A novel switching paradigm for buffer-less WDM networks

Myungsik Yoo † and Chunming Qiao‡
† Department of Electrical Engineering ‡ Department of Compuer Science and Engineering
University at Buffalo (SUNY), Buffalo, New York 14260
‡ Phone: (716)645-3180 ext. 140 ‡ Fax: (716)645-3464
myoo@eng.buffalo.edu, qiao@computer.org

The emergence of Gigabit Ethernet and Terabit switches/routers, coupled with the continuing advances in the DWDM networking technology, provides new and strong incentives to building a flexible, efficient and bandwidth-abundant fiber-optic network infrastructure capable of providing ubiquitous services. Specifically, given that the line speed of the Terabit switches[1] and routers has approached OC-48 (2.5 Gb/s) and may soon reach OC-192 (10Gb/s), it is only natural to provide direct WDM interconnects between these switches and routers. Meanwhile, DWDM systems capable of supporting as many as 128 wavelengths, each operating at OC-192 have recently been announced. By using soliton technology, the distance between required signal regeneration points can be increased to about 6,000 kilometers. As a result, all-optical networks are becoming *feasible*, not just being desirable as a way to support data transparency for the benefit of certain applications.

For a new all-optical environment, both existing switching paradigms, namely circuit switching and packet/cell switching, have inherent shortcomings. Optical circuit switching is based on two-way reservation protocols such as tell-n-wait (TAW) or ATM Block Transfer with Delayed Transmissions (ABT-DT)[2], 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. 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 (in-band control). It can efficiently support bursty traffic due to statistical sharing of the resources. However, while the header is being processed, the payload needs to be buffered, which necessitates the use of fiber delay lines (FDLs) or O/E and E/O conversions along with electronic buffer (e.g., RAM), either of which is difficult to implement in optical networks. In particular, due to variations in the processing time of the header 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.

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 tell-n-go (TAG)[3] or ATM Block Transfer with Immediate Transmission (ABT-IT), in which a data burst (a number of packets/cells) 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: TAG-based and Just-Enough-Time(JET)-based[4]. TAG-based OBS protocol does not use an offset time, that is, a burst is sent immediately after (or almost at the same time as) the control packet with in-band (or out-of-band) control. Accordingly, the data burst has to go through Fiber Delay Lines (FDLs) at each intermediate node while the control packet is being processed. This is shown in Figure 1 (a), where p is the propagation delay per hop, and δ is the header (or control packet) processing time per hop. Note that the coupling between the timing of the burst and the control packet could be as tight as in packet/cell switching.

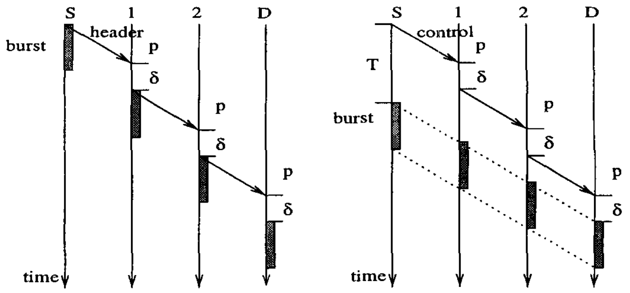

Figure 1: OBS protocols (a) TAG-based (b) JET-based

On the other hand, JET-based OBS uses an offset time, T, between each burst and its control packet to avoid the use of FDLs as shown in Figure 1 (b). The use of the offset time, combined with out-of-band control, results in a loose coupling between the timing of the burst and its control packet. Specifically, using JET, a source sends out a control packet, which is followed by a burst after $T \geq \sum_{h=1}^{H} \delta$, where H is the total number of hops to detination. In addition, JET uses delayed reservation (DR) to efficiently utilize the bandwidth. For example, as shown in Figure 2, 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 $T(i) = T - \sum_{h=1}^{(i-1)} \delta$. Note that JET-based OBS protocols can use FDLs for resolving conflicts between bursts, and in fact, DR results in efficient use of the buffer (i.e., FDLs) as well.

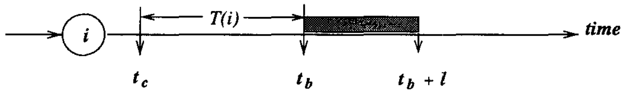

Figure 2: An illustration of the Delayed Reservation (DR)

Other variations of OBS are also possible according to the format of burst and the reservation scheme. 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 made long to reduce the control overhead.

Figure 3 (a) shows the performance comparison between the JET protocol (which uses DR) with a variation called NoDR, which does not uses DR. Assuming that both use the same minimum offset time, two cases in which no FDLs are needed (i.e., $b = 0$) and in which FDLs providing the maximum delay (buffering) time equal to the average burst length are used (i.e., $b = 1$), respectively, are considered in a torus (or meshed-ring) network having 16 nodes. The results indicate that JET which uses DR alone (i.e., $b = 0$) can achieve the same performance as NoDR which uses FDLs alone (i.e., $b = 1$). In addition, JET ($b = 1$) can outperform NoDR ($b = 0$) by about 80%, and the other two by at least 50%. Since the performance improvement of JET ($b = 1$) over NoDR ($b = 0$) is larger than the sum of the improvement of NoDR ($b = 1$) over NoDR($b = 0$) and the improvement of JET($b = 0$) over NoDR ($b = 0$), it shows that the use of DR can improve not only the bandwidth utilization, but also the buffer effectiveness (through intelligent buffer allocation and management).

By applying the concept of using offset time, JET-based OBS can implement priority mechanism in buffer-less WDM networks as illustrated in Figure 4. For example, assume that there are two classes, class 0 (e.g., non-real-time applications) with low priority and class 1 (e.g., real-time applications) with high priority. An *additional offset time*, t_{offset}, will be given to class 1, but not to class 0. Let t_{ai} and t_{si} be the arrival

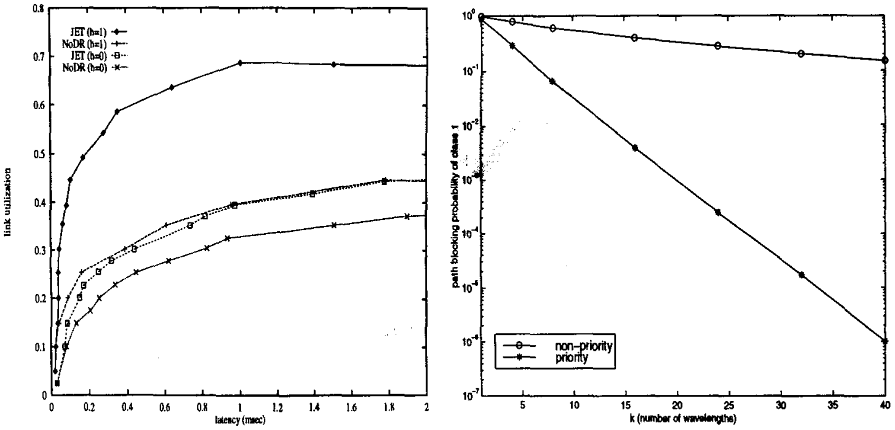

Figure 3: (a) NoDR vs. JET (b) non-priority vs. priority

time of a request and its corresponding burst in class i, respectively. For simplicity, we assume that $T = 0$, and accordingly, $t_{s0} = t_{a0}$ and $t_{s1} = t_{a1} + t_{offset}$. Consider the case $t_{a0} < t_{a1} < t_{a0} + l_0$, where l_0 is the length of burst in class 0. The burst in class 1 would be blocked *had* no t_{offset} been assigned to it. However, such a blocking can be avoided by using t_{offset} so that $t_{s1} = t_{a1} + t_{offset} > t_{a0} + l_0$. More specifically, assume that the burst length of class 0 has an exponential distribution, then as long as t_{offset} is made to be equivalent to three times of the average burst length in class 0, the chance that a burst in class 1 will not be blocked by a burst in class 0 is above 95%. Note that none of the bursts of class 0 arriving later than t_{a1} will block the burst in class 1.

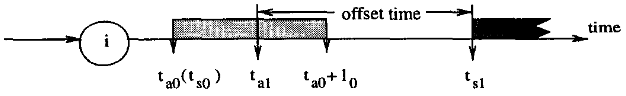

Figure 4: additional offset time for priority

Figure 3 (b) shows the analytic results on the average blocking probablity of a burst in class 1 on a path of 6 hops as a function of the number of wavelengths when offered load on each wavelength is 0.8. If no additional offset time is given to class 1, (i.e., *non-priority*), then both class 0 and 1 have the same high blocking probability. However, by utilizing t_{offset} to prioritize class 1, (i.e., *priority*), the average blocking probability of a burst in class 1 can be as low as 10^{-6} with 40 wavelengths, which is five orders of magnitude lower. Note that when class 1 is given the priority, the blocking probability of class 0 increases. However, the blocked bursts in class 0 can either be buffered, or dropped and then retransmitted later.

Finally, we note that OBS can be used to support IP over WDM by running IP software on top of every optical switch (or cross-connect). A dedicated control wavelength is used to provide the "static/physical" links between these IP entities. Specifically, it is used to support packet switching between (physically) adjacent IP entities which maintain topology and routing tables. To send data, a control packet is routed from a source to its destination, which sets up a connection by configuring all optical switches along the path. Then, a burst (e.g. one or more data IP packets, or an entire message) is delivered without going through intermediate IP entities, thus reducing its latency as well as the processing load at the IP routers.

1. J. S. Turner, Terabit Burst Switching, tech. Rep. WUCS-97-49, CS Dept., Washington Univ. at St. Louis, 1997.
2. ITU-T Rec. I. 371. Traffic control and congestion control in B-ISDN. Perth, U.K. Nov. 6-14, 1995.
3. G. Hudek and D. Muder, Signaling analysis for a multi-switch all-optical network, in Proc. of Int'l Conf. on Communication (ICC), pp. 1206–1210, June 1995.
4. M. Yoo, M. Jeong, and C. Qiao, A high speed protocol for bursty traffic in optical networks, in SPIE Proc., All optical Communication Systems, vol.3230, pp. 79–90, Nov. 1997.