deflection routing after conversion are inefficient as they produce additional load in an already highly loaded network. For medium loads, *ConvDefl* outperforms *Conv* and *ConvDeflFDL* which are all outperformed by *ConvFDL* and *ConvFDLDefl*. For low loads, loss probability of *ConvDeflFDL* drops rapidly as enough network capacity becomes available. *ConvFDL* does not decrease as

fast as all other schemes due to severe contention in few nodes which cannot be resolved locally in the FDL buffer.

In general, the performance of contention resolution schemes is sensitive to both network and buffer dimensioning which should be considered in their analysis. The combination of conversion with FDL buffers yields lower losses than conversion with deflection routing in most cases, however, at the cost of additional buffering. *ConvFDLDefl* yields significant improvements of the burst loss probabilities for all load values without the two-state behavior of *ConvDeflFDL*.

4.3 Contention resolution analysis with partial wavelength conversion capabilities

A cost-conscious alternative to full wavelength conversion, which is costly due to technological limitations, is the use of Partial Wavelength Conversion (PWC) for which there are a number of analytical results in the existing literature. In PWC, there is a limited number of shared converters, and consequently some bursts cannot be switched towards their destination and therefore blocked when all converters are busy despite the availability of free wavelength channels at the destined output port. In PWC, converters may be collected as a single converter pool for converter sharing across all output ports, which is referred to as the Share-Per-Node (SPN) architecture [21]. A simpler architecture allows separate converter pools per output (Share-Per-Link (SPL)) [21]. Recently, a similar architecture is proposed in [22] where a pool of converters is shared for all bursts arriving on the same input link (Share-Per-Input-Line (SPIL)). We have analyzed the case of SPL in [23] by exactly solving the steady-state vector of a continuous-time Markov chain with a block tridiagonal generator using numerically efficient block-tridiagonal LU factorizations. More general arrival processes such as the phase-type arrival processes can be handled using the same framework presented in [23]. Using the analytical model of [23], the burst blocking probability P_b is depicted as a function of r which equals the percentage ratio of the number of converters to the number of channels in a 64-channel optical link with load $\rho = 0.7$ for different values of the squared coefficient of variation, denoted by γ , of the interarrival times. The case of $\gamma = 1$ is for Poisson arrivals; $\gamma < 1$ cases are obtained by using an Erlang- k distribution which has $\gamma = 1/k$ and the case of $\gamma > 1$ is obtained by using an hyper-exponential distribution with two phases, namely the H_2 distribution, and with balanced means. Figure 5 demonstrates that burst blocking probabilities are significantly lower for regularly spaced arrivals (i.e, small γ) for SPL links. It is therefore concluded that the variance, besides the mean, of interarrival times has a substantial impact on burst blocking performance in SPL links and these second order traffic characteristics also

need to be taken into consideration for accurate performance modeling.

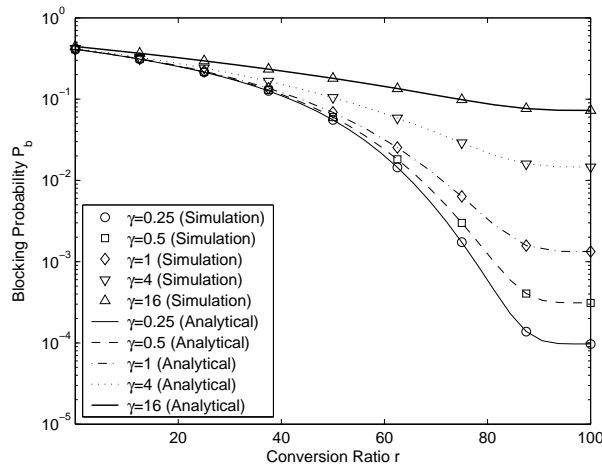


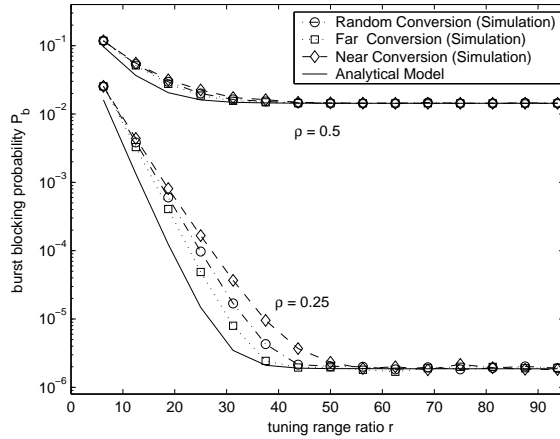
Fig. 5. Burst blocking probabilities as a function of W for different values of γ

4.4 Contention resolution analysis with limited wavelength conversion capabilities

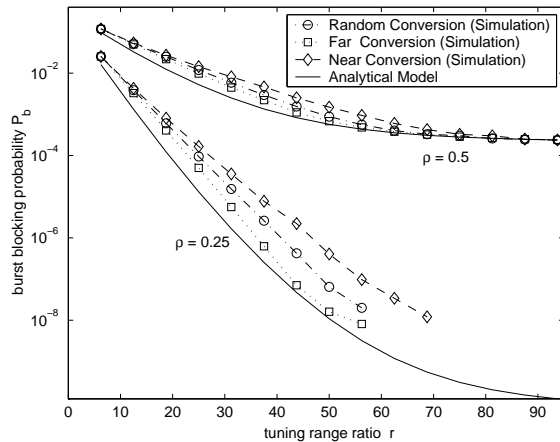
Another issue in wavelength conversion is whether there is a specified range of wavelengths that a given wavelength can be converted to, i.e. tuning range. In limited range wavelength conversion, a burst arriving on a certain wavelength can be converted to a fixed set of wavelengths above and below the original wavelength. The degree of conversion d describes the number of wavelengths in the tuning range and tuning range ratio parameter r is the percentage ratio of d to the number of wavelength channels. We have studied in [24] three wavelength conversion policies for a burst whose wavelength is occupied. A random, or the nearest, or the farthest available wavelength is selected from the tuning range in the so-called random, near, and far conversion policies, respectively. As an example, we use a 32-channel OBS link with W dedicated converters and investigate the performance of the three conversion policies for 4 different pairs of W and ρ values in figure 6. We observe that the “far conversion” policy outperforms the others especially with increased wavelength conversion ratios, moderate tuning range ratios, and at lighter loads. We also provide a modification of the algorithm proposed in [23] as an approximate analytical model and figure 6 demonstrates that the model provides a lower bound for all conversion policies studied.

The analysis of a single OBS link is useful but it comes short of addressing network related issues below:

- capacity wasted in an optical link due to bursts that would be blocked in subsequent hops,



(a)



(b)

Fig. 6. Blocking probabilities for $K = 32$ with respect to the tuning range ratio parameter r for two different values of $\rho = 0.25$ and $\rho = 0.5$ when (a) $W = 15$ and (b) $W = 25$

- the reduced load on an optical link resulting from burst blocking at previous hops.

However, it is generally accepted that exact solutions for blocking probabilities in large OBS networks is not possible and approximate techniques are therefore used. In [25], a reduced-load fixed point approximation model to evaluate blocking probabilities in OBS networks is proposed to cover a variety of OBS policies such as just-enough-time, just-in-time, burst segmentation, and route-dependent priorities. A variant of reduced-load fixed point approximation method is proposed in [26] that addresses OBS networks with PWC-type OBS nodes. Although there is evidence that reduced load fixed point approximations provide acceptable approximations to the burst blocking probabilities for relatively large loads and high degrees of wavelength conversion in mesh networks, work still needs to be done towards performance analysis of more general OBS networks.

5 Traffic analysis for OBS networks

A number of traffic studies [27–29] have shown that IP traffic suffers long-range dependence. It has also been argued that long-range dependence cannot be characterized with a single process. At small time scales, the traffic may even show independent increments, and self-similarity is only observed at medium-large times cales [30]. However, for the specific case of OBS burstifiers, it may happen that only the number of bytes per burstification interval matters and not the packet arrival dynamics. For time-based burstifiers, for instance, only the number of bytes that arrive in a timeout period matter, regardless of the packet arrival process in such interval. Usually, the timeout timescale is large enough to capture traffic long-range dependence. Thus, a fluid-flow model with long-range dependence seems suitable to analyze the burstifier performance. The most common model in this case is the Fractional Gaussian Noise (FGN), which has been shown to model accurately traffic from a LAN [31]. We have successfully used this model to analyze the burstifier performance, and the results are in good agreement with those obtained from real traces of *core* traffic [32]. However, for burst length-based algorithms the packet arrival dynamics do matter. In this case, bursts are constant-size and the burst inter-arrival time is variable, as it is the sum of non-deterministic packet interarrival times.

A simulation experiment was performed with a Fractional Gaussian Noise traffic generator and a timer-based assembly algorithm. This is a self-similar traffic model that accurately portrays traffic from LANs [31]. Figure 7 shows the burst interarrival time survival function (complementary distribution function). While the results show that the burst interarrival time has a hypergeometric distribution, the Poisson approximation is valid for small time scales [32]. It is worth noting that two timeout values were evaluated (2 ms and 4 ms) and that the x-axis shows the normalized burst interarrival time to the timeout value.

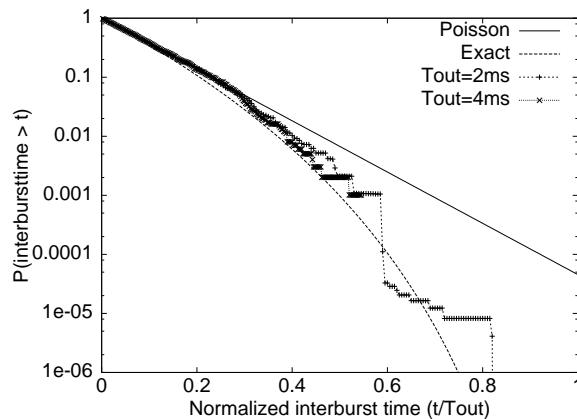


Fig. 7. Burst interarrival time survival function

On the other hand, figure 8 shows the variance-time plot of OBS traffic, where

FEC traffic represents the incoming IP traffic corresponding to the same destination (Forward Equivalence Class). For small time scales the traffic shows independent increments, while self-similarity remains unchanged for large time scales.

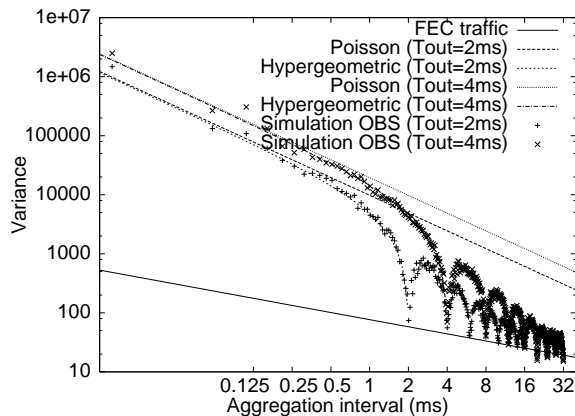


Fig. 8. Variance-time plot

6 TCP issues for a paneuropean OBS network

When OBS is employed in wide area networks some new aspects, discussed below, have to be taken into account for TCP.

Assembly delay. First of all, the burst assembly process adds variable delays to the RTT (round trip time) which affect end-to-end performance. This process must be investigated in relation to the assembly algorithm applied to aggregate TCP segments in optical bursts. The timeout period in time-based or the buffer threshold in burst length-based assembly algorithms can be chosen optimally to achieve high TCP goodput [33],[34],[35]. On the other hand, adaptive burst assembly timeouts can be used to dynamically adjust the burst length and the burstification delay according to the dynamics of the incoming TCP traffic, possibly shaped by TCP congestion control. Such adaptive algorithms can enhance the TCP goodput with respect to static burst assembly algorithms [36].

Burst Losses. Burst losses represent the aspect that mainly influences end to end performance. Burst losses are a consequence of contention in OBS nodes and depend on node architecture, scheduling and contention resolution algorithms [34], [37]. Routing strategies and network dimensioning which avoid traffic concentration on few links allow to reduce contention occurrence. Burst losses also cause multiple segment losses, which leads to correlation in TCP segment losses. The effect of these losses depends on the level of segment aggregation in a burst [38],[39]. The correlated delivery of segments in a burst

allows the congestion window to increase faster than in conventional networks, leading to a correlation benefit that partially compensates the negative effects of burstification delay.

Scheduling algorithms. Some scheduling algorithms, e.g., void filling type algorithms, may increase the delay jitter and may generate out-of-order burst arrivals, in case of employment of Fiber Delay Lines (FDLs). Although the TCP goodput is shown to increase with the square root of the number of packets in a burst belonging to the same TCP flow [40], scheduling of longer bursts is more difficult, which may consequently increase burst loss probability. On the other hand, larger bursts require longer burstification delays, and these factors may in turn decrease the TCP goodput.

Aggregation strategy. If destination-based burstification is employed, it is very likely that, when a burst is lost due to contention, the dropped burst contains segments belonging to many different TCP flows, resulting in synchronization between the congestion control algorithms of the affected flows. With per-flow burstification, a burst loss causes the loss of segments belonging to the same TCP connection only, and the flow synchronization problem is eliminated.

Performance evaluation. The above elements impact on TCP throughput, which can be evaluated by simulation or analysis [38,41,40]. In particular ongoing research within the OBS task-force aims at evaluating end-to-end performance in OBS networks when TCP is used and at determining whether modified versions of TCP bring advantages in optimizing performance.

Some results related to a common reference scenario are here provided to provide an overview of the main research achievements.

Regarding OBS network features, the following parameters are considered: 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). Regarding TCP features, the parameters chosen are: maximum segment size is 512 bytes, minimum retransmission timeout is 0.5 seconds, maximum window size is 65535 bytes, delayed ACK with 0.2 seconds.

Two main performance indicators are considered, namely TCP goodput, defined as the number of bits that a TCP receiver receives per second without counting out-of-order and duplicate packets, and TCP throughput as the gross rate of a TCP source allowed by the network context.

The first evaluation regards the influence of the burstification process on TCP. Depending on the specific aggregation scheme, the number of burstification buffers per egress node, denoted by B , at the ingress node will be different, e.g., $B = 1$ for destination-based, $B =$ number of flows for flow-based and $1 < B <$ number of flows for class, session and group based schemes. In or-

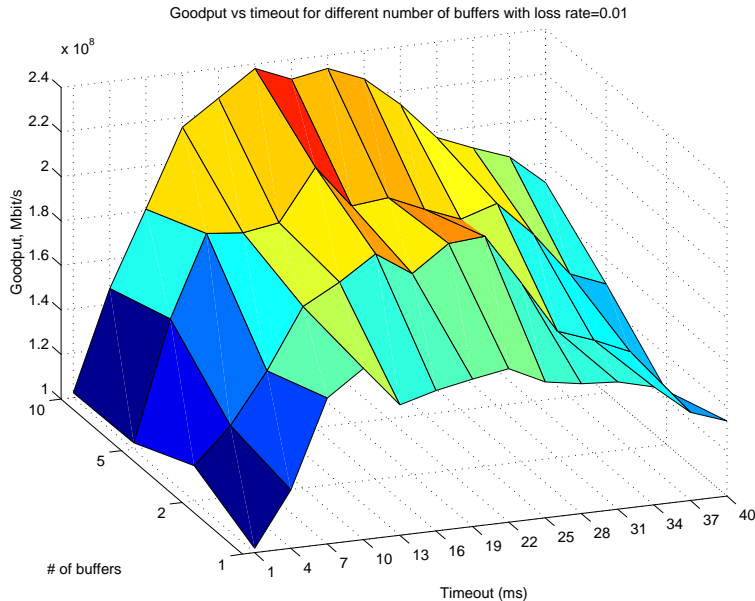


Fig. 9. Average total TCP goodput $p = 0.01$ as a function of B and burstification timeout for time-based burst assembly

der to optimize the TCP goodput the effects of the number of burstification buffers with the time-based burstification scheme are considered. Simulations with 10 TCP flows multiplexed over a common optical channel are performed for TCP NewReno, since this is currently the most widely used version of TCP. The loss of bursts is simulated by intentionally dropping some of the bursts transmitted over the optical link with probability p . The simulations are run for a duration of 200 seconds. The effect of the number of burstification buffers on the goodput is shown in Figure 9 which is obtained using the time-based burstification mechanism for $p = 0.01$. We observe that the average goodput increases as the number of aggregation buffers increase. The maximum average goodput increases by more than 34% with $B = 10$, which corresponds to flow-based burstification, compared with $B = 1$, which corresponds to destination-based burstification. The main reason for this increase is the elimination of the synchronization between the congestion control algorithms of TCP flows when flow-based burstification is used instead of the destination-based burstification. In the same figure, it is shown that goodput can be optimized in relation to the burst assembly timeout. Burst length and burstification delay can be adjusted for matching the dynamics of the incoming TCP traffic. Adaptive algorithms can be applied to enhance the TCP goodput with respect to static burst assembly algorithms.

The second evaluation regards the influence of switch design on TCP throughput. An 8×8 SPN switch architecture equipped with wc shared-per-node wavelength converters is considered. Figure 10 shows the TCP send rate as a function of the offered load A_0 per input wavelength by assuming $wc = 64$ and varying the number N of wavelengths per fiber as a parameter. A pure

time-based burst assembly strategy is used with 3 ms of burstification time. The dropping edge shifts to higher A_0 values as N increases up to a maximum value of $N = 128$, in this case $N = 32$, after which the increased blocking due to converters unavailability reduce the maximum achievable A_0 . In case of lower values of wc , e. g. $wc = 8$, the dropping edge shifts to lower values of A_0 as N increases. This outlines the strong impact of the parameter wc on the TCP send rate and in some situations an increase of N implies a reduction of the send rate. Therefore, the core node design, e.g. the choice of the number of wavelengths N and the number of converters wc , should be properly done to optimize the overall network design.

The third evaluation is to analyze the behavior of different TCP versions in the presence of burst losses and is obtained by simulation. A time-based burst assembly strategy with the destination-based scheme ($B = 1$) is used with 3 milliseconds and 300 milliseconds. To measure the goodput we have considered the FTP transfer of 500 files of 10 Megabytes. The results are shown in figure 11, which represents TCP goodput versus burst loss probability. The measures have been gathered with the 95% confidence interval being less than 5% of the mean.

In the simulated scenario, SACK performs better than Reno and Tahoe for both burst assembly timeouts. In the event of a burst loss, several TCP segments get lost. Thus, out of order TCP segments will arrive at the receiver, which immediately sends duplicate ACKs. When Tahoe is used, every time that three duplicate ACKs are detected by the sender, the first lost segment is retransmitted (fast retransmit), and slow start is performed [42]. On the contrary, when a TCP Reno sender detects three duplicate ACKs, after retransmitting the first lost segment, the fast recovery phase is performed [43], halving the congestion window. In case that several segments are lost, the TCP congestion window can be reduced to a very low value. In this case, the TCP sender will be unable to receive three duplicate ACKs, ending up in the triggering of a TCP retransmission timeout, which leads to slow start. Thus, multiple dropped segments can have a catastrophic effect on TCP Reno performance. TCP SACK [44] overcomes Reno's limitations using selective acknowledgments, which allow the data receiver to inform the sender about all segments that have arrived successfully. To sum up, SACK improves Reno and Tahoe performance in the face of burst losses.

7 Quality of service for OBS networks

Most Quality of Service (QoS) mechanisms for OBS networks implement a relative QoS model where the performance of a class is defined with respect to other classes. This section aims at comparing contention resolution mecha-

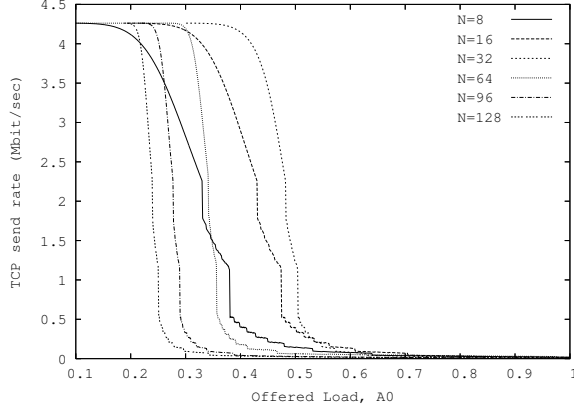


Fig. 10. TCP send rate as a function of the offered load per wavelength, varying the number of wavelengths per link as a parameter, $wc = 64$

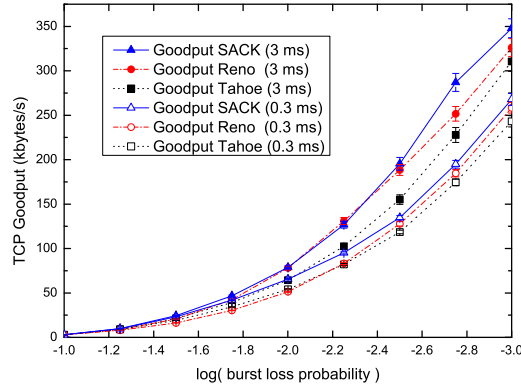


Fig. 11. TCP goodput vs burst loss probability

nisms that adopt such models and that are most studied in the literature. In particular, we study:

- *Offset time differentiation* (OTD) [45] where the idea is to assign an extra offset-time to high priority (HP) bursts, resulting in an earlier reservation.
- *Burst preemption* (BP) [46] which in case of contention overwrites the resources reserved for the low priority (LP) burst. We implement a simple full-preemptive scheme where each HP burst is allowed to preempt at most one LP burst. The preempted LP burst is discarded.
- *Burst Dropping with Wavelength (BD-W) or Buffer (BD-B) thresholds* [47] which provide less transmission resources to LP bursts than to HP bursts. In our implementation, the LP class has access to 50% of the most-indexed wavelengths in the BD-W case and it has access to 50% of the delay lines with shorter length in BD-B mechanisms.
- *Burst Length Differentiation* (BLD) [48] uses shorter burst assembly thresholds for HP bursts than those used to aggregate LP bursts. In our imple-

mentation of BLD, the mean burst length of a LP burst is 4 times longer than the mean burst length of a HP burst.

All the mechanisms are evaluated in the same scenario of a single isolated node. The mean load per wavelength is 0.8 Erlang. The percentage of HP bursts over the total traffic is 30%. The burst length and inter-arrival times are Gaussian distributed. The scheduling algorithm applies the LAUC-VF scheme.

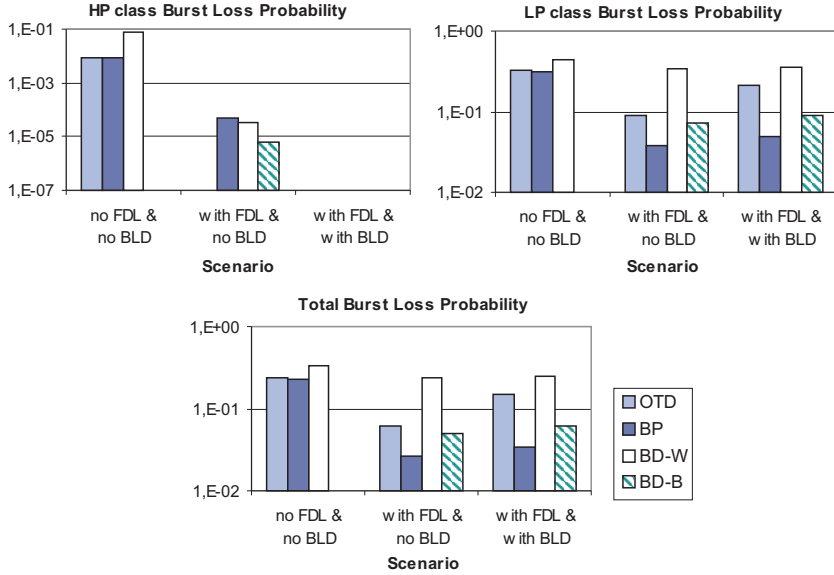


Fig. 12. Comparison of selected QoS mechanisms with and without FDL and BDL technique

Figure 12 presents the results for the different QoS scenarios, namely with and without a feed-forward FDL buffer (of 4 delay lines) and with BLD applied. In the bufferless case, we can observe that both OTD and BP offer the same loss performance for the HP class, however BP shows better behavior for both the LP and the total traffic. This is due to the scheduling, which, according to [49], suffers performance degradation due to the variable offset-times in the OTD mechanism.

Moreover, FDL buffers improve the performance in general. Additional operation of the BLD method may further improve the loss performance of the HP class when used together with other QoS techniques. More specifically, BD-W and BD-B mechanisms may be recommended in order to boost the HP performance of the BP technique since it does not impact significantly the performance of the LP class. Finally, BLD should not be used with OTD because of the sensitivity of the OTD mechanism to scheduling operation that suffers performance degradation if bursts of different lengths are scheduled [49].

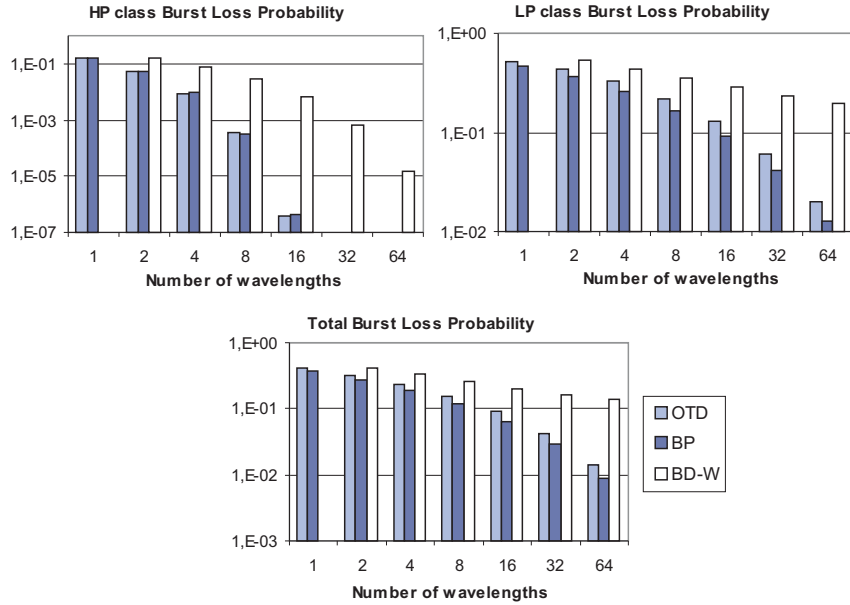


Fig. 13. Comparison of selected QoS mechanisms as a function of the number of wavelengths

In order to improve the effectiveness of QoS differentiation in bufferless scenario we can increase the number of wavelengths provided for the transmission of data bursts. As Figure 13 shows, the improvement of HP class BLP in both OTD and BP mechanisms can be really high. On the other hand, BD-W has the worst performance among all evaluated mechanisms. The reason is that BD-W mechanism has effectively less wavelengths available for burst transmissions on the output port than other mechanisms since it provides a limited number of wavelengths for the LP class.

8 Resilience Issues

As a part of e-photon/ONE, we have studied how survivability differentiation may be provided in a hybrid optical network using component and node redundancy. We aim at providing a set of service classes with different performance and availability characteristics [50].

The hybrid network under study is OpMiGua, which is one of several hybrid optical network architectures proposed in recent literature. Compared to other hybrid optical networks, a feature in the OpMiGua architecture is the division of the traffic into two different service classes with respect to performance: Guaranteed Service Transport (GST) and Statistically Multiplexed (SM) traffic. GST traffic follows fixed wavelength paths through a Wavelength Routed

Optical Network (WRON), resulting in no loss due to contention, while SM traffic is statistical multiplexed through optical packet switches. See [51] for details. The basic logical structure of an OpMiGua node without any fault-tolerance is shown in Figure 14. The component denoted as packet separator (PS) separates the signal to the optical cross connect or the packet switch. This may be realized utilizing e.g. the polarization of the incoming signal.

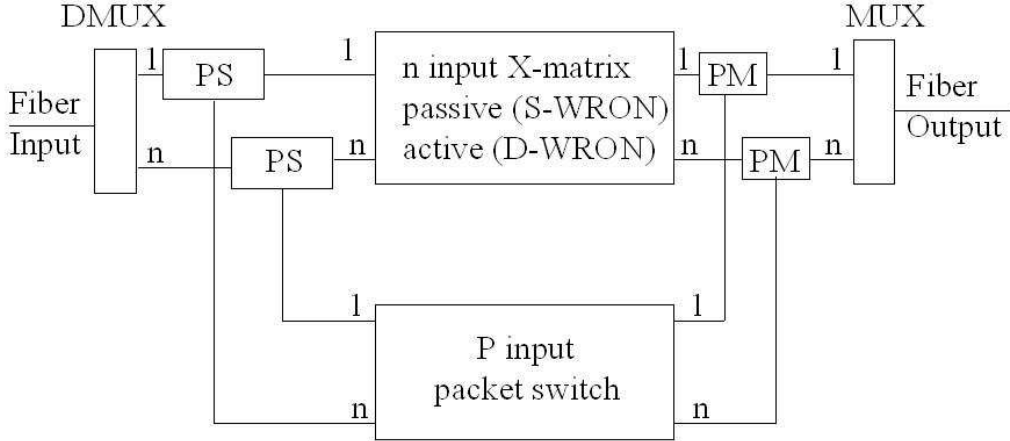


Fig. 14. Principle of an OpMiGua node. PS = Packet Separator. PM = Packet Multiplexer, n = Number of wavelengths. A single fiber input with w wavelengths is shown. The total number of inputs/outputs of the cross coupling matrix and packet switch is n and P respectively. Typically $n = P$, i.e. all inputs may carry both statically routed and circuit switched packets.

We extend the OpMiGua node to improve the availability of the traffic and to achieve availability differentiation by adding node and/or component redundancy. We differentiate the traffic into three service classes:

- Guaranteed service with high survivability traffic stream (denoted GShigh): This service is available for applications needing both high performance and high survivability. Low information loss is necessary because strict real-time demands exclude end-to-end retransmission. This service is business critical. In normal operation this service will use the GST lightpaths defined end-to-end through the network, utilizing the static or dynamic OXCs in the OpMiGua nodes.
- Guaranteed service but with lower survivability traffic stream (denoted GSlow): This service is available for applications needing the same high performance as GShigh, but with lower availability demands. Low information loss is still necessary in normal operation due to strict real-time demands, but the service is not business critical. This service will use the GST lightpaths defined end-to-end through the network.
- Best effort service traffic stream without strict performance or survivability demands (denoted BE): This service is sufficient for most data traffic using the Internet today. BE traffic will be carried by the SM service in an

OpMiGua network.

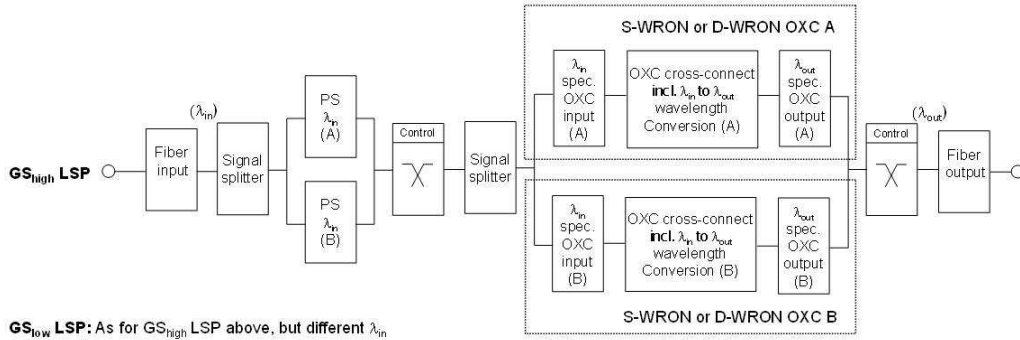


Fig. 15. Availability block diagrams for BE traffic and GS traffic.

We have studied and proposed different node architectures supporting the above traffic classes, see [52] for details. Further, we briefly outline one of these design proposals. We employ an OpMiGua node with duplication of the PS and duplication of a static OXC for S-WRONS. In addition we use a single packet switch handling BE traffic streams only. An availability block structure of the proposed node is given in Figure 15. By employing the principles above, our specifications of the three different service classes in an OpMiGua network can be met: From the OpMiGua architecture, GShigh and GSlow traffic will have higher performance characteristics than BE traffic. Furthermore, the duplicated OXC results in an improved availability for GShigh and GSlow traffic compared to BE traffic. GShigh and GSlow traffic may be differentiated with respect to the availability by employing access restriction to the spare OXC in the case of failure. The duplicated PS ensures a higher availability for both GS and BE traffic.

In future works on this topic, numerical evaluations of the proposed node design will quantify its reliability.

9 Burst switch architectures

Within e-Photon/ONE, we have evaluated the performance of a feed-forward delay based switch architecture scheme, shown in [53][54]. The design consists of a "scheduling" unit with k input/output ports and a $k \times k$ nonblocking space switch. Each branch delays the incoming packets/bursts, assigning incoming bursts to outgoing slots, resolving contention and maintaining burst or packet ordering for the same outgoing link. The problem of scheduling packets through a branch of delay blocks to avoid collisions is a problem of routing a permutation between inputs and outputs in the equivalent Benes network,

where nonoverlapping paths in the network correspond to collision-free transmission through the delay blocks.

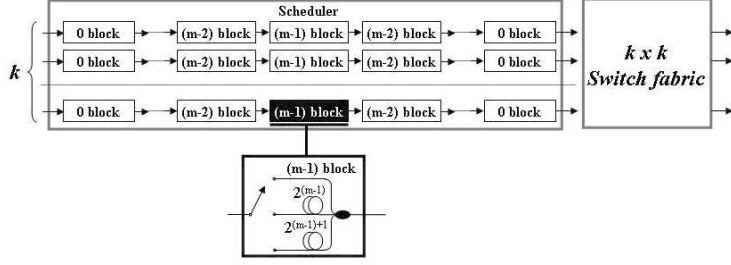


Fig. 16. The Scheduling switch; a feed-forward delay based switch architecture. The Scheduler comprises k branches, each with $2\log_2 T - 1$ delay blocks. The i^{th} delay block consists of one three-state switch and three delay lines of length 0, 2^i and 2^{i+1} packet slots.

Each delay branch consist of $2m - 1$ delay blocks, where $m = \log_2 T$. T is assumed to be a power of 2 and corresponds to the maximum number of sequential slots, belonging to the same or different bursts/packets from all incoming links that request the same output and can be served with no contention. The i^{th} block consists of a three-state (or two 2×2) optical switch and three fiber delay paths, corresponding to delays equal to 0, 2^i and 2^{i+1} slots. To ensure that the packets in the incoming frame can be assigned to any slot in the outgoing frame, the latter must start at least $(3T/2) - 2$ after the incoming frame begins. Such an input queuing scheme can be viewed as implementing an optical packet buffer of depth T [55]. One of the major advantages of the scheduling packet switch is its modular buffering scheme design that can be easily expanded to accommodate more burstiness in the traffic (in a way similar to the way electronic buffers can be expanded in a conventional electronic switch). The cost of the switch, when measured in terms of the elementary switching elements it requires, grows only logarithmically with T [55].

We have also investigated the performance of the switch under a heavy tailed truncated Pareto distribution, which is considered by many researchers to be a good model for bursty traffic in real networks [29]. In our model [56], bursts consists of packet slots that arrive during ON periods and which are separated by idle periods (OFF periods). To generate a Pareto-distributed sequence of ON periods, we have generated a Pareto-distributed sequence of burst (packet train) sizes, followed by Pareto-distributed idle periods. The minimum burst size is 1, corresponding to a single packet arrival. The formula to generate a Pareto distribution is: $X_{\text{PARETO}} = \frac{b}{x^a}$ where x is a uniformly distributed value in the range $(0, 1]$, b is the minimum nonzero value of X_{PARETO} , denoted b_{on} and b_{off} for the packet train and the idle period, respectively, and a is the tail index or shape parameter of the Pareto distribution. We performed computer simulations for a $k=2$ and $k=4$ I/O ports and $a_{on} = 1.7$, $a_{off} = 1.2$ and $X_{min} = 10^{-4}$. In our simulation, we have selected a_{on} to be larger than a_{off} , since

in real traffic, the probability of having extremely large OFF periods is higher than the probability of having extremely large ON periods. Figure 17 shows the corresponding loss ratio results for $T \in [2 \dots 64]$. Burst destinations were evenly distributed.

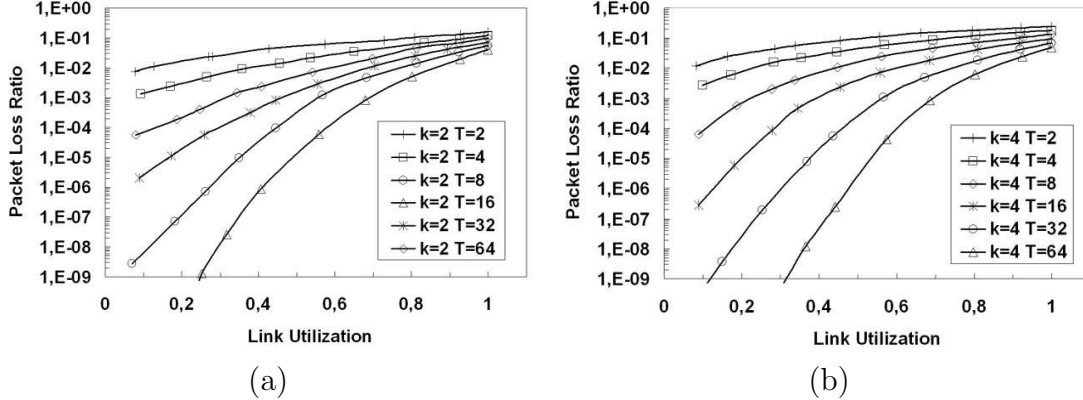


Fig. 17. Switch loss ratio for (a) $k=2$ and (b) $k=4$ versus link utilization for $T \in [2 \dots 64]$. Burst arrivals and idle periods follow a truncated Pareto distribution with a tail index of 1.7 and 1.2 respectively.

10 The physical layer for OBS

As future optical technology moves to 40Gb/s and beyond, networking solutions must be designed to be compatible with these bit rates, in order to reduce the cost per bit [57]. Optical burst switching (OBS) has been introduced as a switching technology based on fast switching requirements, as the relatively slow switch set-up times (milliseconds) are small compared to the data burst duration and therefore throughput is almost unaffected. However, the introduction of new bandwidth on demand services [58] (e.g. Grid services: high resolution home video editing, real-time rendering, high-definition interactive TV and e-health) over OBS implies new constraints for the switching speed and technology requirements, which become particularly important when high speed transmission is considered. Such applications usually involve large number of users that need transmission of relatively small data bursts and possibly with short offset time.

A flexible OBS network must be able to support the small data bursts generated by aforementioned types of applications and services. For example, a burst with 300 ms duration, transmitted at 10Gb/s can be switched by a MEMS based switch typically within 20 ms. Considering only the switching time, the throughput of the system is 93.7%. If the same burst is transmitted at 160Gb/s then its duration is 18.75 ms and routing through the same

switch would decrease the system's throughput to less than 50%. This becomes more severe when smaller bursts with short offset time are treated by the OBS switch. For this reason, the deployment of fast switching technology is essential for future high speed OBS networks where the evolving bandwidth on demand services are supported.

It should be noted though, that the burst control packet/header (BCH) requires intensive and intelligent processing (i.e. QoS, routing and contention resolution algorithms) which can only be performed by specially designed fast electronic circuits. Recent advances in the technology of integrated electronic circuits allow complicated processing of bursty data directly up to 10Gb/s [59]. This sets the upper limit in the transmission speed of the BCH. On the other hand the optical data bursts (they don't need conversion to electronic domain for processing) are those that determine the capacity utilisation of the network. The optical bursts (data burst) can be transmitted at ultra-high bit rates (40 or 160Gb/s), providing that the switching elements can support these bit rates. Faster bursts indicate higher capacity utilisation of the existing fibre infrastructure and significantly improved network economics. The deployment of fast switching assists the efficient bandwidth utilisation but provides an expensive solution when it scales to many input ports. On the other hand, there is no additional benefit for long bursts of data, if fast switching is utilised.

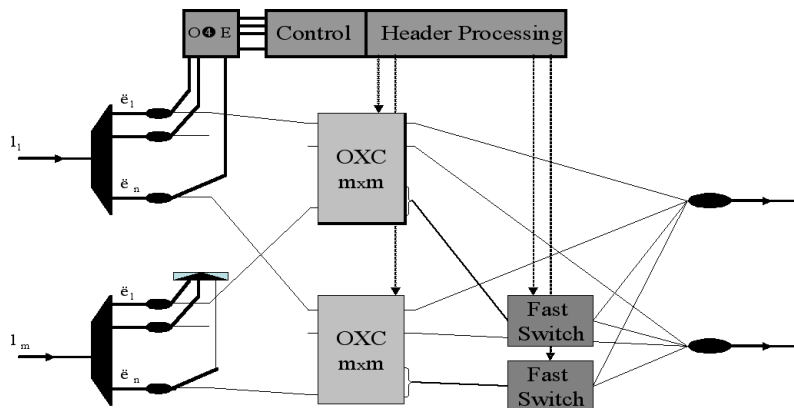


Fig. 18. Architecture that combines slow switching (OXCs) and fast switching elements

10.1 Core optical burst switching node

Therefore, in this section a switch architecture is proposed comprising a combination of fast (e.g. based on semiconductor optical amplifier) and slow (e.g. MEMS-based) switches. The switch architecture is shown in figure 18. The general idea is based on the use of MEM-based OXCs that has a number of output ports connected to a fast optical switches. When a BCH appears, the control mechanism must first recognise if the BCH belongs to a burst with

slow switching requirements (usually long burst) or a burst with fast switching requirements (usually short burst). In the first case the OXC is reconfigured so that when the long burst arrives it is automatically routed to the appropriate output port. In the other case the short bursts are routed directly to the fast switch (through pre-defined paths) and switched immediately to the next node. This architecture requires all the switching paths inside the OXC to be initially connected to the fast switch ports and special design constraints must be considered to avoid collision. The benefit of the proposed scheme is that it reduces the requirements on fast switching and therefore smaller and cost efficient matrices are only required.

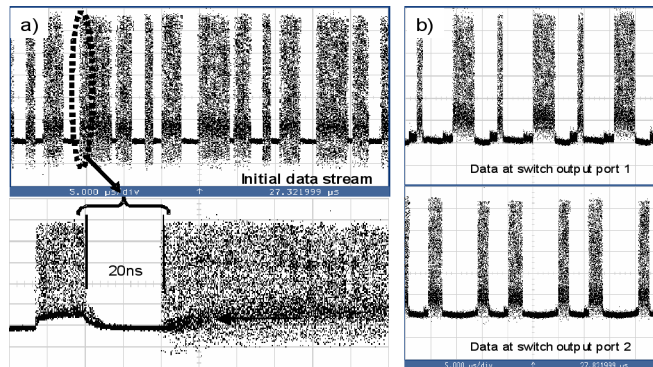


Fig. 19. A fast switching example of asynchronous data. a) The initial data stream and the header in front of the payload. b) The switched data as they appear at two output ports of the switch.

The proposed fast switching mechanism is based on the use of fast active components, like Semiconductor Optical Amplifiers (SOAs). Switching is achieved by converting the signals wavelength and routing it to an output port of a passive routing device (AWG) [60]. This solution is scalable but the bit-rate depends on the utilised conversion technique. However, almost bit rate transparent wavelength conversion schemes have been proposed and fast switching of asynchronous bursty data at 40Gb/s has been demonstrated, with technology scalable to more than 160Gb/s [60]. This solution provides switching in nanoseconds and therefore can almost eliminate the required offset time for the short data bursts, offering increased throughput. Figure 19 shows the fast switching results for short data bursts with short offset time. The BCH in front of the 40Gb/s data burst is extracted and processed and the different bursts of data are routed to two output ports of the switch.

10.2 Edge optical burst switching node

In an OBS network, an ingress edge router able to initiate BCH and also map IP traffic into the optical domain in the form of variable length optical bursts is mandatory. Here, we propose a novel architecture for a wavelength

agile ingress edge OBS router and demonstrate a mechanism that efficiently maps Internet traffic onto optical bursts. The main design considerations are: wavelength agility, traffic aggregation based on the class of service (CoS), variable length optical burst construction, data burst and BCH transmission. The proposed architecture utilises a high-speed hardware platform and a fast tuneable laser and also supports the user network interface (UNI) functionality by interfacing with the control plane.

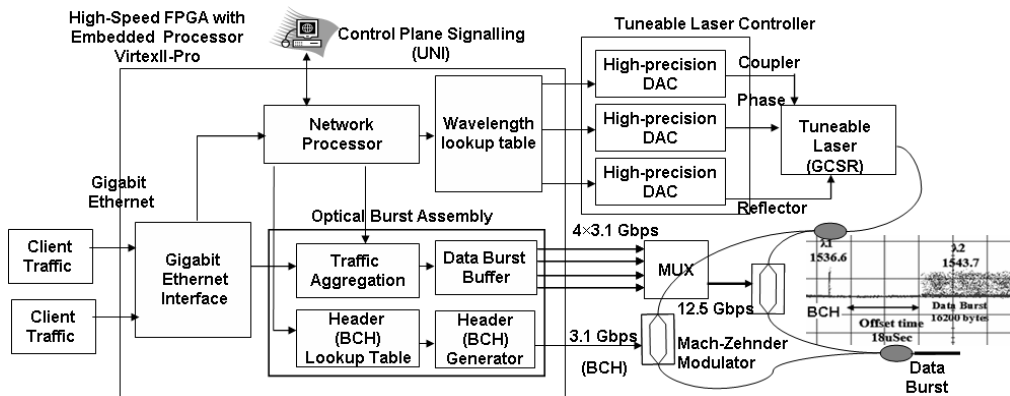


Fig. 20. Architecture of a wavelength agile edge optical burst switching interface

Figure 18 shows the architecture of an edge OBS router and the experimental setup. The proposed architecture comprises the following units: 1- Input interfaces to accept IP traffic through the Gigabit Ethernet links. 2- Traffic aggregation and optical burst assembly unit to generate optical bursts and their associated BCHs. 3- Tuneable laser source and its controller to facilitate wavelength assignment for data bursts and BCHs. 4- User-network signalling and control interface (UNI) to obtain the required information from control plane (i.e. data burst and BCH wavelengths, BCH information and burst transmission parameters like offset time). The Class-based IP traffic aggregation and the agile wavelength assignment constitute the optical burst assembly and transmission mechanism which is a key unit in the edge OBS router. To prototype the proposed optical burst assembly and transmission mechanism, a high-speed field programmable gate arrays (FPGA) platform is deployed with a tuneable wavelength allocation mechanism (figure 20)[61]. In this architecture, IP packets from a traffic generator enter into the edge router through a Gigabit Ethernet input interface. The incoming packets are aggregated in aggregation buffers based on their header contents (address and CoS). The embedded network processor with the help of the proposed traffic classification method performs class-based optical burst assembly. This is achieved by aggregating the incoming packets with the same CoS and the same destination address into the same optical burst. Aggregated IP packets (data bursts) are stored in data burst FIFO buffers prior to transmission. As the hardware platform has maximum speed of 3.125 Gbps at each of the high-speed IO ports, a 4:1 Mux is used to generate 12.5 Gbps payload by combining four 3.125 Gbps output streams in parallel. Before transmission of each aggregated data

burst a BCH is transmitted in front of the data burst at 3.125 Gbps. Also the tuneable laser is set to emit suitable wavelengths for each BCH as well as data burst. Two high-speed lookup tables (i.e. the header and the wavelength lookup tables) stores wavelength information and BCH information respectively, for each CoS and destination address. The wavelength allocation mechanism comprises a fast tuneable grating coupled sampled rear reflector (GCSR) laser and a high-speed hardware controller. The GCSR laser is set to generate the required wavelengths for each data burst and BCH. It can be tuned to any wavelength within the C-band by controlling the currents of the coupler, reflector and phase sections. Fast and accurate laser tuning is facilitated with the use of a high-speed hardware controller combined with three high precision digital to analogue converters (DAC) connected to each laser section [61].

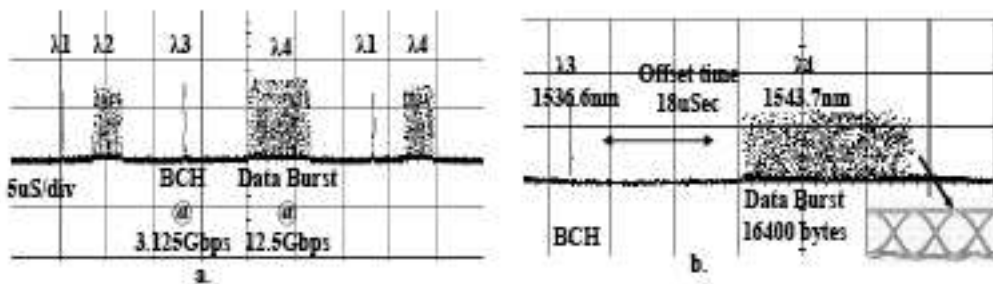


Fig. 21. Experimental result of the OBS transmitter

The advantage of the proposed optical burst assembly mechanism is that it provides an easily reconfigurable and wavelength agile mechanism. This allows the allocation of optical burst transmission parameters from the control plane through the UNI interface. (i.e. suitable wavelengths for each BCH and data burst, variable offset times and also aggregation parameters for each CoS and destination address). In this experiment, transmission of variable length optical bursts with their associated BCHs, in four different wavelengths has been demonstrated. ($\lambda_1=1536.6\text{ nm}$ and $\lambda_3=1546.1\text{ nm}$ for BCHs, $\lambda_2=1543.7\text{ nm}$ and $\lambda_4=1548.5\text{ nm}$ for data bursts). The generated optical bursts are shown in figure 21. Variable offset times, with maximum of $18\mu\text{s}$ (figure 21 b) are considered. The data burst length is also variable and it changes from 3KByte to 16KByte in multiple integrals of 100 bytes (figure 21 a). The results show that 16.3 dB extinction ratio can be achieved by the proposed transmitter.

11 Conclusions

In this paper we have provided an overview of European research on network aspects, such as routing, resilience and contention resolution, together with switch architectures, of OBS networks, within the e-Photon/ONe network of

excellence. Traffic analysis, quality of service, TCP/IP over OBS and physical layer aspects have also been covered. The paper shows that there is an increasing activity in the field within the European Union. The e-Photon/ONE network of excellence provides the framework to coordinate such diverse OBS research activity.

References

- [1] C. Qiao, M. Yoo, Optical burst switching (obs): A new paradigm for an optical internet, *High Speed Networks* 8 (1999) 69–84.
- [2] Y. Xiong, M. Vanderhoute, C. Cankaya, Control architecture in optical burst-switched WDM networks, *IEEE Journal of Selected Areas in Communications* 18 (10) (2000) 1838–1851.
- [3] K. C. et. al, EBRP: A hybrid signaling protocol for efficient burst-level reservations and qos differentiation in obs networks, *Journal of Optical Networking* 5 (3) (2006) 147–158.
- [4] G. P. V. Thodime, V. M. Vokkarane, J. P. Jue, Dynamic congestion-based load balanced routing in optical burst-switched networks, in: *IEEE GLOBECOM*, 2003, pp. 2628–2632.
- [5] F. Ramón-Salguero, J. Andrés-Colas, A. Molins-Jiménez, J. Enríquez-Gabeiras, Dynamic routing strategies to postpone network congestion, *Tech. Rep. TD02043, COST 279* (2002).
- [6] J. Aracil, O. Gonzalez de Dios, J. P. Fernandez-Palacios, Routing strategies for obs networks based on mrdv, in: *Proceedings of the 7th international conference on Transparent Optical Networks 2005*, 2005.
- [7] J. S. Turner, Terabit burst switching, *Journal of High Speed Networks* 8 (1) (1999) 3–16.
- [8] M. Yoo, C. Qiao, QoS performance in IP over WDM networks, *IEEE Journal of Selected Areas in Communications* 18 (10) (2000) 2062–2071.
- [9] S. Yao, B. Mukherjee, S. J. B. Yoo, S. Dixit, All-optical packet-switched networks: a study of contention-resolution schemes in an irregular mesh network with variable-sized packets, in: *Proceedings of SPIE OptiComm 2000, Dallas*, 2000.
- [10] K. Dolzer, C. M. Gauger, J. Späth, S. Bodamer, Evaluation of reservation mechanisms for optical burst switching, *AEÜ International Journal of Electronics and Communications* 55 (1) (2001) 18–26.
- [11] K. Dolzer, C. M. Gauger, On burst assembly in optical burst switching networks—a performance evaluation of just-enough-time, in: *Proceedings of the 17th International Teletraffic Congress (ITC 17)*, Salvador, Brazil, 2001, pp. 149–160.

- [12] H. L. Vu, M. Zukerman, Blocking probability for priority classes in optical burst switching networks, *IEEE Communications Letters* 6 (5) (2002) 214–216.
- [13] C. M. Gauger, Dimensioning of FDL buffers for optical burst switching nodes, in: *Proceedings of the 5th IFIP Optical Network Design and Modeling Conference (ONDM 2002)*, Torino, 2002.
- [14] COST Action 266, Final report: Advanced infrastructures for photonic networks, http://www2.cordis.lu/cost/src/266_indivpage.htm (2003).
- [15] D. Hunter, M. Chia, I. Andonovic, Buffering in optical packet switches, *Journal of Lightwave Technology* 16 (12) (1998) 2081–2094.
- [16] C.-F. Hsu, T.-L. Li, N.-F. Huang, Performance analysis of deflection routing in optical burst-switched networks, in: *IEEE Infocom*, New York, 2002.
- [17] Y. Chen, H. Wu, D. Xu, C. Qiao, Performance analysis of optical burst switched node with deflection routing, in: *IEEE International Conference on Communication (ICC)*, Anchorage, 2003.
- [18] X. Wang, H. Morikawa, T. Aoyama, Deflection routing protocol for burst switching WDM mesh networks, in: *SPIE Terabit Optical Networking: Architectures, Control and Management Issues*, Boston, 2000.
- [19] C. M. Gauger, Optical combination of converter pools and FDL buffers for contention resolution in optical burst switching, *Journal of Photonic Network Communications* 8 (2) (2004) 139–148.
- [20] C. M. Gauger, M. Köhn, J. Scharf, Comparison of contention resolution strategies in OBS network scenarios, in: *International Conference on Transparent Optical networks (ICTON)*, Wroclaw, 2004.
- [21] V. Eramo, M. Listanti, P. Pacifici, A comparison study on the wavelength converters number needed in synchronous and asynchronous all-optical switching architectures, *IEEE Jour. Lightwave Tech.* 21 (2003) 340–355.
- [22] V. Eramo, M. Listanti, Input wavelength conversion in optical packet switches, *IEEE Comm. Letters* 7 (6) (2003) 281–283.
- [23] N. Akar, E. Karasan, Exact calculation of blocking probabilities for bufferless optical burst switched links with partial wavelength conversion, in: *IEEE BROADNETS'04, Optical Networking Symposium*, 2004.
- [24] K. Dogan, N. Akar, A performance study of limited range partial wavelength conversion for asynchronous optical packet/burst switching, in: *International Communications Conference*, Istanbul, Turkey, 2006.
- [25] Z. Rosberg, H. L. Vu, M. Zukerman, Performance analyses of optical burst-switching networks, *IEEE Jour. Sel. Areas in Comm.* 21 (7) (2003) 1187–1197.
- [26] X. Chu, J. Liu, Z. Zhang, Analysis of sparse-partial wavelength conversion in wavelength-routed wdm networks, in: *IEEE INFOCOM'04*, 2004.

- [27] V. Paxson, S. Floyd, Wide area traffic: The failure of Poisson modeling, *IEEE/ACM Transactions on Networking* 3 (3) (1995) 226–244.
- [28] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, On the Self-Similar Nature of Ethernet Traffic (Extended Version), *IEEE/ACM Transactions on Networking* 2 (1) (1994) 1–15.
- [29] Mark E. Crovella, A. Bestavros, Self-similarity in world wide web traffic: Evidence and possible causes, *IEEE/ACM Transactions on Networking* (6) (1997) 835–846.
- [30] T. Karagiannis, M. Molle, M. Faloutsos, A. Broido, A nonstationary poisson view of internet traffic, in: *IEEE Infocom 2004*.
- [31] I. Norros, On the use of Fractional Brownian Motion in the theory of connectionless networks, *IEEE Journal on Selected Areas in Communications* 13 (6) (1995) 953–962.
- [32] M. Izal, J. Aracil, On the influence of self similarity on optical burst switching traffic, in: *Proceedings of GLOBECOM 2002*, 2002.
- [33] X. Yu, J. Li, X. Cao, Y. Chen, C. Qiao, Traffic statistics and performance evaluation in optical burst switched networks, *Journal of Lightwave Technology* 22 (12) (2004) 2722–2738.
- [34] M. Casoni, E. Luppi, M. L. Merani, Impact of assembly algorithms on end-to-end performance in optical burst switched networks with different QoS classes, *Proceedings of 3rd International Workshop on Optical Burst Switching (WOBS 2004)*.
- [35] M. de Vega Rodrigo, J. Gotz, An analytical study of optical burst switching aggregation strategies, *Proceedings of 3rd International Workshop on Optical Burst Switching (WOBS 2004)*.
- [36] X. Cao, J. Li, Y. Chen, C. Qiao, Assembling TCP/IP packets in optical burst switched networks, *Proc. Globecom 2002* 3 (2002) 2808–2812.
- [37] M. Casoni, M. Merani, On the performance of tcp over optical burst switched networks with different qos classes, in: *Proc. of 3rd IEEE International Workshop on QoS in Multiservice IP Networks*, Catania, 2005.
- [38] A. Detti, M. Listanti, Impact of segments aggregation on TCP reno flows in optical burst switching network, *Proceedings of Infocom 2002* 3 (2002) 1803–1812.
- [39] X. Yu, C. Qiao, Y. Liu, TCP implementations and false time out detection in OBS networks, *Proceedings of Infocom 2004* (2004) 774–784.
- [40] X. Yu, C. Qiao, Y. Liu, D. Towsley, Performance evaluation of TCP implementations in OBS networks, *Tech. Rep. 2003-13*, CSE Dept., SUNY.
- [41] J. Padhye, V. Firoiu, D. Towsley, J.F.Kurose, Modeling TCP reno performance: a simple model and its empirical validation, *IEEE/ACM Transactions on Networking* 8 (2) (2000) 133–145.

- [42] R. Stevens, TCP/IP Illustrated, Volume 1: The Protocols, Addison-Wesley, 1994.
- [43] M. Allman, V. Paxson, W. Stevens, TCP congestion control, RFC 2581 (Proposed Standard) (apr 1999).
- [44] M. Mathis, J. Mahdavi, S. Floyd, A. Romanow, TCP selective acknowledgement options, RFC 2018 (Proposed Standard) (oct 1996).
- [45] M. Yoo, C. Qiao, Optical burst switching for service differentiation in the next-generation optical internet, *IEEE Communications Magazine* 39 (2) (2001) 98–104.
- [46] A. Kaheel, H. Alnuweiri, A strict priority scheme for quality-of service provisioning in optical burst switching networks, in: *Proceedings of IEEE Symposium on Computers and Communications (ISCC 2003)*, Turkey, 2003.
- [47] F. Callegati, W. Cerroni, G. Muretto, C. Raffaelli, P. Zaffoni, Qos routing in DWDM optical packet networks, in: *Proceedings of 1st Workshop on Quality of Service Routing (WQoS2004) co-located with the Fifth International Workshop on Quality of Future Internet Services (QoFIS 2004)*, Barcelona, Spain, 2004.
- [48] M. Klinkowski, D. Careglio, S. Spadaro, J. Solé-Pareta, Impact of burst length differentiation on QoS performance in OBS networks, in: *Proceedings of IEEE ICTON 2005*, Barcelona, Spain, 2005.
- [49] J. Li, C. Qiao, J. Xu, D. Xu, Maximizing throughput for optical burst switching networks, in: *Proceedings of IEEE Infocom 2005*, Hong Kong, 2004.
- [50] B. Helvik, *Dependable Computing Systems and Communication Networks*, 2001.
- [51] S. Bjornstad, D. R. Hjelm, N. Stol, A packet switched hybrid optical network with service guarantees, To Appear in *IEEE Journal on Selected Areas in Communications* 24.
- [52] N. Stol, H. Overby, S. Bjornstad, A. Kimsas, A. Mykkeltveit, Differentiated survivability in the opmigua hybrid optical network, To Appear at the 10th Conference on Optical Network Design and Modelling.
- [53] E. Varvarigos, The packing and the scheduling packet switch architectures for almost alloptical lossless networks, *IEEE J. Lightwave Technol.* (16) (1998) 1757–1767.
- [54] K. Vlachos, K. Seklou, E. Varvarigos, Performance evaluation of an optical packet scheduling switch, *IEEE GLOBECOM* (2004) 1857–1861.
- [55] I. C. et al, Cord: Contention resolution by delay lines, *IEEE Journal on Selected Areas in Communications* 14 (5) (1996) 1014–1029.
- [56] K. Vlachos, C. Bintjas, E. Varvarigos, Performance evaluation of an optically interconnected scheduling switch network for pareto traffic, *OSA Journal on Optical Networking* 3 (11) (2004) 760–768.

- [57] M. O. Networks, Myths and realities about 40G optical technology, white paper : www.mintera.com/WhitepaperWEBnew.pdf (2002).
- [58] E. van Breusegem, M. de Levenheer, J. Cheyns, P. Demeester, D. Simeonidou, M. J. O'Mahoney, R. Nejabati, A. Tzanakaki, I. Tomkos, An obs architecture for pervasive grid computing, in: Proc. 1st Conference on broad band network (Broadnets 2004), San Jose, California, USA, 2004.
- [59] J. Gaither, 300-pin msa bit-error rate tester for the ml10g board and rocketphy transceiver, XILINX, Application note: XAPP677 Virtex-II Pro family (Jan. 2004).
- [60] D. Klondis, R. Nejabati, C. Politi, M. O'Mahony, D. Simeonidou, Demonstration of a fully functional and controlled optical packet switch at 40gb/s, in: proc. 30th European Conf. on Optical Comm. PD Th4.4.5, Stockholm, Swede, 2004.
- [61] R. Nejabati, D. Klondis, D. Simeonidou, M. O'Mahony, Demonstration of user-controlled network interface for sub-wavelength bandwidth-on-demand services, in: proc. Optical Fiber Communication Conference (OFC) 2005 paper OWK2, Anaheim, California, USA, 2005.