Performance evaluation of Burst deflection in OBS networks using Multi-Topology routing

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

Abstract – **This paper evaluates the combination of Optical Burst Switching (OBS) and Multi-Topology (MT) routing. Using MT routing, a source router has a choice of sending IP packets on several different paths to the destination. In OBS networks, deflection may reduce burst loss rate. We evaluate a deflection method that is based on MT routing and that ensures that deflected bursts will not loop indefinitely in the network. The performance of the method is evaluated by simulation and compared to two other deflection methods as well as just discarding burst that can not be scheduled. Performance is evaluated in three irregular networks with different topology characteristics. Our main load is IP-packets that arrive according to a self similar process. These packets are assembled into bursts that are transmitted either when the burst buffer is full or a timer expires.**

Keywords: Optical burst switching, Multi-Topology routing, Performance modeling, Burst loss rate, Burst deflection, Self-similar and Poisson arrival processes.

I. INTRODUCTION AND MOTIVATION

In Optical Burst Switched (OBS) networks [1], packets (e.g. IP packets) are assembled into bursts in the optical network ingress nodes, and the complete burst is transmitted either when the burst buffer is full or when a timer expires (hybrid burst assembly). A control packet precedes the burst in the network and reserves resources for the succeeding burst. In this paper "Just Enough Time" scheduling is used [2]. The burst is kept in the optical domain, while the control packet is converted from optical to electrical (and back) in each switch.

When the time slot that a burst needs on the output fiber is not completely available, there is a *contention* on the output line. The simplest approach is then to discard the burst. However, by deflecting the burst and send it out on another line, the burst may later arrive safely at its destination [4].

In general, deflection methods have three main drawbacks [5,6,7]: Some methods deflect bursts in a random direction, which might be counterproductive when considering the destination of the burst. The burst may also return to the point from which it was originally deflected, which may cause indefinite looping and even more contention. Some deflection methods try to deflect the packet on an alternative (loop free) path towards the egress. With connection-oriented routing, this will always be possible (given that the topology is bi-connected). In this paper we assume connection-less routing, and then an alternative loop free path may not be readily available [8].

Multi-Topology (MT) routing is developed within the Internet Engineering Task Force (IETF) [9]. MT routing is used in IP networks so that different streams, different traffic classes or different services (eg. multicast and unicast) can be forwarded in different topology images, and hence take different paths to the network egress node. These different topology images are subsets of the original topology. In each subset (topology) all routers are still present, but some links are removed. However, links should only be removed such that the network is still fully connected. The IETF "Request for Comments" for MT routing ([9]) specifies that a packet is forwarded in one and the same topology from ingress to egress. When used as a method for burst deflection, as will be described in the sequel, this will not be the case, but topology changes must be restricted in order to avoid indefinite looping.

In this paper we evaluate burst deflection in OBS networks based on MT routing by simulating traffic in three networks. The arrival process of bursts into the OBS-networks [10] is made up from self-similar IP-packets, and simulate a hybrid burs assembly method. At the end of the paper we compare these results to results achieved by an arrival process using Poisson distributed bursts.

The deflection method used in this paper was proposed in [20]. The contribution of this paper is a much more thorough discussion and evaluation of its performance.

This paper is organized as follows. In the next section we present MT routing and our deflection method based on MT routing that guaranties freedom from (indefinite) looping in any (bi-connected) topology. In section 3 we describe our performance evaluation method. In sections 4 we compare the performance of the different methods using three different irregular network topologies and self similar IP-traffic. In section 5 we compare our results with Poisson distributed burst arrivals. Finally in section 6 we conclude.

II. MULTI-TOPOLOGY BURST DEFLECTION

In an MT-capable IP-router there is (conceptually) one forwarding table for each topology image. One topology is the original topology, usually called the *default topology*, while in this paper we call the other topologies *backup topologies*. These backup topologies are subsets of the original topology, where some links are removed in each topology, while the network is still fully connected. In order to identify the topologies and the forwarding tables, the original (default) topology/table is numbered 0, and the backup topologies/tables are numbered from 1 and up.

MT routing is developed for shortest path, connection-less forwarding, and we assume that the switches in our OBS networks forward the bursts the same way. However, more sophisticated routing algorithms e.g. based on

 ISBN:978-1-61208-113-7

knowledge about the traffic matrix, may be used instead of shortest path routing (e.g. [11]).

All bursts are initially routed in the default topology. When the control packet that precedes the burst, arrives at a switch, the switch first tries to forward the burst (and the control packet) on the primary output link as decided by the default forwarding table. By installing wavelength converters, the probability of finding an available time slot increases [3]. In this paper we assume full wavelength conversion.

If there is a contention on the primary output link, the burst (and the control packet) is deflected according to one of the backup forwarding tables. As in MT routing, all switches contain one pre-calculated forwarding table for each topology. In order to be able to handle contention on any link, we need each link to be removed from at least one topology. We define a *complete* set of backup topologies as a set of topologies in which all links in the original topology are removed at least once. Figure 1 shows a full (the original network) topology on top left, and a complete set of 3 backup topologies.

We have devised algorithms to find complete sets of backup topologies for a given network [12,13]. The sizes of these sets have been shown to be surprisingly small; we have never come across a (normal) network that needs more than 5 backup topologies. When a complete set of backup topologies are found, each switch calculates one (loop free) forwarding table for each topology. In the network in figure 1 each switch will have four forwarding tables: One default forwarding table (according to the full topology) and three backup forwarding tables (These tables might be optimized to fill less space than four times what is needed for one table). MT routing does not specify what type of traffic maps to the different topologies, although the Type of Service (TOS) field in the IP header might be used if available.

In our burst deflection mechanism based on MT routing, a number in the control packet header tells which topology the burst (and the control packet) is currently forwarded in. Whenever a control packet arrives at a switch, this topology number is extracted and the corresponding forwarding table is used to find the bursts primary output link. If a switch can not send a burst out on a primary link because of contention, it deflects the burst by sending it out in any of the backup topologies that does not contain this link. Because of the way the complete set of backup topologies is constructed, at least one such topology does exist. The number in the control packet header is then set to this new topology number, and the burst is forwarded all the way to the egress in this topology (assuming no more deflections). If the burst experience a second (or third etc.) contention, it may conditionally be deflected once more. However, in order to avoid indefinite looping, we restrict all bursts to be routed in each topology at most once. This is achieved by only allowing a burst to be deflected to a topology with a higher number. When there is no higher numbered topology available, the control packet and the burst is discarded.

All backup topologies are fully connected, and loop-less forwarding tables are precomputed for all topologies. When deflected to another topology, the burst may return back to a node in the network it has visited before. However, because of the restriction that the burst is routed in each topology at most once, the burst will never loop indefinitely in the network.

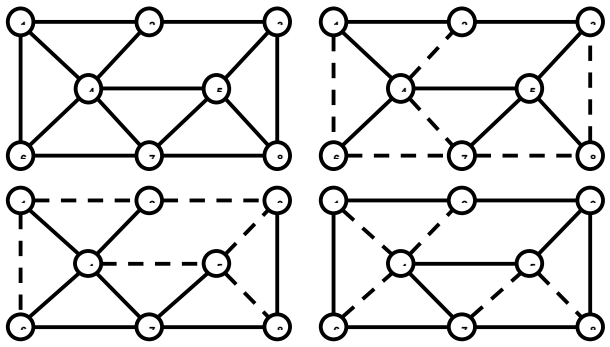

Figure 1. Original network top left, and a complete set of three backup topologies. Removed links are dashed. Notice that all links are dashed at least once.

III. PERFORMANCE EVALUATION

We have implemented a full OBS discrete event simulation model in the J-sim framework [14]. The data sources and burst assembly modules, as well as the OBS-switches and schedulers are built from scratch. Topologies, link propagation times and forwarding tables for the specific scenarios are read from files at system start up time.

The traffic load onto an OBS core network may come from IP-subnets and Ethernets. It is well known that Ethernet and IP traffic exhibit self similar properties [16]. We generate a self similar arrival process using a large number (100) of Pareto sources, with Hurst parameter 0.9, in each ingress node [17]. Whenever a Pareto source starts a new on period, a destination address is chosen according to the probability given by the traffic matrix, and IP packets are generated and sent to the same destination with constant intervals of 10 μs for the duration of the on period.

The size of the IP packets is varied from 80 to 1600 bytes, with a mean of 500 bytes. IP packets are assembled into bursts by a hybrid burst assembly method, meaning that a burst is transmitted when it is full, or a timer expires, whatever comes first (a 2 ms timer value is used in this paper). In this paper fixed burst size (50 000 bytes) is used.

When, at the end of the paper, we generate network load using Poisson distributed bursts, each ingress node runs one Poisson process per optical egress node, generating fixed sized bursts (50 000 bytes). The mean arrival rate is determined by load in the traffic matrix.

While the bursts are kept in the optical domain, and use very short time through a node, the control packet delay used in this paper is 10 μs in each node. The control packet lead time (CPT, i.e. how long ahead of the burst the ingress sends the control packet) is varied from 90 to 200 μs, depending on the diameter of the network (in number of switches). Hence, if a burst loops in the network, it will overtake the control packet (and they both become discarded) in between 9 to 20 hops. All experiments reported in this article are set up with equal capacity links.

 ISBN:978-1-61208-113-7

Each link has 10 channels (lambdas) and each channel has a capacity of 1 Gbit/sec.

We compare deflection based on MT routing (denoted ***Multi-Topology*** in the plots) with two other well known deflection methods: ***Hot Potato*** that chooses an alternative output link at random and ***Second Shortest*** path that tries to output the bursts to the output link where the next switch has the shortest distance to the destination (excluding the primary output link). Notice that for both methods a packet may be deflected back to where it came from, and hence in general these methods can not guarantee freedom from looping. Indefinite looping in the network is only prevented by the fact that when the burst is overtaking the control packet they are both discarded. In the case that the control packet is not able to reserve the needed resources for the data burst at all (deflection is not possible), the burst (and the control packet) is discarded by the switch.

In addition to comparing MT deflection with Hot Potato and Second Shortest, we also compare it with ***Regular*** burst dropping, i.e. when a burst may not be scheduled on the primary output link, it is immediately discarded (no deflection). The performance evaluation is carried out using three realistic and irregular networks with different characteristics; the Pan-European COST 239 network [18] and two networks from the Rocketfuel project from Washington University [15]; the Exodus network and the Sprint US network.

The COST 239 network is a proposed Pan-European core network topology consisting of 11 nodes (European cities) connected by 26 (bidirectional) links. The propagation delays are estimated based on the distances between the cities. The control packet lead time used by the ingress nodes is 90μs. 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 ms. Initial control packet lead time is set to 120 μs. The second network 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 ms. Initial CPT is set to 200 μs. All nodes are ingress nodes (generating traffic), egress nodes and internal switching nodes in the network. The traffic matrix is symmetric all-to-all. For each experiment we have made a one-second run for each load value.

IV. SIMULATION OF IP-TRAFFIC

In this section we report simulation results from running the three deflection methods, Multi-Topology, Hot Potato and Next Shortest as well as no deflection (Regular). The burst arrival process is a hybrid burst assembly of simulated self similar IP traffic.

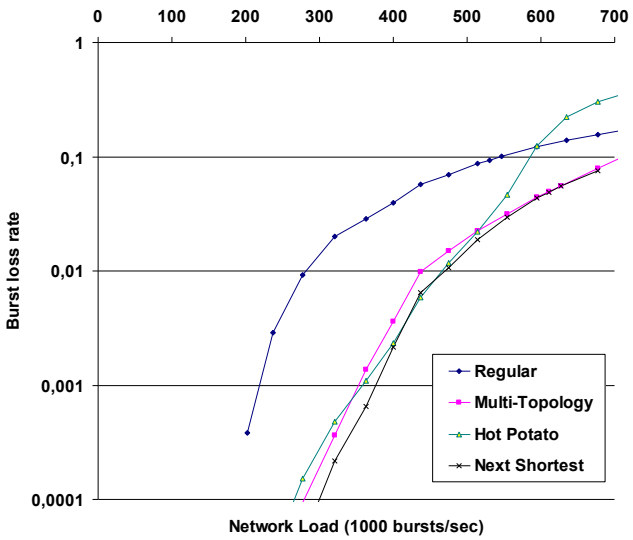

Figure 2. Burst loss rate in the Cost network with increasing network load.

A. *The Cost network*

The results are depicted in figure 2. With 100 Pareto sources per ingress node, the maximum load it is possible to send into the network from the 11 ingress nodes is approximately one M bursts/sec. As long as the total traffic generated is below 150 000 bursts/sec (i.e. each of the 11 ingress nodes generates about 5.5 Gbit/sec), there is no packet loss anywhere in the network.

When the load has increased to 300 000 bursts/sec, the burst loss rate for the Regular method is above 1%, while all the other methods still yield very good results. There is a very distinct change in the increase for the Hot Potato deflection method compared to the other methods at about 3% loss rate. Here the loss rate of this deflection method starts to increase steeply, while the loss rate of Multi-Topology and Next Shortest continue with a much smaller increase.

Also observe that Multi-Topology and Next Shortest perform almost identical for all load values, although Next Shortest seems to always perform slightly better. These methods are the most stable ones, meaning that they perform quite well for all loads.

B. *The Exodus network*

The simulated performance of the Exodus network is depicted in figure 3. Again, notice how Hot Potato deflection performs badly at high loads, and good at low and medium loads.

Also in this network, Next Shortest and Multi-Topology perform about the same, but this time Multi-Topology is mostly the better of the two. Above about 4% loss rate, Multi-Topology also performs better than Hot Potato deflection.

 ISBN:978-1-61208-113-7

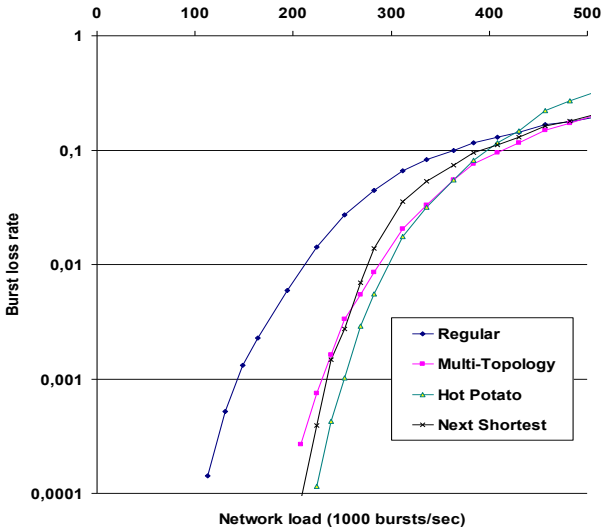

Figure 3. Burst loss rate in the Exodus network with increasing network load.

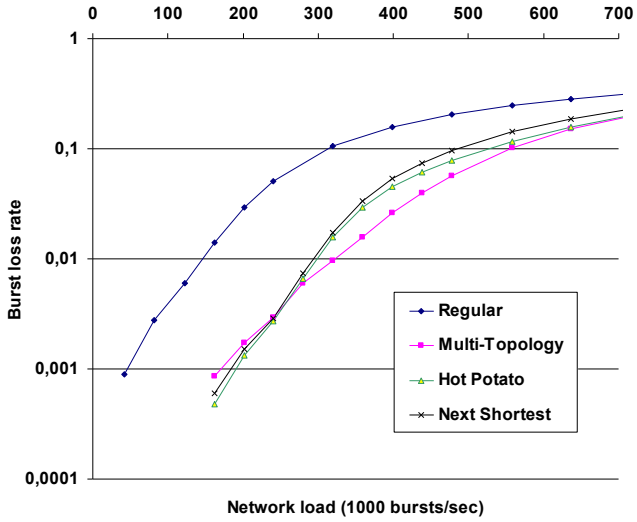

Figure 5. Burst loss rate in the Sprint network with increasing Poisson distributed load.

C. *The Sprint network*

The performance in the Sprint network is seen in figure 4. Here the Hot Potato method is not performing so badly for high loads; in fact it seems that the difference in performance decreases for high loads. In the Sprint network, Multi-Topology deflection is clearly the best method for all load values.

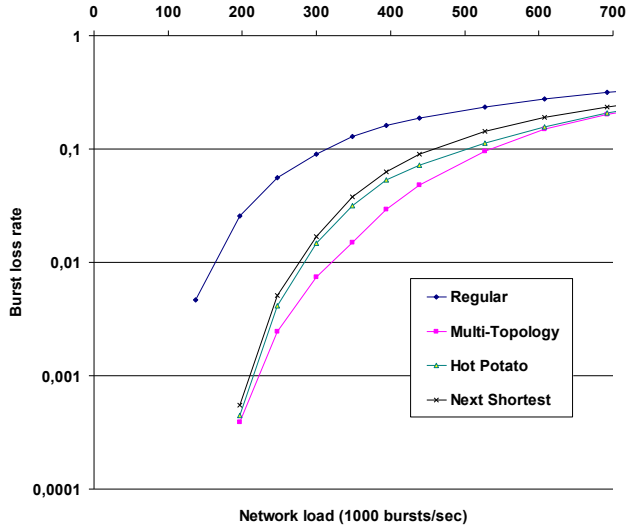

Figure 4. Burst loss rate in the Sprint network with increasing network load.

V. POISSON DISTRIBUTED BURST ARRIVALS

We have also run the experiments described above using Poisson distributed burst arrival processes. Figure 5 shows the burst loss rate in the Sprint network, and figure 6 compares simulated performance of the Multi-Topology method in all the three networks. For low loads, there are only few losses in a second, and then the results are of course less statistical significant.

In our experiments there is not much difference between the performances caused by aggregated self similar traffic and Poisson distributed traffic. First we see this by comparing the plots in the figures 4 and 5. In fact these two plots seem almost identical for burst loss values above 1%. In figure 6 we see that in the Exodus and Sprint networks, Poisson distributed burst loads performs a little better than aggregated self similar traffic, while in the Cost network the situation is reversed. For low loads, however, simulated self similar traffic seems to perform better than Poisson distributed burst load in all three networks, although here we have very little data.

If we assume that a bursty load performs worse (have more burst losses in the network) than a smooth load, and we compare the performance of aggregated self similar traffic (assembled into bursts) with Poisson distributed burst traffic, there is nothing in our experiments that indicates that one of these arrival processes produces smoother burst loads than the other. In future work we will look closer into this problem scenario [10].

 ISBN:978-1-61208-113-7

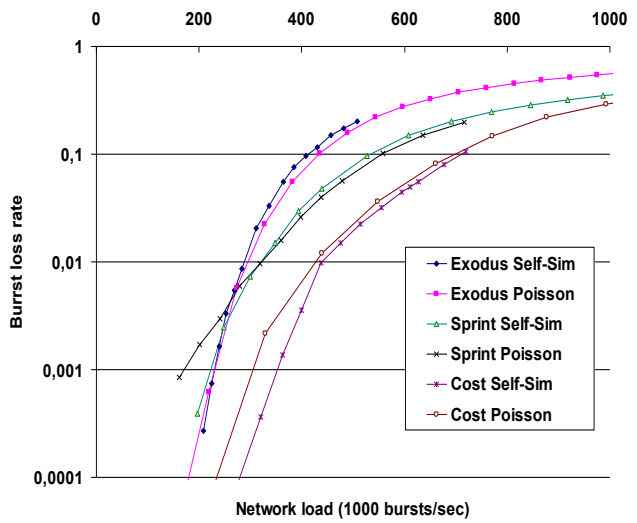

Figure 6. . Simulated performance of Multi-Topology deflection comparing aggregated self similar IP traffic and Poisson distributed bursts in the three networks with increasing network load.

VI. CONCLUSIONS

In this paper we have made a thorough evaluation with realistic traffic of a burst deflection method based upon Multi-Topology (MT) routing. As outlined in the introduction, deflection in OBS networks has been extensively studied before. Also forwarding in sub-graphs of the original optical network has been proposed in order to handle link failures, but, among other things, then the source must know in which sub-graph to forward [19].

MT routing is a novel way to do burst deflection, and hence a way to decrease burst loss probabilities in OBS networks. We have developed a special way to find backup topologies where all links are removed from at least one topology. By pre-calculating shortest paths in these topologies, a burst that can not be scheduled on the primary link, is deflected to an alternative path to the optical egress.

The performance of the MT deflection method is evaluated by comparing it with "Next Shortest" path and "Hot Potato" deflection, as well as with no deflection (just discarding bursts that can not be scheduled on the output link). Three irregular 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.

For high network loads our experiments confirms previous results [5], i.e. that deflection, and in particular Hot Potato deflection, creates more network traffic, and hence makes the burst drop probability higher than Regular routing with immediate dropping of packets when the primary output link is congested.

The arrival processes of bursts have been generated by a hybrid burst assembly method, fed by simulated self similar traffic with variable sized IP packets. At the end of the paper we re-evaluated the performance using a Poisson distributed burst arrival process. The results from these new tests are, for medium and high loads, very similar to the results obtained using self similar IP traffic, and hence strengthen the results from section 4.

Except for very high loads, when Hot Potato routing performs very badly, Multi-Topology seems to be comparable in performance with the two other deflection methods. In the Sprint network, Multi-Topology performs best for all network load values. Next Shortest deflection may (and will in some cases) loop the burst immediately back to the point of congestion, and as long as the original output link is congested, the burst may continue to loop in the network (until discarded when overtaking the control packet). Deflection based on Multi-Topology routing guarantees that such indefinite looping never occurs, and may hence be a viable alternative to other deflection methods in OBS networks.

ACKNOWLEDGEMENTS

Thanks to Audun Fosselie Hansen who took part in the initial design of the OBS simulation model, has provided that backup topologies and programmed most of the algorithm that finds complete sets of backup topologies used in the reported experiments. The author would also like to thank Amund Kvalbein for making the Pareto source module. Our OBS simulation model is based upon another network simulator developed by him and the author. Thanks also to the rest of our colleagues at Simula Research Laboratory.

REFERENCES

[1] Y. Chen, C. Qiao and X.Yu, "Optical Burst Switching (OBS): A New Area in Optical Networking Research," IEEE Network Magazine 18, 16-23 , May/June 2004.

[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] J. Ramamirtham, J. Turner, J. Friedman, "Design of Wavelength Converting Switches for Optical Burst Switching", IEEE Journal on Selected Areas in Communications, Vol. 21, No. 7, Sept. 2003.

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

[5] 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

[6] 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.

[7] 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

[8] Xiaowei Yang and David Wetherall, "Source Selectable Path Diversity via Routing Deflections", In Proceedings SIGCOMM'06, ACM, September 2006.

[9] P. Psenak et al., “Multi-Topology (MT) Routing in OSPF. RFC 4915, IETF, June 2007

[10] G. Hu, K. Dolzer, C. Gauger, “Does burst assembly really reduce the self-similarity?”, in Optical Fiber Communications Conference, OFC2003, vol.86 of OSA Trends in Optics and Photonics Series, Washington, D. C, 2003, pp.124-126.

[11] Jing Teng, Rouskas, G.N., “Routing path optimiza-tion in optical burst switched networks”, Optical Network Design and Modeling, pp: 1-10, Feb. 7-9, 2005

[12] A. Kvalbein, A. F. Hansen, T. Cicic, S. Gjessing and O. Lysne, “Fast Recovery from Link Failures using Resilient Routing Layers”, In 10th IEEE Symposium on Computers and Communications (ISCC 2005), pp 554-560, IEEE Computer Society, 2005.

[13] A. Kvalbein, A.F. Hansen, T. Cicic, S. Gjessing, O. Lysne, “Fast IP Network Recovery using Multiple Routing Configurations”. In proceedings IEEE 25th Annual Conf. on Computer Communications (INFOCOM) May 2006.

[14] 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.

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

[16] Willinger, W., Taqqu, M.S., Sherman, R., Wilson, D.V., “Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level”, IEEE/ACM Transactions on Networking, Vol. 5, No. 1, pp: 71 – 86 , Feb. 1997.

[17] G. Horn, A. Kvalbein, J. Blomskøld, E. Nilsen, “An Empirical Comparison of Generators for Self Similar Simulated Traffic”, Elsevier Performance Evaluation 64(2): 162-190, 2007.

[18] O’Mahony, M.J., ”Results from the COST 239 project. Ultra-High Capacity Optical Transmission Networks”, 22nd European Conference on Optical Communication, pp: 15-19, Sept. 1996

[19] M. T. Frederick, P. Datta, A. K. Somani “Sub-graph routing: A generalized fault-tolerant strategy for link failures in WDM optical networks”, Computer Networks, Vol. 50, No. 2, Feb. 2006, pp. 181-199, Elsevier 2006.

[20] S. Gjessing, “A novel method for re-routing in OBS networks”, In International Symposium on Communications and Information Technologies, 2007. ISCIT '07. Sydney, NSW, October 2007.

 ISBN:978-1-61208-113-7