A novel optical burst switching architecture for high speed networks

Amit Kumar Garg[1] **and R. S. Kaler**[2]

[1]*School of Electronics and Communication Engineering, Shri Mata Vaishno Devi University (J and K), India*
[2]*Department of Electronics and Communication Engineering, Thapar University (Punjab), India*

Received January 10, 2008

A novel optical burst switching (OBS) high speed network architecture has been proposed. To verify its feasibility and evaluate its performance, just-enough-time (JET) signaling has been considered as a high performance protocol. In the proposed architecture, to avoid burst losses, firstly, a short-prior-confirmation-packet (SPCP) is sent over the control channel that simulates the events that the actual packet will experience. Once SPCP detects a drop at any of the intermediate nodes, the actual packet is not sent but the process repeats. In order to increase network utilization, cost effectiveness and to overcome some limitations of conventional OBS, inherent codes (e.g., orthogonal optical codes (OOC)), which are codified only in intensity, has been used. Through simulations, it shows that a decrease in burst loss probability, cost effectiveness and a gain in processing time are obtained when optical label processing is used as compared with electronic processing.

OCIS codes: 060.4250, 060.4259.
doi: 10.3788/COL20080611.0807.

The proliferation of bursty internet protocol (IP) services has fuelled intensive research on the design of flexible and fast reconfigurable optical networks, since transferring the generated variable size payloads (usually over Ethernet frames) over synchronous digital hierarchy (SDH) necessitates wasteful over-provisioning. In an effort to adjust the IP paradigm to optical core networks, the optical burst switching (OBS) architecture has emerged as an attractive solution[1,2].

The basis of the OBS architecture is the separation of control information from the data. Traffic is aggregated in bursts of variable size and a burst header is emitted per burst, to pave the burst's way through the nodes. One wavelength is devoted to the control information which is the only one processed in every node. Data travel from source to destination in the optical domain without undergoing any optical to electrical conversion, which decreases the node processing requirements and thus cost. The main achievement is dynamic traffic multiplexing directly in the optical domain. Its weakness is rather heavy burst loss[3] except at very low utilization levels due to its ambitious on-the-fly switching. Due to the limited buffering capability of OBS, its performance suffers high losses, exactly as happened in Ethernet LANs, which are not tolerable in this high speed environment and drive the interest to slotted solutions[4]. Slotting can reduce the collision probabilities in OBS networks by avoiding the quite large waste of partial collisions. This concept has also been proven in passive optical networks[5] and wavelength division multiplexing (WDM) metro systems[6]. Pioneering OBS proposals, such as Just-Enough-Time (JET) and Just-In-Time (JIT) operate without acknowledgement, i.e., one way reservation. The common trend is the elimination of round-trip waiting time before the information is transmitted (the so-called Tell-And-Go-approach). Thus, OBS network combines the best circuit and packet switching in the optical realm as shown in Table 1, because it has an intermediate granularity compared with circuit and packet switching and can be implemented with mature technology. Furthermore, due to opto-electric-opto conversions, conventional electronic processing of labels becomes limited in speed and becomes the main bottleneck as data bit rates go higher. In order to increase network utilization and to overcome some limitations of OBS, studies have been done in techniques like optical buffering in fiber delay lines (FDL), grooming, deflection-routing, and various burst assembly schemes. However, these approaches lack cost effectiveness[6,7].

In the proposed OBS architecture, slotting has been used to achieve a meticulous reservation mechanism that allows sending payload on the certainty of no collision at any node. So, in this system, the IP packets are aggregated and possibly segmented at the system periphery creating fixed size data packets (slots). This paper analyzes the performance of a proposed network architecture that uses code labels based on incoherent codes, such as optical orthogonal code (OOC)[8], with JET signaling.

Considering that the average burst length is strongly related to the performance of the OBS control channel, optical code labels and optical processing have been applied to make the JET signaling at the OBS control channel. Figure 1 describes the internal architecture of OBS node.

Table 1. Comparison of Optical Switching Schemes

Optical Switching (Paradigm)	Bandwidth Utilization	Latency (setup)	Optical Buffer	Traffic Adaptively
Circuit	Low	High	Not Required	Low
Packet/Cell	High	Low	Required	High
Burst	High	Low	Not Required	High

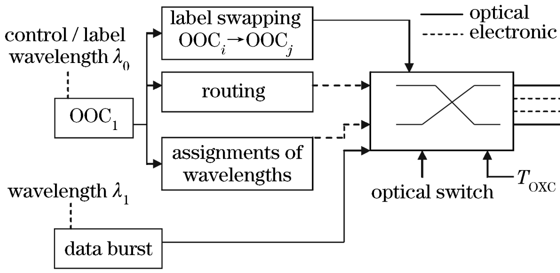

Fig. 1. Internal architecture of OBS node.

The optical processing basically consists of optical correlation of the OOC labels that arrive in the OBS node. In this case, each bit is divided up into n time periods, called chips. The total number of illuminated chips in the code is called the weight w (Hamming). The encoder of each transmitter represents each bit 1 by sending the code sequence, however, a bit 0 is not encoded and is represented by using all zero sequence. The set of OOC sequences is characterized by $(L, w, \lambda_a, \lambda_c)$, where L is the length, λ_a, λ_c are the maximum values of the autocorrelation and cross-correlation, respectively. Considering x and y as two sequences in an OOC, the autocorrelation and cross-correlation are given respectively by

$$\sum_{l=0}^{L-1} x_l x_{l+\tau} = \begin{cases} w, & \text{for} \quad \tau = 0 \\ \leq \lambda_a, & \text{for} \quad 1 \leq \tau \leq L-1 \end{cases}, \quad (1)$$

$$\sum_{l=0}^{L-1} x_l y_{l+\tau} \leq \lambda_c, \quad \text{for} \quad 0 \leq \tau \leq L-1, \quad (2)$$

where τ is the relative delay between two sequences x_l and $y_l \in \{1, 0\}$. Considering the correlation restrictions given by Eqs. (1), (2) and the number of sequences in OOC families with the same weight, the possible lengths are given by

$$L \geq L_{\min} = [Cw(w-1) + 1], \quad (3)$$

where C represents the number of sequences in the OOC family with the same weight.

In the proposed architecture, the scheduling has been considered as "check and go", because before sending the data slot, the switching contentions along the path are checked by a "SPCP" (i.e., a control message instead of the OBS burst header). SPCPs travel in the control channel carrying information not only to prepare the optical path for the data burst switching as burst headers do, but also to be informed of the outcome of scheduling by each node. The SPCP, will return back with the outcome of the reservation attempt. To make this possible, an offset just longer than the round trip time is used. In the event of a negative outcome, the data slot will not be emitted. So, only the queues that receive a positive SPCP will emit one slot to be switched according to the already prepared schedules in each node. SPCPs are not owned by the payload slots that trigger their launch, but are used first in the relevant queues. Thus, the first-in first-out (FIFO) order is always preserved, despite the existence of negative SPCPs. Different queues are maintained per destination. Source routing is employed and knowledge about each route also includes the round trip distance measured in slots which are obtained by ranging after every route update. The entry port of each node includes a synchronizer, which by means of variable delay lines aligns the boundaries of the incoming slots in all input ports and wavelengths. A small guard band of a few tens of nanoseconds provides a safety margin. In the control channel, the transmission occurs in fixed size slots (i.e. control slots) which are the same size of the data slots and are also synchronized to the data slots. Each control slot can accommodate a large number of SPCPs, the format of which is shown in Fig. 2.

Assuming some bits reserved for future use, the SPCP size reaches 64 bits. For a transmission rate of 2.5 Gbps in the control wavelength and 110 μs slots, each control slot has a size of 240,000 bits and can host more than 3,800 SPCPs, making congestion in the control channel extremely unlikely.

Every node acts as a source node for its local traffic and in parallel as a switch for the rest of the traffic. As source, each node is responsible for the generations of SPCPs to serve its local traffic. The SPCPs are managed per queue, with each queue associated with a destination. The SPCP departs as soon as a payload fitting the slot has been prepared. Once a SPCP with a certain T_d (offset delay) value has been launched, the relevant data will depart T_d slots after confirming a SPCP with positive reservation and a new SPCP. The SPCP will either return at planned time $(T_d - 1)$ or earlier, albeit with a negative reservation will be returned. As core, each node is responsible for the handling of transit SPCPs. As the SPCP enters each core node, the node switch learns that T_d slots later a data slot will enter from the same port in the indicated wavelength, also claiming the indicated output, implying a request to reserve a targeted future slot in this output port in the same wavelength (unless conversion is supported in which case this is to be decided by the scheduler).

In contrast with ordinary switches, which performs the next slot scheduling at any time, the nodes in this system are given an early warning by the SPCP, which enables them to execute their future switching actions later on, while simultaneously allowing them to execute their

input wavelength λ	source address	destination address	offset delay	ACK /NACK	SPCP counting reservation number	F/B	ECF
10	10	10	10	1	2	1	8

← 64-bits →

Fig. 2. SPCP format. F/B: forward or backward SP-CP direction; ECF: error correction field.

current switching actions earlier in time. When a new SPCP claims an output, any contentions from other SPCPs entering from other input ports are resolved within the present slot time, considering also any previous claims for the same slots that are already marked in the scheduling log. The latter, of course, have precedence over newer claims, among SPCPs with the same T_d, i.e., those with a higher attempt number (N_a) are preferred in the contentions. Also, if N_a is the same, the ones nearer their destination are favored. In the event that some or all channels are busy at that slot, the SPCP that could not be accommodated are switched around and return back without completing their trip to the other end. Returning with a negative reservation, these SPCPs prompt the dispatch of a new SPCP by the source edge node. No data slot is sent, avoiding loss of data. Thus, positive reservations are used to cause a departure of data and negative ones to cause the issue of a new SPCP.

The round trip time of control slots is larger than that of data slots by the extra time needed in each node for processing. Control is distributed with each node to execute the scheduling and reservations on the basis of local information. Each core switch may employ wavelength converters as well as fiber delay lines in this paper, although a significant improvement is to be expected by adding delay line buffering.

In this paper, Network burst intensity is modeled by queue theory and the burst loss probability is calculated for both optical and electronic label processings. The optical label processing time ($T_{\rm OPL}$) is given by

$$T_{\rm OPL} = \frac{1}{(L-1)/T_{\rm C}}, \quad (4)$$

where $T_{\rm C} = 1/(BL)$ is the chip-period and B stands for bit rate. By using Eq. (3), $T_{\rm C}$ can be written as

$$T_{\rm C} = \frac{1}{B\left[Cw\left(w-1\right)+1\right]}. \quad (5)$$

Equations (4) and (5) show that the optical processing time depends on the bit rate, the number of codes and the code-weight, thus, optical code processing may affect the OBS network traffic characteristics, e.g., burst length. The minimum burst length $(T_{\rm MBL})$[9] that could be transported to the OBS network is given by

$$T_{\rm MBL} = N\left[k - k_{\rm C}\right]T_{\rm P}, \quad (6)$$

where N is the number of fibers in the core-router, k is the number of wavelengths that carry the data bursts, $k_{\rm C}$ is the number of wavelengths for control, and $T_{\rm P}$ is the label processing time. To evaluate the OBS network performance as a function of time processing and burst length, a network utilization efficiency ($\eta_{\rm NUE}$) and economy of wavelength (EW) as performance metric parameters, have been considered

$$\eta_{\rm NUE}(\%) = \frac{T_{\rm burst}}{T_{\rm burst}+T_{\rm P}} \times 100, \quad (7)$$

where $T_{\rm burst}$ is the burst length. For optical processing in order to quantify the number of saved wavelength, EW is given by

$$\mathrm{EW}(\%) = \frac{W_{\rm elec}-W_{\rm opt}}{W_{\rm elec}} \times 100, \quad (8)$$

where $W_{\rm elec}$ and $W_{\rm opt}$ are the number of wavelengths utilized for electronic and optical label processing, respectively.

Also, an output port of an OBS node using JET, behaves as a M/G/K/K loss system. The traffic intensity (ρ) of the queue is given by

$$\rho = \lambda\left[T_{\rm burst} + T_{\rm OXC}\right], \quad (9)$$

where λ is the burst-arrival rate (BAR) in bursts per second and $T_{\rm OXC}$ is optical cross-connect switch-time. Similarly, the burst-loss probability is given by Erlang's formulum[10]:

$$p_{\rm b} = \frac{(1/k!)r^k}{\sum\limits_{m=0}^{k}(1/m!)r^m}, \qquad r = \rho k. \quad (10)$$

The proposed architecture has been evaluated by computer simulation using a 4×4 Torus network topology[2], where all nodes have one edge interface receiving locally generated traffic and four interfaces to core nodes. The load entering every node is destined with equal probability to each other. The routing is static, determined beforehand using the open shortest path flow (OSPF) algorithm. There are 64 wavelengths at 10 Gbps for data and one control wavelength at 2.5 Gbps in every link of the network. The slot size is 100 μs, accommodating 1-Mbit data with $T_{\rm OXC} = 1$ ms. The length of links between nodes is 220 km, corresponding to 11 slots, or 1.1 ms. Hence, the round–trip times are $2\cdot n\cdot 1.1$ ms, where n is the number of hops. Given that the maximum number of hops in the network is 4, the round-trip can reach 8.8 ms at the maximum domain dimension of 880 km. Full wavelength conversion has been assumed in all nodes, but no FDLs for delaying data. In order to provide a realistic pattern for the input traffic of the network, a traffic aggregation unit receiving bursty IP traffic from a large number of sources is simulated, generating 1 Mbit payloads. The traffic aggregation unit in each node turns into slots whenever 1-Mbit data are collected or a 3-ms time-out occurs as described in Ref. [11]. The routing is determined beforehand choosing those shortest paths that create a symmetric distribution of the routes over the links. The loading in the scenarios considered in this paper is symmetric, i.e., all nodes produce the same amount of traffic destined to all other nodes with equal probability.

The results concerning the optical label processing time are shown in Fig. 3, where optical label processing time versus the number of OOC has been illustrated by

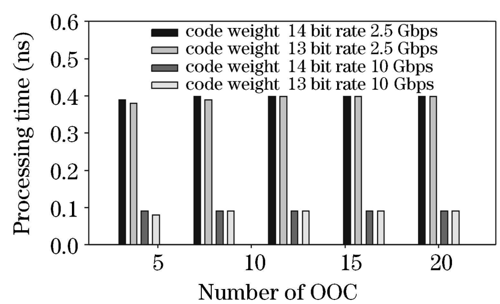

Fig. 3. Processing time of optical label processing versus number of OOC.

considering bit rates $B = 2.5$, 10 Gbps and code-weights $w = 13$, 14. It has been observed that low processing times are obtained with optical label processing. The optical label processing time converges approximately to 0.4 and 0.1 ns for bit rates of 2.5 and 10 Gbps respectively, independent of OOC number and code-weight as expected from Eq. (4). This occurs because the limitation of optical processing is the bit period which is the reciprocal of the bit rate, and the optical label processing time is very short compared with electronic processing of about 50 μs[12].

The optical label processing time may also affect the traffic characteristics of OBS network because the burst length accepted to be transported in the OBS network is a function of label time processing[9]. Table 2 summarizes these results, as expected from Eq. (6), for an OBS network with 64 wavelengths for optical burst data traffic per link along with one control wavelength. From Table 2, it has been observed that for a 10-Gbps bit rate the minimum burst length accepted to be transported is 15 ns and 4 μs for optical and electronic label processing, respectively. Figure 4 shows the utilization efficiency versus the burst-length. Considering that up to 50% of IP traffic consists of packets smaller than 520 bytes, then for a system of 10-Gbps bit rate, packets with duration of 0.417 μs are considered, 50% of these packets have the mean length of 33 ns. Then by using optical label processing, OBS network approaches conventional packet switching granularity and thus network utilization can be increased especially for small bursts composed by IP traffic.

Table 2. Minimum Burst Length Required for Electronic and Optical Label Processing

Processing-Time	Minimum Burst-Length at Bit Rates	
	2.5 Gbps	10 Gbps
Optical (ns)	60	15
Electronic (μs)	13	4

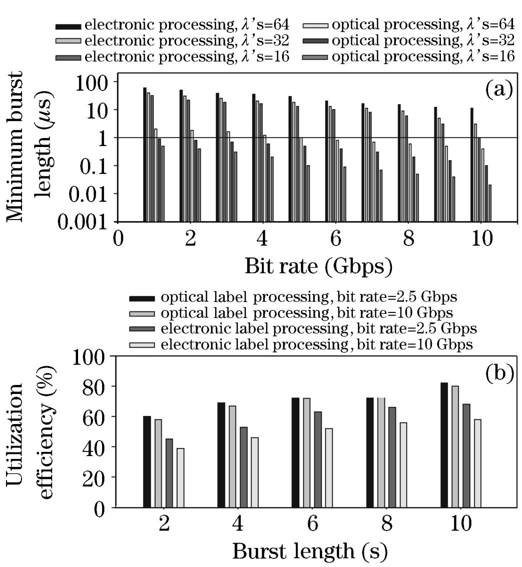

Fig. 4. (a) Minimum burst length versus bit rate as a function of number of wavelengths; (b) utilization efficiency versus the burst length as a function of bit rate.

The effects of bit rate and minimum burst length of different number of wavelengths ($\lambda's$) for transport by OBS network are shown in Fig. 4 (a), as expected from Eq. (6), for both optical and electronic label processing. It has been observed that when optical label processing is utilized, in comparison with electronic processing, the minimum burst length decreases for all number of wavelengths considered. It is also observed that as Eq. (6), the minimum burst length increases, when the number of wavelengths increases and the bit rate decreases.

A processing time of 50 μs[9] has been considered for conventional electronic label processing, whereas optical label processing times are 0.4 ns, 0.1 ns for bit rates of 2.5 and 10 Gbps, respectively. In Fig. 4(b), it is seen that the utilization efficiency for optical label processing is higher than that for electronic processing. However, when the processing time is higher than the burst length, the utilization efficiency of electronic label processing decreases whereas the utilization efficiency of optical processing is practically the same for different bit rates. These results are obtained by considering the burst loss probability equal to zero. This is an effective solution for adjusting burst length. Also, it does not depend on network resources like wavelength, as shown in Eq. (7).

From Fig. 5, it is also observed that burst-loss probability for optical label processing is lower than electronic processing at the same traffic-load (0.6 Erlangs). Using electronic label processing, a burst-loss probability of about 10^{-3} is obtained for 33 and 60 wavelengths (with BAR = 3200 and 6400 bps) whereas 30 and 53 wavelengths are required for optical label processing. This shows an improvement in wavelengths usage of OBS network, which in turn increases the network utilization gain (efficiency). For burst loss probability of 10^{-3}, there is an EW = 9% and 12 % at BAR of 3200 and 6400 bps, respectively, as expected from Eqs. (8) and (10).

Figure 6 shows the average queuing delay over all destinations versus the total offered load expressed as a percentage of the total output capacity of each node. At low loads, as expected from Eq. (9), the queuing delay remains essentially near the round trip time, since each burst departs with the return of the almost always positive, SPCP.

In Figs. 7(a) and (b), the queuing delays for payloads destined to 1-hop and 4-hop destinations respectively are shown for two different system load values (60% and > 65%). For the 60% load, both 1 and 4 hop queues show

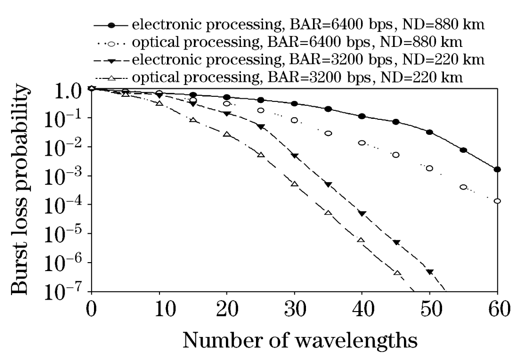

Fig. 5. Burst-loss probability versus number of wavelengths with BAR of 3200 and 6400 bps, network-diameter (ND) 220 and 880 km.

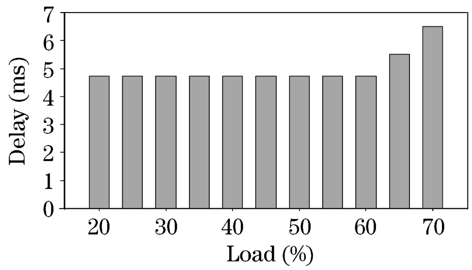

Fig. 6. Average queuing delay versus load.

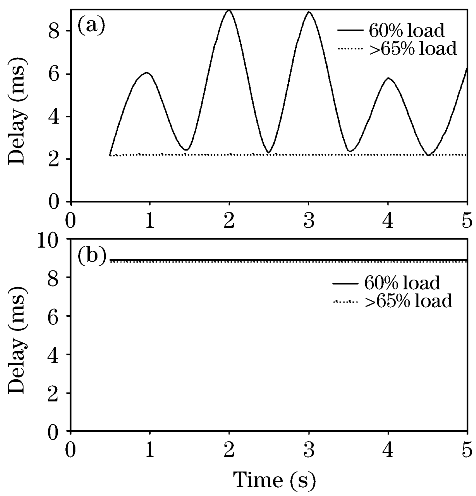

Fig. 7. Queuing delay for (a) 1-hop (b) 4-hop destinations. The delays for 4-hop destinations in 60% and > 65% load coincide in (b).

delay that slightly deviates from the respective round-trip times. For loads larger than 65%, most scouts return with a positive acknowledgement, hence the queuing delay is close to the round trip time since each burst departs as soon as the relevant scout returns back. As load increases beyond 65%, the congestion is first mirrored in 1-hop queues, since the relevant scouts are addressing future slots that are already reserved by higher-hop length queues. As a result, although the 1-hop queue experiences the congestion even for 65% load, the delay of the 4-hop queue remains intact regardless of the increasing load in the system.

Figure 8 shows the result of throughput of the network versus load. It has been observed that the throughput value of proposed OBS architecture is higher than conventional OBS (JIT-based) architecture. This is attributed as throughput gain, which provides an improvement in wavelength usage of OBS network. And the improvement in return increases the network utilization gain. Thus the proposed OBS architecture can use bandwidth more efficiently.

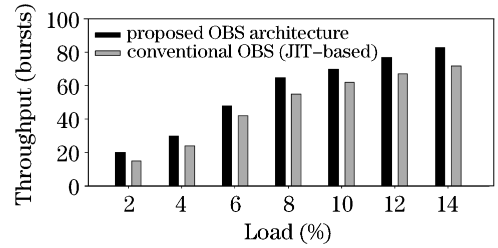

Fig. 8. Throughput versus load.

In conclusion, congestion-free OBS architecture utilizing a SPCP and optical label processing with JET signaling has been proposed. In comparison to conventional OBS architecture and electronic processing, the proposed architecture is the best. The proposed concept produces a selective throttling of traffic according to the specific congestion conditions along each route, which allows a graceful release of traffic from the periphery buffers. This is in contrast to typical OBS where the network is burdened at times of congestion with repeated retransmissions further aggravating the problem. Thus, the system makes exemplary use of the advantages of all optical technology by meticulous delegating switch planning and contention avoidance to control processing, thereby, sidestepping the lack of cheap optical buffering.

A. Garg's e-mail address is garg_amit03@yahoo.co.in.

References

1. J. S. Turner, J. High Speed Network. **8,** 3 (1996).
2. C. Qiao and M. Yoo, J. High Speed Network. **8,** 69 (1999).
3. J. Teng and G. N. Rouskas, IEEE J. Sel. Area. Commun. **23,** (2005).
4. J. Ramamirtham and J. Turner, in *Proceedings of IEEE INFOCOM 2003* **3,** 2030 (2003).
5. J. Y. Wei and R. I. McFarland, Jr., IEEE J. Lightwave Technol. **18,** 2019 (2000).
6. M. Yoo, C. Qiao, and S. Dixit, Proc. SPIE **4213,** 124 (2000).
7. G. C. Hudek and D. J. Muder, in *Proceedings of Int'l Conf. on Communication (ICC)* **2,** 1206 (1995).
8. S. K. Lee, Photonic Network Commun. **6,** 51 (2003).
9. F. Callegati, H. C. Cankaya, Y. Xiong, and M. Vandenhoute, IEEE Communication Magazine **37,** 124 (1999).
10. M. Yoo, C. Qiao, and S. Dixit, IEEE J. Sel. Area. Commun. **18,** 2062 (2000).
11. H.-C. Leligou, K. Kanonakis, J. Angelopoulos, I. Pountourakis, and T. Orphanoudakis, Eur. Trans. Telecomm. **17,** 93 (2006).
12. Y. Sun, T. Hashiquchi, V. Q. Minh, X. Wang, H. Morikawa, and T. Aoyama, IEEE Communication Magazine **43,** S48 (2005).