A hybrid signaling protocol for enhancing performance of OBS networks.

Franz Roeck[1], Kostas Ramantas[2,3], Erich Leitgeb[1] and Kyriakos Vlachos[2,3]

[1] *Institute of Broadband Communication, University of Technology, Graz, Austria*
[2,3] *Computer Engineering and Informatics Department & Research Academic Computer Technology Institute, University of Patras, Rio, Greece (Tel: +30 2610 996990, email: kvlachos@ceid.upatras.gr)*

Abstract— A new hybrid signaling scheme is proposed that combines a two-way with an one-way reservation protocol. The key idea is to synchronize the assembly process with the reservation so as to start simultaneously and hard reserve part of the end-to-end path in two way mode for a duration equal to the burst assembly time. In this way, upon the arrival of the first packet in the assembly queue, the two-way reservation part of the signaling scheme may start simultaneously. Upon the completion of the burst assembly process, burst transmission may start immediately since the path has been successfully established for a certain number of hops. The framework that we propose benefits from the parallel execution of the signaling messages and the assembly process, and exploits the burstification delay to guarantee burst transmission for at least a part of the network path.

***Index Terms*— Optical Burst Switching, Signaling, TAW, JET.**

I. INTRODUCTION

OBS has been introduced to combine the advantages of both packet and circuit switching and is considered a promising technology for the next generation optical Internet, [1]. An OBS network consists of a set of optical core routers and edge routers. An optical burst is constructed at the network edge, by aggregating a number of variable size packets. Each edge router maintains a separate (virtual) queue for each Forwarding Equivalence Class (FEC) to hold the data packets that belong to that FEC until a burst is formed. A FEC is defined from a source-destination pair and optionally from a set of Quality-of-Service requirements. In OBS networks an out-of-band control header, known as the burst header packet (BHP) is transmitted ahead of the burst in order to configure the switches along the burst's route. A number of signaling protocols [2]-[8] for OBS networks have been proposed so far. The signaling schemes found in the literature can be categorized into two main classes: two-way and one-way protocols. In two-way reservation schemes (also called *Tell-and-Wait*), end-to-end connections are fully established before the transmission of any data can start, while resources at intermediate nodes are reserved immediately upon the arrival of the SETUP packet at these nodes, [2]. This guarantees lossless transmission of bursts, at the expense of a high pre-transmission delay (in the order of RTT) and low link utilization.

In one-way reservation schemes (also called *Tell-and-Go*), a setup packet is sent in advance over the path, preceding the arrival of the burst by a small time offset. This minimizes the pre-transmission delay, but can result in high burst dropping probability. A number of one-way reservation schemes have been proposed for OBS networks, including the Just-Enough-Time (JET) [3], Horizon [4] and Just-In-Time (JIT) [5],[6].

Another interesting approach is hybrid signaling protocols that constitute a compromise between one-way and two-way signaling protocols, proposed in [7] and [8]. The *intermediate-node-initiation* protocol, proposed in [7] is a hybrid signaling protocol, where two-way reservation is being carried up to an intermediate node and one-way reservation for the rest of the path. The scheme actually decreases the delay associated with the establishment of the full end-to-end path to a sub-multiple one. Another approach presented in [8] makes use of a two-way reservation protocol and a burst length prediction mechanism to synchronize the assembly with the reservation process, so that both may start simultaneously with the first arrival of a packet.

In this work, we propose a novel resource reservation scheme for OBS networks, based on predictions of the burst length, as in [8], but combing it with an intermediate node initiation protocol as in [7]. In our study, we have used an N-order Normalized LMS (Least Mean Square) filter that provides adequate accuracy and has been previously used in OBS [9],[10]. In the proposed scheme, an intermediate node is selected so that the RTT to that node matches the burstification delay. In this way, the RTT is tuned to the assembly time, in contrast to [8] where it was the assembly time tuned to match the RTT. Thus, upon the arrival of the first packet in the burst assembly queue, the reservation of the two-way part may start immediately based on a burst length prediction.

The rest of the paper is organized as follows. Section II presents the network concept of the proposed hybrid signaling scheme, while Section III presents in detail the burst length prediction mechanism. Finally section IV presents evaluation results based on ns-2 simulations with emphasis given on the performance gains in terms of burst loss and edge delay.

II. NETWORK CONCEPT

OBS networks have been widely associated with one-way signaling protocols for on-demand capacity provisioning with a low delay overhead. However, burst losses in one-way protocols happen due to contention even at light loads and increase fast with the increase of the network load, making it is difficult to guarantee a certain level of QoS to end-users. In addition, assuming that each OBS edge router services concurrently thousands of active TCP connections, QoS support becomes an unrivaled task that requires cross layer (transport, network and physical layer) processing.

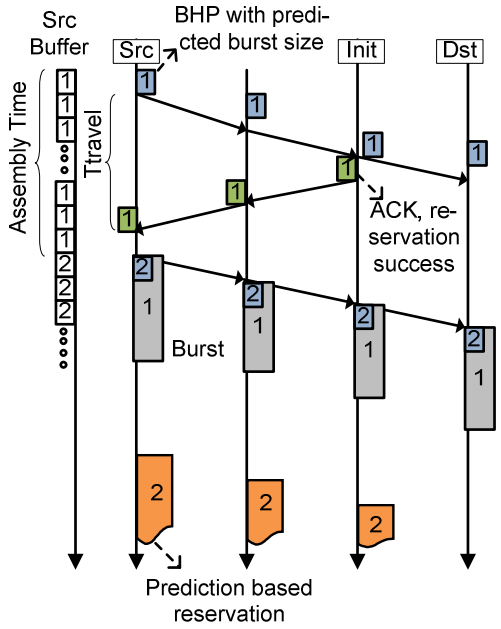

Figure 1: Timing considerations of the proposed scheme during a successful burst reservation.

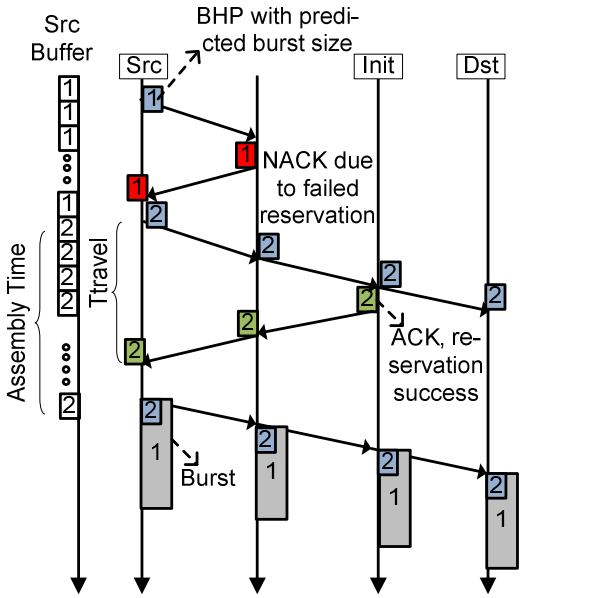

Figure 2: Timing considerations in case of a blocked request in the two-way part of the network path

The aim of this work is to propose a predictive resource reservation protocol for OBS networks, using a hybrid signaling protocol. Hybrid signaling protocols combine two-way and one-way signaling for resource reservation, providing a trade-off between burst losses at the core and edge delay. Signaling in such hybrid schemes is divided in a two-way part, up to an intermediate node (termed *initiator*) and an unacknowledged one-way part until the end-destination. In the two-way part, transmission is guaranteed to be lossless, if the path is established, while in the one-way part, transmission is best-effort. The proposed scheme reduces burst delay, by facilitating the parallel execution of the assembly process and the two-way part of the resource reservation until a certain intermediate node. Upon the arrival of the first packet in the queue, the two-way reservation part of the signaling scheme may start simultaneously based on a prediction of the burst length. Figure 1 illustrates the timing constrains of the proposed scheme in a successful burst reservation, while Figure 2 displays the signaling in the case of a blocked request in the two-way part of the network path.

A. Timing Considerations

In the proposed scheme, the RTT of the two-way part of the network path is tuned to the burst assembly time, through the appropriate selection of the initiating intermediate node. Specifically, the edge router assigns to each queue an assembly timer (T_{MAX}) and then selects the intermediate node by calculating the RTT time that the header packet needs to travel according to:

$$T_{travel} = 2 * \sum_{i=s}^{x}(t_{link\,i}) \leq T_{max} \qquad Eq.1$$

T_{travel} time is the time that the header packet needs to travel to the selected intermediate node, until which the reservation will be two-way and come back to the source. It is calculated by summing up the link-delays ($t_{link\,i}$) along the route as long as the sum is smaller than or equal to T_{max} /2.

The calculation of the intermediate node takes place only once, based on the burst assembly time and can be different for different source-destination pairs. Therefore, for each source-destination pair the exact intermediate node is known a priori and the reservation process can be formalized as follows.

Upon the arrival of the first packet in the queue (see Figure 1), a prediction mechanism estimates the size of the queue, at T_{MAX} time later and immediately transmits a setup (BHP) packet to reserve resources according to that prediction. Packets for that specific destination continue to arrive and are being stored in the same assembly queue. The BHP message (see Figure 1) propagates downstream and reserves resources in two-way mode until the intermediate node and in one-way mode until the final destination. The intermediate node, upon receiving the setup message, generates an acknowledgement message that is sent back to the source node to acknowledge the reservation, while the BHP message continues towards the end destination. In case the reservation fails in any node across the two-way part, a NACK message is generated at that node and is sent back to the source, as shown in Figure 2. In such a case, the BHP message is dropped. In both the two-way as well as the one-way part of the reservation process, delayed reservations are employed.

Upon the reception of the ACK message by the source node, the assembly process stops, and the burst transmission phase is triggered. Since the resource reservation was based on the predicted burst size, there is a possibility of over- or under-estimation of the actual burst size. In the case of an overestimation, all packets in the queue are transmitted as a single burst, simply wasting part of the reserved bandwidth. In the second case however, a subset of packets in the queue is assembled to a burst that corresponds exactly to the predicted burst size. The burst is transmitted, while the extra data gathered (termed *backlog* data) are left in the assembly queue,

and are transferred to the next assembly cycle. In case a NACK message is received at the source, the assembly process is reset and all data gathered are transferred to the next assembly cycle. This is done, so as to truly emulate a two-way reservation mechanism that retries to setup a path until its maximum delay has been reached. In general, the new assembly cycle is started, when the first packet arrives in the queue and ends either with receiving an ACK or a NACK message. In either case, the size of the backlog is added to the predicted burst size of the prediction filter, for determining bandwidth requirements of the next transmission.

It is worth noting here that the assembly time actually determines the length of the two-way reservation part over the end-to-end path. Thus, it can be the case, the proposed scheme to act as a complete two-way scheme, when T_{MAX} is equal to or even larger than the RTT of the path, or like an one-way scheme, when the RTT to the first hop is larger than the assembly time. For large scale, mesh network topologies, the part of the two-way or one-way reservation over the end-to-end path is not constant for all source-destination pairs, as it depends on the links delay across the path. In contrast to [8], in the proposed scheme we employ the same assembly time for all source-destination pairs. Having constant assembly times for all source-destination is preferable, as it leads to smoother burst sizes, making scheduling more efficient. It has been shown [11] that variable burst sizes are more difficult to schedule, leading to throughput losses.

B. Delay Considerations

In this section, we present analytical formulas to derive the average packet delay of the proposed signaling scheme. As mentioned, the proposed scheme uses hybrid signaling combining a two-way signaling protocol until an intermediate node, followed by one-way signaling until the end destination. Due to the parallel execution of the assembly process with the two-way part of the reservation, the total edge delay equals T_{MAX} in addition to the expected retransmissions overhead. The latter is due to the blocked BHP messages that inevitably trigger a new setup attempt. It actually depends on blocking probability p_i on each node across the path, along with the corresponding round trip time, RTT_i to that node. Given that the overall blocking probability across the two-way part of the network path is $p = 1 - \prod_{i=S}^{INT}(1 - p_i)$, where *s* is the source node and *INT* the intermediate one, then the expected delay overhead due to a blocked reservation request is:

$$T_r = \sum_{i=s}^{int} \left(\frac{p_i}{p} * RTT_i \right) \quad Eq.2$$

Thus, the expected edge delay is derived from T_{MAX} plus the retransmission overhead multiplied with the expected number of retransmissions per burst:

$$T_{edge} = T_{max} + \frac{p}{1-p} * T_r \quad Eq.3$$

Finally, the expected end-to-end packet delay is derived by summing the edge delay from Eq.3 with the burst transmission time t_b and the one-way propagation delay that the burst needs to travel from source to destination:

$$T = \frac{T_{edge}}{2} + t_b + \sum_{i=S}^{D} (t_{link\,i}) \quad Eq.4$$

III. BURST LENGTH PREDICTION

For burst size prediction, we have considered an N-order Normalized LMS (Least Mean Square) Linear Predictive Filter (LPF). This filter has been shown to provide high accuracy for a small time complexity of $O(N)$ for the coefficient calculation. This makes it suitable for online short-term predictions, in the order of an assembly timer, which is why it has been chosen for burst size predictions in previous works like [8]. The burst prediction process can be formalized as follows: Let $L_d(k)$ be the length of the k^{th} burst that corresponds to the k^{th} assembly cycle. The length of the next incoming burst is then predicted according to those of the previous *N* bursts by:

$$\check{L}_d(k+1) = \sum_{i=1}^{N} [h(i) \cdot L_d(k-i+1)] \quad Eq.6$$

where, $h(i),\ i \epsilon \{1, \dots N\}$ are the coefficients of the N-order LPF. We update the predictive filter coefficients using an efficient algorithm [12], where the coefficients for the $(k+1)^{th}$ burst prediction are estimated as:

$$\mathrm{h}^{k+1} = \mathrm{h}^{k} + \mu \cdot e(k) \cdot {\mathrm{L_d}}^{k} / \left\| {\mathrm{L_d}}^{k} \right\|^2 \quad Eq.7$$

where h is the current coefficient vector, μ is the filter step-size parameter, which was kept constant to 0.1, $e(k)$ the difference between the actual and the predicted length of the k^{th} data burst and $(\mathrm{L_d})^k$ the vector of the last N real burst sizes**.** Since the prediction accuracy of the LMS filter is not 100%, there is a possibility of over- or under- estimation of the actual burst size. Whereas an overestimation only results in a waste of bandwidth, an underestimation will result in a loss of a part of the burst at the network core due to insufficient resource reservation. Instead of dropping the extra data, we have opted for transmitting them during the next assembly cycle, which however increases burst transmission delay. Thus, it is important to compensate for the underestimation errors as it would clearly enhance performance. This can be done by introducing a correction margin δ, which is added to the predicted burst size. The correction margin δ is estimated as a multiple of the filter's variance, controlled by the "*aggressiveness*" parameter α:

$$\delta = a \cdot \sqrt{\frac{\sum_{i=1}^{N} e^2(k-i+1)}{N}} \quad Eq.8$$

The predictor's output is then modified to $L(k+1) = \tilde{L}_d(k+1) + \delta$. Aggressive reservation, i.e. with a high correction margin, can prevent underestimation errors, [10], thus decreasing the delay induced by transferring excessive

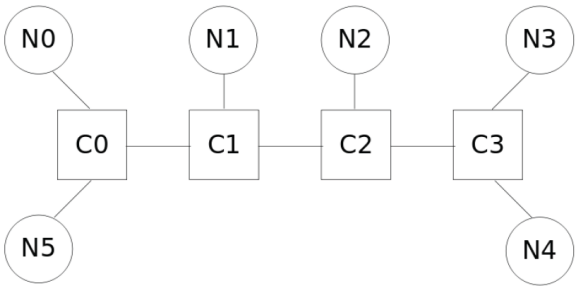

Figure 3: Network topology under study

data to the following assembly cycle. But this comes at the cost of wasting resources, as part of the burst reservations will be overestimated.

IV. Performance Evaluation

The proposed scheme was evaluated in both a simplified topology and NSFnet, using ns-2 simulation platform. The simplified topology used is displayed in Figure 3. It consists of four core nodes connected on a bus topology and six edge-nodes connected to the core. UDP connections are set up from all edge nodes N0, N1, N2, N3 and N5 to destination N4, generating Poisson packet arrivals with an average rate of λ. The delay for all links was set to 2ms. This simple network topology allows for selecting the intermediate node, simply by varying the aggregation time in 4ms steps, creating contention on every node of the network. For example for a 2-hop long two-way reservation, the assembly time should be 8msec, while for a complete two-way should be 16msec or for complete one-way less than 4msec.

The filter's performance is summarized in TABLE I, displaying the Coefficient of Variation (CoV) metric, as well as the underestimation probability (P_u). The latter is defined as the probability that the predicted burst size is smaller than the actual one. Each set of metrics was derived within a simulation cycle for all bursts transmitted, varying parameter α which controls the predictor's aggressiveness. From TABLE I, it follows that the use of the correction parameter decreases the underestimation probability, albeit the increase of the predictor's variance.

TABLE I: Performance of LMS filter for different correction margins, controlled by parameter α.

Correction Margin	CoV	P_u
α=0	0.28	0.40
α=1	0.29	0.11
α=2	0.38	0.03
α=3	0.55	0.01

Using the abovementioned traffic profile and $\alpha = 1$ that better balances prediction error and underestimation probability, we investigated the performance of the proposed signaling protocol in terms of burst drop ratio and edge delay. Our target is to investigate, what is the performance gain (in terms of blocking), when part of the reservation is being done in two-way mode, compared to standard one-way and two-way

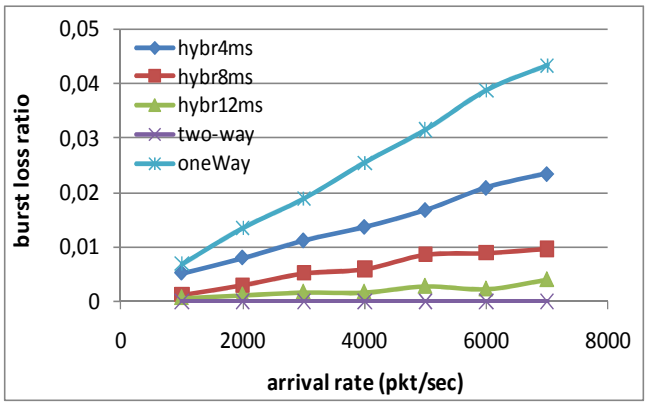

Figure 4: Burst drop ratio versus packet arrival rate for different assembly times and signaling schemes.

signaling. Emphasis was given in achieving packet loss-free operation. In such a case, packets that would be otherwise dropped, due to an underestimation in the burst length or a blocked setup message, are delayed for transmission during next assembly cycles. Thus, actual data losses will only happen in the unacknowledged one-way part of the network path. As regards packet edge delay, as shown in Eq.3 it depends on the assembly timer and the retransmission overhead, induced by blocking at the two-way part of the network path. In what follows we have measured the retransmissions ratio as well as the average packet queuing delay at the network edge.

Figure 4 displays the burst drop ratio of the proposed hybrid signaling scheme versus the arrival rate for different assembly timers, where each timer corresponds to an equivalent number of intermediate hops. For comparison, the performance of two-way and one-way signaling with an assembly timer of 4ms is also displayed. From Figure 4, it can be clearly seen that burst drop ratio of the hybrid scheme continuously decreases with the increase of the number of hops of the two-way part of the path, as losses only happen on the unacknowledged one-way part of the network path. In particular, the net gain for an arrival rate of 7kpacket/sec is 2%, 3.5% and 4% for 1, 2 and 3 hop two-way reservation. The gain in blocking will further increase for even higher traffic loads. To this end, we may argue that even hard reserving a small part of the path, there is a clear gain in the burst loss ratio, which can be significant for high traffic loads.

Figure 5 displays the percentage of retransmissions per burst for different packet arrival rates. In general, retransmissions result in an increased delay overhead, as well as a higher overhead for the control channel, since more than one BHP is sent per data burst. To this end, we may argue that the extra control overhead remains negligible as in all cases of Figure 5, the retransmission ratio remains under 3%.

Figure 6 shows the edge delay, from the time a packet arrives at the buffer, till it is transmitted in a burst. As expected, when moving from complete one-way to complete two-way, there is a standard increase in the delay time as the two-way part of the network path increases. In particular, as expected from Eq. 3, we observe on average an 2msec increase for each intermediate hop added (which translates to 4ms increase in the assembly timer), plus the difference in the retransmission overhead. Overall, we may argue that the

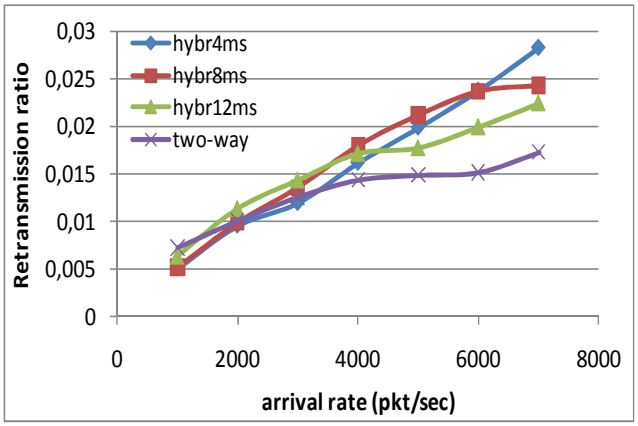

Figure 5: Retransmission of connection requests versus packet arrival rate for different assembly times and signaling schemes.

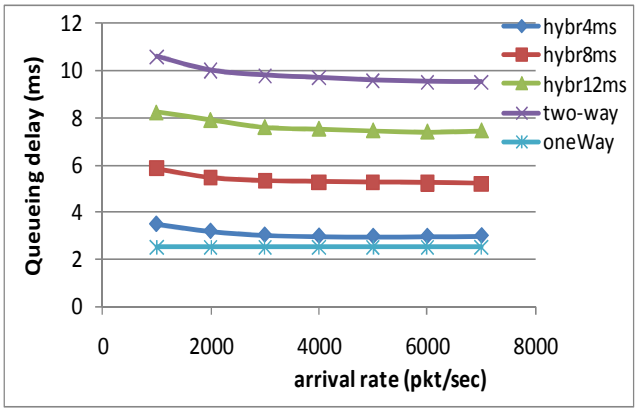

Figure 6: Average packet queuing time versus packet arrival rate for different assembly times and signaling schemes.

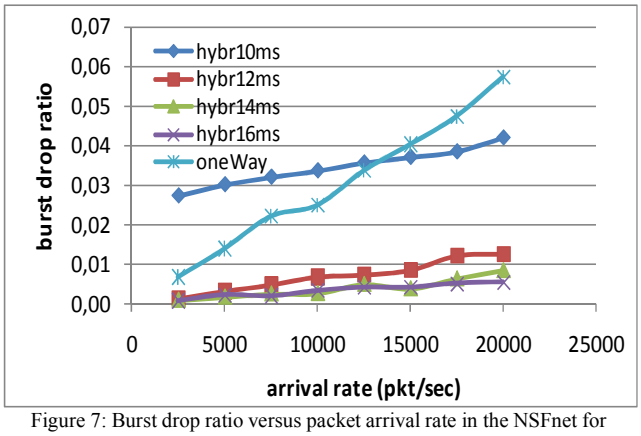

Figure 7: Burst drop ratio versus packet arrival rate in the NSFnet for different assembly times and signaling schemes

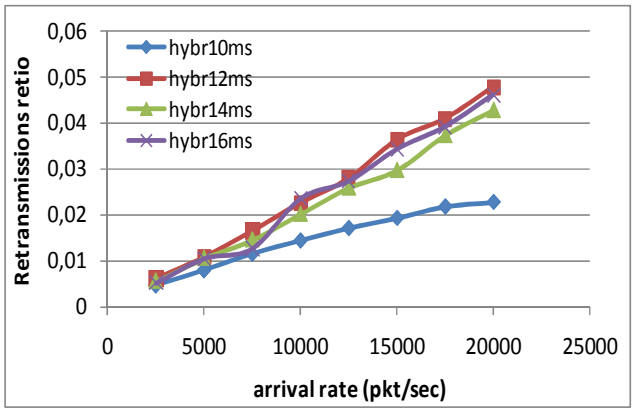

Figure 8: Retransmission of connection requests versus packet arrival rate in the NSFnet, for different assembly times.

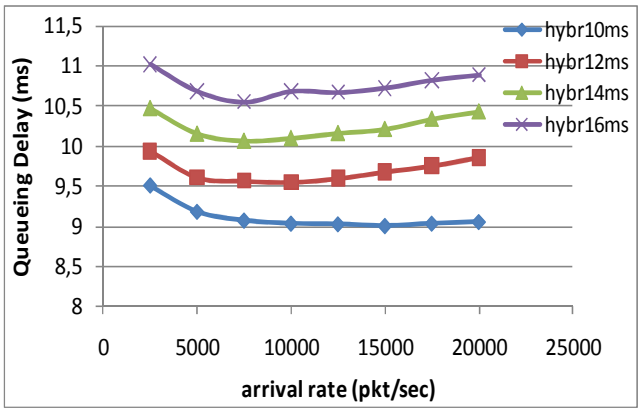

Figure 9: Average packet queuing time versus packet arrival rate for different assembly times in the NSFnet network topology.

proposed scheme successfully captures the trade-off between delay and burst loss, controlled by the selection of the intermediate node. Specifically, the loss probability decreases with decreasing the unacknowledged part on the network path, while the edge delay (queuing delay) increases. However, this extra delay is, in any case predictable and always less than the end-to-end round trip time delay induced by two-way OBS.

For validating the results over a full network topology, we employed the proposed scheme over the NSFnet topology with 8 edge and 6 core nodes. In that case, due to the different link delays in each network path, the choice of an assembly timer does not correspond to a predetermined number of intermediate hops, but varies for each source-destination pair. In the following experiments, we considered one wavelength per link at 1Gb/s capacity, while packet arrival rate, (same for all edges nodes) was varied from 2kpacket to 20kpacket/sec following a Poisson distribution profile.

Figure 7 displays the burst loss ratio of the proposed hybrid scheme, compared to standard one-way and two-way signaling. T_{MAX} is chosen to be 10, 12, 14 and 16msec. From Figure 7, the clear gain in burst loss ratio can be seen. Burst loss ratio significantly decreases with the increase of the two-reservation part of the path and further it increases at a much slower rate with the increase of the packet arrival rate. In particular for an arrival rate of 20kpakcets/sec the gain in burst loss ratio is 2% even when two-way reservation is being performed for one hop only. For the rest of the cases, the gain is higher (4% to 6%).

Finally, Figure 8 and Figure 9 display the number of retransmissions and the edge delay of the hybrid scheme respectively. As expected the increase of the two-way part of the path results in an increase of the edge delay, as in the simple network topology. Additionally we can see that the extra overhead induced by retransmissions is again negligible, since the absolute number of reconnection attempts remains small. In other words the majority of bursts are being sent at most with two connection requests.

V. Conclusions

In this paper, a new hybrid signalling scheme is proposed that combines a two-way with an one-way reservation protocol. The key idea is to hard reserve a part of the end-to-end path using a two-way protocol that will increase the burst forwarding probability. Further with the use of a burst size prediction mechanism, the assembly process can be synchronized with the two-way reservation part so as the reservation of resources to start immediately with the arrival of the first packet in the burst assembly queue. Simulation experiments have shown that when reserving even a small part of the end-to-end path in two-way mode there is a clear performance gain in the burst forwarding probability achieving lower burst loss ratios , while packet delay remains bounded and retransmission requests have a small impact on the edge delay.

Acknowledgements

The work described in this paper was carried out with the support of the BONE-project ("Building the Future Optical Network in Europe"), a Network of Excellence funded by the European Commission through the 7th ICT-Framework. Franz Roeck work was performed during his staying in University of Patras.

References

[1] C. Qiao and M. Yoo, "Optical burst switching (OBS)-A new paradigm for an optical internet," J. High Speed Networks, vol. 8, no. 1, pp. 69–84, 1999.

[2] M. Dueser and P. Bayvel. "Analysis of a dynamically wavelength-routed optical burst switched network architecture", IEEE/OSA Journal of Lightwave Technology, 20:574-585, April 2002.

[3] M. Yoo and C. Qiao, "Just-enough-time (JET): A high speed protocol for bursty traffic in optical networks", IEEE/LEOS Technology Global Information Infrastructure, pp. 26–27, Aug. 1997.

[4] Y. Xiong, M. Vandenhoute, and H. Cankaya, "Control architecture in optical burst-switched WDM networks", IEEE Journal on Selected Areas in Communications, vol. 18, pp. 1838–1851, October 2000.

[5] J. Y. Wei and R. I. MacFarland Jr, "Just-In-Time signaling for WDM optical burst switching networks", IEEE/OSA Journal of Lightwave Technology, vol. 18, pp. 2019–37, Dec. 2000.

[6] I. Baldine, G. N. Rouskas, H. G. Perros, and D. Stevenson, "JumpStart: A just-in-time signaling architecture for WDM burst-switched networks", IEEE Communications, 40(2):82-89, February 2002.

[7] Vinod M. Vokkarane, "Intermediate-node-initiation (INI): A generalized signaling framework for optical burst-switched networks", Journal of Optical Switching and Networking, vol. 4, pp. 20–32, 2007.

[8] Kyriakos Vlachos and Demetris Monoyios, "A Virtual One-Way Signaling Protocol With Aggressive Resource Reservation for Improving Burst Transmission Delay", IEEE/OSA Journal of Lightwave technology, Volume 27, Issue 14, pp. 2869 – 2875, July15, 2009.

[9] D. Morato, J. Aracil, L. A. Dıez, M. Izal, E. Magana, "On linear prediction of Internet traffic for packet and burst switching networks", in Proc. of ICCCN, pp.138–143, 2001.

[10] J. Liu et al. "FRR for Latency Reduction and QoS Provisioning in OBS Networks", IEEE Sel. Ar. in Comm.,vol. 21, pp.1210, Sept. 2003.

[11] Jikai Li, Chunming Qiao, Jinhui Xu, Dahai Xu, "Maximizing Throughput for Optical Burst Switching Networks", IEEE/ACM Transactions on Networking, vol.15, no.5, pp.1163-1176, Oct. 2007

[12] J.R. Treichler, et al. Theory and Design of Adaptive Filters. New York: Wiley, 1987.