Telecommunication Systems 27:2–4, 115–131, 2004

Congestion Control and Contention Elimination in Optical Burst Switching

ABDELILAH MAACH, GREGOR V. BOCHMAN and HUSSEIN MOUFTAH amaach@site.uottawa.ca
School of Information Technology & Engineering, 800 King Edward Ave, Room 5105, P.O. Box 450,
Stn. A, Ottawa, ON, K1N 6N5 Canada

Abstract. Optical burst switching (OBS) is a proposed new communications technology that seeks to expand the use of optical technology in switching systems. However, many challenging issues have to be solved in order to pave the way for an effective implementation of OBS. Contention, which may occur when two or more bursts compete for the same wavelength on the same link, is a critical issue. Many contention resolution methods have been proposed in the literature but many of them are very vulnerable to network load and may suffer severe loss in case of heavy traffic. Basically, this problem is due to the lack of information at the nodes and the absence of global coordination between the edge routers. In this work, we propose another approach to avoid contention and decrease the loss. In this scheme, the intermediate nodes report the loss observed to the edge nodes so that they can adjust the traffic at the sources to meet an optimal network load. Furthermore, we propose a combination of contention reduction through congestion control and bursts retransmission to eliminate completely bursts loss. This new approach achieves fairness among all the edge nodes and enhances the robustness of the network. We also show through simulation that the proposed protocol is a viable solution for effectively reducing the conflict and increasing the bandwidth utilization for optical burst switching.

Keywords: optical network, optical burst switching, contention avoidance, load balancing

1. Introduction

Dense Wavelength Division Multiplexing (DWDM) is a fiber-optic transmission technique [Maach and Bochmann, 11; Strand et al., 15]. It is a multiplexing of many different wavelength signals onto a single fiber to obtain a set of parallel optical channels. Each channel uses a specific wavelength or color. This allows efficient use of the fiber bandwidth and hence, limits the use of additional fibers.

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 DWDM creates an opportunity for more extensive use of optical resources in switching and routing [Listanti et al., 10] in the second generation of optical network systems [Song and Wu, 14; Hunter and Andonovic, 5].

Basically, 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) [Yoo and Qiao, 24; Turner, 17] as a forwarding technique. A burst switching network carries data over DWDM links with several channels per link [Verma et al., 19; Yoo and Qiao, 23]. 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 one 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 [Yao et al., 22], 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 conventional packet switching 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 [Yoo et al., 25; Venugopal et al., 18]. Even worse, unfairness could rise among edges since dropped bursts could belong to an edge node with low traffic.

In this work we propose a protocol that can provide the edge nodes with statistical information on the burst loss rate, in order to adjust the traffic at the edges. This approach aims to control the traffic and keep the network out of congestion. In this scheme, the edge nodes could have an important role in this protocol since they can store a burst or postpone its sending whereas intermediate nodes are only reporting losses. This way one can combine the intelligence of edge nodes with the high switching capacity of intermediate node to efficiently use OBS as a reliable carrier with low loss. Furthermore, using this protocol, the sources that suffer loss will be notified so that they can schedule the retransmission of the dropped bursts.

This helps to keep the performance in an optimum state and balance the load over all the available resources such as the fiber wavelengths and intermediate nodes. Therefore, this protocol aims at reducing the burst loss rate (by controlling the load and avoiding congestion at the optical level). For farther loss reduction, one could combine this approach with other techniques. Nevertheless, there is an opportunity to enhance the performance of optical burst switching and eliminate 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. Indeed, 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.

The retransmission approach relies on the intermediate nodes to 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 the optical burst switching technique and contention problem. Section 3 presents a congestion avoidance and contention reduction technique. Section 4 presents a retransmission approach. 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

Optical burst switching is a technique for transmitting information across the network by setting up the switch and reserving resources only during the time the burst is crossing. 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 at the receiving node, where it is 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.

In an optical network using optical burst switching technique, the edge nodes are able to store and process IP packet whereas the intermediate nodes will perform forwarding according to the egress destination. Data is collected at the edge nodes and aggregated into bursts to be sent through network core. Nevertheless, prior to burst departure, the edge node sends an optical header, which informs each intermediate node of the upcoming data burst so that it can configure its switch fabric in order to switch the burst to the appropriate output port. The control packet (also called Optical Burst Header, OBH) carries pertinent information and is converted to the electrical domain to be processed at each node.

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 [Yoo and Qiao, 24] consists of the use of delayed fiber lines to delay the data burst while the OBH is being processed at an intermediate node.

The routing principle of OBS is similar to the one used by Multi-Protocol Label Switching (MPLS) [Doverspike and Yates, 2; Qiao and Buffalo, 13] in the sense that both OBS and MPLS use a label to forward the data. The MPLS label edge routers (LERs) are substituted by the edge electronic routers and label switching routers (LSRs) are replaced by optical cross-connects (OXCs). An OXC is a path switching element that establishes routed paths for optical channels by locally connecting an optical channel from an input port fiber to an output port on the switch element. This device can move optical signals between different optical fibers, without the need for conversion to the electrical domain.

OBS can take advantage of this similarity and exploits recent advances in the MPLS control plane in terms of routing protocols, traffic management and quality of services. Nevertheless, there are structural differences between LSRs and OXCs. Indeed with the former, the forwarding information is carried explicitly as part of the labels inserted at the beginning of data packets while with the latter the switching information is sent separately within another wavelength. Besides, OXCs do not perform packet level processing in the data plane while the LSRs are datagram devices, which may perform certain packet level operations in the data plane such as buffering, error correction and queuing with different level of priorities. These differences may incur some enhancement to adapt MPLS to the new environment especially to deal with the problems of quality of services and traffic engineering.

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 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. Several methods have been proposed in the literature to decrease the loss rate. Some of these techniques could be implemented in software, such as deflection [Wang et al., 20] routing and segmented bursts [Maach and Bochmann, 11], while the others require specific hardware, such as burst buffering [Chlamtac et al., 1; Turner, 17] and wavelength converters [Yates et al., 21; Turner, 17]. These techniques may reduce the contention, but they all remain sensitive to the traffic load. Indeed according to [Turner, 17], 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 as long as possible. Furthermore, in OBS, the load control could be done only by the edge nodes since they have more intelligence and adequate physical resources such as buffers and can handle both electronic 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.

In the following section, we will focus on an algorithm that controls the load and achieves fairness among all the network edge nodes. They will be able to share the available network capacity while keeping the dropping probability at a low level. In the same time, whenever a burst is dropped, the source node will be notified in order to retransmit the lost burst and hence guarantee delivery, thus avoiding the long retransmission delay of TCP.

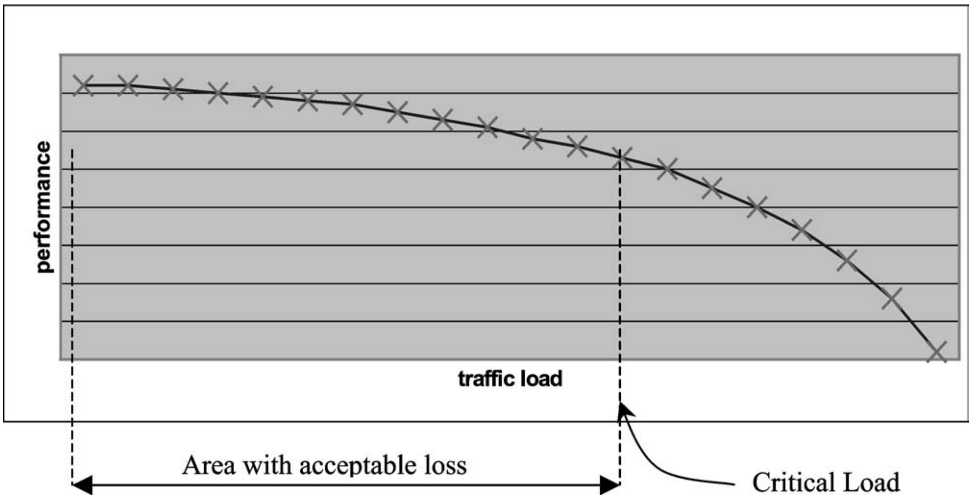
3. Congestion avoidance and traffic shaping in optical burst switching

Regardless of the routing technique used with OBS, 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.



Graph 1. Performance as function of traffic load.

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 edges 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 keeps it out of congestion area. Similar approaches to congestion avoidance [Jain et al., 6; Floyd and Fall, 3; Floyd and Jacobson, 4], 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 [Jin et al., 7] (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 [Laberteaux et al., 8, 9; Padhye et al., 12; Su et al., 16] (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 (bufferless 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.

Another concern is related to fairness. It may occur that some edge nodes flood the network which results in increased burst blocking, also for nodes with low traffic. Fair congestion control is therefore necessary.

Fairness, among all edge nodes, is considered to be achieved if:

- Each edge node is guaranteed the amount of bandwidth proportional to the whole capacity of the network. This is the quota of the edge node.
- Dropping probability of the burst belonging to edges with traffic below their quota should reflect this traffic load. This means that they do not have to pay for the excessive load generated by other edge nodes.
- Each edge gets a fair share of the excess capacity. In case that some edge nodes do not use their full quota, the bandwidth left should be shared equally among those who need more bandwidth.

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 L_i is the traffic load of edge node E_i, then to keep the loss in the acceptable area, the load L_i is constrained by the following formula: $\sum L_i < CL$. 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 ($\sum L_i > CL$) when some edge nodes send more traffic and overload the network.

The critical load (CL_i) of node E_i is defined as the maximum of traffic the node can send through the network in case of heavy traffic. CL_i is the quota assigned to node E_i. The critical load of all the nodes should not exceed the critical load of the network that is $\sum CL_i < CL$.

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 CL_i.

An edge node E_i will behave as follow:

If the load L_i is less than CL_i then E_i will not be involved in the adjustment process. And it can increase its load up to CL_i.

But if the load L_i is more than CL_i, the edge E_i 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 guarantees 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 network state. 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 flows 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. Besides that some of nodes (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.

Statistic reports will be sent by each intermediate node to all the network edges through predefined broadcasting trees established between each intermediate node and the edge nodes. As shown in figure 1, the broadcasting tree is 1 to n.

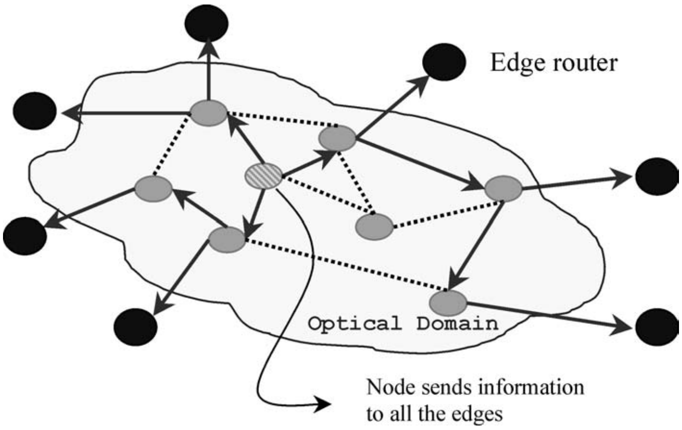


Figure 1. A broadcasting tree from a node to the network edges.

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

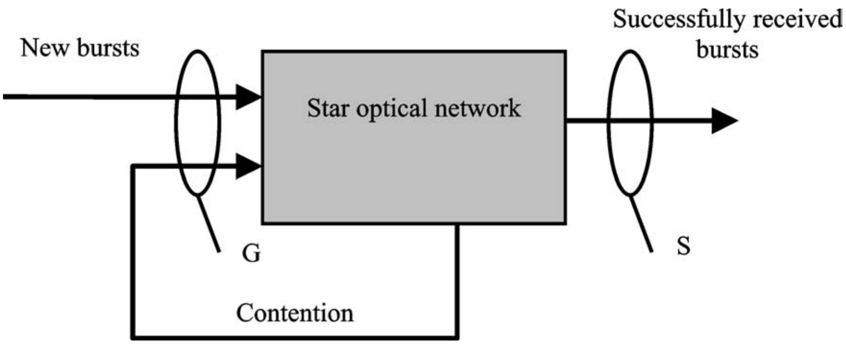


Figure 2. A star optical network with retransmission scheme.

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 2 shows the model we are using in this evaluation; the edge nodes send bursts to the core node, which forwards them to their destinations (if resources are available) or drops them (in case of contention). 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.

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

No contention means there is only one burst or no burst in a period of time. That is the probability

$$[\text{1 burst or no burst in 1 frame time}] = \frac{(G)^0}{0!}\mathrm{e}^{-G} + \frac{(G)^1}{1!}\mathrm{e}^{-G} = \mathrm{e}^{-G} + G\mathrm{e}^{-G}. \quad (1)$$

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

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

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

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 scheme, we perform a number of simulations on a mesh network. In this simulation, we consider a NSFNET topology with 14 nodes as shown in figure 3. In this model, it is assumed that each single fiber has the same number of wavelengths. All the links are bi-directional and wavelength channels are operating at 2.5 Gbps (one wavelength is used for the control channel). The fiber length is shown in figure 3, the propagation delay between two connected node range between 1.5 ms and 14 ms. Also, each node of the network consists of an optical burst switch handling both bypassing and local traffic (locally generated or terminated). A static route was chosen between each pair of nodes using Dijkstra algorithm. The switching time and the processing time of a control packet in each node are set to 5 µs. Also it is assumed that no buffers and no wavelength conversion are used in the 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) and the burst length is exponentially distributed with an average of 40 µs (100 Kb with 2.5 Gbps). Each node is equipped with a burst generator. The inter-arrival time is varied

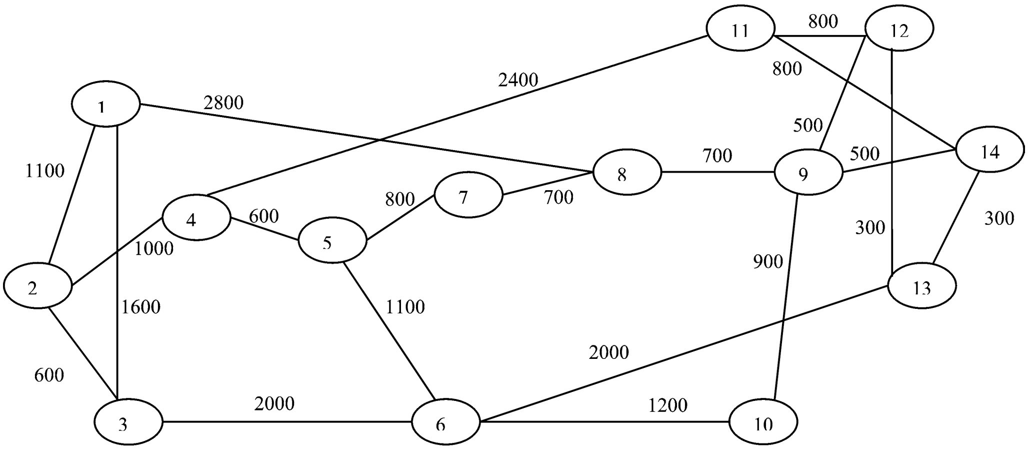
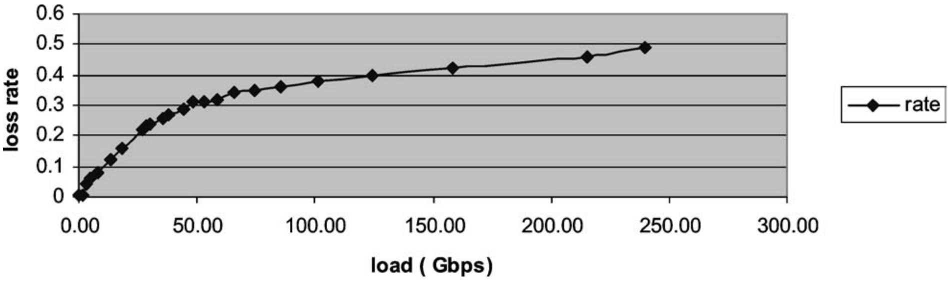


Figure 3. NSFNET topology with 14 nodes.



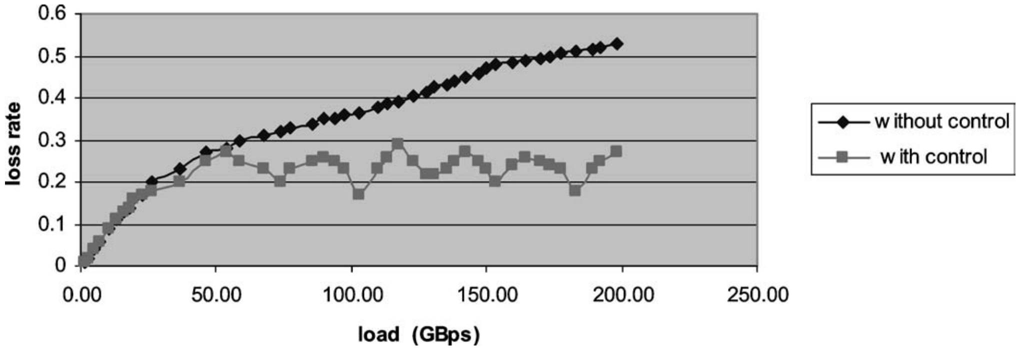
Graph 2. Loss rate as function of load.

and the loss probability is analyzed for each load. Graph 2 shows the loss rate versus the load. As we mentioned before, the loss keeps increasing as the load gets higher. The critical load is a parameter design that determines the loss rate that the network designers are willing 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 100 ms inter arrival time and length of burst exponentially distributed with an average of 40 μs.

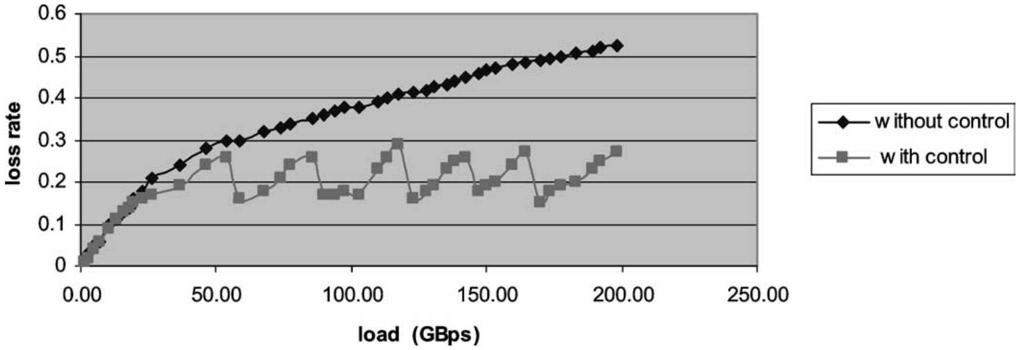
In the second simulation, we test the performance of our proposed scheme against OBS without congestion control. The performance metric we use for this purpose is burst loss rate. In this model, the edge nodes are receiving traffic (they handle both electrical and optical information). The external traffic is feeding the nodes buffers. The collected traffic is then 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 100 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, we investigate two decreasing scheme. The first one consists of decreasing the inter-arrival time by 5 ms (to send more traffic) if the inter-arrival time is larger than 100 ms and the loss rate is higher than the critical one. The second one consists of returning back to the critical load (the node sets the inter-arrival time to 100 ms when the congestion is detected).

Graph 3 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



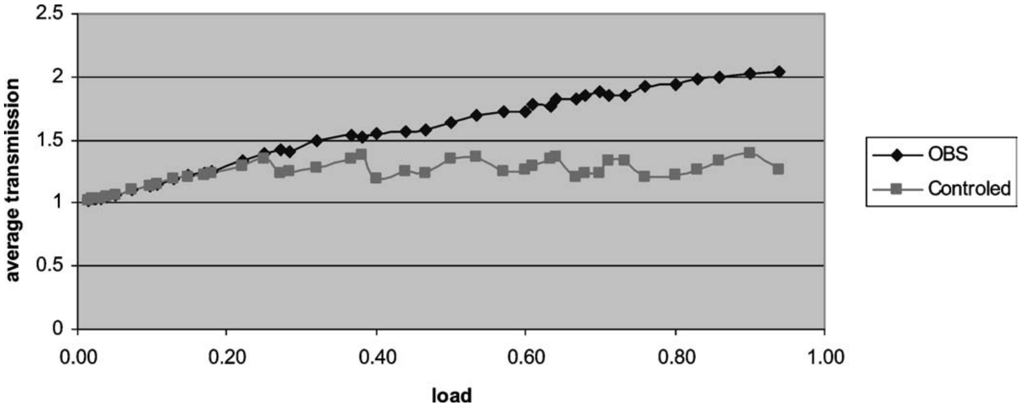
Graph 3. Loss rate as function of load wit and without congestion control.



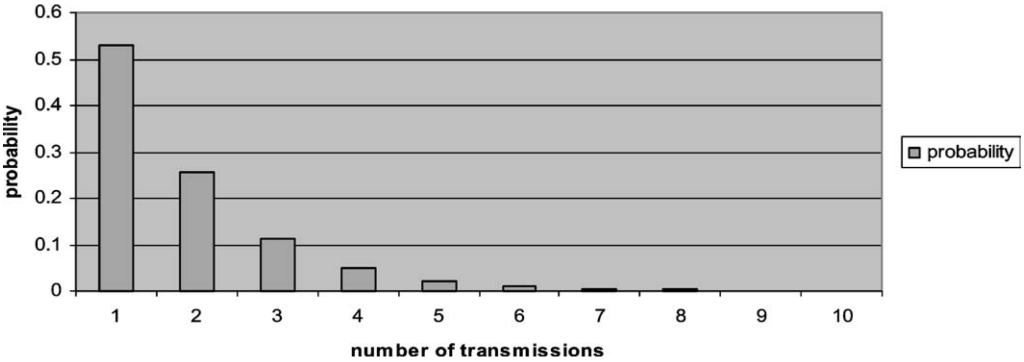
Graph 4. Loss rate as function of load wit and without congestion control.

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 burst drop (in this simulation, a notification is sent by a node when a 3 bursts have been dropped). The notification could be triggered either by a number of dropped bursts or periodically in time. In the former case, an intermediate node can send a negative acknowledgement if the number of loss reaches a given number of bursts (this number define the amplitude of oscillation). In the latter case, all the nodes will sent periodically an acknowledgment to report the loss observed during this period.

Graph 4 shows the loss rates observed in the network. In this simulation, we investigate the scheme that consists of returning back to the critical load (when congestion is detected, all the nodes with traffic load beyond the critical load should return back to the critical load). We observe that the loss is dropped sharply to the critical loss when the congestion is detected and continues to oscillate around the critical loss. Nonetheless, the result is very similar to the previous one where the loss is dropped progressively to the critical loss. And both of them prove that the congestion control technique effectively control the loss and optimize the resource utilization. For a large geographical network, the propagation delay may affect the results. Some notifications take more time



Graph 5. Average number of transmission per burst with and without congestion control.

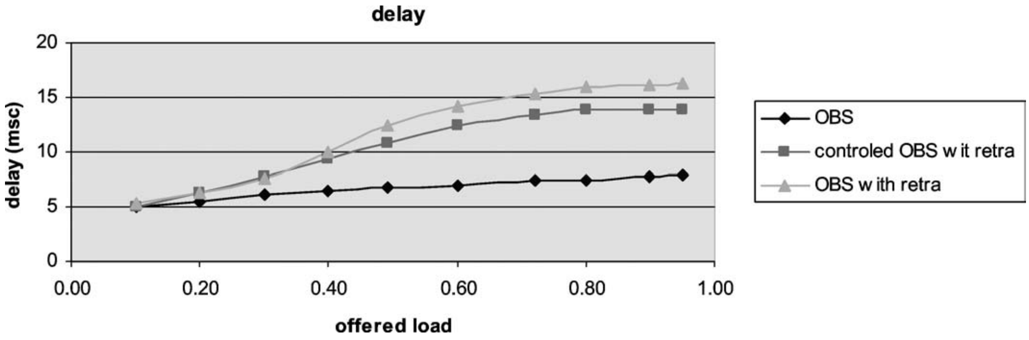


Graph 6. Distribution of the number of retransmissions.

to reach a source. In general, the retransmission scheme is very suitable for metropolitan networks.

We also investigate the average number of retransmissions needed in the average to send a burst using the retransmission scheme. Graph 5 shows the average number of transmissions with or without congestion control. For OBS without congestion 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 nT (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. The retransmission scheme assumes that the source nodes have buffers large enough to store the sent bursts until they reach their destinations.



Graph 7. Average delay per burst.

Graph 6 shows the distribution of the number of retransmissions. These results are for high traffic load with controlled load. The probability decreases quickly; only small number of retransmissions is needed to deliver a burst to its destination. More than 50% the burst reach the destination in one hop. And the probability to have an excessively high number of retransmission is very low.

Graph 7 shows the delivery delay for OBS and OBS with retransmission with or without congestion control. The graph cumulates both queuing and propagation delay. In this simulation all the fiber links have the same length (500 km each, the propagation delay from an edge node to another is 5 ms). The delay of OBS is smaller because there is no retransmission and delay is only for those that reach their destination. In fact the real delay should take into account the retransmission from a source of a dropped burst which will be longer. The delay of OBS with retransmission and congestion control is better than the one without control. This is because with out congestion a burst maybe retransmitted many times before it reaches its destination.

6. Conclusion

In OBS the edge nodes keep sending bursts regardless of the network load and without any global coordination which may overwhelm the network leading to a situation where the contention is very high and the loss is excessively unacceptable. Furthermore in case of contention an intermediate node simply drop a burst and ignore it. The higher layers are therefore in charge of detecting and recovering the loss. This may increase the burden of higher layers and increase the recovery time. And hence resource wasting (since the recovery is performed from farther source). In order to avoid this problem we think that intermediate and edge nodes should be engaged in a global process to keep the loss in an acceptable level and recover from any eventual loss. Such a process aims to enhance the performance of optical burst switching and eliminate a burst loss completely. We propose to reduce contention by controlling the load and avoiding congestion. Basically the intermediate nodes provide the edge nodes with statistic information on the burst loss rate. Which in turn are using this information to adjust their traffic and balance the load over the different wavelengths. This technique is incremented by a retransmission scheme where intermediate nodes notify (by sending a negative acknowledgment to the nodes that the dropped bursts belong to) and report the loss. This way the edge node could retransmit the dropped burst and hence increasing the network robustness and reliability.

Congestion avoidance acts as a traffic shaping and admission control in photonic domain. Nevertheless this scheme could be extended to control the congestion at every node. Every source node will receive statistics from different nodes in order to calculate the loss rate (that every node is suffering). This kind of measurements will allow the source nodes to adjust traffic flows separately taken into account the load of crossed nodes. It will also allow source node to redistribute its traffic over other paths. And hence converge to a global load balancing.

The retransmission scheme relies on buffers at the edge nodes that can hold the sent burst until a destination is reached (no negative acknowledgement received during a period of time). The size of such buffer depends on the network size and links capacity. The retransmission may incur additional delay to a burst (if one or more transmissions are needed). However this delay could be not acceptable by some class of traffic.

References

[1] I. Chlamtac, A. Fumagalli and C.J. Suh, Multi-buffer delay line architectures for efficient contention resolution in optical switching nodes, IEEE Transactions on Communications 48(12) (2000) 2089–2098.

[2] R. Doverspike and J. Yates, Challenges for MPLS in optical network restoration, IEEE Communications Magazine 39(2) (2001) 89–96.

[3] S. Floyd and K. Fall, Promoting the use of en-to-end congestion control in the Internet, IEEE/ACM Transactions on Networking 7(4) (1999) 458–472.

[4] S. Floyd and V. Jacobson, Random early detection gateways for congestion avoidance, IEEE/ACM Transactions on Networking 1(4) (1993) 397–413.

[5] D.K. Hunter and I. Andonovic, Approaches to optical Internet packet switching, IEEE Communications Magazine 19(9) (2000) 116–122.

[6] R. Jain, K.K. Ramakrishnan and D.-M. Chiu, Congestion avoidance in computer networks with a connectionless network layer, Technical Report, DEC-TR-506, Digital Equipment Corporation (1988).

[7] S. Jin, L. Guo, I. Matta and A. Bestavros, A spectrum of TCP-friendly window-based congestion control algorithms, IEEE/ACM Transactions on Networking 11(3) (2003) 341–355.

[8] K. Laberteaux, C. Rohrs and P. Antsaklis, A practical controller for explicit rate congestion control, IEEE Transactions on Automatic Control 47 (2002) 960–978.

[9] K.P. Laberteaux, C.E. Rohrs and P.J. Antsaklis, An adaptive inverse controller for explicit rate congestion control with guaranteed stability and fairness, International Journal of Control 76(1) (2003) 24–47.

[10] M. Listanti, V. Eramo and R. Sabella, Architectural and technological issues for future optical Internet networks, IEEE Communications 38(9) (2000) 82–92.

[11] A. Maach and G.V. Bochmann, Segmented burst switching: Enhancement of optical burst switching to decrease loss rate and support quality of service, in: *Sixth IFIP Working Conf. on Optical Network Design and Modeling ONDM*, Torino, Italy (2002) pp. 69–84.

[12] J. Padhye, J. Kurose, D. Towsley and R. Koodli, A model-based TCP-friendly rate control protocol, in: *Proc. of Internat. Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)* (1999).

[13] C. Qiao and S. Buffalo, Labeled optical burst switching for IP-over-WDM integration, IEEE Communications Magazine 38(9) (2000) 104–114.

[14] S. Song and Z. Wu, A broadband integrated services network architecture based on DWDM, in: *Canadian Conf. on Electrical and Computer Engineering*, Vol. 1 (2000) pp. 347–352.

[15] J. Strand, A. Chiu and R. Tkach, Issues for routing in the optical layer, IEEE Communications Magazine 39(2) (2001) 81–87.

[16] C. Su, G. de Veciana and J. Walrand, Explicit rate flow control for ABR services in ATM networks, IEEE/ACM Transactions on Networking 8(3) (2000) 350–361.

[17] J. Turner, Terabit burst switching, International Journal of High Speed Networks 8(1) (1999) 3–16.

[18] K.R. Venugopal, E.E. Rajanand and K.P.S. Kumar, Performance analysis of wavelength converters in WDM wavelength routed optical networks, in: *Proc. of the Fifth Internat. Conf. on High Performance Computing*, Barcelona, Spain (1998) pp. 239–246.

[19] S. Verma, H. Chaskar and R. Ravikanth, Optical burst switching: A viable solution for terabit IP backbone, IEEE Network Magazine 14(6) (2000) 48–53.

[20] X. Wang, H. Morikawa and T. Aoyama, Burst optical deflection routing protocol for wavelength routing WDM networks, in: *Proc. of SPIE/IEEE OPTICOM 2000*, Dallas, TX, USA (2000) pp. 257–266.

[21] J. Yates, M. Rumsewicz and J. Lacey, Wavelength converters in dynamically-reconfigurable WDM networks, IEEE Communications Surveys (1999) 2–15.

[22] S. Yao, B. Mukherjee and S. Dixit, Asynchronous optical packet-switched networks: A preliminary study of contention – resolution schemes, in: *Proc. of Optical Networks Workshop*, Richardson, TX (2000).

[23] M. Yoo and C. Qiao, A new optical burst switching protocol for supporting quality of service, in: *SPIE Proceedings, All-Optical Networking: Architecture, Control and Management Issues*, Vol. 3531, Boston, MA (1998) pp. 396–405.

[24] M. Yoo and C. Qiao, Optical burst switching (OBS) – A new paradigm for an optical Internet, International Journal of High Speed Networks 8(1) (1999) 69–84.

[25] M. Yoo, C. Qiao and S. Dixit, QoS performance in IP over WDM networks, IEEE Journal on Selected Areas in Communications 18(10), Special Issue on Protocols and Architectures for the Next Generation Optical WDM Networks (2000) 2062–2071.