Available online at www.sciencedirect.com
SCIENCE DIRECT
Optical Switching and Networking 2 (2005) 217–229
Optical Switching and Networking
www.elsevier.com/locate/osn

Rate-controlled optical burst switching for both congestion avoidance and service differentiation☆

H. Boyraz[1], N. Akar*

Electrical and Electronics Engineering Department, Bilkent University, TR-06800 Bilkent, Ankara, Turkey

Received 23 May 2005; received in revised form 21 December 2005; accepted 27 January 2006
Available online 3 March 2006

Abstract

Optical Burst Switching (OBS) has recently been proposed as a candidate architecture for the next generation optical Internet. Several challenging issues remain to be solved to pave the way for the OBS vision. Contention arises in OBS networks when two or more bursts are destined for the same wavelength, and a wide variety of reactive contention resolution mechanisms have been proposed in the literature. One challenging issue in OBS is proactively controlling the traffic flowing through the OBS network so that the network does not stay in a persistent state of contention, which we call the congestion avoidance problem. Another challenging issue is the need for service differentiation, which is common today in electronically switched networks via the use of advanced buffer management and scheduling mechanisms. However, such mechanisms cannot be used in OBS networks due to the limited use, or total absence, of buffering. One of the popular existing approaches to service differentiation in OBS networks is the use of larger offset times for high-priority bursts which, however, increases the delays and may adversely affect application-level performance. In this paper, we propose a feedback-based rate control protocol for the control plane of the OBS network to both address the congestion avoidance and service differentiation issues. Using this protocol, the incoming traffic is dynamically shaped at the edge of the OBS network in order to avoid potential congestion in the burst-switched core. Moreover, the traffic shaping policies for the low and high priority traffic classes are different, and it is possible using the proposed protocol to isolate high-priority and low-priority traffic almost perfectly over time scales on the order of a few round-trip times. Simulation results are reported to validate the congestion avoidance and service differentiation capabilities of the proposed architecture.

Keywords: Optical burst switching; Rate control; Service differentiation; Wavelength conversion; Fiber delay lines

1. Introduction

Optical Burst Switching (OBS) has recently been proposed as a candidate architecture for the next generation optical Internet [1]. The central idea behind OBS is the promise of optical technologies to enable switch reconfiguration in the microsecond/millisecond range, therefore providing a near-term optical networking solution with finer switching granularity in the optical domain [2]. At the ingress node of an IP over OBS network, IP packets destined for the same egress node and with similar quality of service (QoS) requirements are segmented into so-called bursts, which are defined as a collection of IP packets, whereas IP packet reassembly is carried out at the egress OBS node.

☆ Expanded version of a talk presented at the 2005 Conference on Optical Network Design and Modeling (ONDM), Milan, Italy, February 2005.

* Corresponding author. Tel.: +90 312 2902337; fax: +90 312 2664192.
E-mail address: akar@ee.bilkent.edu.tr (N. Akar).
[1] Present Address: Lockheed Martin - Simulation, Training and Support, 12506 Lake Underhill Road, Orlando, FL 32825, USA.

1573-4277/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.osn.2006.01.001

In OBS, the reservation request for a burst is signalled out of band (e.g., over a separate wavelength channel) as a Burst Control Packet (BCP) and processed in the electronic domain. We assume the Just Enough Time (JET) reservation model [1], in which each BCP has offset time information that presents the traversed OBS node with the expected arrival time of the corresponding burst. The offset time, on the other hand, is adjusted at each OBS node on the way to account for the processing/switch configuration time. When the BCP arrives at an OBS node toward the egress node, the burst length and the arrival time are extracted from the BCP and the burst is scheduled in advance to an outgoing wavelength upon availability. Contention happens when multiple bursts contend for the same outgoing wavelength, and is resolved by either deflection or blocking [3]. The most common deflection technique is in the wavelength domain; some of the contending bursts can be sent on another outgoing wavelength channel through wavelength conversion [4]. In Full Wavelength Conversion (FWC), a burst arriving at a certain wavelength can be switched onto any other wavelength towards its destination. In Partial Wavelength Conversion (PWC), there is a limited number of converters, and consequently some bursts would be dropped when all converters are busy despite the availability of free channels on wavelengths different to the incoming wavelength [5]. Other ways of deflection-based contention resolution are in the time domain by sending a contending burst through a Fiber Delay Line (FDL), or in the space domain by sending a contending burst via a different output port so as to follow an alternate route [1]. If deflection cannot resolve contention using any of the techniques above, then a contending burst is blocked (i.e., data is lost) whose packets might be retransmitted by higher layer protocols (e.g., TCP).

Contention resolution policies are considered to be reactive approaches, since they are invoked after contention occurs. There is a vast amount of literature on contention resolution methods for OBS networks, but most of them break down in the case of heavy network load and may suffer from severe losses. A persistent state of burst contention in OBS nodes leading to burst losses is referred to as congestion. An alternative proactive approach to reduce network congestion is by controlling either the rate of traffic injection into the network [6] or by changing the route of the burst [7] so that congestion does not arise. In [8], a feedback-based congestion avoidance mechanism called SFC (Source Flow Control) is introduced. In this proposed mechanism, OBS core nodes send explicit messages to the edge nodes to reduce their transmission rates on congested links by measuring the load at their output ports. In [9], the intermediate nodes report the burst loss information to all edge nodes so that they can adjust their burst injection rates to control the network load.

The differentiated services model, adopted by the IETF, serves as a basis for service differentiation in the Internet today [10]. However, class-based queueing and advanced scheduling techniques (e.g., Deficit Round Robin [11]) that are used for service differentiation in IP networks cannot be used in OBS domains due to the lack of optical buffering with current optical technologies. It would be desirable to develop a mechanism by which operators can coherently extend their existing service differentiation policies in IP networks to their OBS-based networks as well. For example, if the legacy policy for service differentiation is based on packet-level strict priority queueing, then one would desire to provide a service in the OBS domain that would mimic strict priority-based service differentiation. A popular approach for QoS differentiation in OBS networks is to assign sufficiently large extra offset times to high-priority bursts, which then increases their probability of successful reservation at the expense of increased blocking rates for low-priority bursts, thereby providing a method for service differentiation [12]. The drawback of this approach is the increased end-to-end delays that may not be tolerated by some high-priority applications and the sensitivity level of isolation between service classes to the underlying burst assembly policy [13]. Another proposal for QoS differentiation is based on the dynamic allocation of resources (e.g., wavelength channels, FDLs) to high- and low-priority traffic so as to provide preferential treatment for high-priority traffic [14]. The efficiency of such a scheme strongly depends on the adaptivity of the resource allocation mechanism to the actual traffic load distribution and the partitionability of existing resources. Ref. [15] maintains a usage profile for each traffic class at the OBS nodes and implements a preemptive wavelength reservation algorithm to ensure QoS. In [16], the burst assembly procedure and the transmission of a BCP are allowed to be processed in parallel through the so-called Forward Resource Reservation (FRR) scheme, and the flexibility in launching epochs of BCPs is used as a means of delay-based service differentiation. In the alternative active dropping approach of [17], low-priority bursts are intentionally dropped using loss rate measurements to ensure relative loss differentiation. Ref. [18] describes two mechanisms, namely early

dropping and wavelength grouping, and integrates these two mechanisms to support absolute QoS in terms of loss rates in OBS core nodes.

In this paper, we propose a control-plane feedback architecture for connection-oriented OBS networks for both congestion avoidance and service differentiation. The proposed architecture is based on the explicit-rate distributed control mechanism used for ATM networks, for example the ERICA algorithm [19]. In this architecture, we propose that Resource Management (RM) packets, in addition to BCPs, are sent through the out-of-band control channel to gather the available bit rates for high- and low-priority traffic using a modification of the Available Bit Rate (ABR) mechanism in Asynchronous Transfer Mode (ATM) networks [20]. Core OBS nodes, on the other hand, calculate an effective capacity off-line for each of their OBS interfaces based on their contention resolution capabilities. These nodes then employ an online explicit rate algorithm to allocate this effective capacity dynamically in a max–min (maximum–minimum) fair fashion to the high priority OBS connections using that particular link. The remaining capacity, if any, is then allocated again using the max–min fairness principles to low-priority OBS connections. Such a resource allocation mechanism is said to be *prioritized max–min fair*. Finally, the explicit rate fields of RM packets are written appropriately by the core nodes on their way from the destination back to the source. Receiving back the RM packets with information on these two explicit rates for each OBS connection, a scheduler at the ingress node is proposed for arbitration among high- and low-priority bursts destined for different edge nodes. The overall architecture, called Differentiated ABR (D-ABR), provides several contributions to the existing literature:

- D-ABR addresses congestion avoidance and service differentiation within the same unifying architecture;
- service differentiation is achieved without having to use large offset times for high-priority traffic;
- service differentiation is achievable even in case of un-partitionable resources, e.g., a single wavelength system;
- the congestion avoidance capability of the proposed architecture moves congestion away from the OBS domain to the edges of the network, where buffer management to cope with bursty traffic is far easier and less costly.

The paper is organized as follows. Section 2 is devoted to the proposed congestion avoidance and service differentiation architecture for OBS networks. In Section 3, we present the results of our simulation studies using a variety of traffic and topology scenarios. We conclude in the final section.

2. Differentiated ABR (D-ABR) architecture for OBS networks

We envision an OBS network comprising edge and core OBS nodes connected via optical links. An optical link between two OBS nodes is a collection of wavelengths that are available for transmitting bursts. We also assume a number of wavelength channels for the control plane between any two nodes. We assume that the control plane is free of contention. Incoming client packets (e.g., IP, ATM, etc.) to the OBS domain are assumed to belong to one of the two classes, namely High-Priority (HP) and Low-Priority (LP) classes. For the data plane, ingress edge nodes assemble the incoming client packets on the basis of a burst assembly policy (see, for example, [21]) and schedule them for transmission toward the edge-core links. We assume a number of tuneable lasers available at each ingress node for the transmission of bursts. The burst de-assembly takes place at the egress edge nodes. In this paper, we concentrate on a connection-oriented OBS network, in the sense that bidirectional Virtual Connections (VC) are established at the control plane between each source–destination pair for carrying control-plane packets (e.g., BCP, RM, etc.). The optical bursts, on the other hand, follow the same path as their corresponding BCP packets. The bidirectional path that the bursts will follow in the data plane is called a Virtual Lightpath (VL). The offset-time between the burst and its BCP that is required for switch setup is ignored in this paper, since we capitalize only on the D-ABR protocol and its performance in this study. Studying D-ABR with the realistic case of non-zero offset times is left for future research, although we believe that, when the offset times are much smaller than the propagation delays, then their effects will be marginal on system performance. However, such effects remain to be seen in local or regional networks, for which this assumption may not be valid. We also assume that the core nodes do not support deflection routing but have PWC and FDL capabilities on a share-per-link basis [22].

The proposed architecture has the following three central components:

- off-line computation of the *effective capacity* of optical links;
- *D-ABR protocol* and its working principles;
- architecture for the *edge scheduler*.

2.1. Effective capacity

For an optical link, its corresponding Effective Capacity (*EC*) is the amount of traffic in bits/s (b/s) that can be burst switched by the link while meeting the desired QoS requirement in terms of burst blocking probabilities. We propose that an effective capacity is assigned to each of the interfaces (or links) of the OBS node on the basis of its contention resolution capabilities. For effective capacity calculations, we assume an asynchronous (i.e., unslotted) node with a number of OBS interfaces. We assume that the OBS link of interest has K wavelength channels, each channel capable of transmitting at c b/s. Given the burst traffic characteristics (i.e., burst interarrival time and burst length distributions) and a QoS requirement in terms of a desired burst blocking probability P_{loss}, our first goal is to find the *EC* of this optical link, for which we first need a burst traffic model. In our study, we propose that the effective capacity is to be found on the basis of a Poisson burst arrival process with rate λ (bursts/s), an exponentially distributed burst service time distribution with mean $1/\mu$ (s), and a uniform distribution of burst wavelengths. Once the traffic model is specified and the contention resolution capabilities of the optical link are given, one can use off-line simulations (or analytical techniques, if possible) to find the *EC* by first finding the maximum $\lambda_{\max}$ that results in the desired blocking probability P_{loss}, and then setting $EC = \lambda_{\max} c/\mu$.

We note that improved contention resolution capability of the OBS node also increases the effective capacity of its optical links. We study two contention resolution capabilities in this paper, namely PWC and FDL. In PWC, we assume a wavelength converter bank of size $0 < W < K$ dedicated to each output link. Based on the model provided in [5], a new burst arriving at the switch on wavelength w and destined to output line k:

- is forwarded to output line k without using a Tuneable Wavelength Converter (TWC) if channel w is available, else
- is forwarded to output line k using one of the free TWCs in the converter bank and using one of the free wavelength channels selected at random, else
- is blocked.

An efficient computational procedure based on block-tridiagonal LU factorizations is given in [5] for finding the blocking probabilities in PWC-capable optical links, and therefore the *EC* of an optical link can be obtained very rapidly in bufferless PWC-capable links. Besides wavelength conversion, we optionally use FDLs in our numerical experiments for contention resolution purposes. We study the case of L FDLs per output link, where the ith FDL, $i = 1, 2, \ldots, L$, can delay the burst by $b_i = i/\mu$ s. The burst reservation policy that we use is to first try wavelength conversion for contention resolution and, if conversion fails to resolve contention, we attempt to resolve it by suitably passing a contending burst through one of the L FDLs. To the best of our knowledge, no exact solution method exists in the literature for the blocking probabilities in asynchronous OBS nodes supporting FDLs, and therefore we suggest the use of off-line simulation in the latter scenario to compute the *EC* of FDL-capable optical links. The optical link model using PWC and FDLs that we use in our simulation studies is depicted in Fig. 1. We note, however, that the *EC* for more general OBS nodes with more sophisticated architectures can still be calculated using off-line simulations, although such a detailed analysis is outside the scope of the current paper.

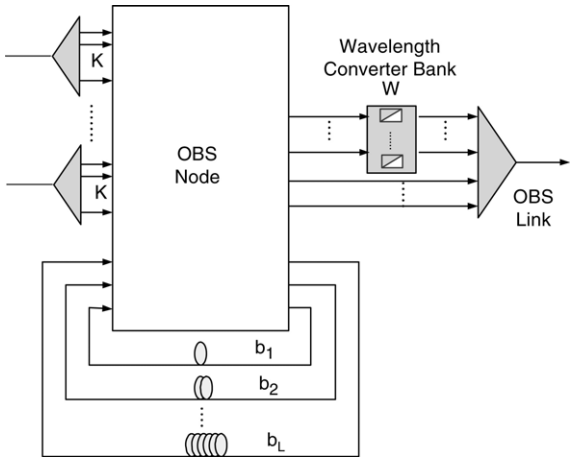

Fig. 1. The general architecture of the OBS node under study.

2.2. D-ABR protocol

The feedback mechanism is key for our congestion avoidance and service differentiation architecture. Our goal is to provide flow control so as to keep burst losses at a minimum, but also to emulate strict priority queueing through the OBS domain. For this purpose, we propose that a feedback mechanism similar to the ABR service category in ATM networks is to be used in OBS networks as well [23].

In ATM Explicit Rate (ER) ABR, for a bidirectional Virtual Circuit (VC) an RM cell is sent by the source towards the destination after the transmission of a configurable number of data packets. The RM cell is then returned back upon its arrival at the destination ATM switch. RM cells have an ER field that is written

```
ABR Capacity ← Target Utilization
* Link Bandwidth
** AT THE END OF AN AVERAGING INTERVAL **
Load Factor ← ABR Input Rate / ABR Capacity
Fair Share ← ABR Capacity / # VCs
** ON THE RECEIPT OF A FORWARD RM CELL **
CCR[VC] ← CCR in RM Cell
** ON THE RECEIPT OF A BACKWARD RM CELL **
ER[VC] ← Max(Fair Share, CCR[VC]/Load Factor)
ER[VC] ← Min(ER[VC], ABR Capacity)
ER in RM Cell ← Min(ER in RM Cell, ER[VC])
```

Fig. 2. The pseudo-code for the ERICA ABR algorithm.

by the ATM switches to provide feedback to the sources on the reverse path of the RM cells. The ER field is indicative of the rate at which the corresponding data cells can be injected into the network. The sources monitor the returning RM cells and adjust their transmission rates according to the ER information in the RM cells. There are a number of proposals for ATM switches to carry out ER calculations; see, for example, the references for existing ABR rate control algorithms [24–26]. In this paper, we choose to test the basic ERICA (Explicit Rate Indication for Congestion Avoidance) algorithm due to its simplicity, fairness, and rapid transient performance [19]. Another reason behind the choice of the basic ERICA algorithm is that it does not use the queue length information as other ABR rate control algorithms do, but this feature turns out to be very convenient for OBS networks with either very limited queueing capabilities (i.e., a limited number of FDLs) or none at all. We leave a more detailed study of rate control algorithms for OBS networks for future work, and we now outline the basic ERICA algorithm and describe our modification to this algorithm next in order to mimic the behaviour of strict priority queueing.

In basic ERICA, at the end of an averaging interval the ATM switch monitors the load for a given link and determines the load factor, the available capacity, and the number of currently active VCs. The load factor is calculated as the ratio of the measured input rate during the averaging interval to the target capacity of the output link, while the latter is the product of the link bandwidth and a target utilization parameter that is chosen to be a fraction of one, e.g., 0.95. A large load factor indicates excessive congestion and a low value implies link underutilization. The goal of the switch is to maintain the network at a unit load factor. Moreover, the fair share of each VC is found by dividing the link capacity by the number of active VCs. On the arrival of a forward RM cell, each VC's Current Cell Rate (CCR) is updated on the basis of the current cell rate field information. On the other hand, upon the arrival of a backward RM cell for a certain VC, the switch calculates the ER of that VC by taking the maximum of the fair share and the current cell rate divided by the load factor. This calculated ER is inserted in the ER field of the backward RM cell if it is less than the ER value in the RM cell. This proposed way of ER calculation is known to lead to an efficient and fair operating point in a simple way, although improvements for this basic scheme are proposed in [19]. We also note that a priori information on propagation delays is not made use of to potentially improve performance in basic ERICA; such more advanced mechanisms for feedback-based rate control are beyond the scope of the current work. The pseudo-code for the basic ERICA algorithm is given in Fig. 2.

In our proposed feedback architecture for OBS networks, the ingress edge node of the bidirectional VL periodically sends RM packets with a period T, in addition to the BCPs, through the control channel. As in ATM ABR, these RM packets are returned back by the egress node to the ingress node using the same route due to the way VLs are established. The central idea behind the basic ERICA algorithm is to simultaneously achieve fairness and high utilization, whereas with our proposed modification we also attempt to provide isolation between HP and LP traffic. The pseudo-code for our proposed ERICA-based algorithm is given in Fig. 3. RM packets in this proposed architecture have two separate explicit rate fields for HP and LP traffic, namely `HP-ER` and `LP-ER`, respectively. We define an averaging interval of length T_a and an ERICA module for each OBS link. The ERICA module maintains two counters to calculate the HP and LP input bit rates, namely `HP Input Rate` and `LP Input Rate`, respectively, using the burst length information in the corresponding BCPs. The `HP Capacity` is set to the `Target Utilization` times the *EC* of the OBS link and the `LP Capacity` is set to `HP Capacity` minus the `HP Input Rate`, since the *EC* of the OBS link is the capacity that the HP traffic can use to achieve a desired burst blocking probability, and the remaining capacity is then to be used for LP traffic. The fair shares for each class, namely `HP Fair Share` and `LP Fair Share`, are then calculated as the ratio of the corresponding capacities to the number of VLs that had sent at least one burst in the corresponding class during the averaging interval. Moreover, for each VL we keep track of two quantities `HP-CBR[VL]` and `LP-CBR[VL]` to denote the Current Bit Rate (CBR) for HP and LP traffic via counting the number of received HP and LP bits for

```
HP Capacity ← Target Utilization * Effective Capacity

** AT THE END OF AN AVERAGING INTERVAL **

for all VL do
  HP-CBR[VL] ← Measured HP Bit Rate for VL
  LP-CBR[VL] ← Measured LP Bit Rate for VL
end for
HP Input Rate ← ∑ HP-CBR[VL]
LP Input Rate ← ∑ LP-CBR[VL]
LP Capacity ← HP Capacity - HP Input Rate
HP Load Factor ← HP Input Rate / HP Capacity
LP Load Factor ← LP Input Rate / LP Capacity
HP Fair Share ← HP Capacity / # VLs with HP Activity
LP Fair Share ← LP Capacity / # VLs with LP Activity

** ON THE RECEIPT OF A BACKWARD RM PACKET **

HP-ER[VL] ← Max(HP Fair Share,HP-CBR[VL]/HP Load Factor)
HP-ER[VL] ← Min(HP-ER[VL],HP Capacity)
if HP-CBR[VL] < HP Fair Share &HP-ER[VL] ≥ HP Fair Share
then
  HP-ER[VL] = HP Fair Share
end if
LP-ER[VL] ← Max(LP Fair Share,LP-CBR[VL]/LP Load Factor)
LP-ER[VL] ← Min(LP-ER[VL],LP Capacity)
if LP-CBR[VL] < LP Fair Share &LP-ER[VL] ≥ LP Fair Share
then
  LP-ER[VL] = LP Fair Share
end if
HP-ER in RM Packet ← Min(HP-ER in RM Packet,HP-ER[VL])
LP-ER in RM Packet ← Min(LP-ER in RM Packet,LP-ER[VL])

** ON THE RECEIPT OF A BACKWARD RM PACKET AT SOURCE **

HP-PBR[VL] ← Min(HP-ER[VL],HP-PBR[VL]*(1+RIF))
LP-PBR[VL] ← Min(LP-ER[VL],LP-PBR[VL]*(1+RIF))
```

Fig. 3. The pseudo-code for the ERICA-based D-ABR algorithm.

the particular VL, respectively, as measured by the OBS node during the averaging interval. On the arrival of a backward RM cell for a given VL, the OBS node calculates the explicit rates for the HP and LP traffic, namely `HP-ER[VL]` and `LP-ER[VL]`, respectively, and these values are insterted in the backward RM packet if they are less than the `HP-ER` and `LP-ER` values in the backward RM packet. To avoid transient overloads, we adopt the explicit rate calculation mechanism (see Fig. 3) based on [19]. When a backward RM packet arrives back at the originating OBS ingress node, the node calculates the Permitted Bit Rates (PBR), namely `HP-PBR[VL]` and `LP-PBR[VL]`, to determine the allowable rates at which bursts of type HP and LP, respectively, can be sent towards the OBS network over the specified VL. The PBR calculation policy uses a multiplicative increase scheme using the parameter `RIF` (Rate Increase Factor) which conservatively updates the corresponding `PBR` in case of an abrupt increase in the available bandwidth with a choice of `RIF` < 1. On the other hand, if there is a drop in the available bandwidth, then in this study we suggest that the response to this drop should be rapid.

We use the term Differentiated ABR (D-ABR) to refer to the architecture proposed in this study that controls the rate of injection of HP and LP traffic towards the OBS network. The distributed D-ABR protocol that we propose distributes the effective capacity of optical links to HP traffic first using max–min fair allocation, and the remaining capacity is then used by LP traffic, still using the same allocation principles; see the definition of max–min fairness and algorithms to find max–min fair allocations in [27].

2.3. Edge scheduler

An ingress edge node maintains two queues, namely the HP and LP queues, on a per-egress basis. Since there are multiple egress edge nodes per ingress, a scheduler at the ingress edge node is needed to arbitrate among all per-egress queue pairs while obeying the rate constraints instructed by the `PBR` values that are described in the previous subsection. The ingress node

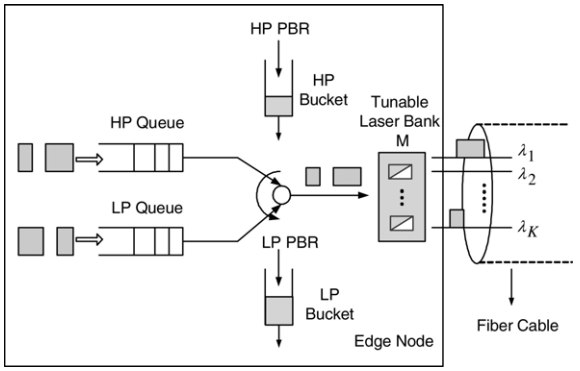

Fig. 4. The structure of the edge scheduler.

structure is presented in Fig. 4 for the special case of a single egress (i.e., single VL). In this case, there are two buckets of size B bytes for HP and LP traffic, which are used for controlling the traffic injection rates. The HP (LP) bucket continuously fills with credits at a rate of HP-PBR (LP-PBR). Let L_h and L_l denote the length of the burst at the head of the HP queue and that of the LP queue, respectively. Occupancy of the HP bucket greater than L_h bytes implies sufficient credits and, upon the availability of a free wavelength channel, this HP burst can be transmitted using one of the M tuneable lasers while draining L_h bytes from the bucket. If either the HP queue is empty or if there are not enough credits for the HP burst at the head of the HP queue, then the LP bucket is checked against L_l bytes to decide if the burst at the head of the LP queue can be transmitted. If either there are no waiting bursts or neither of the credits suffices to make a transmission, then the edge scheduler goes into a wait state until a new burst arrival, a sufficient bucket fill, or the arrival of a backward RM packet. In the case of multiple (say E) egress edge nodes, the E HP per-egress queues are first checked on a round-robin basis to see if their corresponding buckets have enough credits for potential transmission. After the transmission of all rate-compliant HP bursts, if there are still unused wavelength channels then the E LP per-egress queues are served again on a round-robin basis and by monitoring the corresponding bucket occupancies. For details on the scheduler implementation, we refer the reader to [28].

3. Numerical results

In this paper, we have implemented an event-based simulator in our simulations using the C++ programming language to study the effectiveness of the D-ABR protocol. In the first example, we use an OBS multiplexing scenario for proof-of-concept purposes. In the second example, we study a more general topology, the so-called Generic Fairness Configuration-1 (GFC-1), which was proposed in [29] to test the max–min fair allocation capabilities of distributed rate control algorithms. In the final example, we also take into consideration the effects of bursty traffic and electronic buffers at the edge, so as to compare the performance of D-ABR against conventional un-controlled OBS. In this paper, we choose to present the service differentiation capability of the proposed architecture without giving comparisons with other QoS schemes proposed for OBS networks. Our proposed approach achieves service differentiation without having to use large offset times like in the offset-based service differentiation, which is an advantage. However, it is very well known that feedback control protocols react within a few round trip times and, if the traffic mix changes abruptly, then it is not possible to provide perfect isolation between different traffic classes for such short periods of time, i.e., within a few round trip times. In such cases, open-loop techniques such as offset-based service differentiation may be expected to perform better. A comparison of open-loop and closed-loop techniques in the context of service differentiation for a wide range of traffic scenarios is left for future research.

3.1. OBS multiplexer

We study the D-ABR protocol for the simulation topology depicted in Fig. 5. All the links are assumed to have the same propagation delay D. In this study, there are 25 ingress nodes and one single egress node, thus representing an OBS multiplexing system. Each of the fibers has $K = 100$ wavelength channels. The capacity c of each channel is assumed to be 10 Gb/s. The burst lengths are exponentially distributed with mean 20 kB. We set all the HP and LP bucket sizes to $B = 2$ MB and all the HP and LP queues maintained at the ingress nodes are assumed to have infinite storage capacity. The RM cells are sent every $T = T_a$ s. The rate increase factor RIF is set to 1/16. Each of the ingress nodes is connected to the single OBS core node using $M = 4$ tuneable lasers. Traffic sources are classified into five classes, each comprising five ingress nodes where the HP and LP Poisson burst arrival rates are identical within a class. We also vary the traffic demands for each class in Gb/s in time, as given in Table 1. For comparison purposes, we tested four different scenarios, which are described in Table 2. In scenarios A and B, we use $EC = 700$ Gb/s, which is shown to ensure $P_{\text{loss}} \approx 3.2 \times 10^{-5}$ by off-line simulations for an OBS link with $L = 15$ FDLs and $W = 20$ TWCs. In scenario

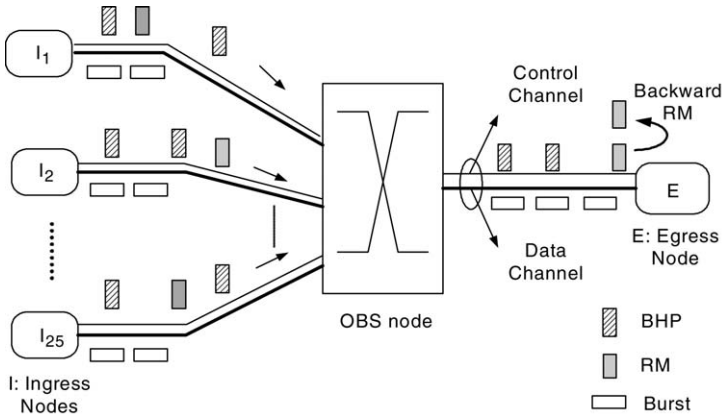

Fig. 5. The simulation topology with 25 ingress OBS nodes multiplexed over a single OBS link with 100 wavelength channels.

Table 1
The burst arrival rates for HP and LP traffic, in Gb/s, for the five traffic classes

	$0 \le t < 150$ s		$150 \text{ s} \le t < 300$ s		$300 \text{ s} \le t < 450$ s	
	HP rate	LP rate	HP rate	LP rate	HP rate	LP rate
Class 1	35	20	35	20	15	20
Class 2	15	5	20	5	20	5
Class 3	18	0	35	0	25	0
Class 4	12	30	12	30	10	30
Class 5	0	25	0	25	0	25

Table 2
Four different simulation scenarios for the OBS multiplexer example

	Scenario			
	A	B	C	D
D (ms)	2	20	2	2
T_a (s)	0.1	1	0.1	0.1
W (# converters)	20	20	20	50
L (# FDLs)	15	15	15	0
EC(Gb/s)	700	700	665	500

C, we employ `Target Utilization` $= 0.95$, so we set $EC = 700 * 0.95 = 665$ Gb/s to further reduce burst losses. We use $EC = 500$ Gb/s in the final scenario D (i.e., 50 TWCs and no FDLs) and this choice of EC yields $P_{\text{loss}} \approx 1.8 \times 10^{-4}$ on the basis of the numerical algorithm presented in [5].

First we study the total number of bursts (HP or LP) dropped in time $(0, t)$ for the four scenarios A–D in Fig. 6. The best performance in terms of dropping rate is achieved with Scenario C, but at the expense of a reduction in throughput, since the EC of the OBS node is set such that the load on the node is lighter. The burst drop rate is generally constant in all the scenarios except for $t = 150$ s, when there is an abrupt rise in the overall traffic demand. This event is followed by a substantial number of blocked bursts, and the blocking performance immediately improves once the D-ABR protocol reaches the steady-state. Since the traffic demand decreases at $t = 300$ s, we do not see any additional burst drops due to traffic change at this instant. We monitor P_{loss} in the interval $160 \text{ s} < t \le 450$ s (i.e., in the steady-state) and these steady-state blocking probabilities are also shown in Fig. 6. The steady-state measured burst blocking probabilities in Scenarios A and B ($P_{\text{loss}} = 8.4 \times 10^{-6}$ and 7.9×10^{-6}, respectively) are less than the desired blocking probability for which the EC was set (i.e., we recall the desired $P_{\text{loss}} \approx 3.2 \times 10^{-5}$). Similar results also hold for Scenario D. The provisioned burst blocking probability was obtained using the Poisson arrival assumption but, with the D-ABR burst shaping protocol, the burst arrival process becomes more regular than Poisson, thus reducing the Coefficient of Variation (CoV) of the arrival process. Such a reduced CoV has an improving effect on burst blocking performance [5] for independent and identically distributed (iid) burst interarrivals, and therefore the results are in accordance with [5]. For iid burst interarrivals, we conjecture that the provisioned QoS under the Poisson assumption provides a lower bound on the measured steady-state blocking performance. Moreover, Scenarios A and B differ from each other in the link delays, which

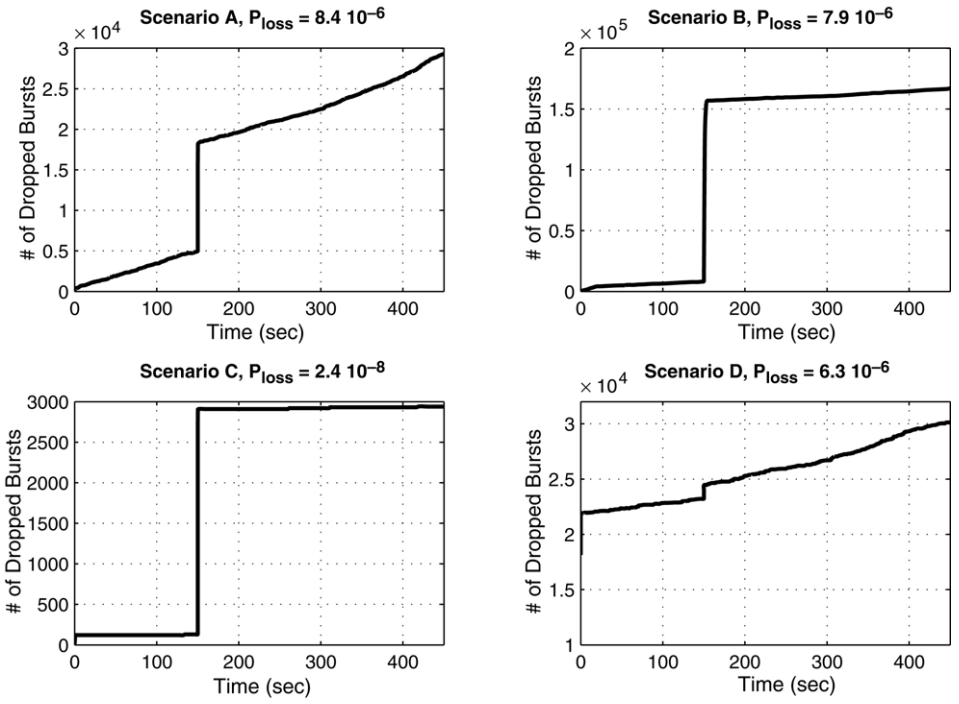

Fig. 6. Total number of dropped bursts at the OBS node in time $(0, t)$ for the scenarios A–D.

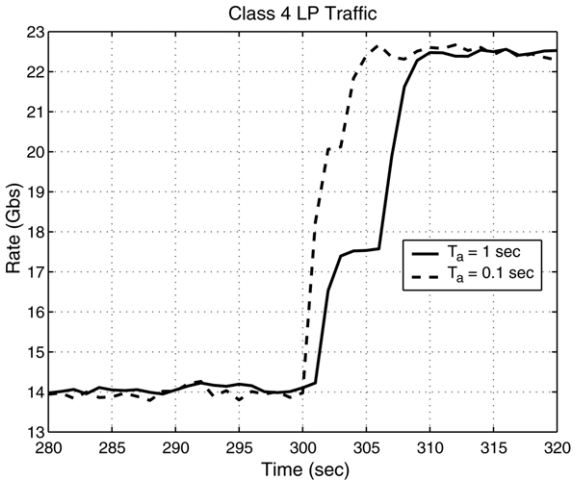

Fig. 7. The transient response of the system upon the traffic demand change at $t = 300$ s in terms of the throughput of Class 4 LP traffic. The solid curve is for Scenario B and the dotted curve is for Scenario A.

does not seem to have much of an impact on the steady-state blocking probability. However, the D-ABR algorithm performance at the instant of abrupt changes (i.e., $t = 150$ s or $t = 300$ s) is significantly better for Scenario A than for Scenario B; note the number of burst drops that take place at $t = 150$ s for these scenarios. The settling time is defined as the time it takes to reach a steady-state in control systems terminology. The RTT (Round Trip Time) is the time delay of the system, which also increases the settling time of the control system. The RTT in Scenario A is much less than that of Scenario B, which explains the difference in the transient response of these two scenarios. As an example, the effective bit rate of LP traffic for Class 4 is depicted before and after $t = 300$ s in Fig. 7. Scenario A, which has a smaller RTT and therefore a smaller ERICA averaging interval T_a, reaches the steady-state much faster than Scenario B.

We also study the service differentiation aspect with the OBS multiplexer example. The HP and LP smoothed throughputs are depicted in Fig. 8 for Scenario D, for which the solid (dotted) line is used to denote HP (LP) throughput. The results demonstrate that the effective capacity of the optical link at the OBS node is distributed using prioritized max–min fair share; we refer to [27] for a max–min fair share calculation algorithm. To show this, we focus on the time interval $0\text{ s} \leq t < 150$ s as an example. In this time interval, the aggregate HP demand is 400 Gb/s $\leq EC = 500$ Gb/s, therefore the max–min share vector for HP traffic, in Gb/s, is (35, 15, 18, 12, 0), where the ith entry of this vector represents the HP throughput of the ith class VLs. If the remaining capacity $EC - 400$ Gb/s $= 100$ Gb/s is allocated to LP traffic on a max–min fair

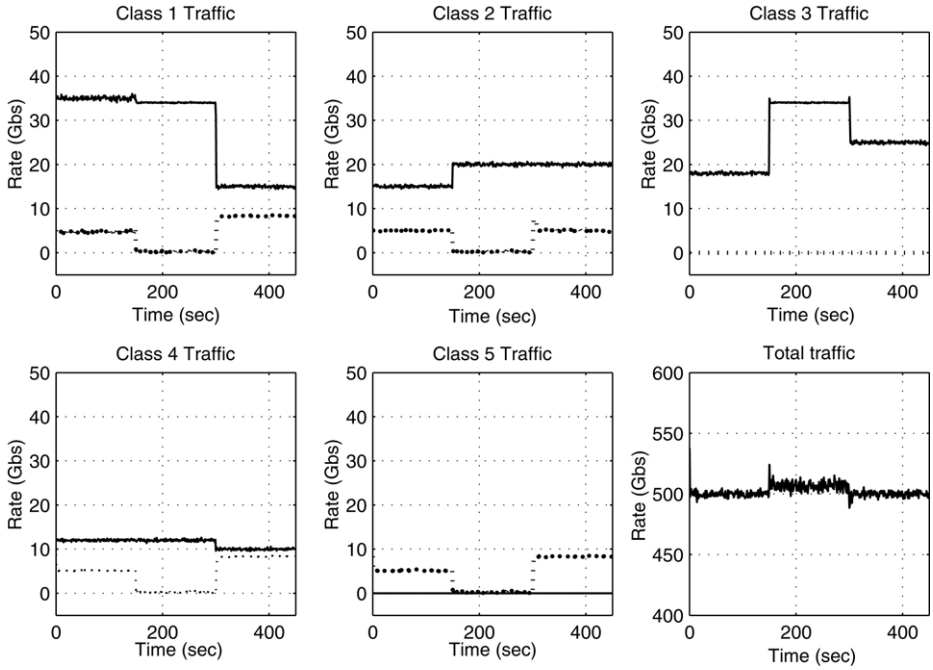

Fig. 8. The HP and LP smoothed throughputs for Scenario D. The solid (dotted) line denotes HP (LP) throughputs.

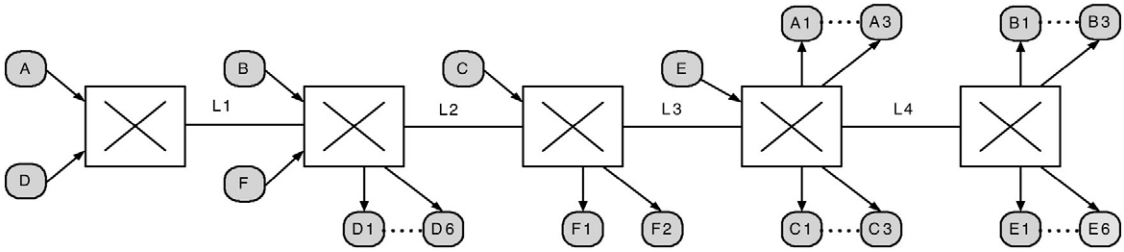

Fig. 9. Generic Fairness Configuration 1 (GFC-1) simulation topology.

share basis, then the max–min fair share vector for LP traffic is found to be (5, 5, 0, 5, 5). Fig. 8 reveals that the max–min fair shares are attainable using the distributed D-ABR protocol proposed in this paper. One can show that this argument is also valid for the other time intervals and scenarios.

3.2. Generic fairness configuration

In this subsection, we use the GFC-1 topology [29]. GFC-1, depicted in Fig. 9, is a five-switch parking lot configuration with multiple bottlenecks and is used to test the max–min fairness feature of the algorithm. In this network configuration, there are six sources with multiple VLs: A, B, and C with three VLs, D and E with six VLs, and finally F with two VLs. As before, each VL carries both HP and LP traffic.

All the links in the given topology are assumed to have the same delay $D = 2$ ms. We assume $K = 100$ wavelength channels and $W = 50$ and $L = 0$ for each link. We assume the capacity c of each channel to be 9.32 Gb/s and the burst lengths to be exponentially distributed with mean 20 kB. The target link load for a burst blocking probability of 10^{-3} is calculated using the numerical algorithm in [5] as 0.536, which results in an effective capacity EC of $100 * 9.2$ Gb/s $* 0.536 = 500$ Gb/s for all the optical links. Each of the ingress nodes is connected to one single OBS core node using $M = K = 100$ tuneable lasers, which enables the sources to use the full capacity of their access links. The averaging interval T_a and the period T of RM packets are both set to 1 ms. The bucket sizes denoted by B are all set to 250 kB, and the RIF is set to 1/16. The VL burst arrivals are Poisson, as in the previous example. We study two different simulation scenarios for this topology, namely scenarios A and B. The traffic demands, the ideal

Table 3
The traffic demands, the ideal max–min fair allocations based on the traffic demands, and the D-ABR results obtained via simulations, all in Gb/s, for the simulation scenario A

VL	High priority			Low priority		
	Traffic demand	Ideal fair share	Simulation results	Traffic demand	Ideal fair share	Simulation results
A (x 3)	25	25	25.10	50	11.11	11.47
B (x 3)	16.67	16.67	16.68	50	22.22	23.02
C (x 3)	25	25	25.01	100	66.67	65.85
D (x 6)	54.17	54.17	54.25	50	11.11	11.29
E (x 6)	41.67	41.67	41.56	50	22.22	22.47
F (x 2)	37.50	37.50	37.39	125	100	98.64

Table 4
The traffic demands, the ideal max–min fair allocations based on the traffic demands, and the D-ABR results obtained via simulations, all in Gb/s, for the simulation scenario B

VL	High priority			Low priority		
	Traffic demand	Ideal fair share	Simulation results	Traffic demand	Ideal fair share	Simulation results
A (x 3)	25	25	25.05	50	22.22	22.41
B (x 3)	16.67	16.67	16.61	50	22.22	22.44
C (x 3)	58.33	58.33	58.24	100	22.22	22.77
D (x 6)	20.83	20.83	20.83	50	38.89	39.08
E (x 6)	25	25	24.92	50	38.89	38.23
F (x 2)	87.50	87.50	87.50	125	33.33	32.85

Table 5
Link burst blocking probabilities for both simulation scenarios A and B

Link	Burst blocking probability	
	Scenario A	Scenario B
$L1$	4.77×10^{-6}	2.73×10^{-6}
$L2$	8.19×10^{-5}	2.06×10^{-4}
$L3$	1.84×10^{-4}	4.51×10^{-4}
$L4$	6.67×10^{-6}	3.32×10^{-6}

max–min fair allocations, and the simulation results obtained using six seconds of simulation runtime are reported for these two scenarios in Tables 3 and 4. Our simulation results clearly demonstrate that the proposed algorithm D-ABR achieves differentiation among HP and LP traffic, and distributes the remaining capacity from HP traffic on a max–min fair share basis among LP flows. The small difference between the simulation results and ideal max–min fair shares can be explained by the stochastic nature of the burst arrival processes. Finally, Table 5 provides the burst blocking probabilities on links $L1, \ldots, L4$. We observe that the steady-state measured burst blocking probabilities in both simulation scenarios and on all links are less than the desired blocking probability for which the EC was set (i.e., we recall the desired $P_{\text{loss}} \approx 10^{-3}$).

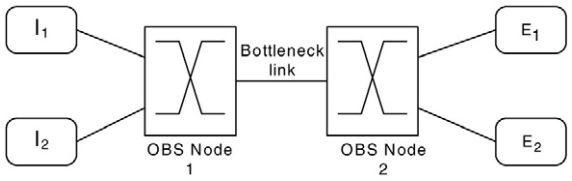

Fig. 10. The two-switch topology to be used for analyzing the effect of bursty traffic on D-ABR performance.

3.3. Bursty traffic

In this example, we use the two-switch topology in Fig. 10 with two ingress and two egress nodes, and a VL is established between each ingress-egress pair, which amounts to four VLs being multiplexed over the single bottleneck OBS link. All the parameters are the same as in the previous example, except that $c = 2.5$ Gb/s and the corresponding $EC = 100 * 2.5$ Gb/s $* 0.536 = 134$ Gb/s for the bottleneck OBS link. For this numerical example, the per-egress HP and LP queues are assumed to have finite storage capacity and are all set to 125 MB, and burst drops at these queues are also taken into account. The HP traffic for each VL is Poisson with rate 8.5 Gb/s, which amounts to an overall HP demand of 34 Gb/s on the bottleneck link. The remaining 100 Gb/s capacity is then to be shared among the LP traffic. We assume that the LP sources are bursty HIGH–LOW sources and that they emit Poissonian bursty traffic with rates $\lambda_0 + \Delta\lambda$ and

$\lambda_0 - \Delta\lambda$ in the HIGH and LOW periods, respectively, where the HIGH and LOW times are identical and deterministic and set to 100 ms. We also set $\lambda_0 = 25$ Gb/s. We assume the worst-case scenario in which all the LP sources are synchronized, i.e., the transition times of each source to HIGH and LOW states are identical. We vary the burstiness parameter $\Delta\lambda$ and observe its impact on the performance on end-to-end burst blocking, taking into consideration the burst losses at the edge queues as well as the OBS bottleneck link. We note that the $\Delta\lambda = 0$ case reduces to the non-bursty Poissonian case, whereas increasing $\Delta\lambda$ implies increased burstiness. Fig. 11 demonstrates the overall gain in terms of end-to-end blocking probabilities as a function of $\Delta\lambda$ in rate-controlled OBS using D-ABR over conventional OBS without flow control. This blocking gain first rises with increasing $\Delta\lambda$ and makes a peak at around $\Delta\lambda = 7.5$ Gb/s with a gain of around 100. The gain then gradually drops to about unity with further increases in $\Delta\lambda$. For small $\Delta\lambda$, the traffic is not bursty and the traffic demand matches the *EC* of the bottleneck link, therefore the differences between the rate-controlled and uncontrolled OBS are minor. With increasing $\Delta\lambda$, epochs of overload begin to arise in which the excess traffic is handled by electronic queueing at the edges. This functionality of edge queues results in a dramatic blocking gain in this regime. However, beyond a certain value of $\Delta\lambda$, the storage capacity of the edge queues fails to absorb all the excess traffic, which in turn steers the gain curve to unity. By increasing queue capacities at the edge node, one can further reduce the burst blocking probabilities in rate-controlled OBS, but at the expense of increasing end-to-end delays. The service differentiation aspect for bursty traffic is provided in Fig. 12 in terms of the burst blocking probability as a function of $\Delta\lambda$. There is an initial drop in the burst blocking probability for small values of the burstiness parameter $\Delta\lambda$ for both classes. We explain this phenomenon with the reduced CoV of the arrival process towards the OBS network when D-ABR takes action. However, this advantage starts to disappear when losses at the edge queues begin to occur with increased $\Delta\lambda$. The blocking probabilities for HP traffic are generally less than 10^{-3}, for which the *EC* was adjusted, but this blocking rate is not achievable for very bursty sources (e.g., $\Delta\lambda = 25$ Gb/s). This observation leads us to believe that the Poisson traffic assumption might be optimistic for very bursty traffic, and there might be a need to refine *EC* calculations for such bursty demands.

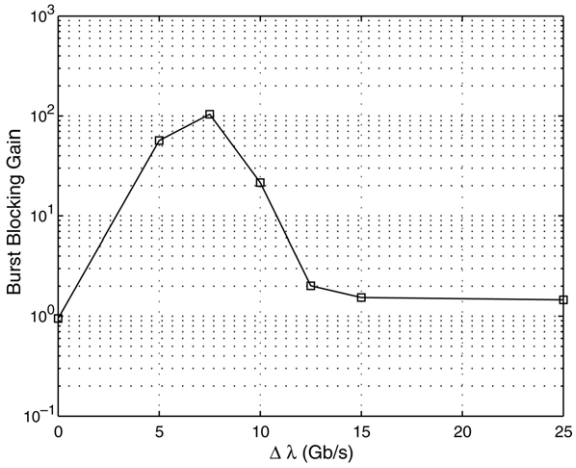

Fig. 11. The overall gain in terms of end-to-end blocking probabilities as a function of $\Delta\lambda$ in rate-controlled OBS using D-ABR over conventional OBS without flow control.

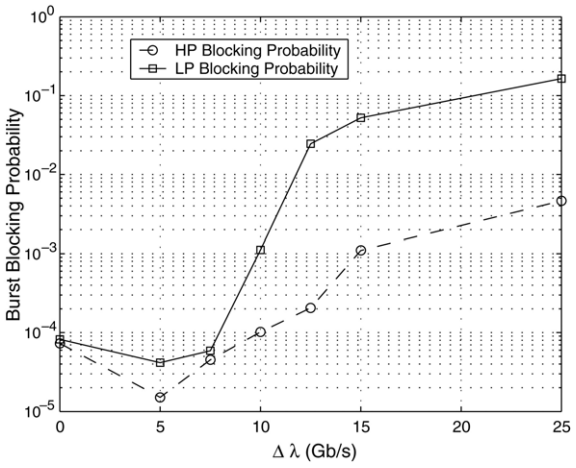

Fig. 12. Burst blocking probabilities as a function of $\Delta\lambda$ for HP and LP traffic for rate-controlled OBS using D-ABR.

4. Conclusions

In this paper, we study a new control-plane protocol, called Differentiated ABR (D-ABR), for congestion avoidance and service differentiation in OBS networks. For non-bursty traffic, we show, using simulations, that the optical network can be designed to work at any desired burst blocking probability by the flow control service of the proposed D-ABR protocol. The proposed architecture moves congestion away from the OBS domain to the edges of the network, where buffer management is far easier and less costly. Consequently, the need for expensive contention resolution elements like TWCs and/or sophisticated FDL structures can be reduced by incorporating D-ABR in the control plane.

Moreover, D-ABR enables strict isolation among high-priority and low-priority traffic throughout the OBS domain. For this purpose, we introduce the concept of prioritized max–min fair allocation for OBS networks and, through a number of simulations, we show that prioritized max–min fair shares are achievable by D-ABR deployment in OBS networks. This feature of D-ABR can help operators to extend their existing strict priority-based service differentiation policies to OBS domains. The benefits in using rate-controlled OBS remain to be seen in the case of more realistic traffic models and, more specifically, TCP traffic, which is left for future research.

Acknowledgements

N. Akar's work is supported in part by The Science and Research Council of Turkey (Tübitak) under project no. EEEAG-101E048 and by the Commission of the European Community IST-FP6 e-Photon/ONe project.

References

[1] C. Qiao, M. Yoo, Optical burst switching (OBS)—a new paradigm for an Optical Internet, J. High Speed Netw. (JHSN) 8 (1) (1999) 69–84.

[2] G.N. Rouskas, L. Xu, Optical packet switching, in: K. Sivalingam, S. Subramaniam (Eds.), Emerging Optical Network Technologies: Architectures, Protocols, and Performance, Springer, Norwell, Massachusetts, 2004, pp. 111–127.

[3] Y. Chen, C. Qiao, X. Yu, Optical burst switching: A new area in optical networking research, IEEE Netw. Mag. 18 (3) (2004) 16–23.

[4] R.A. Barry, P. Humblet, Models of blocking probability in all-optical networks with and without wavelength changers, IEEE J. Sel. Areas Commun. 14 (1996) 858–867.

[5] N. Akar, E. Karasan, Exact calculation of blocking probabilities for bufferless optical burst switched links with partial wavelength conversion, in: 1st Conference on Broadband Networks, BROADNETS'04, Optical Networking Symposium, 2004, pp. 110–117.

[6] S.Y. Wang, Using TCP congestion control to improve the performances of optical burst switched networks, in: IEEE International Conference on Communications, ICC, vol. 2, 2003, pp. 1438–1442.

[7] G.P. Thodime, V.M. Vokkarane, J.P. Jue, Dynamic congestion-based load balanced routing in optical burst-switched networks, in: IEEE GLOBECOM, vol. 5, San Francisco, CA, 2003, pp. 2694–2698.

[8] F. Farahmand, Q. Zhang, J.P. Jue, A feedback-based contention avoidance mechanism for optical burst switching networks, in: Workshop on Optical Burst Switching, WOBS, 2004.

[9] A. Maach, G. von Bochmann, H.T. Mouftah, Congestion control and contention elimination in optical burst switching, Telecommun. Syst. 27 (2–4) (2004) 115–131.

[10] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, RFC 2475: An architecture for differentiated services, status: PROPOSED STANDARD, December 1998.

[11] M. Shreedhar, G. Varghese, Efficient fair queueing using deficit round robin, in: ACM SIGCOMM, 1995, pp. 231–242.

[12] M. Yoo, C. Qiao, S. Dixit, QoS performance of optical burst switching in IP-over-WDM networks, IEEE J. Sel. Areas. Commun. 18 (2000) 2062–2071.

[13] K. Dolzer, C. Gauger, On burst assembly in optical burst switching networks—a performance evaluation of Just-Enough-Time, in: 17th International Teletraffic Congress, Salvador, Brazil, 2001, pp. 149–161.

[14] F. Callegati, G. Corazza, C. Raffaelli, Exploitation of DWDM for optical packet switching with QoS guarantees, IEEE J. Sel. Areas Commun. 20 (1) (2002) 191–201.

[15] C. Loi, W. Liao, D. Yang, Service differentiation in optical burst switched networks, in: IEEE Globecom, vol. 3, 2002, pp. 2313–2317.

[16] J. Liu, N. Ansari, T.J. Ott, FRR for latency reduction and QoS provisioning in OBS networks, IEEE J. Sel. Areas Commun. 21 (7) (2003) 1210–1219.

[17] Y. Chen, M. Hamdi, D.H. Tsang, Proportional QoS over OBS networks, in: GLOBECOM'01, 2001, pp. 1510–1514.

[18] Q. Zhang, V.M. Vokkarane, J.P. Jue, B. Chen, Absolute QoS differentiation in optical burst-switched networks, IEEE J. Sel. Areas Commun. 22 (9) (2004) 1781–1795.

[19] R. Jain, S. Kalyanaraman, R. Goyal, S. Fahmy, R. Viswanathan, ERICA switch algorithm: A complete description, ATM Forum/96-1172, August 1996.

[20] N. Giroux, S. Ganti, Quality of Service in ATM Networks, Prentice-Hall, PTR, 1999.

[21] X. Yu, Y. Chen, C. Qiao, Study of traffic statistics of assembled burst traffic in optical burst switched networks, in: Proc. Opticomm, 2002, pp. 149–159.

[22] V. Eramo, M. Listanti, P. Pacifici, A comparison study on the number of wavelength converters needed in synchronous and asynchronous all-optical switching architectures, IEEE J. Lightwave Tech. 21 (2003) 340–355.

[23] R. Jain, Congestion control and traffic management in ATM networks: Recent advances and a survey, Computer Netw. ISDN Syst. 28 (13) (1996) 1723–1738.

[24] S. Mascolo, D. Cavendish, M. Gerla, ATM rate based congestion control using a Smith predictor: An EPRCA implementation, in: IEEE INFOCOM (2), 1996, pp. 569–576.

[25] A. Kolarov, G. Ramamurthy, A control-theoretic approach to the design of an explicit rate controller for ABR service, IEEE/ACM Trans. Netw. 7 (5) (1999) 741–753.

[26] S. Chong, S. Lee, S. Kang, A simple, scalable, and stable explicit rate allocation algorithm for MAX-MIN flow control with minimum rate guarantee, IEEE/ACM Trans. Netw. 9 (3) (2001) 322–335.

[27] D. Bertsekas, R. Gallager, Data networks, 2nd ed., Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1992.

[28] H. Boyraz, Flow control and service differentiation in optical burst switching networks, MSc Thesis, Ankara, Turkey, Electrical and Electronics Engineering Department, Bilkent University, 2005.

[29] The ATM Forum Technical Committee, Traffic Management Specification Version 4.0, April 1996.