A New Optical Burst Switching Protocol for Supporting Quality of Service *

Myungsik Yoo† and Chunming Qiao‡
† Department of Electrical Engineering
‡ Department of Computer Science and Engineering
State University of New York at Buffalo
Buffalo, New York 14260

ABSTRACT

Optical burst switching (OBS) is a new paradigm proposed to efficiently support the ever-growing broadband multimedia traffic either directly or indirectly (e.g., via IP) over all optical WDM networks. In this paper, we propose a new *prioritized* OBS protocol based on Just-Enough-Time (JET) which can provide Quality of Service (QoS) in buffer-less WDM optical networks. Specifically, we apply OBS to support two traffic classes: *real-time* and *non-real-time*, such that each burst belonging to the former is assigned a higher priority by simply using an *additional* offset time between the burst and its corresponding control (set-up) packet. We analyze the lower and upper bounds on the blocking probability of each traffic class, and evaluate the performance of the proposed *prioritized* OBS protocol through analysis and simulation. We show that *real-time* traffic can achieve a significantly reduced blocking probability by using a reasonable amount of additional offset time. In addition, the overall blocking probability and throughput can be maintained regardless of the *additional* offset time used.

Keywords: Optical Burst Switching, WDM Networks, IP services, QoS and priority, Erlang's loss formula.

1. INTRODUCTION

Over the past several years, all-optical switched networks have received enormous attention in telecommunications and computer communications due to their huge deliverable bandwidth and high degree of data and format transparency. There are two main drivers for optical internetworks. One is the explosion of Internet traffic which has doubled every three to four months, and is expected to keep growing exponentially as more bandwidth intensive applications such as high definition television, video mail and digital audio generate an unprecedented demand for bandwidth on the Internet.

The other driver for optical internetworks is the rapid advance in the Wavelength Division Multiplexing (WDM) optical networking technology, in addition to many successful efforts made on optical components [1] such as optical transmitters, receivers, filters, amplifiers, and add drop multiplexers. Recent experiments showed that using dense WDM [2], it is possible to put 40 or 80 separate wavelengths on a single fiber, each supporting a data rate of OC-48 or OC-192.

Even though the huge bandwidth of all-optical WDM networks could alleviate the effect of the explosion of Internet traffic, challenges remain on how to efficiently provide integrated services to Internet applications in such networks. So far, Asynchronous Transfer Mode (ATM) has been proposed as a solution to support Quality of Service (QoS) according to traffic classes [3–8]. However, the size of ATM cells is too small for the high channel bandwidth of fiber optic, and a plenty of buffer space is required to manipulate elaborate QoS operations. In particular, the buffer not only introduces additional control complexity, but is also difficult to implement in optical networks since

---

* This research is supported in part by a grant from NSF under contract number ANI-9801778

one has to deal with either E/O and O/E conversions in order to use electronic buffers, or optical fiber delay lines (FDLs), which is curbersome with the current technology.

On the other hand, Internet Protocol (IP) as a simple, flexible, robust, and scalable mechanism has been successful in handling Internet traffic over the past couple of decades. Although the recently proposed Multi-Protocol Label Switching (MPLS), which avoids the forwarding overhead of conventional Internet Protocol (IP) routing, is promising, the classless *best effort* nature of IP routing limits its usefulness in supporting future integrated Internet services. Thus, for the next generation optical Internet, a new switching paradigm which can provide different levels of QoS and at the same time efficiently support bursty traffic is highly desirable.

Previously, we have proposed a paradigm called optical burst switching (OBS) and in particular, a protocol called Just-Enough-Time (JET) for supporting bursty traffic in high speed optical backbone networks [9,10]. In this paper, we investigate a new *prioritized* OBS protocol based on JET, which can provide different levels of priorities to different classes of service when running IP directly over the optical networks. Specifically, our study will focus on two classes of services, *real-time* and *non-real-time*. The bursts in the real-time class have a strict bound on delay and delay-jitter, thus requiring a guaranteed low blocking probability. On the contrary, the bursts in the non-real-time class can tolerate delay but require reliable delivery which can be accomplished by buffering and retransmission.

In this study, we assume that no buffer is used, which is highly desirable in all-optical networks. Since blocked bursts have to be dropped in buffer-less optical networks, existing priority schemes such as fair-queueing strategies that require the use of buffer can thus no longer be applied. In order for real-time traffic to achieve a low blocking probability, it should be given a higher priority for resources (bandwidth) reservation. By extending the features of the JET protocol such as the use of Delayed Reservation (DR) and offset time as in the JET protocol, it is possible to efficiently implement different levels of priority in buffer-less optical networks.

The paper is organized as follows. Section 2 gives an overview of OBS and protocols such as JET. In Section 3, our new approach to guaranteed QoS which utilizes a key parameter called offset time will be described. Bounds on blocking probability in each traffic class will be analyzed in Section 4. Section 5 shows numerical results from analysis and simulation, followed by concluding remarks in Section 6.

2. AN OVERVIEW OF OPTICAL BURST SWITCHING (OBS)

In a future optical Internet, it is very important to provide efficient support for bursty traffic because of the nature of broadband multimedia services. Recent studies have shown that, in addition to traffic in a local Ethernet and between remote Ethernets (i.e., WAN traffic), traffic generated by Web browsers, wide-area TCP connections (including FTP and TELNET traffic carried over TCP connections), and variable-bit-rate (VBR) video sources are all self-similar (or bursty at all time scales). More importantly, some have concluded that multiplexing a large number of self-similar traffic streams results in bursty traffic [11–14].

Both existing switching paradigms, namely circuit switching and packet/cell switching, have inherent shortcomings when used to support bursty traffic in optical networks. Optical circuit switching can be based on two-way reservation protocols such as tell-n-wait (TAW) [15] and ATM Block Transfer with Delayed Transmissions (ABT-DT) [16,17], where a source node sends out a control packet to make reservation, then waits for an acknowledgment to come back before transmitting data. It requires at least the end-to-end round trip time (RTT) to set up the path, and is suitable to traffic requiring connections that last much longer than RTT. As an example, at 2.5 Gbps, a burst of 500 Kbytes (or 4,000 average-sized IP packets) can be transmitted in about 1.6 $msec$, and RTT is about 2 $msec$ over a distance of merely 200 Km. Although guaranteed services can be provided to an established connection, the bandwidth utilization will be low if the connection has a relatively shorter duration than RTT, especially when traffic is bursty and the bandwidth supporting the peak rate is reserved for the connection.

An alternative to circuit-switching is packet/cell switching whereby the payload (data) is sent along with its header (control). It can efficiently support bursty traffic due to its nature of statistical sharing of the resources. However, it assumes that buffer space is available at each node. Specifically, while the header is being processed, the payload needs to be buffered, which necessitates O/E and E/O conversions along with electronic buffer (e.g., RAM) or the use of fiber delay lines (FDLs), either of which is difficult to implement in optical networks. Note that such buffering is to compensate for the processing time, not to resolve conflicts. In particular, due to variations in the processing time at the intermediate nodes, all-optical packet/cell switching requires a complicated control and stringent synchronization. Another problem with packet/cell switching is that the size of the payload (especially those ATM cells) is usually too small when considering the high channel bandwidth of optical networks, thus normally resulting in a relatively high control overhead. For example, a single ATM cell will be transmitted in 42.4 *nsec* over OC-192, and hence the processing time has to be below 2 *nsec* to keep the overhead down to 5%.

Optical burst switching (OBS) has been proposed to achieve the balance between the coarse-grained circuit switching and the fine-grained packet/cell switching, thereby combining the best of the two while avoiding their shortcomings. It is based on one-way reservation protocols such as JET [9,10], tell-n-go (TAG) [15,18,19], fast reservation protocol (FRP) [17,20], and ATM Block Transfer with Immediate Transmission (ABT-IT) [16], in which a data burst follows a corresponding control packet without waiting for an acknowledgment. Table 1 compares the three switching paradigms qualitatively.

Optical Switching Paradigms	Bandwidth Utilization	Latency (set-up)	Optical Buffer	Proc./Sync. Overhead (per unit data)	Adaptivity (traffic & fault)
Circuit	low	high	not required	low	low
Packet/Cell	high	low	required	high	high
OBS	high	low	not required	low	high

Table 1. A comparison between three optical switching paradigms

OBS protocols can be roughly classified into two types: JET-based and TAG-based [9,10]. The former uses an offset time, T, between each burst and its control packet, but the latter does not. Specifically, using JET, a source sends out a control packet, which is followed by a burst after $T \geq \sum_{h=1}^{H} \Delta(h)$, where $\Delta(h)$ is the (expected) control delay (i.e., the processing time incurred by the control packet) at hop $1 \leq h \leq H$. Because the burst is buffered at the source (in the electronic domain), no FDLs are necessary at each intermediate node to delay the burst while the control packet is being processed (but such FDLs are necessary when using TAG-based protocols). In addition, JET uses delayed reservation (DR) to efficiently utilize the bandwidth. For example, as shown in Figure 1, the bandwidth on hop i is reserved from the time the burst is expected to arrive, i.e., $t_b = t_c + T(i)$, where t_c is the time the control packet arrives and $T(i) = T - \sum_{h=1}^{(i-1)} \Delta(h)$.

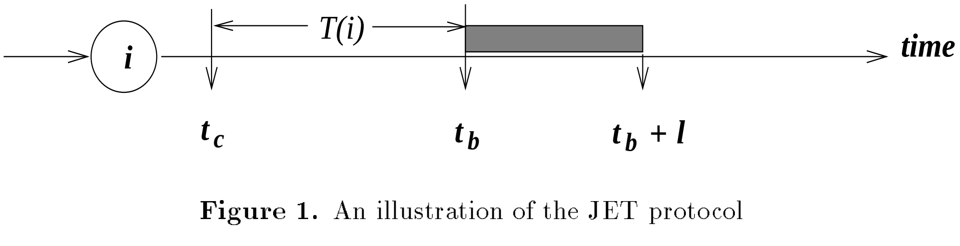

Figure 1. An illustration of the JET protocol

Note that other variations in the format of burst and in the reservation scheme are also possible. For example, one variation is to establish (and then release) an entire session during which several bursts can be transmitted. It is different from circuit switching in that reservation is one-way instead of two-way, although the two may share some disadvantages. Another variation is to use fixed burst length and synchronous switching as in packet switching except that the burst length is long to reduce the control overhead. However, none of these OBS protocols can provide guaranteed QoS.

3. GUARANTEED QOS IN OBS

In this section, we will describe a new JET-based OBS protocol which uses offset time as a simple, yet effective way to provide different levels of priorities. We consider two classes of service: namely class 0 and class 1, respectively. Class 0 corresponds to the *best effort* or *non-real-time* services for applications such as transporting plain data, while class 1 corresponds to *guaranteed* or *real-time* services for applications involving audio and video communications. Since class 1 traffic should be delivered with a strict bound on delay and delay jitter, and thus requiring a low blocking probability, it is given a higher priority for resource (bandwidth) reservation.

Recall that JET-based protocols utilize an offset time to avoid the mandatory use of FDLs. It turns out that offset time can also be used to provide different QoS in buffer-less optical networks. Specifically, in order for class 1 to have a higher priority for bandwidth reservation, an *additional* offset time, denoted by t_{offset}, is given to class 1 traffic (but not to class 0). The value of t_{offset} is a constant, which is usually a multiple of the average burst length as to be discussed later. For simplicity, we assume in the following discussion that t_{offset} is much greater than the base offset time T, and accordingly, will ignore the latter and refer to the former as simply the offset time.

Let t_{ai} and t_{si} be the arrival time and the service-start time for a class i request (where $i = 0, 1$) denoted by $req(i)$, respectively, and let l_i be the burst length requested by $req(i)$. Since we only consider the buffer-less case, a class 0 request, $req(0)$, will try to reserve immediate bandwidth upon its arrival, and will be serviced right away if bandwidth is available (and dropped otherwise). In other words, $t_{a0} = t_{s0}$ when reservation is successful (see Figure 2(b)). However, for a class 1 request, $req(1)$, a delayed reservation is made with an offset time, t_{offset}, and hence, it will be serviced at $t_{s1} = t_{a1} + t_{offset}$ when reservation is successful (see Figure 2(a)). Note that within the same class, requests are serviced on a FCFS basis. However, interactions between the two classes create a complicated queueing effect other than that of simple FIFO.

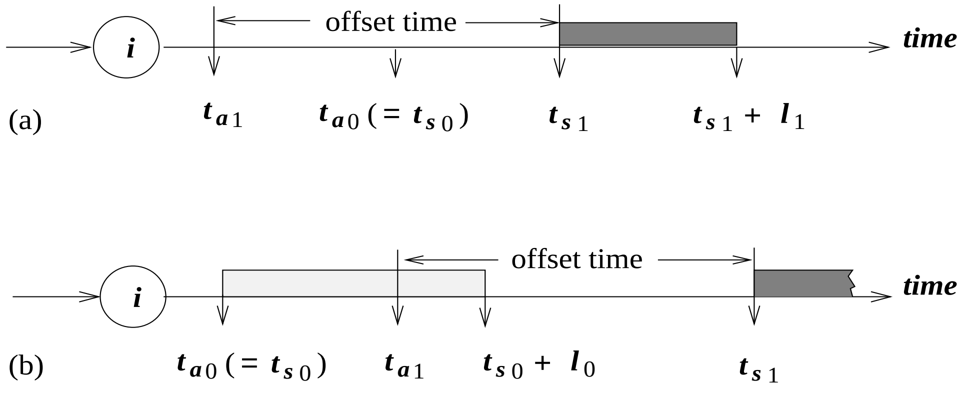

Figure 2. offset time for guaranteed service

Figure 2 illustrates why a class 1 request that is assigned t_{offset} obtains a higher priority for reservation than a class 0 request that is not. Consider the following two situations where conflicts among the two classes of traffic are possible. In the first case as illustrated in Figure 2(a), $req(1)$ comes first and reserves bandwidth using DR, and $req(0)$ comes afterwards. Clearly, $req(1)$ will succeed, but $req(0)$ will be blocked if $t_{a0} < t_{s1}$ but $t_{a0} + l_0 > t_{s1}$, or if $t_{a0} < t_{s1} + l_1$. In the second case, $req(0)$ arrives, followed by $req(1)$ as shown in Figure 2(b). When $t_{a1} < t_{a0} + l_0$,

$req(1)$ would be blocked *had* normal FIFO queue been used, that is, no t_{offset} been assigned to $req(1)$. However, such a blocking can be avoided by using a large enough offset time so that $t_{s1} = t_{a1} + t_{offset} > t_{a0} + l_0$. Given that t_{a1} may equal to $t_{a0} + \sigma$, where $\sigma > 0$ can be very small, t_{offset} needs to be larger than the maximum burst length over all bursts in class 0 in order for $req(1)$ to completely avoid being blocked by $req(0)$. With that much offset time, the blocking probability of (the bursts in) class 1 becomes independent of the offered load in class 0, and only a function of the offered load in class 1. On the other hand, the blocking probability of class 0 will be determined by the offered load in both classes.

4. LOWER AND UPPER BOUNDS ON BLOCKING PROBABILITY

In this section, we consider two OBS protocols, one that distinguishes two classes and assigns a higher priority to class 1, and the other that does not. The former will be referred to as *prioritized* OBS, while the latter, which is just the ordinary OBS, will be referred to as *class-less* OBS to emphasize its difference from the prioritized one. As mentioned earlier, we will assume, for simplicity, that the control delay at each node is negligible and so is the base offset time T. Accordingly, there is little, if any, difference between JET-based and TAG-based OBS protocols.

We will analyze the lower and upper bounds on the blocking probability of each class when using *prioritized* OBS. In order to make such analysis tractable, we consider a system with a single switch and a single output link, and assume that there are two sources for class 0 and class 1, respectively, which generate poisson traffic according to exponential distribution. Let λ_i be the average burst arrival rate and $\mu_i = 1/L_i$ be the average service rate in class i, where L_i is the average burst length in class i. Accordingly, the offered load of class i will be given by $\rho_i = \lambda_i/\mu_i$, and the total offered load will be $\rho = \sum_{i=1}^{n} \rho_i$, where n is total number of classes ($n = 2$ in this analysis).

By assuming that there are k wavelengths on the output link, and that there is no buffering at the switching node, which is capable of wavelength conversion, the blocking probability using class-less OBS can be calculated using the following *Erlang's loss formula* [21].

$$B(k, \rho) = \frac{1/k! \cdot \rho^k}{\sum_{m=0}^{k} 1/m! \cdot \rho^m} \tag{1}$$

The blocking probability of each class using *prioritized* OBS will be dependent on the value of the offset time, t_{offset}, used by class 1. Specifically, let us first examine the probability that $req(1)$ will be blocked by $req(0)$ for a given t_{offset}. Since the burst length is exponentially distributed with an average of L_0, the percentage of the bursts in class 0 whose length is no longer than t is given by the following probability distribution function (PDF) [21].

$$PDF = 1 - e^{-\mu_0 \cdot t} \tag{2}$$

which is plotted in Figure 3, where t is a multiple of the average burst length. Note that such a PDF can be used to determine R, the minimum probability that $req(1)$ will *not* be blocked by $req(0)$. For example, if t_{offset} is set to L_0, then $R = 0.6321$ since the probability that the length of a burst in class 0 is shorter than L_0 (and thus t_{offset}) is $1 - e^{-1} = 0.6321$ (see Figure 2(b) and related discussion). Note that the actual probability that req(1) is not blocked by $req(0)$ may be larger than R since as long as $req(1)$ arrives σ (time units) later than $req(0)$ and σ is not negligible, it will not be blocked by any burst whose length is shorter than than $L_0 + \sigma$.

We may rewrite Equation (2) by replacing PDF with R, and burst length t with offset time value t_{offsst} as follows.

$$t_{offset} = -\frac{\ln(1-R)}{\mu_0} = -\ln(1-R) \cdot L_0 \tag{3}$$

Table 2 shows some of the corresponding values of t_{offset} and R. Note that $R = 1$ can be achieved only with $t_{offset} = \infty$. However, as can be seen in Figure 3 and Table 2, when $t_{offset} = 5 \cdot L_0$, R is over 99%, which means

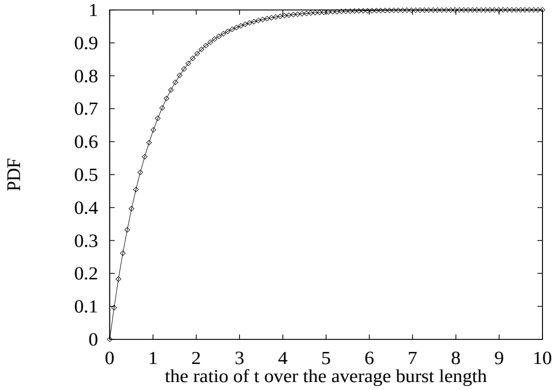

Figure 3. the PDF of exponential distribution $(1 - e^{-\mu_0 \cdot t})$

that class 1 is virtually free of blocking by class 0. In fact, even when $t_{offset} = 3 \cdot L_0$, $R = 0.95$ which is close to 1. Using a larger t_{offset} (e.g. than $5 \cdot L_0$) only increases latency without any significant increase in R.

t_{offset}	$0.4 \cdot L_0$	L_0	$3 \cdot L_0$	$5 \cdot L_0$
R	0.3296	0.6321	0.9502	0.9932

Table 2. offset time and R

We now determine the lower and upper bounds on the blocking probability of class 1, denoted by pb_1. Clearly, it is minimized when $R = 1$, and hence the lower bound on pb_1 is $B(k, \rho_1)$, which is only a function of the offered load in class 1. The upper bound on pb_1, reached when $R = 0$ (or equivalently $t_{offset} = 0$), is $B(k, \rho)$, which is the same as the blocking probability using *class-less* OBS.

We now derive the lower and upper bounds on the blocking probability in class 0, denoted by pb_0, using class-less OBS. It is intuitive that its lower bound will be $B(k, \rho)$, which is achieved when $R = 0$ or $t_{offset} = 0$. However, its upper bound is difficult to obtain through analysis because a class 0 request is blocked not only by other class 0 requests but also by *prioritized* class 1 requests.

Nevertheless, if *conservation law* applies, the overall performance (i.e., blocking probability and throughput) stays the same regardless of the values of R and t_{offset}. Specifically, let the overall blocking probability using *prioritized* OBS be pb, and denote $c_i = \rho_i / \rho$ as the proportion of the overall offered load that belongs to class i. Clearly, $pb = \sum_{i=1}^{2} c_i \cdot pb_i$ and accordingly, we can use the following equation to determine pb_0.

$$pb_0 = \frac{pb - c_1 \cdot pb_1}{c_0} \tag{4}$$

Since $pb = B(k, \rho)$ according to the *conservation law*, one may obtain the upper bound on pb_0 as

$$\frac{\rho \cdot B(k, \rho) - \rho_1 \cdot B(k, \rho_1)}{\rho_0} \tag{5}$$

where $B(k, \rho_1)$ is the lower bound on pb_1 obtained when $t_{offset} = \infty$. Note that for any given t_{offset}, once pb_i, where $i = 0, 1$, are obtained, the overall throughput, S, can be expressed as

$$S = \sum_{i=1}^{n} \rho_i \cdot (1 - pb_i) = \rho \cdot (1 - pb) \tag{6}$$

The above analysis have been verified through simulations as to be discussed next.

5. NUMERICAL RESULTS

In this section, numerical results from both simulation and analysis of the buffer-less single-node system mentioned earlier are presented. These results are obtained assuming that both sources have equal average arrival rate ($\lambda_0 = \lambda_1$) and average service rate ($\mu_0 = \mu_1$), i.e., generate equal amount of offered load ($\rho_0 = \rho_1$). For a given $\rho = \rho_0 + \rho_1$, a larger ρ_0 means a smaller ρ_1 and vice versa. But in either case, the overall performance (i.e., blocking probability and throughput) remains the same if the *conservation law* applies. As an extreme case, when either ρ_0 or ρ_1 is 0, *prioritized* OBS becomes identical to *class-less* OBS regardless of the value of the offset time.

Figure 4(a) compares the blocking probabilities as a function of the total offered load ρ (and their bounds) when $k = 16$ that are obtained through simulation (and analysis). Given that it is impossible to simulate the case where $R = 1$ (or $t_{offset} = \infty$), the top curve shows the value of pb_0 obtained through simulation when $R = 0.95$ (or $t_{offset} = 3L_0$). Note that simulations have also verified that this value is in fact very close to the value of pb_0 when $R = 0.99$ or larger. Compared to the second curve showing the upper bound on pb_0 obtained through analysis (when $R = 1$), one may conclude that the results through simulation and analysis agree with one another fairly well, although the analysis tends to yield a slightly lower value.

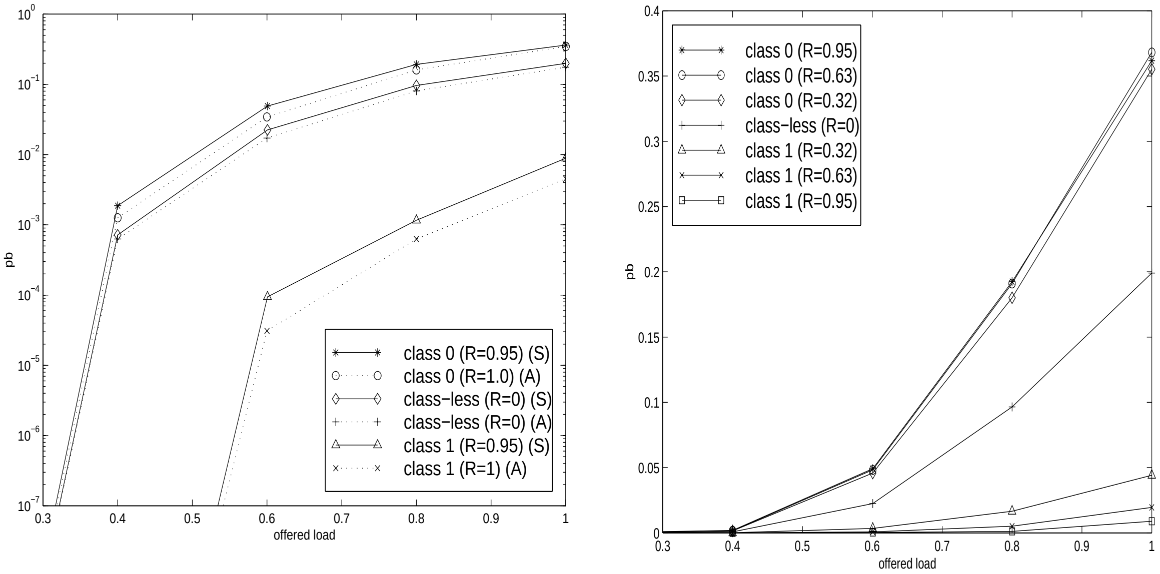

Figure 4. Blocking probability and their bounds through simulation and analysis. (a) in log scale and, (b) in normal scale

Similarly, the bottom two curves show the values of pb_1 obtained through simulation (when $R = 0.95$) and analysis (when $R = 1$), respectively. In addition, the middle two curves show the values of pb obtained through simulation and analysis, respectively, when $R = 0$. Note that, since $R = 0$, there is no difference between *prioritized* OBS and *class-less* OBS, and in addition, such a pb value is both the lower bound for pb_0 and the upper bound for pb_1.

To assess the effect of R (or t_{offset}) on the blocking probabilities, the above simulation results are replotted in Figure 4(b) using a normal scale. In addition, simulation results on pb_i for two other values, namely, $R = 0.32$ (or $t_{offset} = 0.4 \cdot L_0$) and $R = 0.63$ (or $t_{offset} = L_0$), respectively, are also shown. ¿From these results, it is clear that for a given $0 < R < 1$, the curve for $R = 0$ is at the center of the two curves representing pb_0 and pb_1, respectively. As R decreases (i.e., from 0.95 to 0.32), pb_0 decreases but pb_1 increases. The throughput of each class, being proportional to the blocking probability, follows a similar pattern.

The total blocking probability pb and throughput S are shown in Figure 5 (a) and (b), respectively. These results show that unlike the performance of each individual class, the overall performance is almost independent of R (or t_{offset}). In other words, the assumed *conservation law does* hold for pb and S. Note that, although the overall blocking probability improves slightly with R, the overall throughput degrades slightly with R at a high load (i.e., $\rho \geq 0.8$) because more and more long bursts in class 0 are dropped.

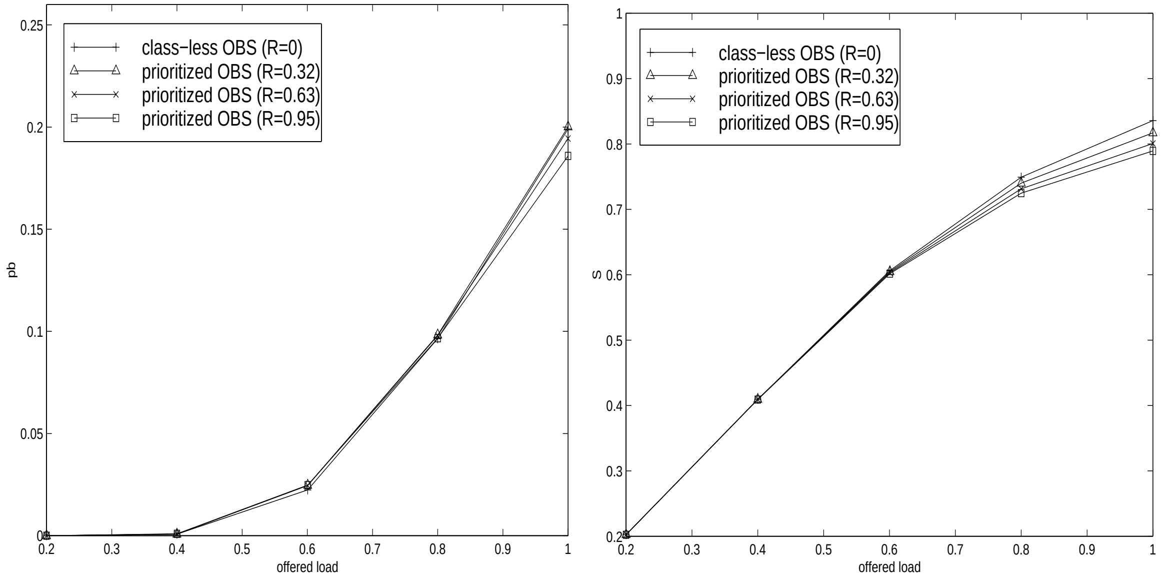

Figure 5. Overall performance (a) total blocking probability and, (b) total throughput

Figure 6 compares the minimum blocking probability of *real-time* traffic using *prioritized* OBS with that of using *class-less* OBS when $\rho = 0.8$. The results in the figure may also be used to determine the number of wavelengths required for a given tolerable blocking probability of *real-time* traffic. Hence, we only consider *real-time* traffic here because it requires a guaranteed low blocking probability in order to meet its delay and delay-jitter bounds. On the other hand, *non-real-time* traffic can tolerate a relatively higher blocking probability, and still achieve a low loss ratio through buffering or retransmission.

In Figure 6(a), the minimum value of pb_1 using prioritized OBS through analysis assuming $R = 1$, which is close to that obtainable by using $t_{offset} = 3 \cdot L_0$, is compared with that of pb through analysis with $R = 0$, the blocking probability encounted by *real-time* traffic if *class-less* OBS is used. Assuming that a burst will encounter the same blocking probability on every hop, the end-to-end blocking probability along a path spanning H hops can be expressed as $1 - (1 - pb)^H$. The values of such an end-to-end blocking probability when $H = 6$ using *prioritized* and *class-less* OBS, respectively, are shown Figure 6(b). As can be seen, with $k = 40$, the one-hop and end-to-end blocking probability of *real-time* traffic can be as low as 10^{-7} and 10^{-6}, respectively, when using *prioritized* OBS. However, they will be 10^{-2} and 10^{-1}, respectively, when using *class-less* OBS. In addition, the improvement in the blocking probability of *real-time* traffic due to *prioritized* OBS will be larger as k increases.

6. CONCLUDING REMARKS

In this paper, we have described a novel paradigm called optical burst switching (OBS) as an efficient way to support bursty traffic (i.e., IP traffic) on top of WDM networks. A prioritized OBS protocol based on Just-Enough-Time (JET) has been proposed to provide QoS. The protocol can work in buffer-less optical networks, which is a desirable feature with the current optical memory technology. It assigns a higher priority to *real-time* traffic simply by making

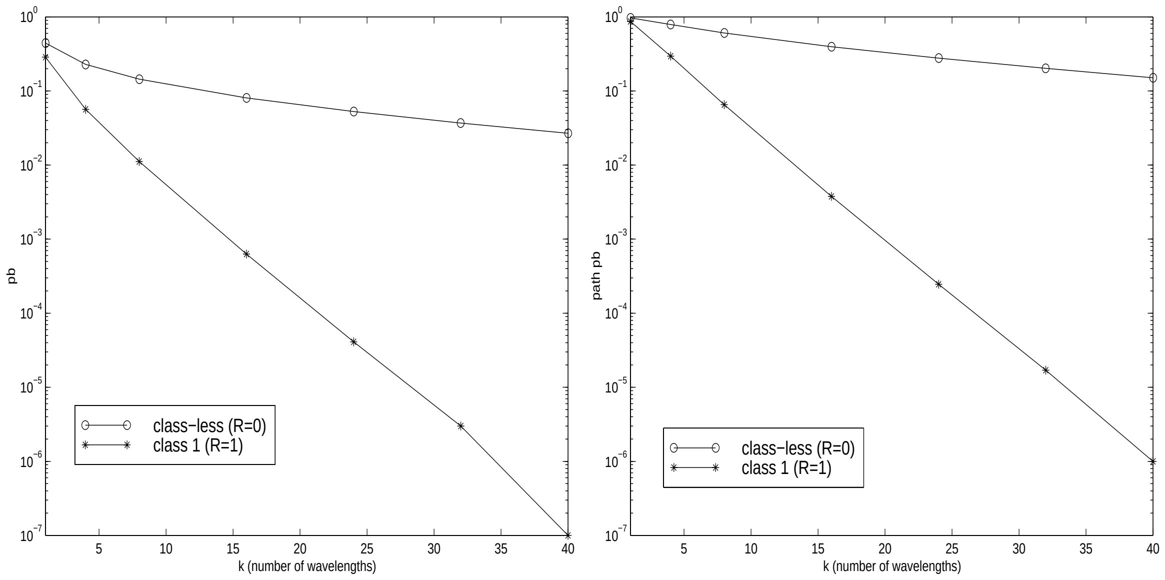

Figure 6. Blocking probability of class 1 (a) one-hop pb and, (b) end-to-end (path) pb

bursts wait for an *additional* offset time after its corresponding control packet is sent. Both simulation and analysis have shown that while the blocking probability of *real-time* traffic can be significantly reduced using the proposed prioritized OBS, the overall network performance can be maintained. Our future research will extend this work to optical networks with a limited optical buffer space provided by fiber delay lines (FDLs). It is expected that by using the buffer (and retransmission), *non-real-time* traffic can be delivered more reliably even when the same high priority is given to *real-time* traffic. The buffer space may or may not be used by *real-time* traffic depending on the delay (and loss ratio) it can tolerate.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Sudhir Dixit and Soracha Nananukul of Nokia Research Center in Boston, MA, for providing valuable suggestions and a friendly working environment during Summer 1998.

REFERENCES

1. M. S. Borella et al., “Optical components for WDM lightwave networks,” *proceedings of the IEEE*, vol. 85, pp. 1274–1307, Aug. 1997.
2. C. Bracket, “Dense wavelength division multiplexing networks: principles and applications,” *IEEE Journal on Selected Areas in Communications*, vol. 8, pp. 948–964, Aug. 1990.
3. M. Laubach, “Classical IP and ARP over ATM,” *RFC 1577*, Jan. 1994.
4. R. Cole, D. Shur, and C. Villamizar, “IP over ATM: A framework document,” *RFC 1932*, Apr. 1996.
5. P. Newnam, G. Minshall, and T. L. Lyon, “IP switching - ATM under IP,” *IEEE/ACM Transactions on Networking*, vol. 6, pp. 117–129, Apr. 1998.
6. E. Guarene, P. Fasano, and V. Vercellone, “IP and ATM integration perspectives,” *IEEE Communications Magazine*, vol. 36, pp. 74–80, Jan. 1998.
7. H. Zhang, “Service disciplines for guaranteed performance service in packet-switching networks,” *Proceedings of the IEEE*, vol. 83, pp. 1374–1396, Oct. 1995.

8. A. Varma and D. Stiliadis, "Hardware implementation of fair queuing algorithms for asynchronous transfer mode networks," *IEEE Communications Magazine*, vol. 35, pp. 74–80, Dec. 1997.
9. M. Yoo and C. Qiao, "Just-enough-time(JET): a high speed protocol for bursty traffic in optical networks," in *IEEE/LEOS Technologies for a Global Information Infrastructure*, pp. 26–27, Aug. 1997.
10. M. Yoo, M. Jeong, and C. Qiao, "A high speed protocol for bursty traffic in optical networks," in *SPIE Proceedings, All optical Communication Systems: Architecture, Control and Network Issues*, vol. 3230, pp. 79–90, Nov. 1997.
11. A. Erramilli et al., "Experimental queueing analysis with long-range dependent packet traffic," *IEEE/ACM Transactions on Networking*, vol. 4, no. 2, pp. 209–223, 1996.
12. W. Leland et al., "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1–15, 1994.
13. V. Paxon and S. Floyd, "Wide area traffic: the failure of poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226–244, 1995.
14. M. Willinger et al., "Self-similarity through high-variability: statistical analysis of ethernet LAN traffic at the source level," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 71–86, 1997.
15. I. Widjaja, "Performance analysis of burst admission control protocols," *IEE proceedings - communications*, vol. 142, pp. 7–14, Feb. 1995.
16. ITU-T Rec. I.371, "Traffic control and congestion control in B-ISDN." Perth, U.K. Nov. 6-14, 1995.
17. P. E. Boyer and D. P. Tranchier, "A reservation principle with applications to the ATM traffic control," *Computer Networks and ISDN Systems*, vol. 24, pp. 321–334, 1992.
18. G. Hudek and D. Muder, "Signaling analysis for a multi-switch all-optical network," in *Proceedings of Int'l Conf. on Communication (ICC)*, pp. 1206–1210, June 1995.
19. E. Varvarigos and V. Sharma, "Ready-to-go virtual circuit protocol : A loss-free protocol for multigigabit networks using FIFO buffers," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 705–718, Oct. 1997.
20. H. Shimonishi, T. Takine, M. Murata, and H. Miyahara, "Performance analysis of fast reservation protocol with generalized bandwidth reservation method," in *Proceedings of IEEE Infocom*, vol. 2, pp. 758–767, 1996.
21. L. Kleinrock, *Queueing Systems, volume 1: Theory.* A Wiley-Interscience Publication, 1975.