On Burst Loss in Optical Burst Switched Networks with Hot Potato Deflection Routing

Stein Gjessing
Dept. of Informatics, University of Oslo & Simula Research Laboratory, P-O. Box 1080, N-0316 Oslo, Norway
steing@ifi.uio.no

Abstract

In this article we investigate the effects of increased Control Packet lead Time (CPT) on the loss probability of bursts in an OBS network with deflection routing. In particular we investigate burst loss caused by bursts overtaking their control packets, and bursts losses caused by blocking. We have made a detailed discrete event simulation model of OBS networks and evaluated the performance of hot potato deflection when bursts are routed connectionless from ingress to egress. To our knowledge such detailed investigations into the effects of increased control packet lead time and causes of burst losses have never been performed before. We conclude that increased control packet lead time gives better performance, in particular for low loads. We also conclude that there might be a point of diminishing return, above which the increase in CPT does not decrease the burst loss enough to justify the larger burst latency.

1 Introduction and motivation

Optical Burst Switching (OBS) is a possible enabling technology for the next generation optical Internet [1]. Using Wavelength Division Multiplexed (WDM) fibers, an OBS network transports the data with minimal delay, by keeping it in the optical domain. In each optical ingress node, IP packets that are destined for the same optical egress node are assembled into one large burst. Just before the burst is ready to be sent, a control packet is built and sent into the network to reserves resources on the links and in the switches on the way through the optical network to the egress node. In order to be able to reserve resources, the control packet is converted from optical to electrical (and back) in each switch, while the succeeding data burst will be switched all optical.

One of the main resources to be reserved is the time slots the data bursts need on the communication fibers between the ingress and the egress routers. In this paper we assume that the control packet reserves "Just Enough Time" on these fibers [2]. In this paper we also assume the availability of wavelength converters in all switches, so that, regardless of the wavelength used to transport a burst on the input fiber, any wavelength can be used on the output fiber. We also assume connection less routing; hence a control packet, containing only an identification of the egress router as its destination address, may find its way to this egress router regardless of where in the network it is.

The control packet is sent ahead of the burst by an amount of time called the Control Packet lead Time (CPT). Since the control packet is converted from optical to electrical and back in each switch, a burst gets closer to its control packet each time they pass a switch. After having passed a number of switches, the burst will overtake the control packet, and if this is not the egress node, both the control packet and the burst are discarded. Hence it is important that the CPT is large enough so that this does not happen.

Even with wavelength conversion, it may be the case that a control packet can not find a slot for a succeeding burst on the preferred output fiber. One possibility is then to drop the control packet (and the burst) and let an end to end protocol take care of the resending of the burst (or resending of the contained IP packets). Two methods, often used in combination, may be used to reduce the probability of burst drop in case of such contention.

One method is to use optical buffers, implemented by Fiber Delay Lines (FDLs). After a round trip in a FDL, the burst may now be scheduled on the original output fiber. Fiber delay lines are not considered in this article. Another method is to route (deflect) the burst out on another fiber than the one on the primary path to the egress node. There are several ways to do this [4]. In the present paper we will analyze the use of so called hot potato deflection routing in irregular networks.

This paper is organized as follows. In the next section we discuss deflection routing in OBS networks and some previous research results regarding this topic. We also state more precisely the topic under consideration in this article. Then in sections 3 and 4 a short description of our simulation environment and the networks that will be simulated, are given. Section 5 details and discusses the simulation results. Finally in section 6 we conclude and point to some topics that needs more research.

2 Deflection in OBS networks

To reduce the burst blocking probability, deflection routing was proposed in [3]. When a burst is deflected, it normally uses a longer path and will hence be routed through more switches. If the CPT is minimal, the burst will then be dropped because the burst will overtake its control packet. One way to prevent this is to use FDLs in the switches, since this will cause the burst, but not the control packet, to be delayed [4].

Kim et al. [7] proposes to increase the CPT in combination with deflection routing. It is shown how the average number of hops increases with increasing load. The consequences of increasing the CPT, or "offset" as they call it, will be studied further in the present paper. We use three different irregular networks of considerable sizes to find results that are as general as possible.

Deflection makes bursts travel longer paths in the network. Hence congestion may become even worse [3]. For higher network loads, deflection is reported to destabilize the network. Because of this, some research find ways to combine deflection with other methods when the load becomes high [5,6].

Hot potato deflection means that whenever there is a contention on the primary output fiber, a random other one (including the one that the burst arrived on) is tried. If this fiber is also unavailable, than all output fibers are tried (in random order) until either one that can accommodate the burst is found, or the burst must be dropped.

Many deflection routing protocols rely on predefined alternative paths. This adds considerable complexity to the routing protocol. In this paper we use connectionless routing with routing based on the address of the egress router. Hence each router needs only one routing table, and there is no need to try to find loop-free alternative paths.

Hot potato routing may create long routing loops when the network is congested. It is the purpose of this paper to analyze exactly what happens when bursts experience contention and are deflected on alternative links. We will let the ingress router send the control packet out with an increased CPT instead of using FDLs in all routers. We will investigate how this increased CPT influences burst loss due to bursts overtaking their control packets. The CPT dictates how long the bursts may loop in the network. When the load gets high, some bursts will obviously be blocked (because there are contentions on all links out of the router), while other bursts will overtake their control packets and then has to be discarded. We hope to use the insight gained to find ways to limit the deflection in case of high loads, because deflection at low and reasonable loads are obviously a good way to reduce burst losses, and may work fine also for higher level protocols such as TCP [8].

3 The evaluation environment

The J-sim framework [9] has been used to implement a full OBS discrete event simulation model. The OBS-switches and the schedulers are built from scratch. Network data for a specific scenario, including topology and link propagation times are read from a file at system start up time.

All traffic is synthetically generated. Each ingress node has as many Poisson processes as the number of optical network egress nodes. A fixed burst size of 50 000 bytes is used in this article. Depending on the traffic matrix, the mean arrival rate of the bursts destined for one and the same egress node is determined. In the experiments used in this article an equal all-to-all traffic matrix between ingress and egress nodes is used. Also all switches in the network act as both ingress and egress nodes.

While the bursts are kept in the optical domain, and use very short time through a node, we assume that the control packet delay is 10 µs in each switch. The control packet lead time (CPT) is varied from 70 to 250 µs, depending on the diameter of the network (in number of switches) and how long a deflection path we want to allow. Hence, if a burst loops in the network, it will overtake the control packet (and become discarded) in between 7 and 25 hops.

All CPTs are the same, irrespective of the source destination pairs. It would be possible to calculate the length of the path (number of switches encountered on the way) to the egress node. However, in the experiments reported in this article we do not calculate such path lengths, and let all bursts in one experiment start out with the same CPT. Also all experiments are set up with equal capacity links. Each link has 10 channels (lambdas) and each channel has a capacity of 1 Gbit/sec.

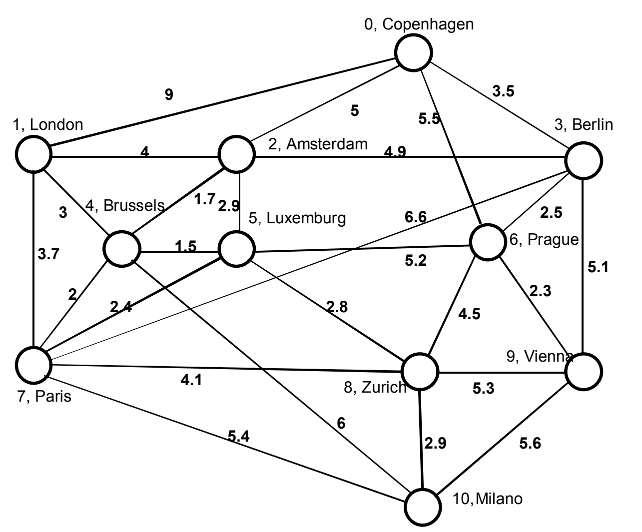

Figure 1. The COST 239 network between 11 European cities. Propagation delay in milliseconds on each link.

4 Performance evaluation

In this paper three networks with different characteristics are used to evaluate hot potato deflection routing with increased CPT; the Pan-European COST 239 network [10] and two networks from the Rocketfuel project from Washington University [11]; the Exodus network and the Sprint US network.

All nodes are ingress nodes (generating traffic), egress nodes and internal switching nodes in the network. The traffic matrix is symmetric all-to-all. This may create bottlenecks that would normally not be seen in a well balanced network. However, the purpose of the experiments is to evaluate burst loss, so we need to observe a varying degree of contention, and hence a well balanced network with few or no burst losses is of less interest. In the three next subsections, the different network topologies that are used in the experiments are described.

4.1 The COST network

The COST 239 network is a proposed Pan-European core network topology consisting of 11 nodes (European cities) connected by 26 (bidirectional) links, as illustrated by Figure 1 [10]. The propagation delays are estimated based on the distances between the cities.

The control packet lead time used by the ingress switch is 70 µs. Since the control packet forwarding time is 10 µs, a burst can be forwarded through 7 switches before it must be discarded because the burst overtakes the control packet. Without deflection, a minimum CPT of 50 µs is needed in this network in order to ensure that all packets reach their destination in the case of no contention at all.

4.2 The Exodus network

The Exodus network is described by the Rocketfuel project and is AS number 3896. By collapsing switches in the same cities, and also collapsing parallel links, we have reduced the network to 17 nodes connected by 29 links. The link latencies vary from 2 to 15 milliseconds. The diameter of this network (in number of switches) is larger than the COST network, and the minimum CPT is set to 100 µs, while 80 µs are needed in order to let all bursts arrive safely in case of no contention.

4.3 The Sprint network

The second network we use from the Rocketfuel project is the Sprint US network (AS 1239). Also this network we have reduced, this time to 45 switches and 95 links. Link latencies vary from 2 to 64 milliseconds. The diameter is even larger, and the minimum CPT used in our experiments is 150 µs, while the minimum needed (with no contention) is 120 µs.

5 Simulation results

In this section we report the simulation results from running hot potato deflection routing with different CPTs on these three networks.

5.1 The COST network

Figure 2, 3 and 4 show the results from the COST network. We have increased the load up to an unrealistically high value in order to also see how hot potato deflection behaves under extreme loads. However, the results for low and medium loads are the ones that are of most practical interest.

Figure 2 shows the number of bursts lost because they overtake their control packets. When the load is very light, there are no or very few bursts that are lost at all. When the load increases, the number of bursts lost by overtaking is increased. Notice that the increase is dramatically higher for the short CPT than for the long CPT. The reason for this is that with a short CPT, the bursts overtakes the control packet much faster. For high loads figure 2 shows few losses caused by bursts overtaking their control packets. The reason for this is that the network is so congested that most bursts are lost because they are totally blocked and can not be scheduled on any of the output links.

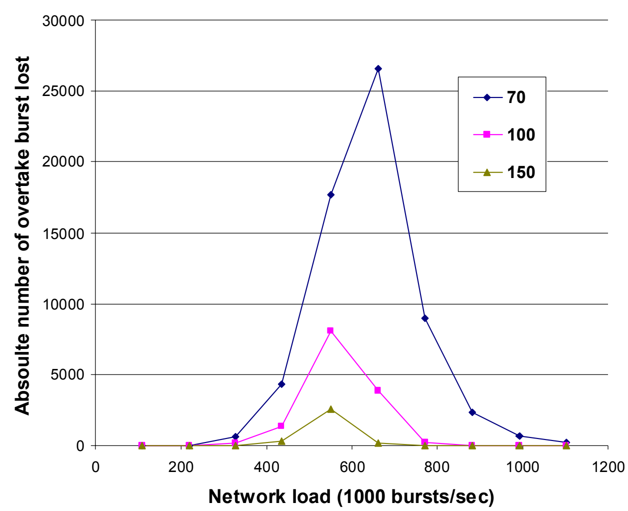

Figure 2. Number of bursts dropped per second because bursts overtake their control packets in the COST network. Three different CPTs values (in µs) and increasing load.

Figure 3 and 4 show the total burst loss probability, called “total xx”. “xx” is the CPT used in µs, and the three CPTs are 70, 100 and 150 µs. Notice that the burst loss probability decreases with increasing CPT. “Regular” burst switching, i.e. no deflection, is shown for reference.

Figure 3 and 4 also show the fraction of bursts that are lost because bursts overtaking their control packet ("overtake xx") for the same CPT values. The "overtake xx"-plots shows the burst loss probability caused by overtaking as a fraction of all burst losses (for the same CPT value). Notice that at low loads the fraction is 1, because all burst losses are caused by bursts overtaking their control packets. Hence at low loads, no bursts are lost because of blocking. At high loads the fraction is 0, meaning that no bursts are lost because of overtaking, i.e. all bursts are lost because of blocking.

At a load of about 435 thousand bursts/sec., a CPT of 150 µs gives a burst loss probability of just above 0.001 (figure 4), of witch almost exactly half (figure 3) is from bursts overtaking their control packets. With the same load, and a CPT of 100 us, the loss probability is 0.003, and almost 90% of the losses are from bursts overtaking their control packets, and with a CPT of 70 µs, the burst loss probability is 0.01 and almost all (96%) are from bursts overtaking their control packets. For reference, no deflection ("Regular") at this load gives a burst loss probability of 0.08.

With a load of 550 thousand bursts/sec, our simulation of the COST network shows that with CPTs of 150 µs, 100 µs, and 70 µs, the total loss probability is respectively 2.7 %, 3.4 % and 4.6 % (with a "Regular" reference of 12.5 %). For higher loads than this (660 thousand bursts/sec and above in our simulations), the burst loss probability soon gets above 20% and is not affected by the CPT. Above this network load, "Regular", i.e. no deflection, is better, because when the deflected bursts continue on in the network they only make the congestion worse. Also notice that at very high loads, all losses are caused by blocking (none by overtaking). This means that a longer CPT would not have any effect on the burst loss probability, because no burst lives long enough in the network to overtake their control packets.

5.2 The Exodus network

We simulated the performance in the Exodus network, and the results are plotted in figures 5 and 6. In particular in figure 6 we see that a CPT of 150 or 200 µs are much better than a CPT of 100 µs for loads less than 400 000 bursts/sec. E.g. at approximately this network load, increasing the CPT from 100 to 150 µs, will yield a burst loss probability decrease of one percent point (from 5.1 % to 4.1 %). A further increase to 200 µs, however, only decreases the burst loss probability to 3.8 %. Notice from the plots that below this load, half of the burst losses or more are caused by bursts overtaking control packets. Also notice that for all load values, a CPT of 100 µs gives much more overtake losses than the two higher CPT values. For loads above 400 000 bursts/sec and a loss probability of approximately 15 %, deflection yields worse results than no deflection ("Regular").

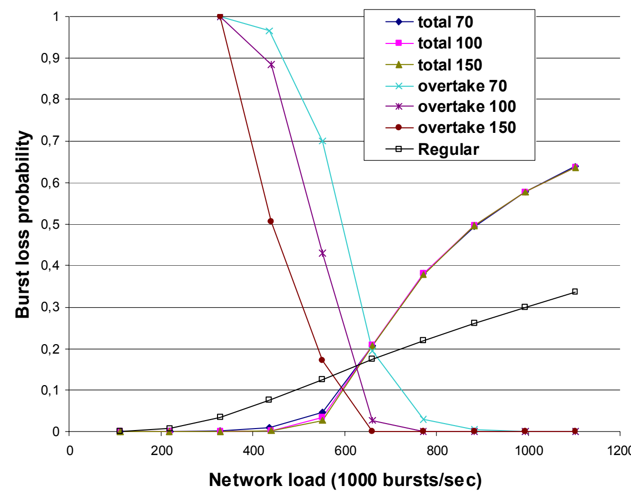

Figure 3. Burst loss probability in the COST network with increasing load and three CPTs. All numbers in the legend are in µs. The "overtake" burst losses are the fraction of bursts lost because they overtake their control packets compared to the total burst loss. Regular is shown for reference.

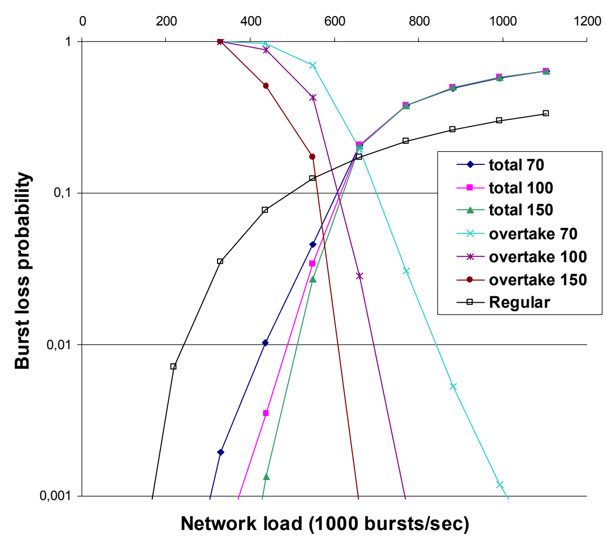

Figure 4. This figure illustrates the same as figure 3, except here the y-axis is logarithmic.

5.3 The Sprint network

The results from running the Sprint network is depicted in figure 7 and 8. With a load around 400 000 bursts/sec., the burst loss probabilities for 150 µs, 200 µs and 250 µs are respectively 5.8 %, 4.5 % and 3.9 %. For both higher and lower load values, the differences in burst loss probabilities between different CPT vales are less than this. A load of approximately 400 000 bursts/sec., is also the value where the fraction of bursts lost by overtaking is highest. In the Sprint network, defection is profitable up to the high burst loss probability of 45 %.

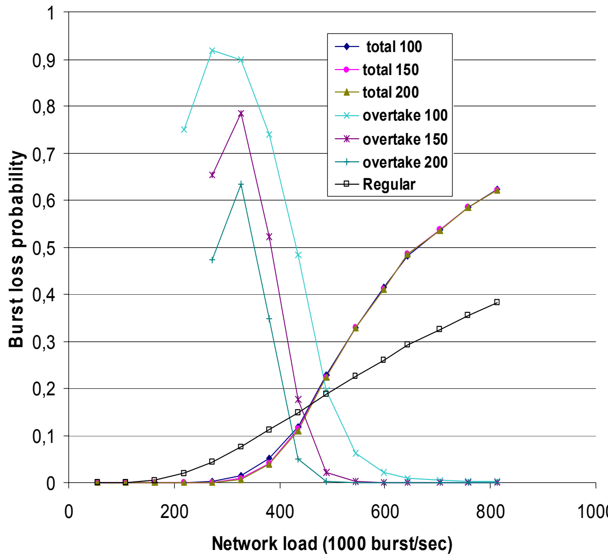

Figure 5. Burst loss probability in the Exodus network with increasing load and three different CPTs. All numbers in the legend are in µs. The "overtake" burst losses are the fraction of bursts lost because they overtake their control packets compared to the total burst loss. Regular shown for reference.

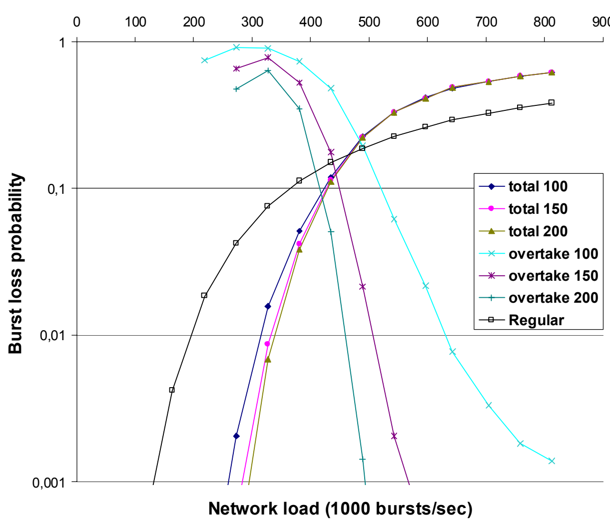

Figure 6. This figure illustrates the same as figure 5, except here the y-axis is logarithmic

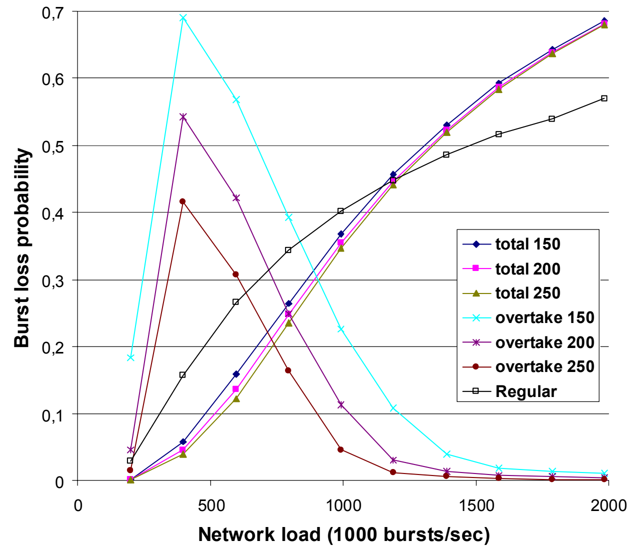

Figure 7. Burst loss probability in the Sprint network with increasing load and three different CPTs. All numbers in the legend are in µs. The "overtake" burst losses are the fraction of bursts lost because they overtake their control packets compared to the total burst loss. Regular shown for reference.

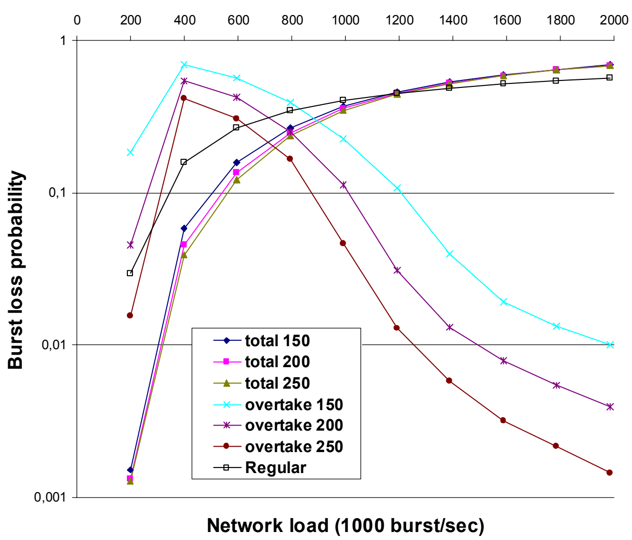

Figure 8. This figure illustrates the same as figure 7, except here the y-axis is logarithmic.

6 Conclusions and Future work

In this paper we have discussed how to reduce the probability of burst loss in Optical Burst Switched (OBS) networks by using hot potato deflection and increasing the control packet lead time (CPT).

Three topologies with different characteristics have been used, with number of switches varying from 11 to 45 and ratio between links and switches varying between 1.7 and 2.3 (i.e. node degree between 3.4 and 4.6). The traffic matrix used has been symmetric, all-to-all, and all links have had the same capacity. The arrival process simulated has been Poisson distributed burst. In order to experience burst loss, the load has been varied, in some cases so much that the burst loss has reached unrealistic and extreme values (almost 70%).

We have observed that with an increase in lead time, the number of bursts that overtake their control packets are dramatically decreased. For low or medium network load values all our simulations show that about half or more of the burst losses are caused by bursts that overtake their control packets. This fraction of burst losses increases with decreasing CPT. However, at high loads, no bursts overtake their control packets because all bursts are blocked.

We have shown that by increasing the CPT and using hot potato deflection, the burst loss probability decreases drastically for low loads. For high loads the CPT value seems to play a lesser role. One of our main results is that there seems to be a point of diminishing return. In all the three networks simulated, there have been three values tested for the CPT. The smallest CPT value used has made the bursts able to travel at least 2 to 5 hops longer than the minimum path from ingress to egress. The next CPT value used has made the bursts able to travel another 3 to 5 hops, and the third and longest CPT value has made the burst able to go even 5 hops longer before they overtake their control packets.

For all the three networks tested it seems that the middle CPT value might give the best trade off between latency and burst loss. Of course, this also depends on how important latency is, because a longer CPT will give a longer latency for all bursts, not only the ones that are deflected. In our example, the middle CPT value is between 80 and 100 µs longer than the absolute smallest CPT needed. Hence, if an added latency of 100 µs is acceptable, compared to no deflection, a burst loss probability of typically 10 times less is possible by using hot potato deflection in a lightly loaded network. It seems that to increase latency another 50 µs (or 5 hops) does not give that good a return. However, also hot potato deflection with a small CPT value works quite well.

In future work we will look at total burst latency, taking both CPT and increased travel time because of deflection into account. We will compare hot potato deflection with other deflection strategies, and also use self similar IP traffic (instead of Poisson distributed bursts) as input, and investigate if this gives other results.

Acknowledgements

Thanks to Audun Fosselie Hansen and Arne Maus who took part in the initial design of the OBS simulator. Thanks also to Amund Kvalbein who helped program another network simulator upon which our OBS simulator is based.

References

[1] Myungsik Yoo, Chunming Qiao, "A novel switching paradigm for buffer-less WDM networks", Proceedings Optical Fiber Communication Conference, San Diego, Ca, USA, pp 177 – 179, Feb. 1999.

[2] Myungsik Yoo, Chunming Qiao, "Just-Enough-Time (JET): A high speed protocol for bursty traffic in optical networks", In Vertical Cavity Lasers, 1997 Digest of the IEEE/LEOS Meetings, pp 26–27, Aug. 1997.

[3] X. Wang, H. Morikawa, T. Aoyama, "Burst optical deflection routing protocol for wavelength routing WDM networks", Proc. SPIE, 2000.

[4] C.-F. Hsu, T.-L. Liu, N.-F. Huang, "Performance Analysis of Deflection Routing in Optical Burst-Switched Networks", Proceedings IEEE INFOCOM, pp 66-73, 2003.

[5] Zalesky, A.; Hai Le Vu; Rosberg, Z.; Wong, E.W.M.; Zukerman, M.; "Modelling and performance evaluation of optical burst switched networks with deflection routing and wavelength reservation", INFOCOM 2004, Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, Vol. 3, 2004, pp 1864 – 1871.

[6] SuKyoung Lee; Sriram, K.; HyunSook Kim; JooSeok Song, "Contention-Based Limited Deflection Routing Protocol in Optical Burst-Switched Networks", Selected Areas in Communications, IEEE Journal on, Volume 23, Issue 8, Aug. 2005 pp 1596 – 1611.

[7] Kim, S., Kim, N., Kang, M., "Contention Resolution for Optical Burst Switching Networks Using Alternative Routing", Proceedings ICC 2002, pp 2678 – 2681.

[8] Schlosser, M., Patzak, E., Gelpke, P., "Impact of Deflection Routing on TCP Performance in Optical burst Switching Networks", Proceedings ICTON 2005, IEEE 2005.

[9] John A. Miller, Andrew F. Seila and Xuewei Xiang, "The JSIM Web-Based Simulation Environment," Future Generation Computer Systems, Vol. 17, No. 2, pp 119-133, Oct. 2000.

[10] O'Mahony, M.J., "Results from the COST 239 project. Ultra-High Capacity Optical Transmission Networks", 22^{nd} European Conference on Optical Communication, pp 15-19, Sept. 1996

[11] http://www.cs.washington.edu/research/networking/rocketfuel/