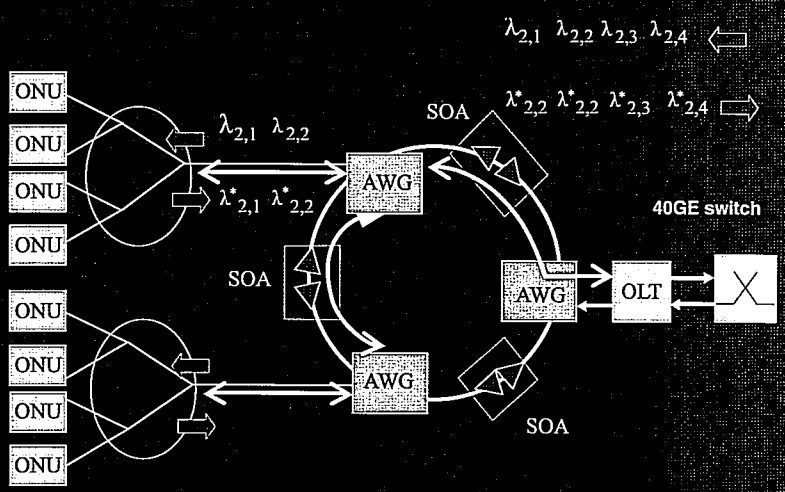
State-of-the-Art Survey

Ioannis Tomkos Maria Spyropoulou
Karin Ennser Martin Köhn
Branko Mikac (Eds.)

LNCS 5412

Towards Digital Optical Networks

COST Action 291 Final Report



Springer

cost

Volume Editors

Ioannis Tomkos
Maria Spyropoulou
Athens Information Technology Centre
19002 Peania-Attica, Greece
E-mail: {itom,mspi}@ait.edu.gr

Karin Ennser
Swansea University
Institute of Advanced Telecommunications
SA2 8PP, Swansea, UK
E-mail: k.ennser@swansea.ac.uk

Martin Köhn
University of Stuttgart
IKR
70569 Stuttgart, Germany
E-mail: martin.koehn@ikr.uni-stuttgart.de

Branko Mikac
University of Zagreb
Faculty of Electrical Engineering and Computing (FER)
Department of Telecommunications
10000 Zagreb, Croatia
E-mail: branko.mikac@fer.hr

Library of Congress Control Number: Applied for

CR Subject Classification (1998): C.2, B.4.3, H.3.4

LNCS Sublibrary: SL 5 – Computer Communication Networks and Telecommunications

ISSN 0302-9743
ISBN-10 3-642-01523-9 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-01523-6 Springer Berlin Heidelberg New York

springer.com

Typesetting: Camera-ready by author, data conversion by Markus Richter, Heidelberg
Printed on acid-free paper SPIN: 12627301 06/3180 5 4 3 2 1 0

Foreword

COST – the acronym for European COoperation in Science and Technology – is the oldest and widest European intergovernmental network for cooperation in research. Established by the Ministerial Conference in November 1971, COST is presently used by the scientific communities of 35 European countries to cooperate in common research projects supported by national funds.

The funds provided by COST – less than 1% of the total value of the projects – support the COST cooperation networks (COST Actions) through which, with € 30 million per year, more than 30,000 European scientists are involved in research having a total value which exceeds € 2 billion per year. This is the financial worth of the European added value which COST achieves.

A "bottom up approach" (the initiative of launching a COST Action comes from the European scientists themselves), "à la carte participation" (only countries interested in the Action participate), "equality of access" (participation is open also to the scientific communities of countries not belonging to the European Union) and "flexible structure" (easy implementation and light management of the research initiatives) are the main characteristics of COST.

As a precursor of advanced multidisciplinary research, COST has a very important role in the realization of the European Research Area (ERA) anticipating and complementing the activities of the Framework Programmes, constituting a "bridge" towards the scientific communities of emerging countries, increasing the mobility of researchers across Europe and fostering the establishment of "Networks of Excellence" in many key scientific domains such as: biomedicine and molecular biosciences; food and agriculture; forests, their products and services; materials, physical and nanosciences; chemistry and molecular sciences and technologies; earth system science and environmental management; information and communication technologies; transport and urban development; individuals, societies, cultures and health. It covers basic and more applied research and also addresses issues of pre-normative nature or of societal importance.

More information is available at: http://www.cost.esf.org/.

ESF provides the COST Office through an EC contract. COST is supported by the EU RTD Framework programme.

8. Green, P.E.: Optical Networking Update. IEEE Journal on Selected Areas in Communications 14(5), 764–779 (1996)
9. Chlamtac, I., Ganz, A., Karmi, G.: Lightpath Communications: An Approach to High Bandwidth Optical WANs. IEEE Transactions on Communications 40(7), 1171–1182 (1985)
10. Poor, H.: An Introduction to Signal Detection and Estimation. Springer, New York (1985)
11. Wauters, N., Demister, P.: Design of the Optical Path Layer in Multiwavelength Cross-Connected Networks. IEEE Journal on Selected Areas in Communications 14(5), 881–892 (1996)
12. Ramaswami, R., Sivarajan, K.N.: Routing and Wavelength Assignment in All-Optical Networks. IEEE Transaction on Networking 3(5), 489–500 (1995)
13. Banerjee, D., Mukherjee, B.: A practical Aproach for Routing and Wavelength Assignment in Large Wavelength-Routed Optical Networks. IEEE Journal on Selected Areas in Communications 14(5), 903–908 (1996)
14. Ali, M., Elie-Dit-Cosaque, D., Tancevski, L.: Enhancements to Multi-Protocol Lambda Switching (MPlS) to Accommodate Transmission Impairments. In: GLOBECOM '01, vol. 1, pp. 70–75 (2001)
15. Ramamurthy, B., Datta, D., Feng, H., Heritage, J.P., Mukherjee, B.: Impact of Transmission Impairments on the Teletraffic Performance of Wavelength-Routed Optical Networks. IEEE/OSA J. Lightwave Tech. 17(10), 1713–1723 (1999)
16. Tomkos, I., et al.: Performance Engineering of Metropolitan Area Optical Networks through Impairment Constraint Routing. OptiComm (2004)
17. Zsigmond, S., Németh, G.Á., Cinkler, T.: Mutual impact of physical impairments and grooming in multilayer networks. In: Tomkos, I., Neri, F., Solé Pareta, J., Masip Bruin, X., Sánchez Lopez, S. (eds.) ONDM 2007. LNCS, vol. 4534, pp. 38–47. Springer, Heidelberg (2007)
18. Bergano, N.S., Kerfoot, F.W., Davidson, C.R.: Margin measurements in optical amplifier systems. IEEE Photon. Technol. Lett. 5, 304–306 (1993)
19. Agrawal, G.P.: Fiber-Optic Communication Systems. Wiley, New York (1997)
20. Ramamurthy, B., Datta, D., Feng, H., Heritage, J.P., Mukherjee, B.: Impact of Transmission Impairments on the Teletraffic Performance of Wavelength-Routed Optical Networks. IEEE/OSA J. Lightwave Tech. 17(10), 1713–1723 (1999)
21. Chen, C.J.: System impairment due to polarization mode dispersion. In: Proc. Optical Fiber Conference and Exhibit (OFC), paper WE2-1, pp. 77–79 (1999)
22. Kissing, J., Gravemann, T., Voges, E.: Analytical probability density function for the Q factor due to pmd and noise. IEEE Photon. Technology Letters 15(4), 611–613 (2003)

7 Performance Issues in Optical Burst/Packet Switching

D. Careglio (chapter editor), J. Aracil , S. Azodolmolky , J. García-Haro, S. Gunreben, G. Hu, M. Izal, A. Kimsas, M. Klinkowski, M. Köhn, E. Magaña, D. Morató, P. Pavón-Mariño, J. Perelló, J. Scharf, S. Spadaro, I. Tomkos, A. Tzanakaki, and J. Veiga-Gontán

Abstract. This chapter summarises the activities on optical packet switching (OPS) and optical burst switching (OBS) carried out by the COST 291 partners in the last 4 years. It consists of an introduction, five sections with contributions on five different specific topics, and a final section dedicated to the conclusions. Each section contains an introductive state-of-the-art description of the specific topic and at least one contribution on that topic. The conclusions give some points on the current situation of the OPS/OBS paradigms.

7.1 Introduction

Optical Burst Switching (OBS) [84] and Optical Packet Switching (OPS) [16] have arisen as an alternative to low-flexible wavelength switching network and are still gaining considerations in the research community.

The principal design objective for an OBS/OPS network is that aggregated user data is carried transparently as an optical signal, without O/E/O conversion. This optical signal goes through the switches that have either none or very limited buffering capabilities. Besides, the control information is carried separately from the user data either in time (OPS) or in space (OBS). In such a network the wavelengths are temporally utilised and shared between different connections. It increases network flexibility and its adaptability to the bursty characteristics of IP traffic.

An OBS/OPS network consists of a set of electronic edge nodes and optical core nodes connected by WDM links (see Fig. 7.1). At the edge nodes, client packets of the same forwarding equivalence class are assembled into containers (called bursts in OBS and packets in OPS). This process is usually called burstification or packetisation. After transmission through the network towards their destination the containers are disassembled at the egress and the original client packets are forwarded to the client network. Each container is composed of a data payload (usually also referred simply as burst or packet) and a header packet (HP). The HP is generated when the burstification process is finished and carries all the information necessary to discriminate the burst or packet inside the network, like for instance, the traffic class or its length. Inside the network the control informa-

I. Tomkos et al. (Eds.): COST 291 – Towards Digital Optical Networks, LNCS 5412, pp. 189–235, 2009.

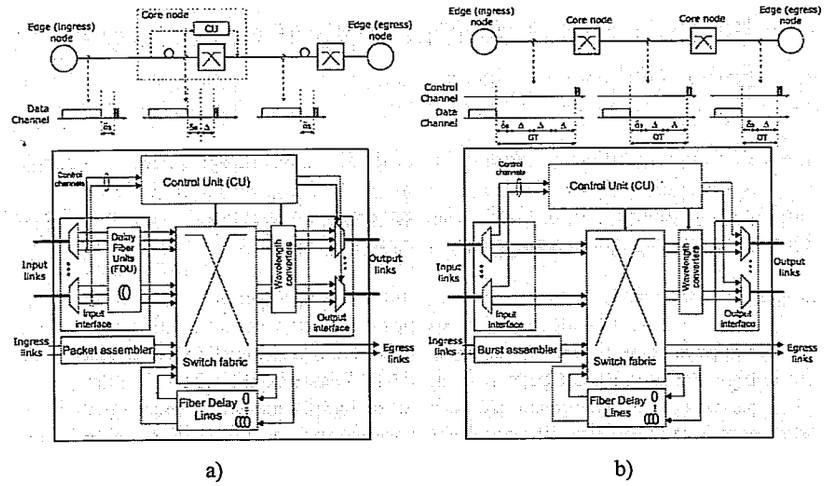


Fig. 7.1. a) OPS node and network architecture, b) OBS node and network architecture. δ_s is the switching time, Δ is the processing time, and OT is the offset time (only for OBS).

tion is processed electronically, whilst the data payload is transmitted all-optically, without optical to electrical conversion.

It has to be mentioned that in the case of OBS network, two different signalling protocols have been proposed adapting the ATM block transfer (ABT) standard designed for burst-switching ATM networks [47]:

- *Tell-and-Wait* (TAW) signalling based on delayed transmission [29]. The TAW protocol, which is recognised sometimes as a two-way signalling protocol, performs an end-to-end resources reservation with acknowledgment in advance of the burst transmission.
- *Tell-and-Go* (TAG) signalling based on immediate transmission [84]. The TAG protocol operates with a one-way signalling and it allocates transmission resources on-the-fly, a while before the burst payload arrives to a node.

The majority of research attentions are put on the one-way signalling model since two-way signalling protocols may present some concerns on the latency produced during the connection establishment process. For this reason this chapter only focus on an OBS network adopting the TAG signalling scheme, which is also the solution adopted in OPS networks.

According to this scheme, each core node must process on the fly the control information. In OPS network (Fig. 7.1(a)), the HP is usually time separated from the optical packet by a guard-time in the order of tens of nanoseconds which helps the extraction of the HP from the optical packet. In OBS network (Fig. 7.1(b)), the HP is delivered to the core node with some *offset time* prior to its burst data payload. While in the OBS network, the offset time is introduced in order to give time for both processing the control information and reconfiguring the switching matrix, in OPS this delay time is supplied by the fibre delay unit introduced at the input interface which delays the arrive of the optical packets.

Once received at the core nodes, the HP is processed in an electronic *controller*. The controller performs several functions, among others the burst *forwarding* and *resources reservation*. The forwarding function, which is related to the network *routing*, is responsible for determination of an output link (port) the data container is destined to. The resources reservation function makes a booking of a wavelength in the output link for the incoming data container. In case the wavelength is occupied by another burst a *contention resolution* mechanism, if exists, is applied. In case no resources are available for the incoming data, it is lost. After the data transmission is finished in a node the resources can be released for other connections.

Briefly, the main differences between OPS and OBS are:

- OPS uses short data containers (optical packets in the order of one to tens of microseconds), the HP (the control information) is attached at the head of the data packets and therefore both (control and data) use the same channel (i.e., in-band control), and finally the switching and control elements must be able to operate very fast (less than one microseconds).
- OBS uses large data containers (optical bursts in the order of tens to thousands of microseconds), the HP is transmitted out-of-band in a separate channel than the data bursts (but a close time relationship is required between control and data), and less time demanding are required for switching and control elements (tens to hundreds of microseconds).

It has to be noticed that the time demanding of the switching and control operations is a consequence of the length of the data containers; shorter data containers require faster operations in order to service the faster arrival rate and to optimize the utilisation of the channel capacity.

In summary the OBS/OPS paradigms support highly dynamic traffic in future networks. By switching on a burst/packet level in the optical data plane it provides on the one hand a much greater flexibility than a network based on circuit switching. With processing of information in the electrical domain, they avoid on the other hand severe technological challenges as for example optical signal processing.

The rest of the chapter summarises the research activities on OPS and OBS carried out by the COST 291 partners in the last 4 years. In the following, we include five sections with contributions on five different specific topics, namely OBS/OPS performance (Section 7.2), burstification mechanisms (Section 7.3), QoS provisioning (Section 7.4), routing algorithms (Section 7.5) and TCP over OBS networks (Section 7.6). Each section contains an introductive state-of-the-art description of the specific topic and at least one contribution on that topic.

Section 7.7 concludes the chapter with some discussions on the current situation of the OPS/OBS paradigms.

Some other aspects such as interoperability with control plane, physical layer constraints, burst switch architectures, test-beds implementation and verification, are not discussed in this chapter. A survey on OBS networks covering some of these issues is presented in [3].

7.2 OBS/OPS Performance

7.2.1 Introduction and State-of-the-Art

Two operations mainly determine the performance of the OBS/OPS networks: resource reservation and contention resolution.

The resources reservation process concerns the reservation of resources necessary for switching and transmission of data containers from input to output port. The resource reservation starts from the setup and finishes after the resource release. Both resources setup and release can be either explicit or estimated. Different resources reservation algorithms have been proposed adopting the above rules:

- *Just-In-Time* (JIT) [100] – performs an immediate resource reservation. It checks for the wavelength availability just at the moment of processing of header packet.
- *Horizon* [96] – performs estimated setup and resources release. It is based on the knowledge of the latest time at which the wavelengths are currently scheduled to be in use.
- *Just-Enough-Time* (JET) [105] – performs estimated setup and resources release. It reserves resources just only for the time of data transmission.

JET is one of the most efficient mechanisms, with improved data loss probability when comparing to other algorithms. A disadvantage is its high complexity compared to the O(1) runtime of Horizon and JIT [14].
The search of the resources can be based on several policies being the simplest ones based on random or round-robin. More advanced policies [101] are:

- *Latest Available Unscheduled Channel* (LAUC), which is a Horizon-type algorithm, keeps a track of the latest unscheduled resources and searches for a wavelength with the earliest available allocation;
- *Void-Filling* (VF), which is a JET-based algorithm, keeps a track of the latest unused resources and allows putting a data container into a time gaps before the arrival of a future scheduled one. VF algorithms achieve better performance than Horizon-based ones, however, at the cost of high processing complexity.

The resources available for the reservation depend on the capabilities of the nodes. Indeed, in case two or more containers pretend to use the same resource, a contention resolution must be applied. Two factors complicate the contention resolution: unpredictable and low-regular traffic statistics, and the lack of optical random access memories. The contention can be resolved with the assistance of following mechanisms:

- *Wavelength conversion* (WC) [20] – converts the frequency of a contending data container all-optically to other, available wavelength;
- *Deflection routing* (DR) [11] – forwards a data container spatially, in the switching matrix, to another output port;

- *Fibre delay line* (FDL) *buffering* [16] – operates in time domain and resolves the contention by delaying the departure of one of data containers by a specific period of time.

In case none of mechanisms can resolve the contention, the data container is dropped.

The wavelength conversion is natural way to resolve contention. A drawback of this mechanism, however, is high cost of WC devices, especially, in case of a full-wavelength conversion, which is performed in wide frequency range. Some solutions make use of limited or shared wavelength conversion capabilities (e.g., [26]).

Application of deflection routing is almost cost-less since no additional devices are necessary for this mechanism. On the other hand, it was shown that deflection routing can improves network performance under low and moderate traffic loads whilst it may intensify data losses under high loads [110]. Another drawback that has to be managed properly is the out-of-order arrival.

Even if one of the principal design objectives was to build a buffer-less network, the application of FDL buffering is considered as well. Both feed-forward and feed-back FDL buffer architectures can be used [45]. In [32] it was shown that combined application of FDL buffering with WC can significantly reduce data loss probability. Some of these results are illustrated in Section 7.6.

Several analytical studies have been proposed to model the behaviour of the resource reservation and contention resolution in OBS/OPS nodes (e.g., [2,9]). Section 7.2.2 studies the accuracy on the use of balking models to analytically estimate the blocking probabilities in OBS nodes that use Fibre Delay Lines (FDLs).

Section 7.2.3 compares the two different switch architectures for OPS nodes, namely Input-Buffered Wavelength Routed (IBWR) switch and Output Buffered (OB) switch.

To enhance the performance of the OBS networks, some hybrid approaches have been proposed employing more than one switching paradigm like Optical Burst Transport Network (OBTN) [34], Overspill Routing in Optical Networks (ORION) [97] or Optical Migration Capable Networks with Service Guarantees (OpMiGua) [6]. Section 7.2.4 presents a comparison between a generic OBS node and the OpMiGua node by means of a qualitative and quantitative analysis. In order to achieve a maximum of comparability both models are chosen as similar as possible and especially are fed with identical traffic.

7.2.2 On the Use of Balking for Estimation of the Blocking Probability for OBS Routers with FDL Lines

Burst blocking probability is the primary performance measure for OBS networks. Typical approach to reduce blocking probability is increasing the time during which an incoming request can be satisfied. This is usually made by storing the packet to be served in memory waiting for delivery at a later time. But since optical buffering is not available at the moment, nor it is a foreseeable technology that

will appear in the close future, optical switch designers resort to alternate solutions such as the Fibre Delay Lines (FDLs). Due to the limited delay availability, a buffered burst may be dropped if the output port/wavelength occupation persists when the burst is to exit the FDL.

Typical approaches to this system assume N input and output ports c wavelengths per port and full wavelength conversion capability. Let us assume that the c wavelengths of an output port are occupied (namely the output port is blocked). An arrival to the system that finds the output port blocked will not enter an FDL *if the delay provided by the FDLs is not large enough to hold the burst during the system blocking time*; namely if the output port residual life is larger than the delay provided by the fibres. A queuing system in which arrivals decide on whether to enter the system based on the system state (number of users, current delay, etc) is called a *balking* system or a system with discouraged arrivals [37]. For instance, an *M/M/c/K* system falls within this category, since arrivals will not enter the system if K customers are already inside it.

We describe the system as a continuous-time discrete Markov chain that represents the number of bursts in the output port (c servers and FDLs). The balking model incorporates the probability that a burst is dropped, i.e. the probability that a burst does not enter the system because the FDL is too short to hold the burst for the system residual life, into transition rates of states with index $i >= c$, as shown on Fig. 7.2.

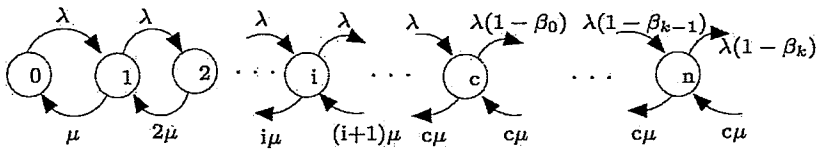


Fig. 7.2. *{Xt, t > 0}*, number of bursts in the output port

The probabilities β_k in a system with FDLs of length L are $\beta_{n-c} = P(T_n > L)$. They depend on T_n the residual life of state n which is the sum of the residual life of the blocked state and the departure time of every previous burst in the FDL. T_n can be calculated as a close expression for a Poisson-distributed arrival and burst length system. From this expression the steady state probabilities π_n for every state can be expressed as seen in (2). This is the model that has been proposed in [9,67].

On [72] we describe simulations performed to check the model on scenarios of 10 Gbps wavelengths in number c from 8 to 128. The burst average size was set to 15 kBytes, which is the average file size in the Internet as reported by [27], yielding a transmission time $E[X]$ = 12.288 μs. Switching times will be assumed to be negligible, since SOA-based switches achieve switching times in the vicinity of nanoseconds [15,68,71]. Finally, each simulation run consists of 10^8 burst arrivals.

We compared simulation results to theoretical results from the model. We found discrepancies in blocking probability *P(blocking)* versus the maximum FDL delay normalized by the time to transfer average burst $D_{max}/E[X]$ (see Fig. 7.3(a)). For low delay values it can be approximated accurately by the Erlang-B formula as expected. However, as D_{max} increases, theoretical blocking probability differs from simulation results.

The hypothesis of the balking model is checked to explain the discrepancy. The discrepancies can be traced to the calculation of β_k. It turns out that the probabilities β_k don't accurately model simulated values and this translates to theoretical state probabilities π_n which don't fit simulated values either. See Fig. 7.3(b) for example comparisons of theoretical β_k and π_n against simulation observed values, for a number of wavelengths equal to 64. Both values (β_k and π_n) take part in product form on the calculation of the loss probability. Fig. 7.3(b) also shows this product $\beta_k\ \pi_n$. The discrepancy in the discouraged arrival probability and state probabilities happen precisely for high occupancy states with small probabilities of occurrence. However, those are the states where losses take place. Therefore, the deviation from the analytical to the real values in that region of the state-space produces the misbehaviour of the loss probability shown in Fig. 7.3(a).

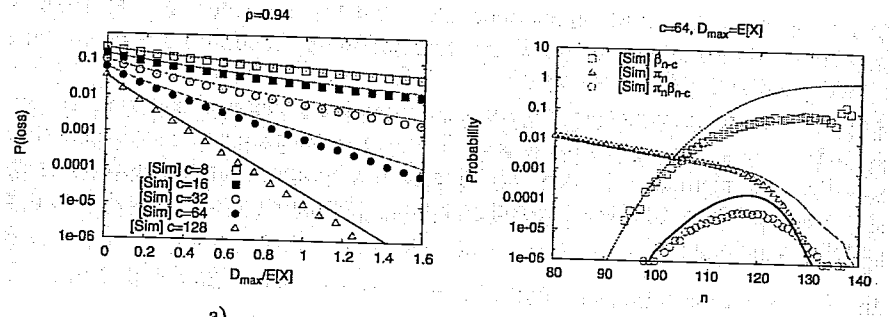


Fig. 7.3. Comparison of simulated and theoretical values, a) Burst dropping probability versus normalized FDL length, b) Comparison between the state probabilities (π_n) and the discouraged arrival probabilities (β_k)

Results in [72] show that the discrepancy between analytical and empirical results become more significant as the loss probability is decreased. Hence, the model becomes less accurate for realistic systems of WDM technology, with a higher output degree (number of wavelengths) and lower losses.

Thus we have shown that balking model accuracy depends on the ratio between fibre delay and service time. If the ratio is large then the balking model is not accurate to derive the blocking probability. On the other hand stronger discrepancies between analytical and simulation results are observed as the number of wavelengths per port increases. But precisely, the foreseeable technological evolution is towards hundreds of wavelengths.

7.2.3 A Performance Comparison of Synchronous Slotted OPS Switches

This contribution surveys the work in scheduling design and performance evaluation of OPS switching architectures, for synchronous slotted traffic. This means, switching nodes where traffic is composed of fixed size optical packets, which are aligned at switch inputs by means of synchronization stages. Results for fixed size and aligned traffic are a performance upper bound, when compared to the asynchronous and/or variable length traffic.

Two types of switching fabrics are studied: Input-Buffered Wavelength-Routed switches [116] (Fig. 7.4) and the OPS switching fabrics able to emulate output buffering (i.e. the KEOPS switch [38], the Output-Buffered Wavelength-Routed switch [116] or the space switch [15]).

IBWR switch is a more cost-effective and scalable architecture, when compared to output buffered fabrics, at a cost of a lower performance because of internal contention. The schedulers included in the comparison are:

- IBWR switch: The IBWR switch is evaluated with two parallel schedulers: (i) I-PDBM [86] scheduler which does not preserve packet sequence, and (ii) OI-PDBM scheduler, which preserves packet sequence at a cost of adding a further performance penalty [36]. Both of them are improvements to the Parallel Desynchronized Block Matching scheduler (PDBM), presented in [79]. PDBM-like schedulers allow a practical implementation which permits a response time independent from switch size.
- Output-buffered switches: For the output-buffered switches and synchronous traffic, the scheduler in [80] is used. This scheduler preserves packet sequence with no performance penalty, yielding to the optimum throughput/delay performance. Output-buffered switches are a performance upper bound for other OPS switching fabrics.

In [36, 79] the performance of IBWR and output buffered fabrics are evaluated under correlated and uncorrelated traffic, for different switch sizes. The results obtained show that the performance of the IBWR switch when packet order is not

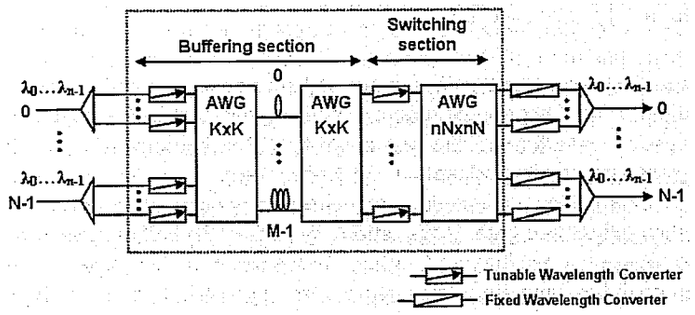


Fig. 7.4. Input-Buffered Wavelength-Routed switch (IBWR).

preserved (I-PDBM scheduler) is very close to the optimum performance given by output-buffered fabrics. A minor loss of performance appears when OI-PDBM scheduler is applied, which preserves packet order. Nevertheless, this performance loss is negligible even at medium and high loads, when the number of wavelengths per fibre is close to 32 or higher (that is, in Dense WDM networks). As an example, in most of the occasions, the same number of Fibre Delay Lines where required in IBWR switches and in output-buffered OPS architectures to achieve the target loss probability of 10^{-7}.

We conclude that the results endorse the application of the IBWR architecture in OPS networks, as a feasible competitor against less scalable output-buffered OPS architectures.

7.2.4 A Performance Comparison of OBS and OpMiGua Paradigms

While in the previous section aspects of OBS have been discussed, in this section OBS is compared with a hybrid optical network architecture named Optical Migration Capable Networks with Service Guarantees (OpMiGua) in order to determine which architecture is better suited for a given scenario. After introducing OpMiGua, we discuss qualitative differences and present results of a quantitative performance evaluation.

7.2.4.1 Optical Migration Capable Networks with Service Guarantees

OpMiGua inherently separates two different traffic classes [6]. High requirements concerning packet loss and jitter are granted by the so called Guaranteed Service class Traffic (GST). Traffic of this class is aggregated into bursts and transported in a connection oriented manner along preestablished end-to-end light paths and is given absolute priority. This ensures that there are no losses due to contention and delay jitter is minimized.

The other class with looser requirements is.Statistically Multiplexed (SM) traffic. This is handled without reservations via packet switching. Losses due to contention and delay jitter due to buffering or deflection routing are allowed. Despite this inherent separation both traffic classes use sequentially the capacity of the same wavelength.

The architecture of a basic OpMiGua node is shown in Fig. 7.5. After entering the node on a wavelength SM and GST packets are separated in the optical domain according to a specific label, e.g., polarization. While GST packets are forwarded to a circuit switch, SM packets are directed to a packet switch. After traversing the respective switches GST and SM packets directed to the same output wavelength have to be multiplexed. Thus, by inserting SM packets in-between the gaps created by subsequent GST packets, the resource utilization is increased.

In order to maintain the absolute prioritization of GST packets, the switching decision for SM packets in the depicted scenario is aware of interfering GST packets on the output wavelengths within a sufficiently large time window [7].

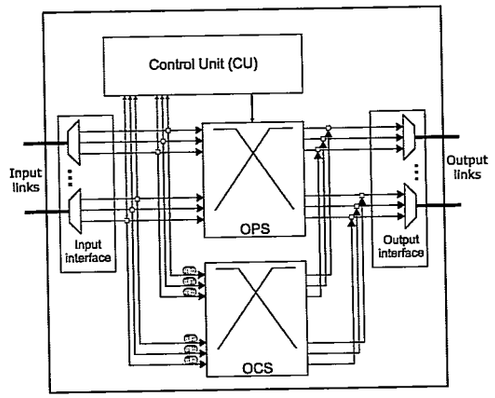


Fig. 7.5. OpMiGua node architecture

In the following we assume the packet switch as well as the circuit switch to be all-optical with full wavelength conversion but without any buffering. Also, we assume that the GST class is used for high priority (HP) and the SM class for low priority (LP) traffic. For OBS, we assume QoS differentiation for two traffic classes, i.e. high priority and low priority, by Offset Time Differentiation (see Section 7.4.2.1 for details on his behaviour).

7.2.4.2 Qualitative Comparison of OBS and OpMiGua

Comparing the two architectures, two main differences can be seen, that have an impact on the system performance. First, while in OBS all traffic is aggregated into bursts at the network ingress, in OpMiGua only the HP traffic is aggregated. Second, while in OBS all traffic shares all wavelengths, in OpMiGua each HP packet is transported on an end-to-end wavelength and only LP traffic can use all wavelengths – in the ingress as well as each core node.

In terms of delay, for reasonable load the delay of HP traffic is comparable in OBS and OpMiGua whereas the delay of LP traffic is higher in OBS. In OpMiGua, the delay of HP is due to three factors: delay in burst assembler, delay in each core node to have absolute priority of HP over LP, and delay due to the serialization of HP bursts into limited number of wavelengths; LP traffic is not aggregated in OpMiGua, thus it is only marginally delayed at the network ingress while the delay in core nodes depends only on the realization of the switching. In OBS, both HP and LP traffic classes are aggregated – thus delayed – and need to be delayed by the offset time; in contrast, the use of all wavelengths for HP bursts may reduce their waiting time.

In terms of delay jitter, it depends on the node architecture – e.g., whether processing delay is compensated by delay lines or by offset times – as well as on contention resolution strategies – whether FDLs and deflection routing is applied

or not. Both aspects have impact on HP traffic as well as on LP traffic. Accordingly, in the network the delay jitter of HP traffic is usually higher in OBS than in OpMiGua whereas the delay of LP traffic is almost comparable.

In terms of network capacity, as in OpMiGua high priority traffic is only circuit switched, direct end-to-end wavelengths are necessary for each node pair exchanging HP traffic. Thus, a full mesh of wavelength channels is needed under the assumption that every node exchanges HP traffic with each other. In contrast, in an OBS network the lower bound is a single wavelength.

7.2.4.3 Quantitative Comparison of OBS and OpMiGua

Our approach for a quantitative comparison of the two architectures OBS and OpMiGua is to use simulation scenarios as similar as possible, which especially includes the traffic offered to both models. Traffic offered to the OBS and OpMiGua node is generated statistically identical traffic on packet level and fed afterwards to an architecture specific aggregation unit, which aggregates HP and LP packets if needed.

One commonly used metric for evaluation of architectures like OBS and OpMiGua is the packet or burst loss probability, which has the disadvantage of not considering differences in the length of lost units. We choose instead the bit loss probability (BLP) as metric, which specifies the lost traffic volume in comparison to total traffic. We consider for this metric both traffic classes in OBS and OpMiGua. However, in OpMiGua, HP traffic does not contribute to this metric as it is by definition lossless.

For the simulations we select a basic single node scenario with n incoming and outgoing fibres and w wavelengths per fiver. Traffic of both priority classes is equally distributed on all wavelengths with S giving the share of HP traffic with respect to the total traffic. Also, the traffic offered to the n output fibres is uniformly distributed. In case of OpMiGua each wavelength carries one HP connection. Packets are generated with exponentially distributed interarrival times and trimodal distributed length [17].

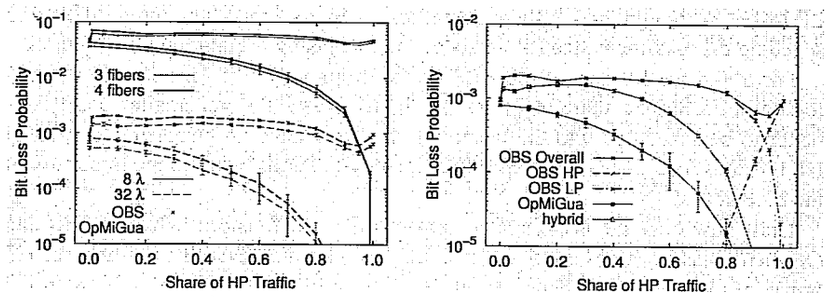


Fig. 7.6. a) BLP vs. S at load 0.6, b) BLP vs. S for n=4 and w=32 at load 0.6.

Traffic is aggregated per wavelength with a size threshold equivalent to a burst duration of 150 μs and a time threshold of 5 ms [51] (see section 7.3 for more details on burstification processes). The additional QoS offset of HP bursts in OBS we chose such that it is bigger than the maximum LP burst duration. This result in an absolute prioritization, but HP bursts may still be lost due to contention among themselves. Finally we use Just-Enough-Time (JET) and LAUC-VF as signalling and scheduling algorithm, respectively. For further details on the model, please refer to [89].

The dependency of BLP and S is shown in Fig. 7.6(a) for a fixed load of 0.6 in scenarios with 3 and 4 fibres and 8 and 32 wavelengths per fibre. At load 1 the mean generated traffic amount per time is equivalent to the maximum transmission capacity of the system. It can be seen that the BLP drops with increasing number of wavelengths. Furthermore the number of fibres has only a very small influence. Last but not least the BLP for OpMiGua is lower than that for OBS.

However, there are obvious differences in the behaviour of OBS and OpMiGua. The BLP of OpMiGua is monotonically decreasing with increasing S. This seems reasonable as the share of lossless HP traffic increases. Fragmentation of the available phases of output wavelengths due to HP traffic is not a real problem for the small LP packets.

All OBS curves show the same basic behaviour, but this is totally different to OpMiGua. Therefore it is exemplarily explained for the scenario n=4 and w=32, which is also depicted in Fig. 7.6(b). Furthermore, BLP is broken down into the parts caused by losses of LP and HP traffic ("OBS-LP" and "OBS-HP").

BLP for S=0 and S=1 should be nearly identical in case of OBS as the offset does not matter anymore if all bursts belong to the same traffic service class. The simulations clearly confirm this expectation.

For very small values of S the completion of HP bursts is mainly triggered by the timeout criterion, which results in small bursts. These small bursts fragment the phases during which a maximum size LP burst can be scheduled. This scheduling is not always possible and in comparison to S=0, where this fragmentation does not occur, the BLP is higher.

In the range S=0.2 to S=0.8 the BLP stays rather constant and originates only of LP losses. Although the LP share decreases it becomes more and more difficult to schedule the maximum size LP bursts due to increasing occupation by HP bursts.

For S>0.8 the LP part of the BLP traffic drops very fast. Besides the obvious reason of decreasing share of LP traffic, the LP bursts also get smaller and by this better to be scheduled into the voids. On the other hand an increasing amount of HP traffic is lost. These two trends in opposite directions result in the minimum of the BLP at 0.95.

Until now only the accumulated impact of the differences between OBS and OpMiGua has been observed and it is unclear to which extend the smoother HP traffic of OpMiGua influences the BLP. Therefore the OBS node is fed with HP traffic having the same characteristics like in case of OpMiGua. Nevertheless this hybrid scenario is rather theoretical, as it is impossible to guarantee this lossless HP traffic within an OBS network scenario.

The resulting BLP can also be seen in Fig. 7.6(b). While this BLP shows at small S more similarities to OBS, it finally behaves like OpMiGua and goes to zero. The sharp increase for S>0 is not as big as for OBS. The reason is that in this scenario less HP bursts are produced. However these bursts are longer as the HP traffic amount is still the same. Remaining differences to OpMiGua, which are in the order of one magnitude, are due to the aggregation of LP traffic.

7.2.4.4 Conclusions

With OBS and OpMiGua we compared two transport network architectures with QoS support for two traffic classes. Based on the current technological development status OBS has less stringent requirements, as switching is done on a bigger granularity.

With respect to delays, the predominant part (besides propagation) originates from aggregation in ingress nodes. Here OpMiGua might have a disadvantage in case of very bursty high priority traffic. On the other hand in OBS high priority traffic has an additional delay due to the offset between header control packet and burst.

Furthermore, for the investigated scenario OpMiGua is better suited. Although traffic generated for both models is statistically identical, traffic fed to the nodes itself shows differences due to absence of LP traffic aggregation and one single destination per wavelength for HP traffic in case of OpMiGua. Observed performance advantages of OpMiGua are caused by these two factors and the difference generally increases with higher HP traffic share.

7.3 Burstification Mechanisms

7.3.1.1 Introduction and State-of-the-Art

The architecture of a typical OBS edge router is depicted in Fig. 7.7. The switching unit forwards incoming packets to the burst assembly units. The packets addressed to the same egress node are processed in one burst assembly unit. There is one designated assembly queue for each traffic class.

Burstification (also known as burst assembly) algorithms can be classified as timer-based (e.g., [35,102]), size-based (e.g., [76,98]), and hybrid timer/size-based (e.g., [107]). In the timer-based scheme, a timer starts upon the arrival of the first packet to an empty queue, i.e. at the beginning of a new assembly cycle. After a fixed time (T_{Thr}), all the packets arrived in this period are assembled into a burst. In the threshold-based scheme, a burst is sent out when enough packets have been collected in the assembly queue such that the size of the resulting burst exceeds a threshold of S_{Thr} bytes. In the *hybrid* algorithm, a burst can be sent out when either the burst length exceeds the desirable threshold or the timer expires.

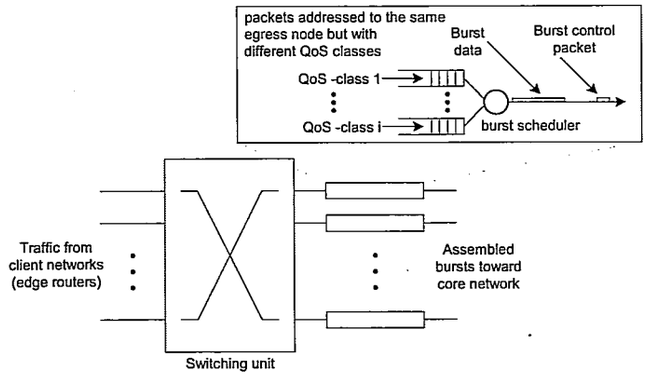


Fig. 7.7. Architecture of an OBS edge node.

Recently, it has been shown that the use of fixed thresholds in burstification algorithms may lead to some performance degradation since they are not flexible enough to take into account the actual traffic situation. In fact, considering that incoming traffic is in general strongly correlated traffic such as TCP or long-range-dependent traffic [23,54,61], the burstification processes based on fixed thresholds are not able to respond to the traffic changes accordingly. Several adaptive burstification algorithms have been proposed to ameliorate this situation [12,85] which can better respond to traffic changes and can provide better performance.

One example is illustrated in Section 7.3.2 where the case of a timer-based burstfication algorithm is analyzed. Given a burstifier that incorporates a timer-based scheme with minimum burst size, bursts are subject to padding in light-load scenarios. Due to this padding effect, the burstifier normalized throughput may be not equal to unity. The results, obtained using input traffic showing long-range dependence, motivate the introduction of adaptive burstification algorithms, which choose a timeout value that minimizes delay, yet they keep the throughput very close to unity.

On the other hand, the burstification, which is executed at the edge nodes, can substantially change the client traffic characteristics and lead to significant improvements to the network performance if the long-range dependence is alleviate. A number of recent publications have studied the traffic characterization of the burstification. The statistics for the size and interarrival time of bursts from the assembly are investigated in [22, 59]. The impact of burstification on the self-similarity level of the data traffic is studied in [42, 48, 103, 107]. A complete analysis is investigated in Section 7.3.3 where the impact of timer- and size-based burstification algorithms on the self-similarity level of the output traffic is reported. Both static and adaptive algorithms are examined and the performance impact of the burstification algorithms in terms of burst assembly delay and its jitter is assessed. The study has shown that the burst assembly mechanism at the OBS

edge router reduces the self-similarity level of the output traffic and that this reduction depends on the parameters of the algorithm. The results reveal that the proposed adaptive burst assembly algorithm performs better comparing to its non-adaptive counterpart.

7.3.2 Delay-Throughput Curves for Timer-Based OBS Burstifiers with Light Load

OBS proposals are in part motivated by the inability to switch optical paths fast enough to be done on a per-packet basis. This problem is solved by gathering bursts of packets to be switched to the same destination, but to keep a low enough rate of switching a minimum burst size use to be proposed as well. This leads to padding short bursts in order to keep this minimum size in timer-based burst gatherers. Padding will not be likely to occur in medium to heavily loaded OBS networks using a timer-based burstifier. However, a light load scenario will potentially produce many bursts with a number of packets below the minimum burst size and padding will be necessary. But load fluctuations do happen in highly loaded networks, during weekends or due to different busy hours at different geographical locations and light-load epochs will be observed[1]. The light-load will imply that when the timer expires, all packets awaiting transmission in the burst assembly queue are transmitted along with a padding space that will add load to the network. Even if this load is not significant in the link that is generating the burst it increases also load at other links and thus it should be quantified.

On [49] this effect is analyzed. The incoming traffic (bytes per time interval) is modelled by a Fractional Gaussian Noise (FGN), which has been shown to model accurately traffic from a LAN [74]. Note that in order to calculate the throughput only the number of information bytes per burst matters and not the packet arrival dynamics. Precisely, the FGN is a fluid-flow model that provides the number of bytes per time interval only. While the small timescale traffic fluctuations are not captured by the model, the long-range dependence from interval to interval is indeed accurately portrayed.

According to our previous results in [48], for a timer-based burstifier, it turns out that the traffic arriving per time interval T_0 is a Gaussian random variable X with mean $\mu = \mu' T_0$ and standard deviation $\sigma = \sigma' T_0^H$ (being μ', σ' and H the mean, standard deviation and Hurst parameter of the traffic arrival process at one time unit time slots).

The throughput of a given burstifier is defined as the ratio between the information bits and the total bits transmitted. If the minimum burst size is b_{min}, the throughput will equal unity whenever $X > b_{min}$ and $E[X]/b_{min}$ if $X < b_{min}$. By using

[1] See for instance http://loadrunner.uits.iu.edu/weathermaps/abilene/ for daily variation of traffic in an Internet

a convenience variable $Y = min\{b_{min}, X\}$ the throughput can be expressed as $\rho = E[Y]/b_{min}$ and we derive in [49] an expression for ρ depending on input traffic parameters.

$$\rho = b_{\min}^{-1}\left(\mu - \sigma\lambda(-\alpha)\right)\varphi(\alpha) + \left(1 - \varphi(\alpha)\right) \tag{7.1}$$

where $\varphi(x) = 1/\sqrt{2\pi}\, e^{-1/2x^2}$ and $\varphi(\alpha) = \int_{-\infty}^{x} \varphi(t)dt$ are the PDF and distribution function of a normalized Gaussian random variable.

To quantify the extra load that enters the OBS backbone because of the added padding we define a new convenience variable $Z = max\{b_{min}, X\}$ that denotes the bits generated by the burstifier. Z is a truncated Gaussian variable from which we derive (in [49]) an expression for the input rate to the OBS core introduced by the burstifier

$$R = \frac{E[Z]}{T_0} = T_0^{-1}\left[b_{\min}\,\varphi(\alpha) + \left(\mu + \sigma\lambda(\alpha)\right)\left(1 - \varphi(\alpha)\right)\right] \tag{7.2}$$

Equations (1) and (2) are validated against high speed traffic from Abilene-I data set. The Abilene-I data set traces contain traffic from two OC-48 links, collected at US core router nodes and are provided by NLANR[2]. For the example we use 10 minutes worth of traffic from a 2.5Gbps link as a real-world traffic source for the burstifier. The trace selected shows an average traffic rate around 480Mbps which, assuming a 10Gbps wavelength in the OBS port, makes the utilization factor be approximately equal to 0.05. Fig. 7.8 shows equations compared to the burst process that would be generated by burstifiying the Abilene-I trace with several T_0 and b_{min} values. Similar results are obtained with synthetic FGN traffic generated with Random Midpoint Displacement algorithms that allows us to have results for broader H parameter range (Abilene-I traces have H values between 0.7 and 0.8.

Results show a negative gradient of the throughput with both the coefficient of variation (instantaneous variability) and Hurst parameter (long-range dependence). However, there is a timeout value that makes such gradient be equal to zero (as can be seen on Fig. 7.8). Such timeout value depends on the minimum burst size, the traffic load and, to a lesser extent, it also depends on the long-range dependence parameter H and the coefficient of variation c_v.

The above observation leads us to seek for an expression that provides the timeout value (T_0) for which the delay throughout curves flatten out to unity. This is beneficial to maximize the throughput at the minimum delay cost and also to decrease the network load. For the Abilene-I trace considered the increased traffic load due to padding is shown in Fig. 7.8. The effect of choosing a wrong timeout

[2] http://pma.nlanr.net/Traces/long/ipls1.html

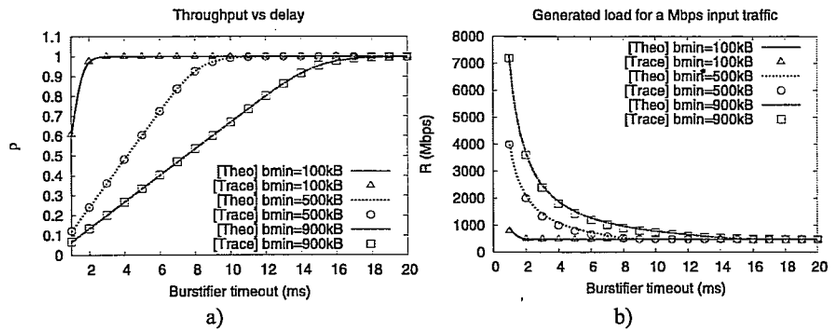


Fig. 7.8. Throughput-delay curve and input traffic to the OBS network for the Abilene-I trace.

value is very significant not only for the throughput, but also for the generated load to the OBS network.

Concerning the change rate of the traffic moments, other proposals based on link state estimation assume that the network load remains stable in timescales of minutes [92]. If that is the case, one could devise an adaptive burstifier that would offer minimum delay and maximum throughput for any given input traffic stream. The timeout value rate of change would be in the scale of minutes, which seems reasonable from a practical implementation standpoint.

In [49] we propose three different adaptive timeout algorithms and compare them for different values of the Hurst parameter H and coefficient of variation c_v. The proposed algorithms are trade-off of complexity versus accuracy. The simplest (L-estimate) requires to estimate the burstifier load $\hat{\mu}'$ and set timeout $T_0^L = {}^{b_{\min}}/_{\hat{\mu}'}$. The chosen T_0^L is the number of sampling intervals needed to fill on average a size of b_{min} at the estimated rate. The basic assumption is that the influence of the second moment and H parameter is negligible.

Using estimators for first and second moments of the traffic arriving to the burstifier we can build more accurate algorithms (LV-estimate) or (LVH-estimate) using also estimations of H parameter of the arrival process. T_0^{LVH} or T_0^{LV} are chosen as the solutions of the nonlinear problem of minimizing T_0 subject to the condition that equation (1) gives throughput values above a desired threshold (i.e. $\rho\left(\hat{\mu}', \hat{\sigma}', \hat{H}'\right) > 0.95$).

Our trace-driven analysis of the Abilene backbone shows that, for most cases of real Internet traffic, first moment estimation is enough to provide a timeout value very close to the optimum. Thus, an adaptive timeout algorithm can be easily incorporated to timer-based burstifiers, with a significant benefit in burstification delay and throughput.

7.3.3 Performance Evaluation of Adaptive Burst Assembly Algorithms in OBS Networks with Self-Similar Traffic Sources

In this work the self-similarity level of the traffic both before and after the execution of a parameterized hybrid and adaptive burstification algorithm is analyzed. The burstification algorithm is an improvement of the one presented in [12].

In order to model the realistic input traffic from the client networks, the arriving and aggregated traffic is made of superposition of fractal renewal point process as it actually describes the self-similar web requests generated by a group of users [88]. The detailed model is described in [5].

Regarding the traffic volume measurement, the approach presented in [42] is adopted focusing on packet and burst-wise measurements because the packet-wise and burst-wise analysis is important on the performance of the electronic control units in core routers. The quantitative values for the Hurst parameter estimation are reported for the proposed adaptive burstification algorithm. The performance of the OBS edge node in terms of delay and delay jitter is also investigated.

7.3.3.1 Adaptive Burstification Algorithm

Within an OBS edge router, the incoming packets (e.g. IP packets) from the client networks will be forwarded to respective queues based on the destination address of egress OBS edge router and possibly the QoS parameters, where the burstification algorithms are used to generate the burst control packet and the data burst. Then the burst control packet and the optical burst will be scheduled to the transmitter and sent out to the core network.

The packet length distribution used in our study has been reported in [101] and has been modified to ignore the packets with size larger than 1500 bytes. The average packet length of the modified distribution is 375.5 bytes and reflects the realistic predominance of small packets in IP traffic.

The main disadvantage of such static burstification algorithm is that it does not take into account the dynamism of traffic and therefore they cannot respond to the traffic changes. This adversely impacts the network performance. Therefore adaptive burst assembly algorithms are proposed to ameliorate this situation. The main idea in these burstification algorithms is to adaptively change the value of the T_{Thr} and S_{Thr}. If we assume that the network uses a static routing algorithm, then according to the link capacity, for each burst assembly queue inside the edge router, we have the following inequality:

$$\sum_{i=1}^{N} \frac{avgBL_i}{T_{Thr,i}} \leq Bandwidth \tag{7.3}$$

where $avgBL_i$ represents the average burst length in the i^{th} burst assembly queue, and the bandwidth of the link is given by *Bandwidth*. Since from (3) the value of T_{Thr} changes with the value of average burst length, we have to infer the value of

$avgBL_i$ from the traffic history. One possible approach is to take into account both the previous value of average burst length and the current sampled value ($SavgBL_i$) as expressed in the following expression [12]:

$$avgBL_i \leftarrow w_1\, avgBL_i + w_2\, SavgBL_i \tag{7.4}$$

where w_1, w_2 are two positive weights ($w_1 + w_2 = 1$). Based on (3) and (4) the two threshold values for the adaptive burstification algorithm are computed as follows:

$$T_{Thr,i} = \alpha \frac{avgBL_i\, N}{Bandwidth} \tag{7.5}$$

$$S_{Thr,i} = \begin{cases} avgBL_i & if\ avgBL_i > \beta\, E\left[L_p\right] \\ \beta\, E\left[L_p\right] & otherwise \end{cases} \tag{7.6}$$

where α, β are burst assembly factors and $E[L_p]$ is expected packet length. In order to synchronize this adaptive burst assembly algorithm with the changes in TCP/IP traffic, we set the value of $w_2 > 0.5$. This will put more weight on the recent burst size. More specifically, when a long burst is sent out (high value of $AvgBL_i$) it is very probable that TCP will send out more packets in the sequel. Therefore it is better to increase the value of both time and size thresholds to deal more efficiently with the incoming traffic. Similarly as soon as the TCP traffic is terminated or initiating a slow start stage, by giving higher weight to w_2, we also dramatically decrease the time and size threshold values. The results that we will present in next section are obtained by setting $w_1 = 0.25$, $w_2 = 0.75$. More details of this adaptive burstification algorithm is presented in [5]. Note that $Tmin_{Thr}$ is given by the following equation:

$$Tmin_{Thr} = \frac{\beta\, E\left[L_p\right] N}{Bandwidth} \tag{7.7}$$

We put a lower limit on T_{Thr} in order to keep the assembly period within a reasonable range and to prevent the burst length decreasing by too much.

7.3.3.2 Numerical Results

The simulation scenario consists of 12 client networks connected to an OBS edge router via a 10 Gbps link. The link between the OBS edge router and the core network is running at 40Gbps. The burst assembly algorithm is implemented within the OBS edge router. The incoming IP packets will be forwarded to the assembly queue associated with its egress edge router. We have defined three levels of traffic load (ρ) at the edge router: 0.3 (light load), 0.5 (medium load) and 0.7 (heavy load), which corresponds to 332889, 554816, and 776742 packets per second, respectively.

The simulation records each packet arrival at the OBS edge router regardless of the source client network and all the incoming packets comprise the aggregated input traffic. The twelve client networks are divided into four groups and the Hurst parameter of each group is set at H=0.7, 0.75, 0.80, and 0.85 respectively. Among the well-known Hurst parameter estimators, i.e. aggregated variance, R/S plot, periodogram, local Whittle and wavelet techniques [18], the wavelet analysis is used because it is robust to many smooth trends, non-stationarities, and high frequency oscillations [91]. In our simulation scenarios we have assumed that there is only one quality of service (QoS) class supported and the destination address of each IP packet is randomly selected from N egress edge routers within the core network. Thus there are N assembly queues in the OBS edge router. We choose N=1, 10, 20 in our simulation. All the traffic processes are measured at the time-scale of 100 μs. The simulation time is 6 seconds and owing to the sufficiently large queues in the OBS edge router, no packet loss is assumed.

The effect of adaptive burstification algorithm on the self-similarity level of the output traffic and burstification delay and delay jitter is studied in the following scenario. The T_{Thr} parameter of the burstification algorithm is estimated dynamically and the S_{Thr} parameter is also evaluated dynamically in favour of larger values for average burst length. Fig. 7.9 depicts the estimated Hurst parameter of the both aggregated input traffic and the optical output traffic, which is injected from edge router to the core network ($N = 1, N = 10$).

It can be seen that the burst-wise output traffic, which is the result of adaptive burstification algorithm, exhibits much lower level of self-similarity in terms of estimated Hurst parameter. In order to compare the hybrid and adaptive burstification algorithms in term of their effects on the self-similarity level of the output traffic, we set up another simulation scenario and the estimated Hurst parameter is depicted in Fig. 7.10. We have to mention that in order to make our comparison unbiased, we have focused on the distribution of bursts, which are generated according to time or size constraints and we have set the parameters for both hybrid [103] and adaptive algorithms in a way that both algorithms generate the same percentage of time-constrained and size-constrained bursts. In other words, both algorithms behave similarly as far as the distribution of bursts is concerned.

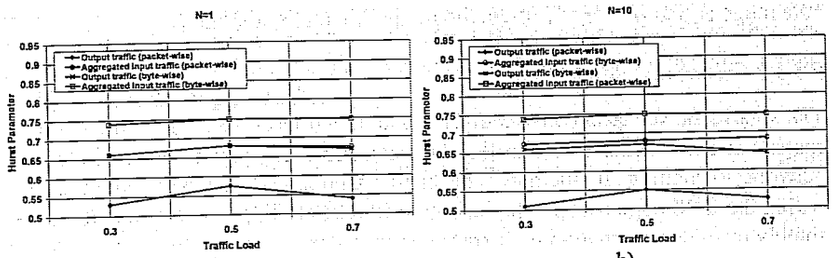


Fig. 7.9. Hurst parameter of input and output traffic for different values of load, a) $N = 1$, b) $N = 10$.

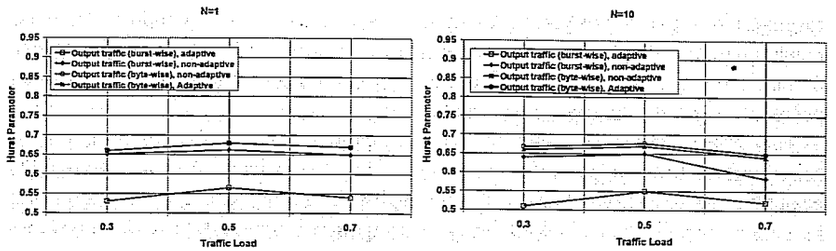


Fig. 7.10. Estimated H of output traffic for different values of load, a) N = 1, b) N = 10.

It can be observed that the byte-wise self-similarity level for both algorithms remains the same. However the burst-wise output traffic for the adaptive burstification algorithm, as expected, exhibits lower level of self-similarity in comparison to the non-adaptive (hybrid) burstification algorithm. This is due to the dynamic feature of algorithm, which adapts the value of both T_{Thr} and S_{Thr} to match dynamically with the incoming traffic.

It can be seen that the adaptive burstification algorithm performs noticeably better that its non-adaptive counterpart. In other words the burstification delay and its jitter in the adaptive algorithm are lower than the same metrics for the non-adaptive (hybrid) algorithm. This observation is valid mainly due to the mechanism that is employed in burstification algorithm. In the adaptive algorithm the T_{Thr} parameter is determined based on the weighted average of burst lengths. Thus T_{Thr} parameter tries to adapt itself according to the computed average burst length and also the recent value of burst length. Furthermore we also enforced a S_{Thr} in our burstification algorithm, which not only put a limit on burstification delay but also tries to synchronize with TCP/IP traffic as much as possible.

Summarizing, the obtained results show that the burstification algorithm at the OBS edges can be used as a traffic shaper to smooth out the burstiness of the input traffic as indicated by the noticeable reduction in the Hurst parameter. Comparing the traffic shaping capability, the adaptive outperform the non-adaptive (hybrid) algorithm in terms of reduction in Hurst parameter, burst assembly delay and burst assembly delay jitter.

7.4 QoS Provisioning

7.4.1 Introduction and State-of-the-Art

This section addresses the problem of quality of service (QoS) provisioning in OBS networks. The lack of optical memories results in quite complicated operation of OBS networks, especially, in case when one wants to guarantee a certain level of quality for high priority (HP) traffic. Indeed the quality demanding appli-

cations, like for instance real-time voice or video transmission, need for dedicated mechanisms in order to preserve them from low priority (LP) data traffic. In particular, the requirements concern to ensure a certain upper bounds on end-to-end *delay*, *delay jitter*, and *burst loss probability*.

The delays arise mostly due to the propagation delay in fibre links, the introduced offset time, edge node processing (i.e., burstification) and optical FDL buffering. The first two factors can be easily limited by properly setting up the maximum hop distance allowed for the routing algorithm. Also the delay produced in the edge node can be imposed by a proper timer-based burstification strategy. Finally the optical buffering, which in fact has limited application in OBS, introduces relatively small delays. Regarding the jitter, it depends on many factors and it is more complicated to analyze; nonetheless, since the delay can be easily bounded, its variations could be also limited accordingly. In this context the burst loss probability (BLP) metric is perhaps of the highest importance in OBS networks that operate with one-way signalling.

In a well-designed OBS network the burst losses should arise only due to resources (wavelength) unavailability in a fibre link. The probability of burst blocking in the link strongly depends on several factors, among others on the implemented contention resolution mechanisms, burst traffic characteristics, network routing, traffic offered to the network and relative class load. Since this relation is usually very complex the control of burst losses may be quite awkward in OBS networks.

Several components can contribute to QoS provisioning in OBS networks. In general, they are related to the control plane operation, through signalling (e.g., [21]) and routing (as e.g., in [10]) functions, and to the data plane operation both in edge nodes (e.g., [106]) and in core nodes (e.g., [52,53,112]). See Fig. 7.11 for a classification of the QoS mechanisms.

Although, a great number of QoS mechanisms have been proposed for OBS networks, still, only a few works study their comparative performance. In [112] some QoS scenarios with two different burst drooping principles applied, namely, a wavelength threshold-based and an intentional burst drooping are analyzed. Finally, the evaluation of different optical packet-dropping techniques is provided in [77]. In this direction Section 7.4.2 makes an extension to these

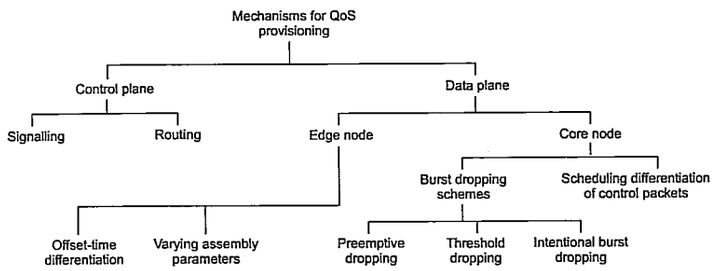


Fig. 7.11. Categories of QoS mechanisms in OBS networks.

studies. In particular, the performance of most frequently referenced QoS mechanisms, namely *offset time differentiation*, *full burst preemption* and *wavelength threshold-based dropping* are compared.

One of the more effective solutions, the *burst segmentation* mechanism [99], is analyzed in Section 7.4.3. The fact that a burst is composed by several packets makes it possible to drop *part* of a burst, so that the remaining packets may continue transmission in subsequent hops. Consequently, the use of burst segmentation provides significant throughput advantages.

7.4.2 Performance Overview of QoS Mechanisms in OBS Networks

7.4.2.1 Frequently Referenced QoS Mechanisms

In this study we focus on three mechanisms:

- *Offset time differentiation* (OTD), which is an edge node-based mechanism [106]. It assigns an extra offset-time to HP bursts in order to favour them during the resources reservation process (see Fig. 2.12a). The extra offset time, when properly setup, allows to achieve an absolute class isolation, i.e., the probability to block a HP class burst by a LP class burst is either inconsiderable or none.
- *Burst preemption* (BP), which is a core node-based burst dropping mechanism [52]. In case of the burst conflict, it overwrites the resources reserved for a LP burst by a HP one; the pre-empted LP burst is discarded (see Fig. 7.12b). In this work we consider a full preemption scheme, i.e., the preemption concerns the entire LP burst reservation.
- *Burst Dropping with Wavelength threshold* (BD-W), which is a core node-based burst dropping mechanism [112]. It provides more wavelength resources in a link to HP bursts than to LP bursts, according to a certain threshold parameter (see Fig. 7.12c). If the resource occupation is above the threshold, the LP bursts are discarded whilst the HP bursts can be still accepted.

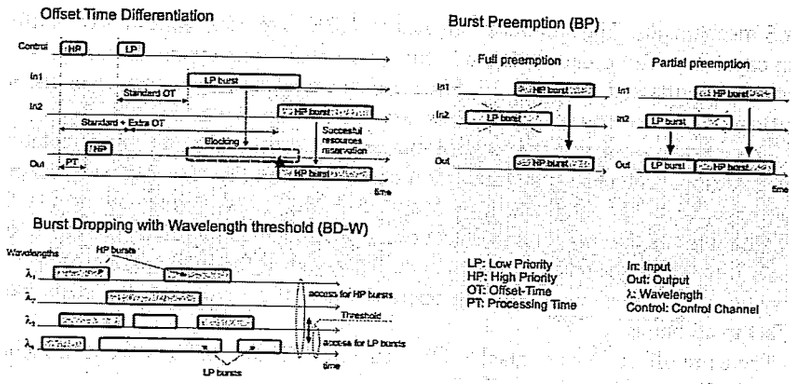


Fig. 7.12. The principle of operation of selected QoS mechanisms.

In order to gain some insight into the mechanism behaviour let us assume a Poisson burst arrival process and i.i.d. burst lengths. Under such an assumption, a burst loss probability in a link can be modelled with the Erlang loss formula (see e.g., [87]).

Both OTD and BP can be characterized by absolute class isolation. In the former, the extra offset time assures that the contention of HP bursts is only due to other HP burst reservations. In the latter, a HP burst can pre-empt whatever LP reservation and the loss of HP bursts is again only due to the wavelength occupation by other HP reservations. In both cases HP bursts compete among themselves in access to the resources and thus the HP class BLP can be estimated as $BLP_{HP} = Erlang(\alpha_{HP}\rho, c)$, where α_{HP} and ρ denote, respectively, the HP class relative load and the overall burst load and c the number of wavelengths.

The behaviour of BD-W depends greatly on its threshold (T_w) selection. Indeed, if $T_w = 0$ (i.e., no resources available for LP bursts), there is only HP class traffic accepted to the output link. Although, the mechanism achieves its topmost performance with regard to HP class and BLP_{HP} is the same as in OTD and BP, still, the LP class traffic is not served at all and $BLP_{LP} = 1$. Notice that in both OTD and BP the LP class traffic still has some possibilities to be served, in particular, if there can be found a free wavelength, not occupied by any earlier HP reservations (the OTD case), or the LP burst is not preempted (the BP case). Now, if we provide some wavelength resources for LP class traffic (i.e., $T_w > 0$), the performance of HP class will be worsening as long as HP bursts will have to compete with LP bursts. In the extreme case $T_w = c$, there is no differentiation between traffic classes and BD-W behaves as a classical scheduling mechanism. Accounting on this analysis, BD-W might require some regulation mechanisms in order to adjust the threshold value according to the required class performance and actual traffic load conditions.

7.4.2.2 Numerical Results

We set up an event-driven simulation environment to evaluate the performance of QoS mechanisms. The simulator imitates an OBS core node with no FDL buffering capability, full connectivity, and full wavelength conversion. It has 4 x 4 input/output ports and $c = \{4, 8, 16, 32, 64\}$ data wavelengths per port, each one operating at 10Gbps. The switching times are neglected in the analysis.

The burst scheduler uses a void filling-based algorithm. In our implementation, the algorithm searches for a wavelength that minimizes the time gap which is produced between currently and previously scheduled bursts. We assume that the searching procedure is performed according to a round-robin rule, i.e. each time it starts from the less-indexed wavelength. To avoid in the analysis the impact of varying offset times on scheduling operation (see [62]) we setup the same basic offset to all bursts.

The extra offset time assigned to HP bursts in OTD is equal to 4 times of the average LP burst duration. Each HP burst is allowed to preempt at most one LP burst if

no free wavelength is available in BP. The preemption concerns a LP burst the dropping of which minimizes the gap produced between the preempting HP burst and the rest of burst reservations. We establish $T_w = 0.5c$ in BD-W so that LP class bursts can access at most the half of all the available wavelengths simultaneously.

The traffic is uniformly distributed between all input and output ports. In most simulations the offered traffic load per input wavelength is $\rho = 0.8$ (i.e., each wavelength is occupied in 80%) and the percentage of HP bursts over the overall burst traffic, also called HP class relative load α_{HP}, is equal to 30%.

The burst length is normally distributed (see e.g., [107]) with the mean burst duration $L = 32$ μs and the standard deviation $\sigma = 2\ 10^{-6}$. In further discussion we express the burst lengths in bytes and we neglect the guard bands. Thus the mean burst duration L corresponds to 40 kbytes of data (at 10Gbps rate). The burst arrival times are normally distributed with the mean that depends on the offered traffic load and the standard deviation $\sigma = 5\ 10^{-6}$.

We evaluate both a data loss probability, i.e., an effective lost of data due to the burst loss, and effective throughput, which represents the percentage of data burst served with respect to overall data burst offered.

All the simulation results have 99% level of confidence.

The results of BLP_{HP} presented in Fig. 7.13(a) confirm the correctness of theoretical argumentation provided in the previous section. In particular, we can see that the performance of both OTD and BP is similar without respect to the number of wavelength in the link.

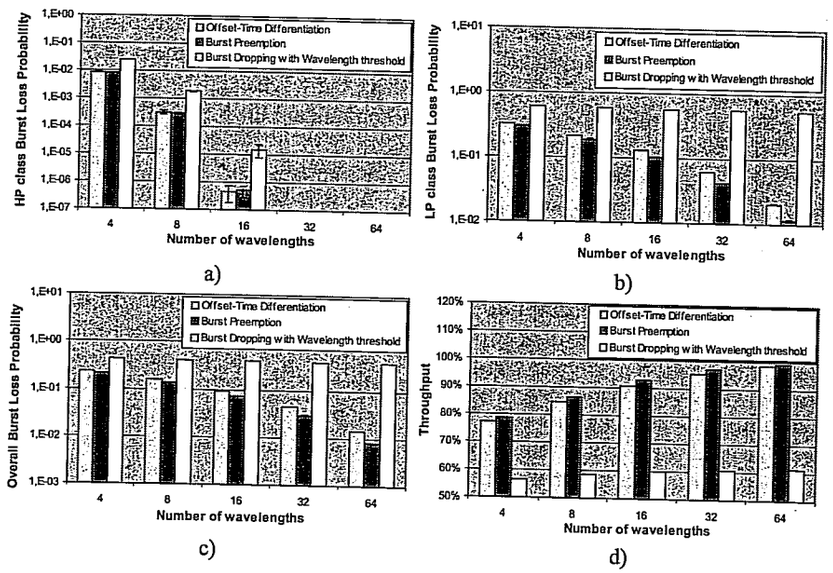


Fig. 7.13. Performance of QoS mechanism vs. link dimensioning (ρ =0.8, α_{HP}=30%), a) HP class BLP, b) LP class BLP, c) overall BLP, d) effective data throughput.

Regarding BLP_{LP} and the overall burst loss and throughput performance (Fig. 7.13(b)-(d)), the results are slightly in the favour of BP when comparing to OTD. The explanation can be found in [62], where it is shown that the scheduling operation may be impaired by the variation of offset-times, the feature which is inherent to OTD mechanism.

Finally, we can see that the BD-W mechanism exhibits very poor performance. The reason is that BD-W has effectively fewer wavelengths available for the burst transmissions than the other mechanisms, whilst at the same time it attempts to serve the same volume of burst traffic.

7.4.2.3 Discussion and Conclusions

Both OTD and BP can be distinguished by their high performance.

Although, OTD is characterized by a relatively simple operation, as long as it does not require any differentiation mechanism in core nodes, still, this mechanisms may suffer from extended delays due to extra offset times. Also, the management of extra offset times with the purpose of providing absolute quality levels might be quite complex in the network.

On the other hand, there exist several proposals that extend the functionality of BP mechanism. Particular solutions focus on providing absolute quality guarantees to individual classes of service [112], improving resources utilization [99], and supporting a routing problem [64]. An inconvenient overhead in the data and control plane due to the preemption operation can be overcome with the assistance of a preemption window mechanism [57].

Finally, we can see that BD-W offers very low overall performance in the studied scenario. It may be advisable to use this mechanism only in the networks of a large number of wavelengths in the link, where the wavelength threshold parameter could be relatively high (in order to accommodate the LP traffic efficiently) and could adapt accordingly to traffic changes.

Concluding, the BP mechanism seems to be an adequate mechanism for QoS differentiation in OBS networks, thanks to its high performance characteristics and advantageous operational features.

7.4.3 Evaluation of Preemption Probabilities in OBS Networks with Burst Segmentation

In case of partial overlapping of two contending bursts, there is no need to drop the entire burst as in the case of full burst preemption; with burst segmentation, either the head of the incoming burst or the tail of the burst in service can be dropped. It has been shown that the burst segmentation technique provides significant throughput benefits and allows for a higher flexibility in quality of service allocation, by placing packets either towards the burst tail or head [99]. In the case of two contending bursts *with the same priority* a proposed solution (in [99]) is to drop the less amount of data. If the residual length of the burst in service is larger

than the incoming burst length, then the burst in service wins the contention. The incoming burst is dropped (either entirely or partially -head-). If the residual length of the burst in service is smaller than the incoming burst length then the incoming burst wins the contention. The burst in service is segmented and the tail is dropped.

On [70] the preemption probability, or probability that the incoming burst wins the contention, is evaluated, within the same priority class. This probability is relevant for OBS network engineering for a twofold reason. First, since the incoming burst and the burst in service contend for the same resources, it is likely that they both follow the same route. Thus, due to tail dropping upon preemption, packet disordering may occur. Second, optical networks are limited by the so-called "electronic bottleneck". If preemption occurs, the optical switch must drop the tail of the burst in service and then switch the contending burst to the corresponding wavelength. This implies a processing cost not only in the optical domain but also in the electronic domain. Actually, additional signalling must be created to re-schedule bursts in the downstream nodes. Another control packet called "trailer" [99] is sent as soon as preemption happens in order to update scheduling information for the rest of OBS switches. Since this implies a processing cost, the likelihood of preemption becomes a relevant issue in OBS network performance.

Switching time is assumed to be negligible in comparison to the average burst length. Burst arrival can be assumed to be Poisson regardless of the possible long-range dependence of incoming traffic, as we have discussed in [48], but burst size will depend on the burst gathering algorithm and traffic input characteristics so several input size distributions will be considered.

Let's call (t_0, l_0) the arrival time and burst size of the first burst to arrive in a busy period, and (t_i, l_i) to subsequent bursts in the same busy period. We show in [70] that if there is a burst (t_*, l_*) in that busy period that wins the contention and preempts the first burst the time distribution of L_* is shifted to larger values in comparison to l_0. Intuitively, the preempting burst has a larger probability of high service times, in comparison to the burst in service.

$$P(l_* > x) > P(l_0 > x) \qquad \forall x > 0 \tag{7.8}$$

From this theorem it turns out that preemption is less likely to occur for the burst that wins the contention than for the first burst in a busy period (burst 0). Hence, *the preemption probability reaches a maximum with the first burst in a busy period* and this probability is given by

$$P(L > A) = \int_0^\infty P(L > x)\, dF_A(x) \tag{7.9}$$

where A is a random variable that provides the residual life of the server (wavelength). We derived it for several usual incoming burst length distributions in [70]

and provided closed expressions for $P(L>A)$ for the case of exponential and Pareto-distributed burst lengths.

We verified by simulation upper bound preemption probabilities for several burst length distribution. For example, Fig. 7.14 shows the cases for Pareto and Gaussian distributions. Note as the utilization factor decreases, the busy periods tend to be shorter. Thus, the system behaviour is closer to the best case that was assumed for the upper bound derivation, i.e. pre-emption of the first burst in a busy period. Hence, the upper bound becomes closer to the simulation results.

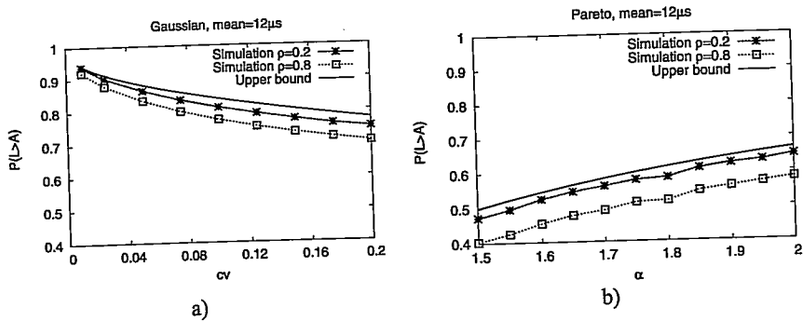


Fig. 7.14. Preemption probability, a) Gaussian and b) Pareto.

We have shown preemption probabilities are highly dependent on the burst length distribution. Hence, for the same traffic load, the burst assembly algorithm has a strong impact on the burst segmentation dynamics in the optical network core.

7.5 Routing Algorithms

7.5.1 Introduction and State-of-the-Art

In this section we concern on the problem of routing in optical burst switching networks (OBS). OBS architectures without buffering capabilities are sensitive to burst congestion. An overall burst loss probability (BLP) which adequately represents the congestion state of entire network is the primary metric of interest in an OBS network.

In general, routing algorithms can be grouped into two major classes: *non-adaptive* (when both route calculation and selection are static) and *adaptive* (when some dynamic decisions are taken) [93]. In static routing the choice of routes does not change during the time. On the other hand, adaptive algorithms attempt to change their routing decisions to reflect changes in topology and the current traffic. Adaptive algorithms can be further divided into three families, which differ in the information they use, namely *centralized* (or global), *isolated* (or local), and *distributive* routing. *Single-path* or *multi-path* routing corresponds, respectively, to

the routing scenarios with only one or more paths between each pair of nodes available. If the decision of path selection in multi-path routing is taken at the source node, thus such routing is called *source routing*. A special case of multi-path routing is *deflection* (or alternative) routing. Deflection routing allows selecting an alternative path at whatever capable node in case a default primary path is unavailable.

Static shortest path routing based on Dijkstra's algorithm is the primary routing method frequently explored in OBS networks (e.g., [107]). In such routing, some links may be overloaded, while others may be spare, leading to excessive burst losses. Therefore several both non-adaptive and adaptive routing strategies, based on deflection, multi-path or single-path routing, have been proposed with the objective of the reduction of burst congestion.

Although deflection routing can improve the network performance under low traffic load conditions, still it may intensify the burst losses under moderate and high loads [110]. Indeed the general problem of deflection routing in buffer-less OBS networks is over-utilization of link resources, what happens if a deflected path has more hops than a primary path. Hence, since first proposals were based on the static route calculation and selection (e.g., [41]), in the next step the authors proposed an optimisation calculation of the set of alternative routes (e.g., [60,65]) as well as an adaptive selection of paths (e.g., [19]). The assignment of lower priorities to deflected bursts is another important technique which preserves from excessive burst losses on primary routes [11].

Multi-path routing represents another group of routing strategies, which aim at the traffic load balancing in OBS networks. Most of the proposals are based on a static calculation of the set of equally-important routes (e.g., [81]). Then the path selection is performed adaptively and according to some heuristic [75,95] or optimised cost function [66,94]. Both traffic splitting [4,63] and path ranking [46,104] techniques are used in the path selection process.

The network congestion in single-path routing can be avoided thanks to a proactive route calculation. Although most of the strategies proposed for OBS networks consider centralised calculation of single routes [111], still some authors focus on distributed routing algorithms [31,44]. Both optimisation [63] and heuristic [28] methods are used.

In literature they are present other routing strategies that give support to network resilience by the computation of backup paths [13,44] and to multicast transmission by duplicating [43,50].

In terms of network optimisation, since an overall BLP has a non-linear character [87], either linear programming formulation with piecewise linear approximations of this function [94] or non-linear optimisation gradient methods [40] can be used. Section 7.5.2 focuses on a multi-path source routing approach and applies a non-linear optimization of BLP with a straightforward calculation of partial derivatives to improve OBS network performance.

7.5.2 Optimization of Multi-Path Routing in Optical Burst Switching Networks

In a non-linear optimization problem we assume that there is a pre-established virtual path topology consisting of a limited number of paths between each pair of source-destination nodes. Using a gradient optimization method we can calculate a traffic splitting vector that determines the distribution of traffic over these paths. In order to support the gradient method we propose straightforward formulas for calculation of partial derivatives.

7.5.2.1 Routing Scenario

We assume that the network applies source-based routing, so that the source node determines the path of a burst that enters the network. Moreover, the network uses multi-path routing where a burst can follow one of the paths given between the pair of source-destination (S-D) nodes. We assume each node is capable of full wavelength conversion and thus there is no wavelength-continuity constraint imposed on the problem.

Selection of path p is performed according to a traffic splitting factor x_p. Constraints on the traffic splitting factor are the following: 1) x_p should be non-negative and less or equal to 1, and 2) the sum of traffic splitting factors for all paths connecting given pair of S-D nodes should be equal to 1.

The reservation (holding) times on each link are i.i.d. random variables with the mean equal to the mean burst duration. Bursts destined to given node arrive according to a Poisson process of (long-term) rate specified by the demand traffic matrix. Thus traffic offered to path p can be calculated as a fraction x_p of the total traffic offered between given pair of S-D nodes.

Here vector $\boldsymbol{x} = (x_1, \ldots, x_P)$, where P means the number of all paths, determines the distribution of traffic over the network; this vector should be optimized to reduce congestion and to improve overall performance.

7.5.2.2 Formulation and Resolution Method

A loss model of OBS network based on the *Erlang fixed-point approximation* was proposed in [87]. In particular, the traffic offered to link e is obtained as a sum of the traffic offered to all the paths that cross this link reduced by the traffic lost in the preceding links along these paths.

The formulation of [87] may bring some difficulty in the context of computation of partial derivatives for optimization purposes. Therefore in [56] we propose a simplified non-reduced link load model where the traffic offered to link e is calculated as a sum of the traffic offered to all the paths that cross this link. The rationale behind this assumption is that under low link losses, observed in a properly dimensioned network, the model in [87] can be approximated by our model [56]; in [56] we can see that the accuracy of the simplified model is very strict for losses below 10^{-2}.

Having calculated the traffic offered to each link, the main steps of the network loss modelling include the calculation of burst loss probabilities on links, given by the Erlang loss formula, loss probabilities of bursts offered to paths, and the overall burst loss probability B (see [56] for detailed formulae).

From this network loss model we define cost function $B(\boldsymbol{x})$ to be the subject of optimization. The optimization problem is formulated as to minimize $B(\boldsymbol{x})$ subject to the constraints imposed on the traffic splitting factor (discussed in SubSection 7.5.2.1). Since the overall BLP is a non-linear function of vector $\boldsymbol{x}$ the cost function is non-linear as well. A particularly convenient optimisation method is the Frank-Wolfe reduced gradient method (algorithm 5.10 in [83]); this algorithm was used for a similar problem in circuit-switched networks [40].

Gradient methods need to employ the calculation of partial derivatives of the cost function. The partial derivative of $B(\boldsymbol{x})$ with respect to x_p, where p means a path, could be derived directly from the network loss formulae by a standard method involving resolution of a system of linear equations. Such a computation, however, would be time-consuming.

Therefore instead in [56] we provide a straightforward derivation of the partial derivative that is based on the approach previously proposed for circuit switched networks [55]. We have managed to simplify the model described in [55] and make the calculation of partial derivatives straightforward, not involving any iteration. The calculation of gradient in our method, therefore, is not longer an issue.

It can be shown numerically that objective function $B(\boldsymbol{x})$ is not necessarily convex. Nevertheless, under moderate traffic loads we have observed that several repetitions of the optimization program always give us the same (with a finite numerical precision) near-optimal value of B.

7.5.2.3 Numerical Results

We evaluated the performance of our routing scheme in an event-driven simulator. In order to find a splitting vector $\boldsymbol{x}$ specifying a near-optimal routing we used solver *fmincon* for constrained nonlinear multivariable functions available in the *Matlab* environment. Then we applied this vector in the simulator.

The evaluation is performed for NSFnet (15 nodes, 23 links) and EON (28 nodes, 39 links) network topologies; different numbers of wavelengths (λs) per link are considered, each transmitting at 10Gbps. The optimized routing (OR) is compared with two other routing strategies: a simple shortest path routing (SP) and a pure deflection routing (DR). We consider 2 shortest paths per each source-destination pair of nodes; they are not necessarily disjoint. In SP routing only 1 path is available. Uniform traffic matrix and exponential burst inter-arrivals and durations are considered. All the simulation results have 99% level of confidence.

In Fig. 7.15 we show B as a function of offered traffic load for different routing scenarios. We see that the optimized routing can achieve very low losses, particularly, when compared with the shortest path routing. Analytical results ('OR-an' in the figure) correspond very well to simulation results. The optimization takes about

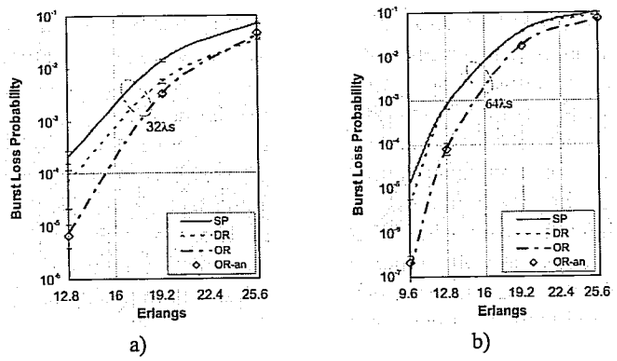


Fig. 7.15. Comparison of routing schemes a) NSFnet, b) EON.

23s and 1800s for NSFnet network (of 420 paths) and EON network (of 1512 paths), respectively, when using a non-commercial *Matlab* solver on a Pentium D, 3GHz computer.

7.5.2.4 Conclusions

In this Section we have proposed a non-linear optimization method for multi-path source routing problem in OBS networks. In this method we calculate a traffic splitting vector that determines a near-optimal distribution of traffic over routing paths. Since a conventional network loss model of an OBS network is complex we have introduced some simplifications. The references formulae for partial derivatives are straightforward and very fast to compute. It makes the proposed non-linear optimization method a viable alternative for linear programming formulations based on piecewise linear approximations of the cost function.

The simulation results demonstrate that our method effectively distributes the traffic over the network and the network-wide burst loss probability can be significantly reduced compared with the shortest path routing.

7.6 TCP over OBS Networks

7.6.1 Introduction and State-of-the-Art

TCP is today the dominant transport protocol in Internet, and it is expected to continue to be used. As TCP is not specifically designed for a particular technology, modifying the standard TCP can lead to a performance optimization in specific environments. In this direction, a great amount of novel TCP versions have been proposed for mobile networks, wireless mesh networks, and high speed networks such as optical switched networks. The development of TCP of the last years has covered three major topics: 1) making TCP more suitable for high speed environments and grid computing applications (e.g. Fast TCP and High Speed TCP), 2) making TCP more robust for non congestion events (e.g. TCP-NCR, TCP-PR and TCP-Aix), and 3) TCP for special environments and applications e.g. for wireless networks.

In the context of OBS network, the problem of having TCP has been widely studied in the literature. The design of novel specific TCP implementation are considered in [89,114,115]. Nonetheless, there is no consensus in whether a completely new TCP version is needed, and which new TCP version should be standardized within all the proposals. In order to have some benchmarking reference, TCP Reno 0 and TCP Sack [30] are generally considered as they are the most popular versions in current networks.

From the performance point of view, the main focus is put on the effect of the burstification delay on TCP behaviour [24,108,109]. In fact, the burstfication process can increase the value of the Round Trip Time and thus decrease the TCP throughput accordingly. At the same time, the high bandwidth delay product of the OBS networks contributes to enlarge faster the congestion window than in current networks and thus increase the TCP throughput [24,25]. As a consequence, the (timer and/or size) thresholds in the burstification process become important trade-offs to achieve high TCP performance [108].

On the other hand, TCP over OBS suffers of the so called False Timeout effect [109]. Due to the bufferless nature of OBS core network and the one-way signalling scheme, the OBS network is subject of random burst losses, even at low traffic loads. The random burst loss may be falsely interpreted as network congestion by the TCP layer, which is therefore forced to timeout and to decrease the sending window. Some mechanisms based on burst retransmission are proposed to alleviate the false timeout effect (e.g., [113]).

The effect of the packet reordering is addressed in Section 7.6.2 where a layered framework to measure the reordering introduced by contention resolution strategies in OBS networks is presented. In particular, characterization is based on the reordering metrics proposed by the IETF IPPM Working Group. The obtained results are twofold. First, they quantify the impact of burst reordering on TCP throughput performance, and secondly, they give insight into solving burst reordering by well dimensioned buffers.

7.6.2 Burst Reordering Impact on TCP over OBS Networks

7.6.2.1 Introduction

In this work, we follow a layered approach to study the viability of OBS as a carrier technology for TCP. Firstly, we quantify the introduced reordering at the OBS layer. With such purposes, we apply the reordering metrics presented by the IETF in [73], which provide us extensive information. First, they quantify the buffer size that should be placed at edge nodes to solve reordering at the OBS layer. This would permit the sending of already ordered packets to the IP layer, so that burst

reordering would remain transparent to TCP. Second, in the case that reordering is left to the TCP layer, they provide information about the violation of the DUP-ACK threshold due to reordering, which allows TCP performance estimation.

It is widely known the effect of packet loss on TCP. In TCP Reno [1], the sender of a TCP session is notified of a packet loss by means of duplicate acknowledgments. In this context, the TCP fast retransmit algorithm is invoked once the duplicate acknowledgment threshold (DUP-ACK) is reached. As a result, the missing packet is retransmitted and the sender's congestion window is halved, which decreases TCP throughput significantly. A similar situation occurs whether a packet becomes reordered. Note, that in the event of reaching the DUP-ACK threshold, TCP may consider a reordered packet as lost, even though it is only delayed and it would later be received.

For the sake of generality, we quantify reordering in OBS networks under several contention resolution strategies. As in [33], we deal with Conv, Defl and FDL basic strategies and combinations of them. Because the order of application of each strategy is essential, combined strategies are named by a concatenation of the former's acronyms. In particular, performance of ConvFDL, ConvDefl, ConvFDLDefl and ConvDeflFDL is also here evaluated.

7.6.2.2 *Scenario under Study*

With evaluation purposes, we implement the 16-node COST 266 reference network [69]. For simplicity, all links have the same length of 200 km, which introduces a link propagation delay of 1 ms. Network resources are dimensioned according to a static traffic demand matrix, obtained from a 2006 European population model [69]. Particularly, a total demand of 9.9 Tbps is offered to the network which corresponds to 990 Erlangs for a 10 Gbps line rate. Then, wavelength capacity is distributed in the network, so that shortest path routing leads to equal blocking probabilities on all links (i.e., dimensioning according to the Erlang model [58]). In this context, different network load situations can be achieved by overdimensioning wavelength capacity by a given factor (denoted as *overdimensioning factor* in the figures).

Regarding traffic characteristics, the burst departure process follows a Poisson process and burst length is exponentially distributed with mean 100 kbit [32]. In turn, in OBS nodes, the number of add/drop ports is unlimited and the switching matrix is non-blocking. Besides, the delay for burst control packet processing is compensated by a short extra FDL of appropriate length at the input of the node.

With contention resolution purposes, we assume one FDL per node with a certain number of wavelengths. The length of this FDL equals the mean burst transmission time, defined as the time needed to transmit an average sized burst (i.e., 10 μs for 100kbit bursts over 10Gbps data links). Note that the wavelengths of this FDL are shared and the number of wavelength converters per node is unlimited. Nonetheless, if all wavelengths are occupied upon burst arrival in the FDL, the burst is discarded.

7.6.2.3 *Simulation Results*

In this section, we evaluate the performance of the strategies *Conv*, *ConvFDL*, *ConvDefl*, *ConvFDLDefl* and *ConvDeflFDL* in an OBS scenario. We focus not only on burst loss probability, but also on introduced reordering, which harms TCP performance as well. Note that a complete characterization of reordering becomes important, especially when assessing a protocol's viability over a given network. With such an objective, the IETF IPPM working group has recently standardized a set of metrics [73] to characterize reordering effects in generic packet networks (e.g., OBS networks). In this section, three of them are selected and further quantified.

Specifically, we evaluate reordering ratio, reordering extent and 3-reordering ratio, which provide a broad view of reordering in the scenario under study. To this end, we measure burst reordering between each demand source-destination pair and we provide global network statistics. Note, that if no wavelength conversion would have been feasible in the network, our conclusions on reordering would still be valid, as Conv is applied first in all schemes. In the evaluation, we assume 8 wavelengths in the FDLs mainly due to cost and hardware integration issues. Moreover, to avoid unnecessary load and high propagation delays in the network, we limit the number of deflections to 1. Previous works demonstrate that the improvements due to further deflections are marginal [32], as long as a reasonable amount of flexibility is allowed in the network. The results have been obtained using the event-driven simulation library IKRSimlib [8].

As can be seen in [82], for high and medium loads ConvDeflFDL introduces the highest reordering, followed by ConvFDLDefl, ConvDefl and ConvFDL. However, towards low loads, all strategies behave similarly in terms of reordering ratio. Particularly, the introduced reordering ratio by Conv alone was there not evaluated. In fact, when applying this strategy all bursts travel along the same path and no buffering is used. Therefore, no reordering is introduced. Concerning burst loss probability, it was distinguished that for high loads the performance of all the strategies that use deflection routing (i.e., ConvDefl, ConvFDLDefl and ConvDeflFDL) is poor, as they overload an already highly loaded network. Nonetheless, towards lower loads, deflection (alone or combined) decrease burst loss probability rapidly, as enough network resources become available. The majority of studies coincide that in a realistic OBS scenario, burst blocking probabilities should range from 10^{-3} to 10^{-6}. Particularly, in this operating range, all strategies introduced the same reordering to the network. However, ConvDeflFDL provided the best performance regarding burst loss probability. At a first sight, this leads to the conclusion that this strategy may provide the best compromise between burst losses and introduced reordering.

In addition, we analyze the possibility to restore burst order directly at the OBS layer. Then, already ordered packets could be sent to the IP layer, so that burst reordering would remain transparent to TCP. With these purposes, a possible solution is the placement of buffers on a per flow basis at OBS edge nodes. Such buff-

ers would store incoming out-of-order bursts, waiting for the expected one to be received. In this context, the reordering extent metric provides information about the mean extent to which bursts are reordered. Therefore, this gives an idea of these buffers' size.

As depicted in [82], deflection routing technique introduces large extents, in the order of one thousand. In fact, deflected bursts transverse at least one more hop than those that go through the direct path. This accounts for an additional propagation delay of 1 ms, which is two orders of magnitude greater than the mean burst transfer time (10 μs in our scenario). Conversely, the use of buffering such as in ConvFDL introduces relatively low extents. Hence, these strategies would enable the restoration of the burst order directly at the OBS layer by means of small buffering capacities. It is noteworthy that towards low loads, the introduced extent by combined strategies tends to the former (e.g., towards low loads, ConvDeflFDL tends to ConvDefl). This is due to the fact that, in a low loaded network, contentions can be solved in the first attempt in most situations.

Until now, we have quantified the reordering ratio and introduced reordering extent for each contention resolution strategy under consideration. While the former provides a general view of what happens in the network, the latter evaluates the possibility to restore order directly in the OBS layer. Note that this information provides understanding about the origins of reordering and evaluates specific solutions to restore it. However, it does not illustrate the direct implication of reordering on TCP. It is our goal now to quantify the n-reordering burst ratio. To this end, we assume n = 3, which matches TCP Reno operation [1].

Referring again to [82], it was shown that 3-reordering ratio increases along with the overdimensioning factor. This could be due to several reasons. For low loads, deflected bursts have more possibilities to succeed, which would increase 3-reordering ratio. Moreover, for higher loads, since more reordering exists, this could decrease 3-reordering. For instance, let us assume a reordered burst. It may happen, that the following ones become also reordered, which could cause this one not to be 3-reordered. Further looking at the obtained results there depicted, it can be seen that buffering technique introduces less 3-reordering ratio than deflection, outperforming ConvFDL all the remainder strategies. For better illustration, absolute 3-reordering ratio was also evaluated in [82]. In fact, it quantifies the ratio of received packets, which become 3-reordered or more. The obtained results presented a behaviour inline with the reordering packet ratio. For high loads, differences between the strategies can be appreciated, outperforming ConvFDL the remainder ones. However, towards lower loads, in a more realistic OBS scenario, all strategies behave equally.

7.6.2.4 Impact of Burst Reordering on TCP Performance

In this section, we quantify the impact of burst reordering on final TCP throughput. Taking into account the already measured 3-reordering ratio at the burst layer in [82], we derive a worst case situation for 3-reordering packet ratio. Then, considering both burst reordering and burst loss pernicious effects, we provide a new figure of merit, called P_{FR}, which quantifies the probability to invoke *fast retransmit* algorithm in TCP Reno. Finally, as the key point of this work, we estimate the theoretical TCP throughput over the scenario under study, which allows us to conclude on its viability.

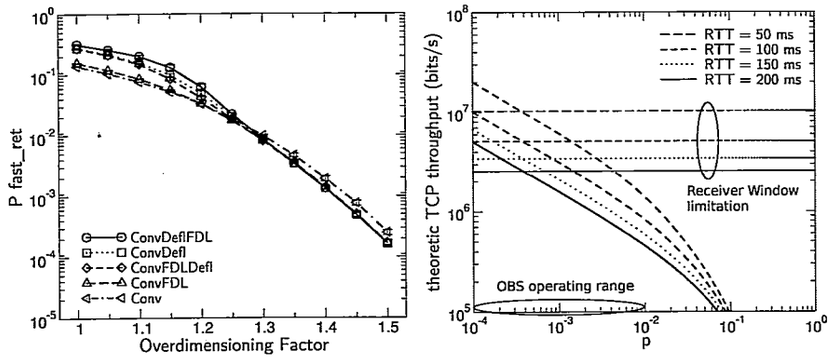
For the n-reordering packet ratio, and according to the definition presented in [73], only the first packet contained in an n-reordered burst is considered as n-reordered. Intuitively, this leads to think that an upper bound for the n-reordering packet ratio is given when exactly 1 packet of the TCP flow under study is contained in each burst. One should remind, that the n-reordering packet ratio is measured on a per TCP flow basis. Therefore, only those packets belonging to the TCP flow under study are considered (bursts can be composed of more packets arriving from different TCP sessions). To validate this intuitive assumption, analytical derivations were provided in [82]. Particularly, it was concluded that the following equation must hold to ensure the worst case assumption

$$P(N_r \geq n_r) \geq \frac{1}{n_p} P\left(N_r \geq \left\lceil \frac{n_r}{n_p} \right\rceil\right), \qquad n_p, n_r \in N. \tag{7.10}$$

where $P(N_r \geq n_r)$ denote the Complementary Cumulative Distribution Function (CCDF) of a burst to become at least n_r-reordered and n_p stays for the number of packets of the same TCP flow per burst. In [82], it was obtained the CCDF of the burst n-reordering ratio for each strategy under study. Indeed, for $n_r = 3$ and $n_r \in \mathrm{N}$, the gathered results accomplished in equation (10). This demonstrates that the assumption of having one packet of the same TCP flow under study to be contained in each burst truly contemplates the worst case scenario for the 3-reordering ratio.

This analysis allows us to estimate a worst case for the final TCP throughput, supposing that TCP runs over the network under study. According to the conclusion above, we assume that 1 packet per TCP flow is contained in each burst. In such a case, 3-reordering packet ratio coincides with the already measured 3-reordering burst ratio. Furthermore, we consider that upon contention a burst is entirely dropped. Thus, packet loss probability P_L equals to burst loss probability P_B. Note that if the receiver does not use selective acknowledgments and the sender uses the basic congestion control presented in [1], reordering has the same effect as packet loss. In fact, reordered packets which exceed the DUP-ACK threshold also trigger the fast retransmit algorithm (i.e., as if they would have been lost). Hence, whether $P(N_r^* \geq n_r)$ identifies the CCDF function of a packet to become at least n_r-reordered, the probability to invoke fast retransmit algorithm can be stated as $P_{FR} = P(N_r^* \geq n_r) + P_L$.

In Fig. 7.16(a), we depict the upper bound for P_{FR} for a DUP-ACK threshold set to 3. In particular, it is obtained as $P_{FR} = P(N_r \geq 3) + P_B$, using the results presented in [82]. As seen, for high loads, Conv and ConvFDL lead to better results, due to the lower reordering they introduce. However, for lower loads, all combined strategies provide similar performance. This is due to the fact that along this



a) b)

Fig. 7.16. a) Probability to trigger TCP fast retransmit (worst case scenario), b) Theoretic TCP throughput (bits/s) according to the model proposed in [78].

range, 3 reordering ratio dominates in front of P_L. The fact that Conv alone provides substantially worse performance demonstrates the need for additional contention resolution in OBS networks.

Up to now, several analyses have been proposed in the literature to model the steady state throughput of a TCP connection. In [78] a model, which considered both congestion avoidance phase and retransmissions caused by time out, is developed and an approximated formula for the throughput B_{TCP} of a TCP session is given.

Fig. 7.16(b) illustrates, for different RTT values, the theoretical TCP throughput according to this model. Mainly, it depicts B_{TCP} and the limitation due to the receiver limitation window, both function of p (the total packet loss probability along the path, or P_{FR} since, in this scenario, 3-reordering has the same effect as packet loss). In this way, given a certain p, the theoretical TCP throughput will be the minimum of both curves.

As mentioned earlier, OBS networks are usually dimensioned to achieve burst loss probabilities ranging from 10^{-3} to 10^{-6}. Looking at Fig. 7.16(a), a network dimensioned to achieve these values (from the results presented in [82], overdimensioning the network by 1.25 - 1.35) would experience P_{FR} values from 10^{-2} to 10^{-3}, depending on the strategy used. Observing now Fig. 7.16(b), we find that, for these p values, the performance of TCP is highly affected by the reordering introduced at the OBS layer. In fact, to assure the proper performance of TCP, p should be lower than 10^{-3}, so that the limiting factor would be the receiver advertised window, rather than the reordering introduced in the network. This demonstrates that reordering should be also considered when dimensioning an OBS network for TCP traffic. As seen, its impact on TCP is much more significant than P_L in the range of operation of typical OBS networks. Moreover, as far as TCP performance is concerned, almost all combined contention resolution strategies under study behave similarly. Although we mentioned earlier that *ConvDeflFDL* may outperform the remainder, such improvements are hidden by the fact that 3-reordering dominates in front of P_L.

7.6.2.5 *Conclusions*

In this section, we propose a layered framework to quantify the impact of burst reordering on TCP performance. First of all, we measure the reordering introduced by several contention resolution strategies. With such purposes in mind, we use the packet reorder metrics proposed by the IETF. Two different approaches to tackle reordering in an OBS scenario have been highlighted and subsequently evaluated. On the one hand, reordering can be solved directly at the OBS layer, by means of well dimensioned buffers. On the other hand, reordering can be left to higher layers, expecting this one to be solved by them.

For the former strategy, we quantify the size of the buffers which should be placed at OBS edge nodes on a per flow basis. Following this line, we find that deflection routing prohibits this solution, since the introduced extents are extremely high. Conversely, we demonstrate that buffering introduces significantly lower extents, which would, a priori, enable this strategy.

For the latter strategy, we focus on its impact on final TCP Reno performance. We propose a new figure of merit, named P_{FR}, which considers not only the pernicious effects from packet loss, but also the ones from caused by reordering. This allows us to conclude, based on the model proposed by [78] that the usual OBS operating range fits no more. On the contrary, network should be dimensioned taking into account not only burst loss probability, but also burst reordering introduced by contention resolution.

7.7 Conclusions

Ten years ago, the growth of the Internet and its bursty statistical characteristic were the main drivers to develop innovative data-centric optical transport networks. In this context the optical burst switching (OBS) and optical packet switching (OPS) technologies were proposed as promising network solutions overcoming the typical inefficiency of the circuit switching network. In fact they were designed with the aim of optimising the utilisation of the WDM channels by means of fast and highly dynamic resource allocation based on a statistical multiplexing scheme.

These ten years of research activities in OPS and OBS covered different, extensively and heterogeneous topics: novel switch architecture with no, partial or full wavelength conversion, multi-switching architecture, efficient scheduling algorithms, routing with traffic engineering capability, mechanisms to support QoS, novel TCP mechanism to enhance the random loss behaviour of the OPS/OBS networks, protection and restoration mechanisms, etc. Some of these topics have

been reviewed in this chapter. An important issue which is a hot topic of current research activity is the deployment of control plane in OBS/OPS networks. As a solution, some studies have initiated to consider a common control plane based for example on the generalised MPLS protocol (GMPLS). Having a common control plane might be desired, in particular, in the context of coexistence of different switching technologies and of the network migration towards all-optical networks. Therefore, the loop can be closed allowing the continuous deployment of optical circuit switching, OBS and OPS.

Nonetheless, nowadays OPS/OBS are still not feasible since the majority of the required optical devices are not commercially available or even not proved in laboratory. This situation creates some slowdown interest in these fields. To move up and gain insight into OPS/OBS, a more strict cooperation between interdisciplinary areas is desired: researchers in photonic material, optical communication and optical networking should dedicate efforts in defining clear requirements, recommendations and guidelines and proposing viable solutions.

References

1. Allman, M., Paxson, V., Stevens, W.: TCP Congestion Control. RFC 2581 (1999)
2. Almeida jr., R.C., Pelegrini, J.U., Waldman, H.: A Generic-traffic Optical Buffer Modeling for Asynchronous Optical Switching Networks. IEEE Communications Letters 9(2), 175–177 (2005)
3. Aracil, J., et al.: Research in Optical Burst Switching within the e-Photon/ONe Network of Excellence. Optical Switching and Networking 4(1), 1–19 (2007)
4. Argos, C.G., de Dios, O.G., Aracil, J.: Adaptive Multi-path Routing for OBS Networks. In: 9th IEEE International Conference on Transparent Optical Networks (ICTON), Rome, Italy, pp. 299–302 (2007)
5. Azodolmolky, S., Tzanakaki, A., Tomkos, I.: Study of the Impact of Burst Assembly Algorithms in Optical Burst Switched Networks with Self-Similar Input Traffic. In: 8th IEEE International Conference on Transparent Optical Networks (ICTON), Nottingham, UK, pp. 35–40 (2006)
6. Bjørnstad, S., Hjelme, D.R., Stol, N.: A Highly Efficient Optical Packet Switching Node Design Supporting Guaranteed Service. In: European Conference on Optical Communication (ECOC), Rimini, Italy, pp. 110–111 (2003)
7. Bjørnstad, S., Hjelme, D.R., Stol, N.: A Packet-Switched Hybrid Optical Network with Service Guarantees. IEEE Journal on Selected Areas in Communications (Supplement on Optical Communications and Networking) 24(8), 97–107 (2006)
8. Bodamer, S., et al.: IND Simulation Library 2.3 User Guide Part I: Introduction (2004), `http://www.ikr.uni-stuttgart.de/INDSimLib`
9. Callegati, F.: Optical Buffers for Variable Length Packets. IEEE Communications Letters 4(9), 292–294 (2000)
10. Callegati, F., Cerroni, W., Muretto, G., Raffaelli, C., Zaffoni, P.: QoS Routing in DWDM Optical Packet Networks. In: WQoSR2004 co-located with QoFIS 2004, Barcelona, Spain (2004)

11. Cameron, C., Zalesky, A., Zukerman, M.: Prioritized Deflection Routing in Optical Burst Switching Networks. IEICE Transaction on Communications E88-B(5), 1861–1867 (2005)
12. Cao, X., Li, J., Chen, Y., Qiao, C.: Assembling TCP/IP Packets in Optical Burst Switched Networks. In: IEEE Global Communications Conference (Globecom), Taipei, Taiwan, pp. 2808–2812 (2002)
13. Chen, Q., Mohan, G., Chua, K.C.: Route Optimization for Efficient Failure Recovery in Optical Burst Switched Networks. In: IEEE HPSR 2006, Poznan, Poland (2006)
14. Chen, Y., Turner, J.S., Mo, P.-F.: Optimal Burst Scheduling in Optical Burst Switched Networks. IEEE/OSA Journal of Lightwave Technology 25(8), 1883–1894 (2007)
15. Chiaroni, D., et al.: First Demonstration of an Asynchronous Optical Packet Switching Matrix Prototype for Multi-Terabit-Class Routers/Switches. In: 27th European Conference on Optical Communication (ECOC), Amsterdam, Netherlands, pp. 60–61 (2001)
16. Chlamtac, I., et al.: CORD: Contention Resolution by Delay Lines. IEEE Journal on Selected Areas in Communications 14(5), 1014–1029 (1996)
17. Claffy, K., Miller, G., Thompson, K.: The Nature of the Beast: Recent Traffic Measurements from an Internet Backbone. In: International Networking Conference (INET), Geneva, Switzerland (1998)
18. Clegg, R.G.: A Practical Guide to Measuring the Hurst Parameter. In: Thomas N. (ed.) 21st UK Performance Engineering Workshop, School of Computing Science Technical Report Series, CS-TR-916, University of Newcastle (2005)
19. Coutelen, T., Elbiaze, H., Jaumard, B.: An Efficient Adaptive Offset Mechanism to Reduce Burst Losses in OBS Networks. In: IEEE Global Communications Conference (Globecom), St. Louis, MO, USA (2005)
20. Danielsen, S., Hansen, P., Stubkjear, K.: Wavelength Conversion in Optical Packet Switching. IEEE/OSA Journal of Lightwave Technology 16(12), 2095–2108 (1998)
21. de Miguel, I., González, J.C., Koonen, T., Durán, R., Fernández, P., Tafur Monroy, I.: Polymorphic Architectures for Optical Networks and their Seamless Evolution towards Next Generation Networks. Photonic Network Communications 8(2), 177–189 (2004)
22. de Vega Rodrigo, M., Götz, J.: An Analytical Study of Optical Burst Switching Aggregation Strategies. In: 3rd International workshop on Optical Burst Switching (WOBS), San Jose, CA, USA (2004)
23. de Vega Rodrigo, M., Spadaro, S., Remiche, M.-A., Careglio, D., Barrantes, J., Götz, J.: On the Statistical Nature of highly-aggregated Internet Traffic. In: 4th International Workshop on Internet Performance, Simulation, Monitoring and Measurement, Salzburg, Austria (2006)
24. Detti, A., Listanti, M.: Impact of Segments Aggregation on TCP Reno Flows in Optical Burst Switching Networks. In: IEEE Infocom, New York, NY, USA, pp. 1803–1812 (2002)
25. Detti, A., Listanti, M.: Amplification Effects of the Send Rate of TCP Connection through an Optical Burst Switching Network. Optical Switching and Networking 2(1), 49–69 (2005)
26. Dogan, K., Akar, N.: A Performance Study of Limited Range Partial Wavelength Conversion for Asynchronous Optical Packet/Burst Switching. In: IEEE International Conference on Communications (ICC), Ankara, Turkey, pp. 2544–2549 (2006)

27. Downey, A.: Evidence for long-tailed Distributions in the Internet. In: ACM Sigcomm Internet Measurement Workshop, San Francisco, CA, USA (2001)
28. Du, Y., Pu, T., Zhang, H., Quo, Y.: Adaptive Load Balancing Routing Algorithm for Optical Burst-Switching Networks. In: OSA Optical Fiber Communication Conference and Exhibit (OFC), Anaheim, CA, USA (2006)
29. Duser, M., Kozlovski, E., Kelly, R.I., Bayel, P.: Design Trade-offs in Optical Burst Switched Networks with Dynamic Wavelength Allocation. In: 26th European Conference on Optical Communications (ECOC), Munich, Germany (2000)
30. Floyd, S., Mahdavi, J., Mathis, M., Podolsky, M.: An Extension to the Selective Acknowledgement (SACK) Option for TCP. RFC 2883 (2000)
31. Gao, D., Zhang, H.: Information Sharing based Optimal Routing for Optical Burst Switching (OBS) Network. In: OSA Optical Fiber Communication Conference and Exhibit (OFC), Anaheim, CL, USA (2006)
32. Gauger, C.: Performance of Converter Pools for Contention Resolution in Optical Burst Switching. In: Optical Networking and Communications (OptiComm), Boston, MA, USA, pp. 109–117 (2002)
33. Gauger, C., Köhn, M., Scharf, J.: Comparison of Contention Resolution Strategies in OBS Network Scenarios. In: 6th International Conference on Transparent Optical Networks (ICTON), Wroclaw, Poland, pp. 18–21 (2004)
34. Gauger, C., Mukherjee, B.: Optical Burst Transport Network (OBTN) – A Novel Architecture for Efficient Transport of Optical Burst Data over Lambda Grids. In: IEEE High Performance Switching and Routing (HPSR), Hong Kong, P.R. China, pp. 58–62 (2005)
35. Ge, A., Callegati, F., Tamil, L.: On Optical Burst Switching and Self-Similar Traffic. IEEE Communication Letters 4(3), 98–100 (2000)
36. González-Castaño, F.J., Rodelgo-Lacruz, M., Pavón-Mariño, P., García-Haro, J., López-Bravo, C., Veiga-Gontán, J., Raffaelli, C.: Guaranteeing Packet Order in IBWR Optical Packet Switches with Parallel Iterative Schedulers. Accepted for publication to European Transactions on Telecommunications journal
37. Gross, D., Harris, C.M.: Fundamentals of Queueing Theory, 2nd edn. John Wiley and Sons, Chichester (1985)
38. Guillemot, C., et al.: Transparent Optical Packet Switching: the European ACTS KEOPS Project Approach. IEEE Journal of Lightwave Technology 16(12), 2117–2134 (1998)
39. Gunreben, S., Hu, G.: A Multi-layer Analysis on Reordering in Optical Burst Switched Networks. IEEE Communications Letters 11(12), 1013–1015 (2007)
40. Harris, R.J.: The Modified Reduced Gradient Method for Optimally Dimensioning Telephone Networks. Australian Telecommunications Research 10(1), 30–35 (1976)
41. Hsu, C., Liu, T., Huang, N.: Performance Analysis of Deflection Routing in Optical Burst-Switched Networks. In: 21st IEEE Infocom, New York, NY, USA (2002)
42. Hu, G., Dolzer, K., Gauger, C.: Does Burst Assembly really Reduce the Self-Similarity? In: OSA Optical Fiber Communication Conference and Exhibit (OFC), Atlanta, March 2003, pp. 124–126 (2003)
43. Huang, X., She, Q., Zhang, T., Lu, K., Jue, J.P.: Small Group Multicast with Deflection Routing in Optical Burst Switched Networks. In: 5th International workshop on Optical Burst Switching (WOBS), San Jose, CA, USA (2006)
44. Huang, Y., Heritage, J., Mukherjee, B.: Dynamic Routing with Preplanned Congestion Avoidance for Survivable Optical Burst-Switched (OBS) Networks. In: OSA Optical Fiber Communication Conference and Exhibit (OFC), Anaheim, CL, USA (2005)

45. Hunter, D.K., Chia, M.C., Andonovic, I.: Buffering in Optical Packet Switches. IEEE/OSA Journal of Lightwave Technology 16(12), 2081–2094 (1998)
46. Ishii, D., Yamanaka, N., Sasase, I.: Self-learning Route Selection Scheme using Multipath Searching Packets in an OBS Network. OSA Journal of Optical Networking 4(7), 432–445 (2005)
47. ITU-t Recommendation I.371, Traffic control and congestion control in B-ISDN (2000)
48. Izal, M., Aracil, J.: On the Infuence of Self Similarity on Optical Burst Switching Traffic. In: IEEE Global Communications Conference (Globecom), Taipei, Taiwan, pp. 2308–2312 (2002)
49. Izal, M., Aracil, J., Morató, D., Magaña, E.: Delay-Throughput Curves for Timer-based OBS Burstifiers with Light Load. IEEE/OSA Journal of Lightwave Technology 24(1), 277–285 (2006)
50. Jeong, M., Qiao, C., Vandenhoute, M.: Distributed Shared Multicast Tree Construction Protocols for Tree-shared Multicasting in OBS Networks. In: 11th International Conference on Computer Communications and Networks (ICCCN), Miami, Florida, USA, pp. 322–327 (2002)
51. Junghans, S.: Pre-Estimate Burst Scheduling (PEBS): An Efficient Architecture with low Realization Complexity for Burst Scheduling Disciplines. In: International Conference on Broadband Networks (Broadnets), Boston, Massachusetts, USA, pp. 1124–1128 (2005)
52. Kaheel, A., Alnuweiri, H.: A Strict Priority Scheme for Quality-of Service Provisioning in Optical Burst Switching Networks. In: 8th IEEE Symposium on Computers and Communications (ISCC), Turkey (June 2003)
53. Kaheel, A., Alnuweiri, H.: Quantitative QoS Guarantees in Labeled Optical Burst Switching Networks. In: IEEE Global Communications Conference (Globecom), Dallas, TX, USA (2004)
54. Karagiannis, T., Molle, M., Faloutsos, M., Broido, A.: A Nonstationary Poisson View of Internet Traffic. In: IEEE Infocom, Hong Kong, P.R. China (2004)
55. Kelly, F.P.: Routing in Circuit-Switched Networks: Optimization, Shadow Prices and Decentralization. Advanced Applied Probability 20, 112–144 (1988)
56. Klinkowski, M., Pioro, M., Careglio, D., Marciniak, M., Sole-Pareta, J.: Non-linear Optimization for Multipath Source-Routing in OBS Networks. IEEE Communications Letters 11(12), 1016–1018 (2007)
57. Klinkowski, M., Careglio, D., Morató, D., Solé-Pareta, J.: Preemption Window for Burst Differentiation in OBS. In: OSA Optical Fiber Communication Conference and Exhibit (OFC), San Diego, CA, USA (2008)
58. Köhn, M., Gauger, C.: Dimensioning of SDH/WDM Multilayer Networks. In: 4th ITG Symposium on Photonic Networks, pp. 29–33 (2003)
59. Laevens, K.: Traffic Characteristics inside Optical Burst Switched Networks. In: Opticom, Boston, MA, USA, pp. 137–148 (2002)
60. Lee, S., Kim, H., Song, J., Griffith, D.: A Study on Deflection Routing in Optical Burst-Switched Networks. Photonic Network Communications 6(1), 51–59 (2003)
61. Leland, W., Taqqu, M., Willinger, W., Wilson, D.: On the Self-Similar Nature of Ethernet Traffic (extended version). IEEE/ACM Transactions on Networking 2(1), 1–15 (1994)
62. Li, J., Qiao, C., Xu, J., Xu, D.: Maximizing Throughput for Optical Burst Switching Networks. In: IEEE Infocom, Hong Kong, P.R. China, pp. 1853–1863 (2004)

63. Li, J., Mohan, G., Chua, K.C.: Dynamic Load Balancing in IP-over-WDM Optical Burst Switching Networks. Computer Networks 47(3), 393–408 (2005)
64. Li, J., Yeung, K.L.: Burst Cloning with Load Balancing. In: OSA Optical Fiber Communication Conference and Exhibit (OFC), Anaheim, CL, USA (2006)
65. Long, K.-P., Yang, X., Huang, S., Chen, Q.-B., Wang, R.: An Adaptive Parameter Deflection Routing to Resolve Contentions in OBS Networks. In: Boavida, F., Plagemann, T., Stiller, B., Westphal, C., Monteiro, E. (eds.) NETWORKING 2006. LNCS, vol. 3976, pp. 1074–1079. Springer, Heidelberg (2006)
66. Lu, J., Liu, Y., Gurusamy, M., Chua, K.: Gradient Projection based Multi-path Traffic Routing in Optical Burst Switching Networks. In: IEEE High Performance Switching and Routing workshop (HPSR), Poznan, Poland (2006)
67. Lu, X., Mark, B.L.: Performance Modeling of Optical-Burst Switching with Fiber Delay Lines. IEEE Transactions on Communications 52(12), 2175–2182 (2004)
68. Ma, X., Kuo, G.-S.: Optical Switching Technology Comparison: Optical MEMS vs. other Technologies. IEEE Communications Magazine 41(11), 16–23 (2003)
69. Maesschalck, S., et al.: Pan-European Optical Transport Networks: An Availability-based Comparison. Photonic Network Communications 5, 203–225 (2003)
70. Magaña, E., Morató, D., Izal, M., Aracil, J.: Evaluation of Preemption Probabilities in OBS Networks with Burst Segmentation. In: IEEE International Conference on Communications (ICC), Seoul, Korea (2005)
71. Masetti, F., et al.: Design and Implementation of a Multi-terabit Optical Burst/Packet Router Prototype. In: OSA Optical Fiber Communication Conference and Exhibit (OFC), Anaheim, CA, USA (2002)
72. Morató, D., Aracil, J.: On the Use of Balking for Estimation of the Blocking Probability for OBS Routers with FDL Lines. In: Chong, I., Kawahara, K. (eds.) ICOIN 2006. LNCS, vol. 3961, pp. 399–408. Springer, Heidelberg (2006)
73. Morton, A., Ciavattone, L., Ramachandran, G., Shalunov, S., Perser, J.: Packet Reordering Metrics. RFC 4737 (2006)
74. Norros, I.: On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks. IEEE Journal on Selected Areas in Communications 13(6), 953–962 (1995)
75. Ogino, N., Arahata, N.: A Decentralized Optical Bursts Routing based on Adaptive Load Splitting into Pre-calculated Multiple Paths. IEICE Transactions on Communications E88-B(12), 4507–4516 (2005)
76. Oh, S., Kang, M.: A Burst Assembly Algorithm in Optical Burst Switching Networks. In: OSA Optical Fiber Communication Conference and Exhibit (OFC), Anaheim, CA, USA, pp. 771–773 (2002)
77. Overby, H., Stol, N.: QoS Differentiation in Asynchronous Bufferless Optical Packet Switched Networks. Wireless Networks 12(3), 383–394 (2006)
78. Padhye, J., Firoiu, V., Towsley, D.F., Kurose, J.F.: Modeling TCP Reno Performance: A Simple Model and its Empirical Validation. IEEE/ACM Transactions on Networking 8(2), 133–145 (2000)
79. Pavon-Marino, P., Garcia-Haro, J., Jajszczyk, A.: Parallel Desynchronized Block Matching: A Feasible Scheduling Algorithm for the Input-Buffered Wavelength-Routed Switch. Computer Networks 51(15), 4270–4283 (2007)
80. Pavon-Mariño, P., Gonzalez-Castaño, F.J., Garcia-Haro, J.: Round-Robin Wavelength Assignment: A new Packet Sequence Criterion in Optical Packet Switching SCWP Networks. European Transactions on Telecommunications 17(4), 451–459 (2006)

81. Pedro, J.M., Monteiro, P., Pires, J.J.O.: Efficient Multi-path Routing for Optical Burst-Switched Networks. In: 6th Conference on Telecommunications (ConfTele), Peniche, Portugal (2007)
82. Perelló, J., Gunreben, S., Spadaro, S.: A Quantitative Evaluation of Reordering in OBS Networks and its Impact on TCP Performance. In: 12th conference on Optical Network Design and Modeling (ONDM 2008), Vilanova i la Geltru, Spain (2008)
83. Pioro, M., Medhi, D.: Routing, Flow, and Capacity Design in Communication and Computer Networks. Morgan Kaufmann, San Francisco (2004)
84. Qiao, C., Yoo, M.: Optical burst switching (OBS) – a new Paradigm for an Optical Internet. Journal of High Speed Networks (Special Issues on Optical Networks) 8(1), 69–84 (1999)
85. Rajaduray, R., Ovadia, S., Blumenthal, D.J.: Analysis of an Edge Router for span-constrained Optical Burst Switched (OBS) Networks. IEEE/OSA Journal of Lightwave Technology 22(11), 2693–2705 (2004)
86. Rodelgo-Lacruz, M., Pavón-Mariño, P., González-Castaño, F.J., García-Haro, J., López-Bravo, C., Veiga-Gontán, J.: Enhanced Parallel Iterative Schedulers for IBWR Optical Packet Switches. In: Tomkos, I., Neri, F., Solé Pareta, J., Masip Bruin, X., Sánchez Lopez, S. (eds.) ONDM 2007. LNCS, vol. 4534, pp. 289–298. Springer, Heidelberg (2007)
87. Rosberg, Z., Vu, H.L., Zukerman, M., White, J.: Performance Analyses of Optical Burst Switching Networks. IEEE Journal on Selected Areas in Communications 21(7), 1187–1197 (2003)
88. Ryu, B., Lowen, S.: Fractal Traffic Models for Internet Simulation. In: 5th IEEE International Symposium on Computer Communications (ISCC), Antibes, Juan les Pins, France, pp. 200–206 (2000)
89. Scharf, J., Kimsas, A., Köhn, M., Hu, G.: OBS vs. OpMiGua – A Comparative Performance Evaluation. In: Proceedings of the 9th International Conference on Transparent Optical Networks (ICTON 2007), Rome, pp. 294–298 (2007)
90. Shihada, B., Ho, P.-H., Zhang, Q.: A Novel False Congestion Detection Scheme for TCP over OBS Networks. In: IEEE Global Telecommunications Conference (Globecom), Washington, DC, USA, pp. 2428–2433 (2007)
91. Stoev, S., Taqqu, M.S., Park, C., Marron, J.S.: On the Wavelet Spectrum Diagnostic for Hurst Parameter Estimation in the Analysis of Internet Traffic. Computer Networks 48, 423–445 (2005)
92. Tan, S.K., Mohan, G., Chua, K.C.: Link Scheduling State Information based on Set Management for Fairness Improvement in WDM Optical Burst Switching Networks. Computer Networks 45(6), 819–834 (2004)
93. Tanenbaum, A.S.: Computer Networks, 2nd edn. Prentice-Hall, Englewood Cliffs (1988)
94. Teng, J., Rouskas, G.: Traffic Engineering Approach to Path Selection in Optical Burst Switching Networks. OSA Journal on Optical Networking 4(11), 759–777 (2005)
95. Thodime, G., Vokkarane, V., Jue, J.: Dynamic Congestion-based Load Balanced Routing in Optical Burst-Switched Networks. In: IEEE Global Communications Conference (Globeom), San Francisco, CA, USA (2003)
96. Turner, J.S.: Terabit Burst Switching. Journal of High Speed Networks 8(1), 3–16 (1999)
97. Van Breusegem, E., Cheyns, J., Colle, D., Pickavet, M., Demeester, P.: Overspill Routing in Optical Networks: a new Architecture for Future-proof IP over WDM Networks. In: Optical Networking and Communications (OptiComm), Dallas, TX, USA, pp. 226–236 (2003)

98. Vokkarane, V.M., Haridoss, K., Jue, J.P.: Threshold-based Burst Assembly Policies for QoS Support in Optical Burst-Switched Networks. In: Optical Networking and Communications (OptiComm), Boston, MA, USA, pp. 125–136 (2002)
99. Vokkarane, V.M., Jue, J.P.: Prioritized Burst Segmentation and Composite Burst Assembly Techniques for QoS Support in Optical Burst-Switched Networks. IEEE Journal on Selected Areas in Communications 21(7), 1198–1209 (2003)
100. Wei, J.Y., McFarland, R.I.: Just-in-time Signalling for WDM Optical Burst Switching Networks. IEEE Journal of Lightwave Technology 18(12), 2019–2037 (2000)
101. Xiong, Y., Vandenhoute, M., Cankaya, H.: Control Architecture in Optical Burst Switched WDM Networks. IEEE Journal on Selected Areas in Communications 8(10), 1838–1851 (2000)
102. Xu, J., Qiao, C., Li, J., Xu, G.: Efficient Channel Scheduling Algorithms in Optical Burst Switched Networks. In: IEEE Infocom, San Francisco, CA, USA, pp. 2268–2278 (2003)
103. Xuw, F., Yoo, B.S.J.: Self-Similar Traffic Shaping at the Edge Router in Optical Packet Switched Network. In: IEEE International Conference on Communications (ICC), New York, NY, USA, pp. 2449–2453 (2002)
104. Yang, L., Rouskas, G.N.: Adaptive Path Selection in Optical Burst Switched Networks. IEEE/OSA Journal of Lightwave Technology 24(8), 3002–3011 (2006)
105. Yoo, M., Qiao, C.: Just-enough-time (JET): A High Speed Protocol for Bursty Traffic in Optical Networks. In: IEEE/LEOS Summer Topical Meetings, Montreal, Canada, pp. 26–27 (1997)
106. Yoo, M., Qiao, C., Dixit, S.: Optical Burst Switching for Service Differentiation in the Next-Generation Optical Internet. IEEE Communications Magazine 39(2), 98–104 (2001)
107. Yu, X., Li, J., Cao, X., Chen, Y., Qiao, C.: Traffic Statistics and Performance Evaluation in Optical Burst Switched Networks. OSA/IEEE Journal of Lightwave Technology 22(12), 2722–2738 (2004)
108. Yu, X., Qiao, C., Liu, Y., Towsley, D.: Performance Evaluation of TCP Implementations in OBS Networks. Tech. Rep. 2003-13, CSE Dept., SUNY, Buffalo (2003)
109. Yu, X., Qiao, C., Liu, Y.: TCP Implementations and False Timeout Detection in OBS Networks. In: IEEE Infocom 2004, Hong Kong, P.R. China, pp. 774–784 (2004)
110. Zalesky, A., Vu, H., Rosberg, Z., Wong, E., Zukerman, M.: Modelling and Performance Evaluation of Optical Burst Switched Networks with Deflection Routing and Wavelength Reservation. In: IEEE Infocom, Hong Kong, P.R. China, pp. 1864–1871 (2004)
111. Zhang, J., et al.: Explicit Routing for Traffic Engineering in Labeled Optical Burst-Switched WDM Networks. In: IEEE International Conference on Computational Science (ICCS), Krakow, Poland (2004)
112. Zhang, Q., Vokkarane, V.M., Jue, J.P.: Biao Chen: Absolute QoS Differentiation in Optical Burst-Switched Networks. IEEE Journal on Selected Areas in Communications 22(9), 1781–1795 (2004)
113. Zhang, Q., Vokkarane, V.M., Wang, Y., Jue, J.P.: Analysis of TCP over Optical Burst-Switched Networks with Burst Retransmission. In: IEEE Global Telecommunications Conference (Globecom), St. Louis, MI (2005)
114. Zhang, Y., Li, L., Wang, S.: TCP over OBS: Impact of Consecutive Multiple Packet Losses and Improvements. Photonic Network Communications (in print)

115. Zhang, Y., Wang, S., Li, L.: B-Reno: A New TCP Implementation Designed for TCP over OBS Networks. In: Future Generation Communication and Networking (FGCN), Jeju Island, Korea, pp. 185–190 (2007)
116. Zhong, W.D., Tucker, R.S.: Wavelength Routing-based Photonic Packet Buffers and their Applications in Photonic Packet Switching Systems. Journal of Lightwave Technology 16(10), 1737–1745 (1998)