Merit Based Scheduling in Asynchronous Bufferless Optical Packet Switched Networks

Tor K Moseng[*], *Harald Øverby*[**], *Norvald Stol*[***]
Department of Telematics, Norwegian University of Science and Technology (NTNU)
O.S. Bragstadsplass 2E, N-7491 Trondheim, Norway
[*] Tel: 91821061, E-mail: torkjeti@stud.ntnu.no,
[**] Tel: 73594350, Fax: 73596973, E-mail: haraldov@item.ntnu.no,
[***] Tel: 73592133, Fax: 73596973, E-mail: Norvald.Stol@item.ntnu.no

Abstract
This paper describes how Merit Based Scheduling (MBS) can be used to increase the average throughput in an asynchronous bufferless Optical Packet Switched (OPS) network. Basically, MBS means that packets are forwarded based on its merit, which includes priority, number of resources taken, number of resources left etc. We show that using MBS increases the average throughput, but only in a limited degree, dependent on the system load and number of wavelengths per fiber.

Keywords: Service differentiation, optical packet switching, merit based scheduling, simulations.

1 Introduction

Sending several data streams along a fiber increase the fiber's throughput considerably, which is what Wavelength Division Multiplexing (WDM) does. The fiber's potential was not fully realized until the invention of WDM [1]. Electronical networks (i.e. electronical switches with optical fibers) need optical-electronic-optical (OEO) conversions at each node in the network. OEO conversions increase the cost to the system and reduce the transparency. Avoiding OEO conversions would be beneficial. The goal of future core networks is therefore to transport the data on all-optical WDM networks.

The evolution of the WDM networks began with Wavelength Routed networks (WR), which operate by setting up circuit switched connections between end nodes. However, the scarce number of wavelengths in a fiber restricts the number of users. The data traffic today is mostly asynchronous Internet traffic, and is different from the traditional voice traffic. This traffic has a bursty nature and a WR network's quasi-static nature prevents it from efficiently supporting the constantly changing user traffic [1]. In OPS, the traffic is carried in optical packets along with in-band control information. A steady increasing number of applications use IP as their network protocol. Transmitting IP directly over WDM is regarded as a future solution [2, 3] and OPS seems like a suitable solution to realize this.

Optical Burst Switching networks (OBS) have granularity between WR and OPS. In OBS packets with similar characteristics (e.g. same destination and QoS demands) are assembled into bursts at the core network's edge node, and each burst is switched as a single entity through the core network. The burst entities can vary in size according to the network configuration parameters. Both in the OPS and OBS networks the control

information is processed in the electronical domain, while the data stays in the optical domain throughout its stay in the core network.

Since OPS and most OBS schemes send data without waiting for any setup acknowledgement, blocking may occur. Blocking is resolved in the wavelength domain, time domain or space domain. Buffers are more costly to realize in optical networks than in electronical networks. Electronical buffers need OEO conversions, while optical buffers are limited [4]. Optical buffering is realized using fiber delay lines (FDLs), where the signal traverses a long fiber and experiences a fixed propagation delay. The blocking probability (in a bufferless system: loss probability) then becomes increasingly important as a measure of the network's performance. Contention resolution schemes must aim at solving blocking situations to lower the loss probability. These schemes can either drop a random packet or drop according to a packet's priority. A packet's priority depend on parameters such as time in the network, pre-defined priority according to the real-time nature of the application, resources used, resources left to use or a packet's merit. A packet's merit is the focus of this article. The proposed OPS scheme Hop Increasing Priority (HIP) will use merit based scheduling (MBS). A packet's merit may be based on a mixture of the priority parameters mentioned. We focus in this paper on the number of resources a packet uses in the network to lower the overall resource usage and loss probability.

The simulation model prioritizes the packets based on their merit. Other prioritizing schemes have been proposed in the literature. One studied scheme is the Just-Enough-Time (JET) scheme [5]. [6] states to that the traditional JET scheme experiences increasing loss probability for increasing number of hops made – thus making it undesirable in multiple hop networks. [6] proposes a scheme that is based on an extended JET scheme. The unwanted loss property of JET in multiple networks is avoided by using an extra offset time (OT) between the packet and its header to prioritize packets with many hops in a multiple hop network. Our scheme, which also avoids the unwanted loss properties of JET, utilizes the Preemptive Drop Policy (PDP) described in [7] – a higher prioritized packet may preempt a lower prioritized packet. Compared to JET, the PDP is less resource demanding in the switches. In [8] the contending packets may according to their priority be segmented and deflected. Here, the lower prioritized packet is dropped in its entirety because packets cannot be segmented.

The rest of the paper is organized as follows: Section 2 presents the proposed HIP scheme. Section 3 presents a performance analysis of the HIP scheme. Finally, section 4 concludes the paper.

2 The Hop Increasing Priority scheme

2.1 Merit Based Scheduling
Merit based scheduling (MBS) means to prioritize and schedule packets based on their merit. A packet's merit is based on several parameters, such as the packet's time in the network, the application's real-time demand, the number of resources used in the network and the numbers of resources left to use on its path in the network. The goal of the MBS is to differentiate the packets and treat the packets fairly – so the parameters

used in MBS must therefore be set according to some fair rules. Figure 1 illustrates the many possibilities for MBS.
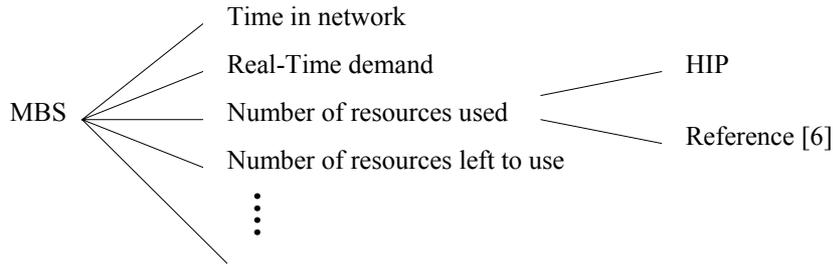

Figure 1: MBS possibilities

Figure 1 shows that MBS is the general method to schedule packets. The packet's merit, which is used during scheduling, can be based on several options as mentioned above and illustrated in the figure. Both [6] and our HIP scheme consider the number of resources used to differentiate the packets. Unlike [6], which uses an OT-based scheme, we use the PDP in this paper.

The JET scheme experiences increasing blocking probability for each hop performed (i.e. resource used). This has the undesirable effect of wasting resources and increasing the network's overall loss probability. When a packet uses a resource, it blocks other packets from using the same resource, which in bufferless systems leads to packet losses. Therefore, the resources must be used in an effective way to avoid resource waste. Basing the packet's merit on the packet's resource consumption in the network will not only utilize the network resources effectively, but also treat the packets fairly. The packets that have traveled farthest (i.e. performed most hops) will be prioritized over new arriving packets. To see how this can be beneficial consider the packets A and B. Assume that packet A has traveled a time δ seconds and crossed ε core nodes (i.e. used ε resource units) on its path in the network. A new packet B arriving from an edge node nearby will be much younger than δ and has crossed only one core node. Contention in a node may demand retransmission of one of the two packets. If packet A is retransmitted, its application would be delayed by an extra time δ and the resource consumption in the network would increase by another ε units. If packet B is retransmitted, its application would not be much extra delayed and the resource consumption would increase by only one unit. By differentiate the packets from this point of view, the delay and resource usage would decrease in the network.

2.2   Service Differentiations with the Preemptive Drop Policy
The HIP scheme will use the Preemptive Drop Policy (PDP) to solve contentions among packets. The merit gives the packet's priority and will thus allow a packet to preempt another. Figure 2 illustrates the four different scenarios between two contending packets. No buffers or deflections are used in the scenarios considered – forcing PDP to drop the blocked/preempted packet.
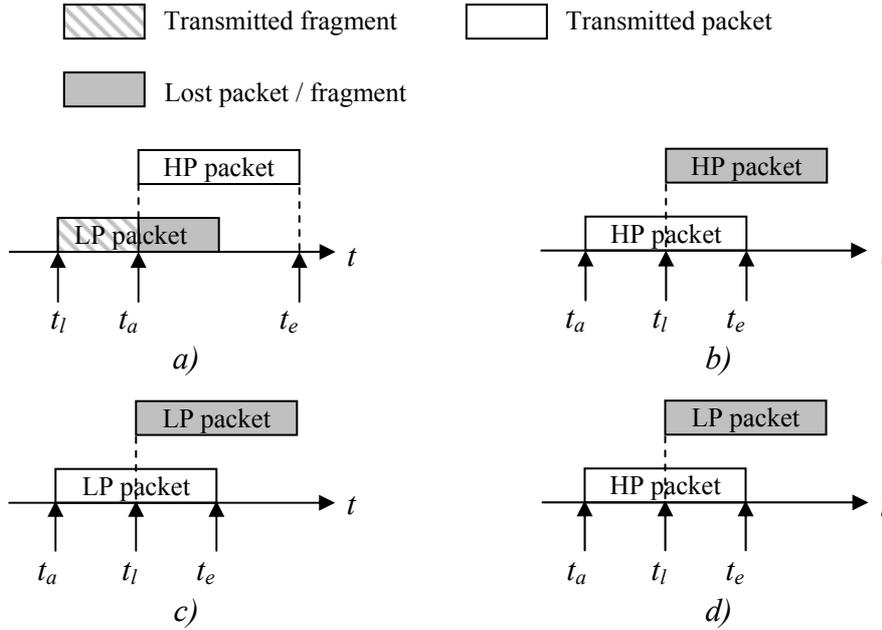
Figure 2: Four different preemption scenarios

In figure 2 a), a low priority (LP) packet is in transmission when a high priority (HP) packet interrupts it. A LP packet will in the HIP scheme be a packet with lower merit than an HP packet. The HP packet forces the LP packet to abort its transmission and takes over the resource. The transmitted LP fragment from time t_l to t_a must be detected at the receiving node to avoid being transmitted further in the network. This can be performed by utilizing a checksum in the header [7]. In figure 2 b) and c), a packet tries to preempt another packet of equal priority. When arriving at time t_l, the packet in transmission has $t_e - t_l$ left of its resource usage, so if no FDLs are available (bufferless system or a full buffer) the arriving packet is dropped. In figure 2 d) an LP packet tries to preempt an HP packet. The arriving LP packet will then be dropped. For an analytical approach to the PDP, it is referred to [7].

3 Performance analysis

3.1 Simulation model
In order to investigate the loss probability and the resource waste in the proposed HIP scheme, a simulation model consisting of 10 core network nodes in a ring topology is used. Each node is connected to a source and a destination, as seen in figure 3. The source and destination represent external networks that generate and receive packets, respectively. The sources generate packets with empirical distributed packet sizes, after an Internet study in [9], from a Poisson process.

A network node (i.e. router/switch) has two inbound and two outbound links – one in each direction (see figure 4). The node takes a packet from an inbound link and switches the packet to either an outbound link or to its external destination network according to a static routing table. Full wavelength conversion is assumed. A packet's routing information is contained in its header. This necessitates an optical-electronic-optical (OEO) conversion of the header. The control logic in figure 4 processes the header and sets up the switching matrix according to the header's information. Meanwhile the

packet is delayed using an FDL. The FDL's length equals the processing delay of the header plus the switch set-up time.
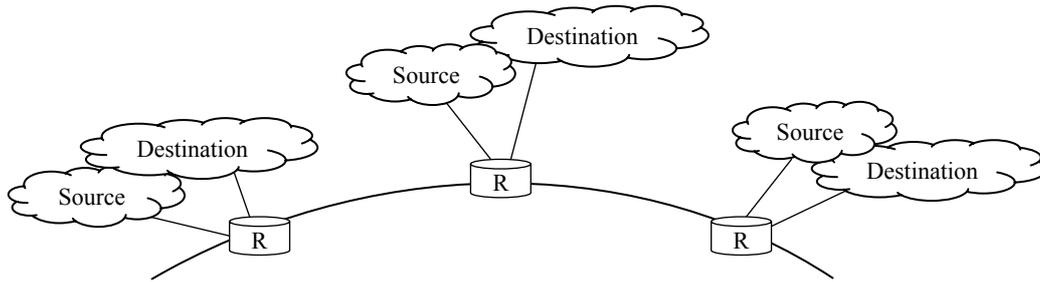

Figure 3: The network configuration for the simulation study

The processing time of the header and the switch set-up time were in the simulations set to zero. Since all packets experience this delay it will not influence the outcome we look for in the simulations. FDLs are therefore logically placed in the node, while omitted from the implementation. After the header has been processed, the packet is transmitted if a wavelength is available or preemption is possible. A merit parameter (i.e. priority parameter) in the header is increased for every resource used, thus giving the packet's priority. The control logic handles the preemption. If the packet has reached its destination router, the packet is put in an electrical buffer for delivery to the destination (see figure 4). The choice of using an electrical buffer is that an electrical buffer is less costly to realize and can store an unlimited amount of data, which is beneficial when sending data from a high speed optical network to a lower speed network – access networks are usually lower speed networks. So when a packet arrives at its destination router, it will not compete with transit packets and do not suffer any loss.
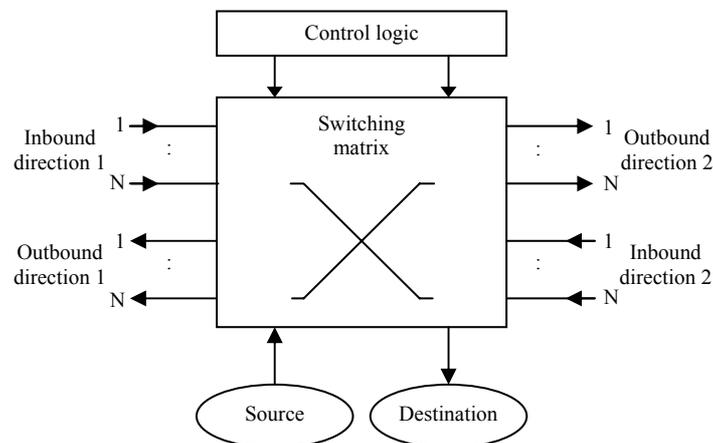

Figure 4: The switch architecture

Each link is assumed consisting of four WDM wavelengths operating at a 1 Gbps data rate. Each source transmits an equal amount of packets to all destinations. The switching nodes experience therefore the same offered load (symmetric traffic). The load is a normalized system load. With 10 nodes, the packets may traverse six nodes – thus use six resources. However, the packets will only fight for up till five resources. As mentioned, when a packet is reaching its destination router, it will be placed in an electrical buffer – making the packets fight for zero to five resources according to their

destination. The packets are therefore at hop count level zero to four (HCL 0 to HCL 4), and named as a zero-hop (i.e. used no resources) to a four-hop packet (i.e. used four resources) according to its HCL. For comparison a best-effort network is also simulated with the same parameter values as the HIP network.

3.2   Simulation Results

All results given in this chapter represents middle values of 10 simulations performed for each load level. On each graph a 95 % confidence interval is illustrated with horizontal error-bars for every point.

Figure 5 shows that the HIP network performs overall increasingly better than the best-effort network for an increasing load. This is because a best-effort network drops a packet that has performed several hops in the network with the same probability as a new packet. Such a long distance packet may on its path have blocked other packets from being served by the switch, and the resources will thus be wasted compared to HIP. By differentiating the packets according to their HCL, the HIP scheme avoids this effect and increases the throughput.
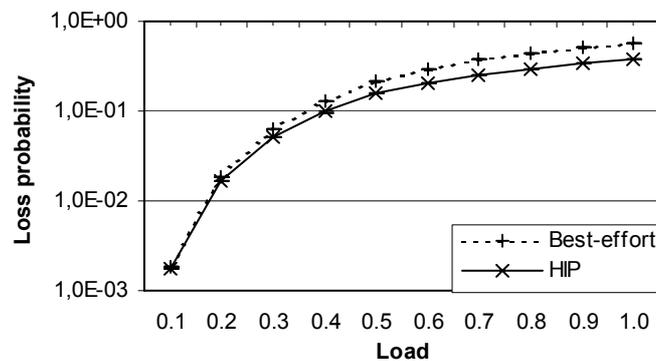

Figure 5: Total loss probability in HIP vs. best-effort

The average resource usage for dropped packets (i.e. the resource waste) will according to the above reasoning be lower in the HIP network, as shown in figure 6. The resource waste will decrease for an increasing load. This can be explained by knowing that the resource waste is the average number of resources a packet uses before it's dropped (i.e. average number of hops performed before being dropped). When the load increases the average number of hops decreases before the packet is dropped – thus lowering the resource waste.
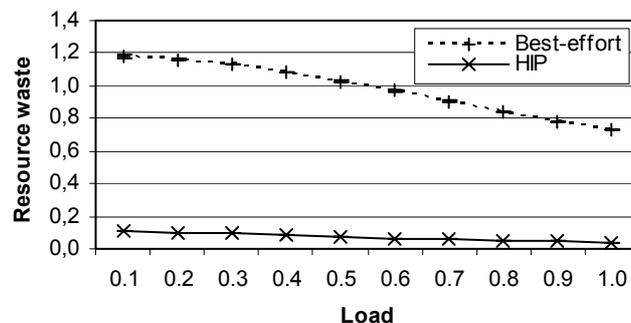

Figure 6: Resource waste in HIP vs. best-effort

By looking at each HCL in the HIP network, the higher levels experience lower loss compared to the lower levels. Figure 7 illustrates that the higher HCLs experience lower loss probabilities (better QoS) than the lower HCLs. HCL 0 will in fact experience even higher loss probability than the per hop loss probability in the best-effort network. Figure 8 shows the cumulative loss probability for the HIP network compared to the best-effort network at load 0.3. After the first hop the packets in the HIP network experience almost no loss. This can be seen as some sort of admission control. Only the zero-hop packets that don't disturb the network's admitted traffic are allowed into the network. When a packet is admitted it receives very good service and experiences a loss probability close to zero.
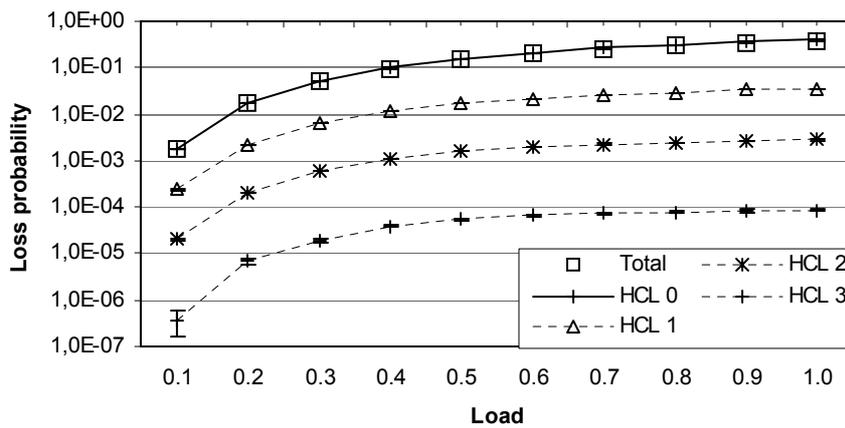

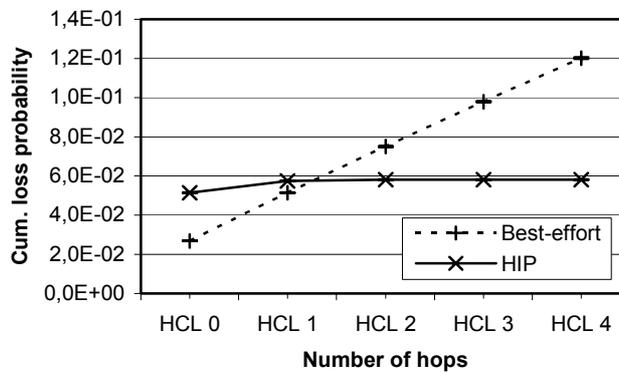*Figure 7: Loss probabilities for each HCL in the HIP network*

Figure 8: Cumulative loss probability for load 0.3

The HIP network was also simulated with 8 and 16 wavelengths for load 0.5 to give a glimpse at the performance for more wavelengths. These simulations showed that the difference between the best-effort and the HIP network was decreasing when the wavelength number was increased. In figure 9 a), the simulation results are given for 4, 8 and 16 wavelengths at load 0.5. When more wavelengths were available, a differentiating scheme like HIP is less needed. The resource waste difference between the two networks is, however, increasing. Figure 9 b) gives the simulation values for the resource waste. While the best-effort network keeps the resource waste quite stable, the HIP network decreases the resource waste. With sixteen wavelengths per fiber, practically all lost packets in the HIP network are zero-hop packets. This fact also

illustrates that a sixteen wavelength HIP network performs even stricter admission control than a four wavelength HIP network.
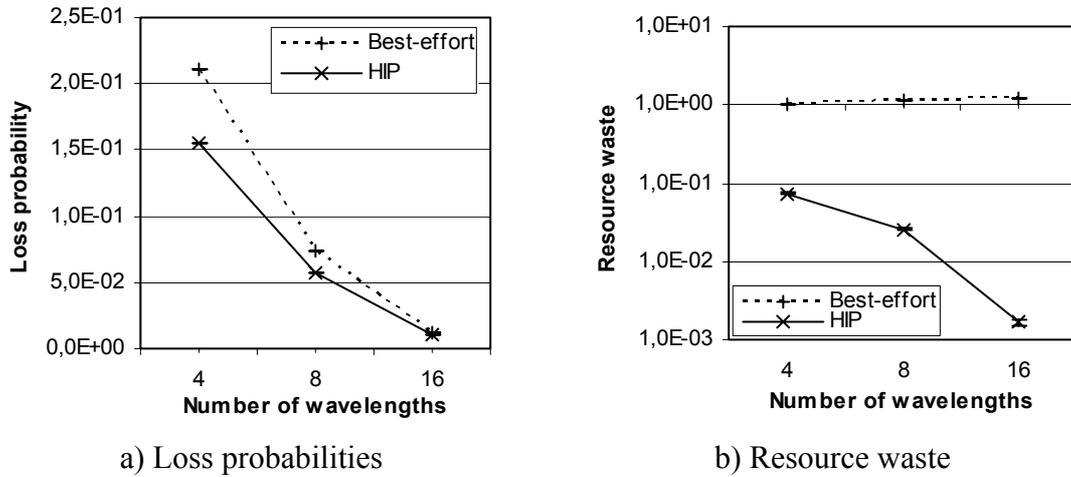

a) Loss probabilities   b) Resource waste

Figure 9: The HIP network with more wavelengths at load 0.5

3.3 Simulation model enhancement

As an expansion to the one class HIP network, two service classes were introduced initially in the HIP network, with 20% being high priority packets. The two classes network was simulated for the loads 0.2, 0.4 and 0.6 at different isolation degrees with the same parameter values as the original one class HIP network.

The question were how to differentiate between the packets; should a low priority packet with many hops be prioritized before a high priority packet with few hops, or should the high priority packets be prioritized no matter what? The solution chosen was to give the high priority packets an initially increase in the header's merit parameter. At isolation degree one, the increase was one initial hop, and so on until isolation degree five that corresponds to full isolation between the low and high priority packets. At isolation degree one, a low priority packet with two hops in the network has a merit parameter equal to two. This packet can then preempt a high priority packet with no hops (merit parameter equal to one). So for increasing isolation degree, the high priority packets receive better isolation, and experience thus lower loss probabilities. The loss probabilities for the high priority packets can be seen in figure 10. Figure 10 shows that an increasing isolation degree lowers the loss probability, and thus increases the throughput, for the high priority packets. The low priority packets' curves are not shown for better illustration, but lay above the curve for isolation degree one in the figure.
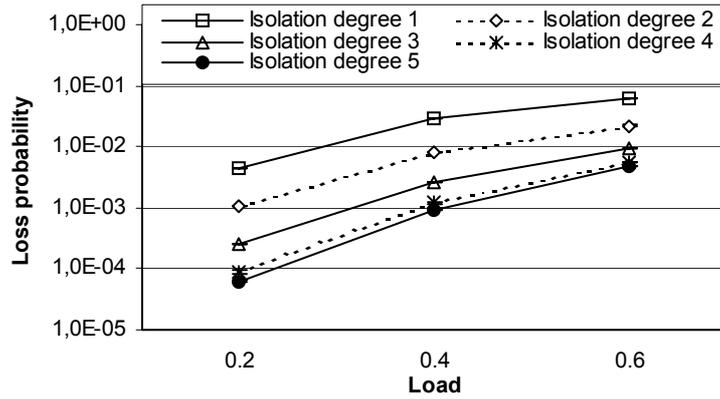
Figure 10: Total loss probability for the high priority packets

3.4   Future research

Whether the increased throughput (though relatively lower for more wavelengths), lower resource waste and the admission control function can justify the increased complexity compared to a best-effort network is yet to be studied and open for future research. A comparison study between our proposed scheme and the scheme proposed in [6] may be performed to compare the complexity and switch performance.

Also, future research could involve alternative ways to handle admission control. One of the results found in this paper is the admission control function resulting from the HIP scheme. Only packets arriving from the external networks that don't interfere with the core network's traffic is admitted into the network. After a packet is admitted the loss probability is close to zero. This is quite beneficial in large OPS networks, where the loss probability of arriving packets could operate as feedback information. An edge node receives packets in an electronical buffer and forwards the packets to the core optical network. The optical part of the edge node (the part with optical transmission equipment) sees the loss probability and could generate feedback information to the edge node's electronical buffer. The feedback information could regulate the electronical buffer's transmission rate to the optical part of the edge node. By reducing the network congestion probability in the edge node, congestion notification traffic from inside of the core network could be reduced considerably. This has to our knowledge not been studied in the literature, but could be included in future research.

Figure 11 illustrates the feedback possibility. An edge node, which is a part of a large optical core network, is enlarged on the left side of the figure. The edge node receives traffic from other core nodes that need processing and rerouting. It also receives traffic from one or several access networks that arrives in the electronical part of the node. In this electronical part, functions like assembling and buffering can be provided. The traffic destined for the core network will be transmitted to the optical part and sent into the core network. If the packets experience high losses, a feedback control mechanism in the optical part could inform the electronical part to decrease its transmission rate to the optical part of the node. Also, since buffering is easier realized in the electronical domain, such an approach seems quite beneficial.
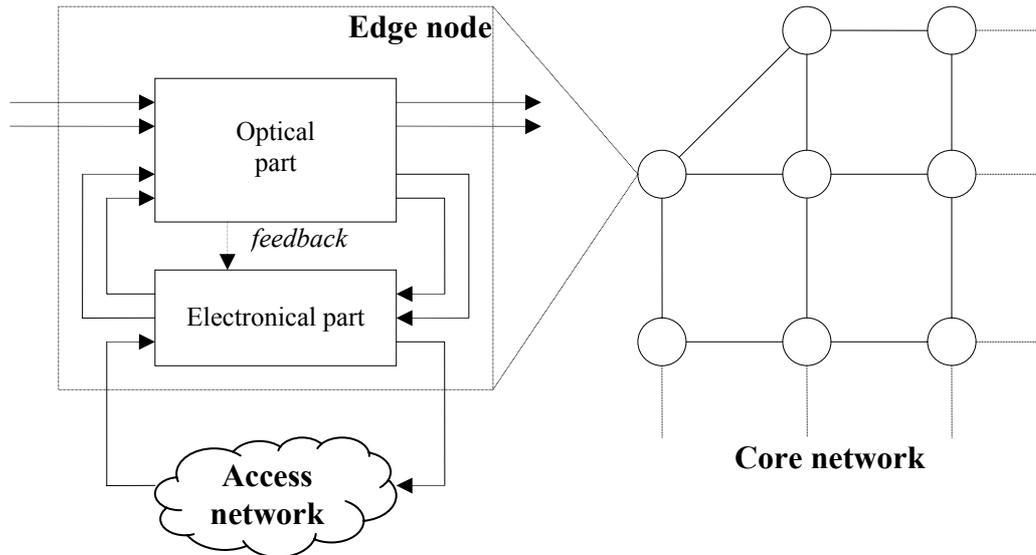
Figure 11: Using loss probabilities as feedback information

4 Conclusion

In this article, MBS was presented, and it was explained how MBS can lower the blocking probability (loss probability in bufferless systems) and reduce the resource waste. Without using any OT – thus lowering the synchronization complexity, MBS together with PDP is the proposed HIP scheme. HIP differentiates the packets based on the packets' resource consumption. A packet that has traveled far in the core network has lower loss probability than new packets. The simulation model is an asynchronous bufferless network, consisting of 10 nodes in a ring topology. The simulation results showed that the proposed HIP scheme performs better in terms of loss probability and resource waste than a best-effort network. The throughput is thus increased and the network resources used more effectively. This is especially true for high loads and few wavelengths per fiber. The proposed scheme also performs admission control. Only new packets that don't block already admitted packets are allowed into the core network.

References

[1]   Tzvetelina Battestilli and Harry Perros, ”*An Introduction to Optical Burst Switching*”, IEEE Optical Communications Magazine, vol.41, August 2003

[2]   Chunming Qiao, ”Optical Burst Switching (OBS) for IP / WDM Integration”, Slides from the University at Buffalo (SUNY), January 2004

[3]   Malathi Veeraraghaven and Mark Karol, "Using WDM technology to carry IP traffic", 34th Annual Conference on Information Sciences and Systems, Princeton, NJ, Mars 15-17, 2000, URL (04.03.2004): http://www.ece.virginia.edu/~mv/pdf-files/ip-over-wdm.pdf

[4]   Martin Nord, Steinar Bjørnstad and C.M. Gauger, ”*OPS or OBS in the Core Network*”, COST 266 / IST-OPTIMIST, Budapest, Hungary, February 3rd 2003

[5]   Chunming Qiao and Myungsik Yoo, *"Optical Burst Switching (OBS) – A New Paradigm for an Optical Internet"*, Journal of High Speed Network, vol.8, no.1, pp.69-84, 1999

[6]   Byung-Chul Kim, You-Ze Cho, Jong-Hyup Lee, Young-Soo Choi and Doug Montgomery, *"Performance of Optical Burst Switching Techniques in Multi-Hop Networks"*, Proceedings of Globecom 2002, vol.3, pp. 2772-2776, 2002

[7]   Harald Øverby, *"A Study on Service Differentiation in Bufferless Optical Packet/Burst Switched Networks"*, In Proceedings of Norsk Informatikk Konferanse (NIK), pp. 105-116, November 24-26, 2003, Oslo, Norway

[8]   Vinod M. Vokkarane and Jason P. Jue, *"Prioritized Routing and Burst Segmentation for QoS in Optical Burst-Switched Networks"*, Proceedings, IEEE/OSA Optical Fiber Communication Conference (OFC) 2002, Anaheim, CA, WG6, pp. 221-222, Mars 2002

[9]   Cooperative Association for Internet Data Analysis (CAIDA), *"Packet sizes and sequencing"*, March 12th 1998, URL (29.05.2004): http://www.caida.org/outreach/resources/learn/packetsizes/