Robust optical burst switching

Abdelilah Maach, Gregor v. Bochmann, Hussein Mouftah
School of Information Technology & Engineering, University of Ottawa Canada
161 Louis Pasteur(CBY A617), PO. Box 450,Stn A,Ottawa,Ont,K1N 6N5 – Canada

Abstract

Contention is inherent to optical burst switching; this may lead to some burst loss, which could be fatal for some kind of applications. In this paper we propose a combination of contention reduction through congestion control and bursts retransmission to eliminate completely bursts loss. The simulation results indicate that this scheme can transform an optical burst switching to a robust burst forwarder. Simulation results also show that the retransmission technique is particularly suitable for metropolitan or local area network where the additional delay incurred by the retransmission is negligible.

Keywords: Optical network, Optical burst switching, contention reduction, burst retransmission.

1 Introduction

Optical technology has been used for a long time to carry information in fibers; however, the rapid growth of the Internet and the progress being made in Dense Wavelength Division Multiplexing (DWDM) [1,2] creates an opportunity for more extensive use of optical resources in switching and routing [3] in the second generation of optical network systems [4,5].
The novel idea of this kind of networks is to keep the information in the optical domain as long as possible. This allows the system to overcome the limitations imposed by the electronic processing and opto-electronic conversion, leading to high-speed data forwarding and high transparency. In this architecture, electronic switches are replaced by optical switches that can handle the optical information. In this paper we will be interested in optical burst switching (OBS) [1,6,7] as a forwarding method. A burst switched network carries data over DWDM links with several channels per link [8,9]. At the same time, at least one channel per link is reserved to carry control information, which is processed in the electrical domain. In OBS, data packets are collected into bursts according to their destination and class of service. Then a control packet is sent over the specific optical wavelength channel to announce an upcoming burst. The control packet, called also optical burst header OBH, is then followed by a burst of data without waiting for any confirmation. The OBH is converted to the electrical domain at each node to be interpreted and transformed according to the routing decision taken at the nodes, and pertinent information is extracted such as the wavelength used by the following data burst, the time it is expected to arrive, the length of the burst and the label, which determines the destination. This information is used by the switch to schedule and set-up the transition circuit for the coming data burst. However, the main concern is burst blocking, which may occur when two or more bursts arrive at the same time and try to leave through the same output, using the same wavelength. This problem, also known as contention [1], is inherent to the OBS technique, due to absence of buffers and storage in the intermediate nodes.
The basic differences between an optical network and a packet switched network are the techniques used to forward information at the network nodes as well as the layers involved in the routing process. Indeed, in the packet-switching network, the switches have the capacity to store and process information. In addition, an intermediate node can participate in managing and monitoring the network. Therefore, with this distributed architecture, the network can face difficult situations (in terms of load and congestion) and regulate the network load by using explicit methods to control the flux and regulate the load. However, in optical burst switching all intelligence resides in the edge nodes, which are at the same time the buffer and the processor of the network, whereas the intermediate nodes are used to forward messages according to their destination with no global coordination. Burst paths are determined at the edges according only to static information such as physical topology and the physical features of switches. This lack of information at the edge nodes (the global state of a network is unknown) may drift the network to an overloaded state where the intermediate nodes are experiencing more contentions. And hence leading to a large waste of bandwidth due to an excessive drop of bursts [10]. Besides that the dropped bursts are simply ignored by the network and rely on higher protocol to recover and retransmit the dropped information which may increase the wasted resources and the delivery delay (since a retransmission should be performed from a source to a destination).
In this work we propose to enhance the performance of optical burst switching and eliminate a burst loss completely. As a first line of defense we propose to reduce

contention by controlling the load and avoiding congestion. In the second step, we retransmit the dropped bursts. These two steps are complementary since the retransmission would be useless if the loss rate is very high. In deed if the loss rate is very high one could retransmit the same burst many times, which may increase the average number of retransmissions and hence the delivery delay increases.

In order to control the load and avoid congestion, the intermediate nodes provide the edge nodes with statistic information on the burst loss rate. Using this information one could adjust the traffic at the edge nodes. In this scheme the edge nodes could have an important role since they can store a burst or postpone its sending whereas intermediate nodes are only reporting losses. Furthermore the intermediate nodes will notify (by sending a negative acknowledgment to the node that the dropped burst belongs to) and report the loss. This way the edge node could retransmit the dropped burst and hence increasing the network robustness and reliability.

The rest of this paper is organized as follows: Section 2 presents contention problem in OBS. Section 3 presents a congestion avoidance and contention reduction technique. Section 4 presents the retransmission scheme. Section 5 presents simulation results and analysis that prove the efficiency of our proposed scheme. Section 6 concludes this work

2 Contention in optical burst switching

In OBS, the data enters the optical cloud via an edge router where it is aggregated and converted to an optical burst to be sent through the core network. The principle is similar to the one used in conventional packet switching network, however the information is separated into two parts; a header and a payload. The main goal of this separation is to minimize the opto-electrical conversion and avoid the limitation incurred by the electronic technologies such as the processing time and conversion. The header is converted to the electrical domain, where it will be processed and converted back to the optical domain. The payload is simply switched in the optical domain according to the information transported by the header. In this technique, the concept of the packet is replaced by a burst; this constitutes an interesting step towards an all-optical network where the largest part of the information remains in the optical domain.

The OBS technique may use an offset between the OBH and its corresponding burst. This offset is calculated by the edge to cover all the processing time through all the switches crossed by the burst. This assumes that the source knows the number of hops needed to reach the destination and the processing time at each node. Another alternative [6] consists of the use of delayed fiber line to delay the data burst while the OBH is being processed at an intermediate node.

Basically OBS is designed to avoid the long end-to-end setup times of conventional virtual circuit configuration with no need for memory at intermediate nodes. However the major problem is the contention, which may occur when one or more bursts arrive at the same time, at an optical cross-connect (OXC), and try to leave through the same output port, using the same wavelength. Contention is inherent to the OBS technique, which basically assumes that the network is bufferless. This feature makes it quite different from the packet switching networks. Indeed, with the electronic switches, the contention is resolved by the store and forward mechanism, which simply keeps the messages in the memory of the switch and postpones their forwarding until the contended output gets free. The contention could affect tremendously the network performance in terms of loss ratio and delivery rate.

To meet QoS requirements such as bounded delay or guaranteed delivery, contention is a key concern. Usually deploying more fibers at the same link decreases considerably contention. However this solution may be expensive especially for large networks. Several methods have been proposed in the literature to reduce the loss rate. Some of these techniques can be implemented in software, such as deflection [11] routing and segmented bursts [1] while others require specific hardware, such as burst buffering [7] and wavelength converters [7]. These methods may reduce the contention, but they all remain sensitive to the traffic load. Indeed according to [7] it is clear that even in ideal networks, where the switches use a number of buffers and can perform wavelength conversion, contention still occurs when the load gets higher. This means that the best way to deal with the contention problem is to control the traffic and keep the load in an optimal range. Furthermore, in OBS, the load control could be done only by the edge nodes since they have more intelligence and they have physical resources such as buffers and can handle both electrical and optical information. Unfortunately, they do not have enough information to adjust their throughput accordingly. No global state is available and the edge nodes are sending data bursts without any coordination.

3 Congestion avoidance and traffic shaping in optical burst switching

In order to reduce contention, the load is a determinant element, since a heavy traffic affects the performance and increases the burst loss-rate. The contention directly affects the network performance. Indeed each burst dropped means a wasted bandwidth, increased delivery delay and decreased throughput. This means that the global efficiency and performance of the global network depends on the loss rate, and hence the performance falls as the load gets higher.

Graph 1 shows a performance (in terms of delivery rate) as a function of traffic load. The graph represents only the performance pattern; the curve shape may depend on the network connectivity and the physical resources such as the number of channels by fiber and switches capacity. Each network has its own curve and it is completely characterized by this performance graph.

According to this graph the delivery rate keeps decreasing with the load, until it becomes excessively low. One can divide the traffic load into two ranges:

The area where the loss is acceptable. The critical load (CL) is the upper limit of this area. The CL itself depends on the maximum acceptable loss rate and the physical topology of the network.

Contention area where the loss is too high.

In this work we propose an approach to keep the load in the acceptable area and make sure that all the edge nodes contribute fairly to this load. The basic idea of this technique is that the edge nodes receive statistical reports (concerning the loss inside the network) that help to calculate the network performance, and hence determine from the loss-load relationship the current traffic load. Therefore by learning from this statistical data, each node increases or reduces its throughput. These statistical reports could be used by the edge nodes to monitor and control the whole network. A statistics distributor protocol could be implemented, as an extension in a control plan, using the same wavelength used to carry the burst headers.

This approach aims to control the traffic and keep it out of the congestion area. Similar approaches to congestion avoidance [12], have been considered in the literature for TCP/IP packet switched networks and asynchronous transfer mode (ATM). Congestion control is a recovery mechanism that helps a network to get out of a congestion state, whereas congestion avoidance scheme allows a network to operate in a safe area. Many solutions have been proposed in the literature to practically control congestion, the most popular are window flow-control and rate flow control. In the windows flow-control scheme (used by TCP), the destination specifies a limit on the number of packet that could be sent by the source. This limit is increased and decreased by the destination dynamically during the whole session to regulate a data flow. In rate flow-control scheme [13] (used by ATM) the destination or the network may ask a source to decrease its rate. Besides that, ATM uses other sophisticated mechanisms to control congestion including traffic shaping and admission control as well as resource reservation. Regardless of the efficiency of these mechanisms, all of them perform congestion control in the electrical level where some resources are available especially buffers and storage spaces that contribute actively in the control process. The idea of optical congestion control is to push some of these functions to the optical domain where a new constraints (buffer-less network) and new challenges rise. Performing congestion avoidance and congestion control in the optical domain increases the performance (in terms of loss rate) of optical burst switching and improves resource utilization.

To avoid congestion and achieve fairness all the edge nodes should adjust their sending traffic continually according to the feedback received from the intermediate nodes.

If we assume that Li is the traffic load of edge node Ei, then to keep the loss in the acceptable area, the load Li is constrained by the following formula: Σ Li < CL. CL is the critical load and is calculated empirically to meet the network requirements in terms of loss.

According to this formula, a global coordination is needed to meet the optimal conditions. Unfairness may occur with heavy traffic (Σli > CL) when some edge nodes send more traffic and overload the network.

The critical load (CLi) of node Ei is defined as the maximum of traffic the node can send through the network in case of heavy traffic. Cli is the quota assigned to node Ei. The critical load of all the nodes should not exceed the critical load of the network that is Σ CLi < CL.

This traffic control scheme could be performed by the edge nodes by the following algorithm:

Let LR be the loss rate, this value is calculated by the edge using the information received from the intermediate nodes. Indeed the intermediate nodes report the loss observed and the number of bursts delivered correctly.

Let CLR be the critical loss rate, this is the loss observed when the network load is in the critical load CL.

The critical load for each edge node is CLi

An edge node Ei will behave as follow:

If the load Li is less than CLi then Ei will not be involved in the adjustment process. And it can increase its load up to CLi.

But if the load Li is more than CLi, the edge Ei must do the following:

- Decreases its load if LR > CLR
- Increases its load if LR < CLR (if needed of course)
- Keeps the same load if LR = CLR

This algorithm guaranties a minimum bandwidth to each edge node. Nonetheless, when a spare of bandwidth is available, (if some edge nodes are not using their full quota) the other edge nodes can share it. They will be notified as the loss ratio is below the critical lost, thereby they can increase their load progressively until the loss ratio becomes equal to the critical loss. On the other hand if some of the edge nodes (with low traffic) increase their load, those with high traffic will give up their advance in terms of used bandwidth and if necessary they will return back to the critical load. The critical load is taken for granted for all the edge nodes.

This algorithm is a simple coordination between the different nodes of the network. Based on the report sent by the intermediate nodes, the edge nodes will measure the network efficiency. For a simple implementation, a single variable is enough to maintain the global stat, this variable is updated whenever the edge nodes receive a report, in general all the nodes receive the same information and hence they have the same value of loss rate. But for more details about the network status, the edge nodes could maintain the status of each node; in this case the edge nodes will calculate the traffic load at each node according to the report received from this node and adjust different flow separately

The information used by this algorithm is sent by the intermediate nodes using a statistic report distribution protocol.

In this protocol, all the intermediate nodes will broadcast, to the edge nodes, the number of dropped bursts and some of them (those directly connected to the edge nodes) will broadcast the number of successful forwarded bursts. This accounting information will help the edge nodes to determine in which range the network is running, thereby they can redress and rectify the situation.
The broadcasting may be performed either synchronously or asynchronously
Synchronously: each station can periodically send its report to all the edge nodes.
Asynchronously: at specific events (whenever a burst or a given number of bursts are dropped) the intermediate node will send its report to all the edges.
We think that the second technique is more suitable to measure the drop. First, there is no need for broadcasting information if there is no drop. Second, with no control information received the edge nodes assume that the network load is in the acceptable loss area.

4 Burst retransmission approach

The congestion avoidance reduces the contention and improves resource utilization. However, bursts may still suffer some losses (with small and limited loss rate). Loss sensitive applications may not tolerate this loss. Therefore strict measures should be taken to eliminate the loss completely.
In this work we propose to retransmit the dropped bursts and make sure that a sent burst is correctly delivered to its destination. In the pure OBS there is no control at the intermediate nodes; the burst is simply ignored in case of contention. The recovery is performed by higher protocols. However, in OBS with retransmission, both the intermediate and edge nodes are involved in the process. Indeed, the edge node should keep a copy of a sent burst until its delivery and the intermediate node should notify and send a negative acknowledgement (in case of contention) to the concerned node with pertinent information (burst identification).
The implementation of this retransmission scheme requires additional information; besides the label and other information related to a burst (burst length, arrival time etc) one needs the sequence number of a burst (it could be carried in the burst header control).
- The source node sends a burst, keeps a copy and sets a timer (the only delay is the propagation time since a burst is not stored in its way to its destination. therefore the source knows exactly the arrival time of the burst; a timer is set to a round-trip from a source to a destination)
- If the source receives a negative acknowledgement it retransmits the burst and repeats the same process
- If no acknowledgement is received during the timer life, the node assumes that the burst has reached its destination and removes the local copy.
Some parameters are crucial for the feasibility of such scheme; one of them is the buffer size of the edge nodes, especially for a very wide network where a round trip could be very significant and hence one may need to store many bursts during a life of a timer. Another parameter is the delivery delay, which could increase with the number of retransmissions. The network size also affects the delivery delay.
This scheme is more suitable for relatively small network (metropolitan or local area networks). In deed the size of the buffer is acceptable and the propagation delay is short and does not incur long delay in case of many retransmissions.
In order to keep the delivery delay acceptable one should control the average number of retransmissions. By controlling a load and avoiding congestion the loss rate could be decreased and consequently the number of retransmissions is reduced.
In order to evaluate the retransmission scheme we used an optical star system. In fact, a star topology is relatively simple and represents an attractive and versatile architecture that could be used to build other complex architectures.
Figure 1 shows the model we are using in this evaluation; the edge nodes send bursts to the core node, which forward them to their destinations (if resources are available) or dropped. In the latter case a notification is sent to the burst source node.
This model has the following assumptions:
Bursts have fixed length of one time unit normalized
G is the expected number of transmissions and retransmission attempts (from all edge nodes) per time unit
S is the number of successful received bursts. It is also the network throughput.
The Offered (new and retransmitted bursts) load is modeled as a Poisson process with rate G.

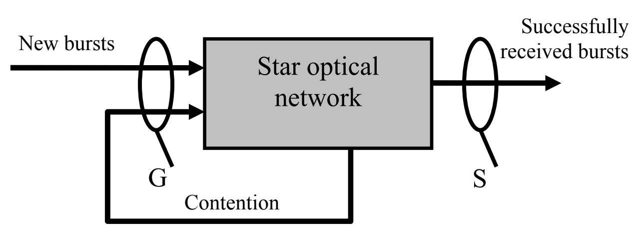


Fig. 1 A star optical network with retransmission scheme

According to this model the probability [k bursts generated in t frame times] = $\frac{(G\_t)^k}{K!} .e^{-Gt}$

No contention means there is only one burst or no burst in a period of time. That is the probability [1 burst or no burst in 1a frame time] = pf

$$pf = \frac{(G)^0}{0!} .e^{-G} + \frac{(G\_)^1}{1!} .e^{-G} = e^{-G} + G .e^{-G}$$

The contention probability is p=1- ($e^{-G} + G .e^{-G}$)
(1)

The probability to transmit a burst in exactly n transmissions is $p_n = p^{n-1}$.(1-p).

The approximate average number of transmissions of a burst $\overline{N}_r$ is given by $\overline{N}_r = \sum_{n=1}^{\infty} n \, . \, p_n =$

$$\sum_{n=1}^{\infty} n \, . \, p^{n-1} .(1\text{-p}) \qquad \text{that is } \overline{N}_r = \frac{1}{1-p} \qquad (2)$$

It is clear according to formula (2) (also intuitively) that the number of retransmissions increases with the loss rate.

5 Simulation results and analysis

In order to evaluate the performance of the proposed congestion avoidance and retransmission scheme, we performed a number of simulations on a star network. In this simulation we consider a star topology with 8 nodes besides the core node as shown in figure 2. In this model, it is assumed that each single fiber has the same number of wavelengths. All the links are bi-directional, wavelength channels are operating at 2.5 Gbps (one wavelength is used for the control channel). We assume that all the fibers have the same length. The edge nodes can send traffic to all the other edge nodes and receive as well. The core node forward bursts to their destinations. The switching time and the processing time of a control packet in the core node are set to 5 µs. Also it is assumed that no buffers and no wavelength conversion are used in the core node.

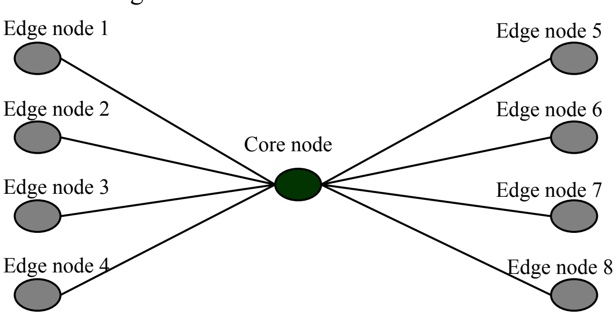
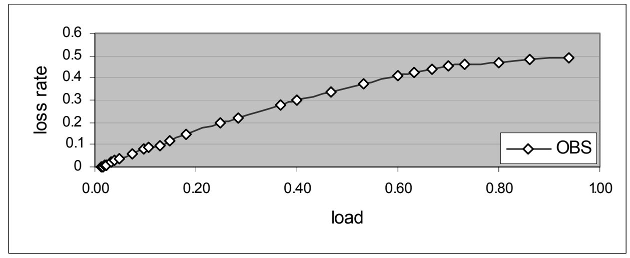


Fig. 2 Star topology with 8 edge nodes

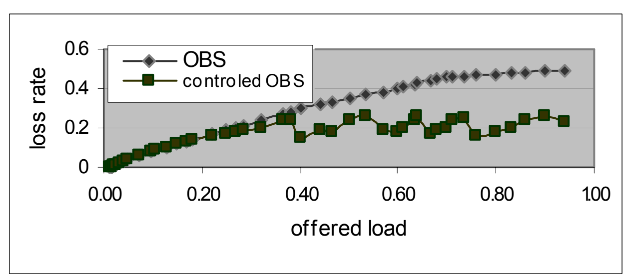
First, in order to determine the critical load for this network, we consider a simulation where each node generates bursts according to a Poisson distribution (burst arrival) where the burst length is 40 µs (100Kb with 2.5 Gbps). Each node is equipped with a burst generator. The inter-arrival time is varied and the loss probability is analyzed for each load. Graph 1 shows the loss rate versus the load. As we mentioned before the loss keeps increasing as the load gets higher (this result is conforming to the formula (1)). The critical load is a parameter design that determines the loss rate that the network designers are welling to accept. In this simulation the critical loss considered is 20%. It corresponds to a generation of burst in each node as Poisson arrival distribution with 140 ms inter arrival time



Graph1. Loss rate as function of load

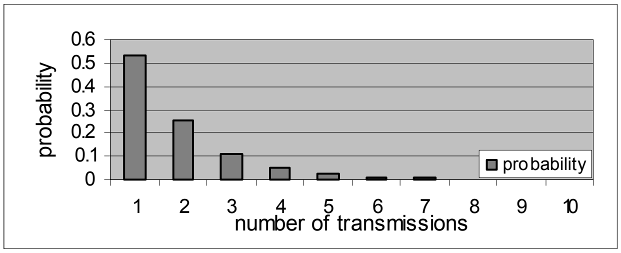
In the second simulation we test the performance of the proposed congestion control scheme against OBS without congestion control. The performance metric we use for this purpose is burst loss rate. In this model, the edge nodes are receiving traffic (they handle both electrical and optical information). The external traffic is feeding the nodes buffers. This in turn is aggregated into bursts to be sent to the core network. In the case of OBS without congestion control the burst are assembled using Poisson distribution the inter-arrival time average is increased or decreased to reduce the buffer length. Whereas, in case of OBS with congestion control the inter-arrival time is adjusted according to the statistics received from the network and the buffer size. The external traffic feeds all the nodes. However, in this simulation we divide the nodes into three categories; those who receive data with the same rate the whole session, those with increased rate and those with decreased rate. Initially, the burst generator in every node is operating with an inter-arrival time corresponding to the critical load (this is for OBS with congestion control). The destination of each burst is selected at random from a uniform distribution among all the other nodes.

The burst generation is Poisson distributed with exponential burst length. Initially the inter-arrival time, of all nodes, is 140 ms. when a node has more traffic and the critical loss is below the critical one, it could decrease the inter-arrival time of its burst generators by 5 ms to send more traffic. In this simulation the inter-arrival time is decreased by 5 ms is case of the inter-arrival time is larger than 140 ms and the loss rate is higher than the critical one.



Graph 2. Loss rate as function of load wit and without congestion control

Graph 2 shows the loss rate with and without congestion control with progressive adjustment (when the loss rate is higher than the critical one all the nodes with sending traffic larger than the critical load decrease their load by increasing the inter-arrival time of their generator by 5 ms). The loss of optical burst switching with congestion control keeps the loss lower (around the critical loss). The oscillation observed is due to the fact that the nodes sent their report only after a certain number of bursts are dropped (in this simulation, a notification is sent by a node when a 3 bursts have been dropped).



Graph 3. Average number of transmissions per burst with and without congestion control

We also investigate the average number of retransmissions required to send a burst using the retransmission scheme. Graph 3 shows the average number of transmissions with or without congestion control. For OBS without control the number or retransmission increases as the load increases. However for OBS with congestion control the average number is around a constant value which is below 1.5. These results are conforming to the formula (2).

The delay increases linearly with the number of retransmissions. A burst retransmitted n times needs n*T (T is a round trip delay). For a very wide network T maybe very significant. Therefore n should be very small to keep the delivery delay acceptable. However in local or metropolitan network the propagation delay is relatively small. In this context the retransmission scheme is very efficient and avoids returning back to the source of data (in case of a dropped burst) or higher protocol to recover

6 Conclusion

In this paper, we proposed a loss-free optical burst switching scheme. This technique aims to cope completely with the loss and guaranty a burst delivery. First we reduced the contention by controlling the load and avoiding congestion. The contention reduction relies on the intermediates node to send statistics about the loss inside the network and on the edge nodes to adjust their traffic accordingly. Since the proposed traffic control dos not eliminate the loss completely we propose another extension that aims to retransmit all the dropped bursts. A source is notified if one of its bursts is dropped and proceeds to its retransmission. The simulation results show that the combination of these two techniques leads to a robust optical burst switching where bursts suffer no loss.

7 References

[1] A.Maach, G.V.Bochmann “Segmented Burst Switching: Enhancement of Optical Burst Switching to decrease loss rate and support quality of service,” Sixth IFIP Working Conference on Optical Network design and modeling, Torino, Italy February 2002, pp. 69 - 84

[2] John Strand, Angela Chiu and Robert Tkach, “Issues for Routing in the Optical Layer”, IEEE Communications Magazine,Vol. 39, No. 2, Feb 2001, pp. 81 - 87

[3] M. Listanti, V.Eramo, R. Sabella, “Architectural and Technological Issues for Future Optical Internet Networks”, IEEE Communications Vol. 38, No, 9 Sept 2000, pp. 82 - 92

[4] Song. S., Wu. Z.” A broadband integrated services network architecture based on DWDM” Electrical and Computer Engineering, Canadian Conference, Vol. 1, May 2001, pp. 347 - 352

[5] D. K. Hunter, I. Andonovic “Approaches to Optical Internet Packet Switching” IEEE Communications Magazine, vol. 19, No. 9, Sept 2000, pp. 116 - 122

[6] M.Yoo, C.Qio, “Optical Burst Switching [OBS]- A New Paradigm for an Optical Internet,” Int’l. J. High Speed Networks, vol. 8, No.1, 1999, pp. 69 - 84

[7] Jonathan Turner, “Terabit Burst Switching,” Int’l. J. High Speed Networks, Vol. 8, No.1, 1999, pp. 3 - 16

[8] S.Verma, H.Chaskar and R.Ravikanth, “Optical Burst Switching: a viable solution for terabit IP backbone,” IEEE Network, Vol. 14, No. 6, , Nov 2000, pp. 48-53

[9] M.Yoo and C.Qiao “A New Optical Burst Switching Protocol for Supporting Quality of Service,” SPIE Proceedings, All-Optical Networking: Architecture, Control and Management Issues, Vol. 3531, Nov 1998, pp 396-405

[10] M. Yoo, C. Qiao, S. Dixit, "QoS Performance in IP over WDM Networks," IEEE JSAC, Special Issue on Protocols and Architectures for the Next Generation Optical WDM Networks, Vol. 18, No. 10, Oct 2000 pp. 2062 - 2071

[11] Y. Chen, H. Wu , D. Xu and C. Qiao, “Performance Analysis of Optical Burst Switched Node with Deflection Routing”, In Proceedings of IEEE ICC, Vol. 2, 2003, pp 1355-1359

[12] Floyd, S. and V. Jacobson: 1993, “Random Early Detection Gateways for Con- gestion Avoidance”. IEEE/ACM Transactions on Networking Vol. 1, No 4, pp 397 - 413.

[13] C. Su, G. de Veciana, and J. Walrand. “Explicit rate flow control for ABR services in ATM networks”. IEEE/ACM Transactions on Networking, Vol. 8, No. 3, June 2000 pp. 350 – 361