

Distributed Contention Avoidance in Optical Burst-Switched Ring Networks

Joan Triay and Cristina Cervelló-Pastor
Department of Telematics Engineering
Technical University of Catalonia
Av. Canal Olímpic, 15
08860, Castelldefels, Spain
Email: {joan.triay; cristina}@entel.upc.edu

Abstract—Optical Burst Switching (OBS) is a promising technology for future optical networks. Due to its less complicated implementation using current optical and electrical components, OBS is seen as the first step towards the future Optical Packet Switching (OPS). Regarding network topologies, the ring is extensively used on current metropolitan environments, and this does not seem it will change in the foreseeable future. In this paper we propose a new medium access protocol for Optical Burst-Switched Networks over metropolitan rings that enables the nodes to access the channel in a distributed way. This new protocol avoids burst blockings in transit nodes, turning itself into an efficient and simple burst contention avoidance mechanism. The use of this protocol maximizes the throughput of the network without deteriorating excessively other parameters such as end-to-end delay or ingress queues.

I. INTRODUCTION

Future Metropolitan Area Networks (MAN) are expected to support an increasing dynamic and demanding amount of data traffic [1]. In this field, the use of optical switching network technologies is an opportunity to efficiently manage the new bandwidth demands in a flexible and scalable way, alleviating the network of the problems and limitations caused by present technologies, such as Synchronous Digital Hierarchy (SDH).

In the future optical network context, ring and bus topologies may still be promising candidates for optical switched networks. Ring is a well-known and popular topology in metro networks. This topology provides facilities to ensure failure resiliency, dynamic fairness and low cost of deployment. The resilience and scalability is principally achieved by introducing bidirectionality and interconnection of rings.

Optical Burst Switching (OBS) [2] has been proposed as a practical paradigm for the optical Internet using enabling optical technologies. One of its motivations is to enhance the performance of current optical circuit switching technologies offering statistical multiplexing of bursts. In OBS, packets are assembled into large bursts and transmitted after an offset time. This offset permits a control packet to reserve resources along the path, so that, subsequently, the burst can be transmitted and switched entirely in the optical domain.

In this paper we present a new distributed medium access protocol for optical burst-switched ring networks. This new protocol provides very good performance in terms of burst blockings. In fact, it makes true the objective of a loss free

network, at the expense of introducing a small extra delay to access the optical channel.

The remainder of the paper is organized as follows. Section II reviews the use of some access protocols for WDM ring networks. The operation of the distributed medium access protocol is introduced in Section III. Section IV discusses some important features that we need to take into account to use the proposed distributed protocol in OBS networks. Section V analyzes the network performance and discusses the results extracted from simulations. Finally, Section VI summarizes the main conclusions of the paper.

II. PROTOCOLS FOR OPTICAL BURST SWITCHED RINGS

With the emerging of the Optical Burst Switching technology, it was just a question of time trying its deployment over optical WDM rings. Various research works have proposed the use of WDM metro ring networks with OBS [3][4][5].

In [3], the authors proposed an OBS ring network composed of N nodes, where each node has an assigned and dedicated home wavelength to transmit its bursts. A fixed-tuned transmitter and tunable receiver (FT-TR) system is necessary for implementing the ring. As long as each node makes use of a reserved wavelength, no collision occurs on the channel. However, collisions on reception may still occur. This protocol didn't provide a scalable and bandwidth-efficient solution for MANs.

A token approach was introduced in [4]. The approach was a distributed OBS metro ring architecture referred to as LightRing. This was a multi-token Media Access Control (MAC) protocol designed to ensure the bandwidth-efficient and loss-free transmission of data bursts. Nodes get access to the channel taking the token that circulates around the ring.

The use of MAC protocols with token schemes has been continuously reused [5][6]. In the Collision-free Optic-burst Ring Network (CORNet) [5], multiple nodes share the same drop wavelength when receiving data bursts, but each node can use any wavelength for transmission. In the architecture, a new Multi-token Access Control Algorithm (MACA) is proposed for arbitrating the accessibility to each data wavelength.

III. DISTRIBUTED ACCESS IN OPTICAL BURST-SWITCHED RINGS

Following we introduce a new MAC protocol for high-speed optical burst-switched networks. The protocol, named as Distributed Access for Optical Burst-Switched Rings (DAOBSR), takes some references from the IEEE 802.6 Distributed Queue Dual Bus (DQDB) Metropolitan Area Network (MAN) standard [7]. DQDB can not be applied AS IS over an optical burst-switched network since both technologies extremely differ in the way control and data planes are defined. Besides, bursts can be various orders of magnitude greater than the DQDB frames defined in the protocol [7].

In the field of distributed medium access protocols, the main candidates are “empty-slot” protocols [6]. These protocols switch fixed-lengths packets synchronously. By giving priority to packets in-transit, the contention to access the medium is resolved by inserting packets in the network only when there are free slots on a given wavelength of the ring. The wavelength information availability can be obtained from an out-of-band control channel, or from in-band packet headers.

The applicability of our approach to OBS relies on two important issues: the definition and instantiation of optical buses around the ring, and the management of slot synchronization.

A. Operation of the protocol

In this section we briefly describe the protocol operation. We consider two unidirectional control channels, the downstream or forward channel, from the head node to the tail node, and the upstream or reverse channel, from the tail to the head node. Both channels use the same path, one in each direction. With these channels we control the burst transmission on the data channel using a distributed FIFO queue.

The distributed queue access protocol operates as follows. Each node holds two counters, a request counter (RQ_C) and a count-down counter (CD_C), for each priority (if different traffic priorities are defined). A node can be either idle or active. In the idle state, the node has nothing to transmit, while in the other state, the node is active and waiting for its turn to transmit a single burst queued in its ingress buffer.

During the idle state, the node keeps track of the requests generated by the downstream stations by monitoring the request (REQ) bit of each request packet traveling on the reverse bus; and the busy (BUSY) bit in the header of each control packet traveling on the forward bus. For every REQ bit set to one, which implies a request of transmission by any of the downstream nodes, the request counter RQ_C is incremented by 1. For every BUSY bit set to zero, which implies an unoccupied slot that will fulfill one of the downstream requests, RQ_C is decremented by 1 if it is greater than 0. If it was already 0, it is not decremented. The current value of RQ_C thus indicates the number of downstream nodes that have a burst waiting to be transmitted.

Whenever a node generates a burst for transmission, it becomes active. The content of the RQ_C is copied onto the count-down counter CD_C , and then RQ_C is reset to 0. The CD_C is a measure of the outstanding requests queued

for transmission before the node gets its turn to transmit. In fact, with this action, the node has placed itself in a global distributed queue with first-come-first-served (FCFS) discipline. Following a similar monitoring scheme as in the idle state, the CD_C is decremented by 1 for every available slot on the forward channel, and the RQ_C is incremented by 1 for every request on the reverse channel. When CD_C reaches 0, the node prepares itself for transmission using the next available free slot on the forward channel. This indicates the node that it has reached the head of the queue and therefore it is allowed to transmit. Following the transmission of its queued burst, the node reverts back to the idle state, and its position in the overall queueing environment is again described by the RQ_C value.

IV. PROTOCOL IMPLEMENTATION ON OPTICAL BURST SWITCHING

The implementation of a distributed medium access control protocol like the one introduced in the previous section for the Optical Burst Switching technology is not straightforward. There exist some considerations that need to be taken into account, such as the instantiation of the buses, the reservation protocol operation in a distributed approach, the use of a slotted network, etc. In the next subsections we describe more in detail each of these issues.

A. Implementing slotted OBS

To ensure a proper operation of DAOBSR we need an efficient slot synchronization technique. The head of the bus (HoB) must periodically transmit control packets announcing the occupation of each time slot. And the same for the tail of the bus (ToB), which is responsible of transmitting request packets on the reverse channel.

In this case, we propose a self-synchronizing network architecture, which is able to perform successfully without the use of electronics using simple all-optical circuits [9]. Each node provides the necessary clock signals that are self-extracted from the incoming control packets. Two preamble bits are inserted in front of each control packet to assist clock extraction at each node. In order to account for the clock acquisition time of the self-synchronization stage, we consider one additional guardband bit between the preamble and the payload of the control packet.

In a metropolitan environment where network links are on the order of tens or hundreds of kilometers, and for the case of a channel capacity of 1 Gbps, the slot size should be of tens of kbytes. We consider a burst length equal to the slot size minus the guard times.

B. Bus instantiation

Bus instantiations around a ring can be done under two scenarios, each one with their advantages and disadvantages: a static pre-planned scenario or a dynamic planning. In this protocol evaluation we have chosen to statically generate the buses under the assumption that the network carries traffic flows distributed uniformly around the ring. This solution is

easier to implement, but it lacks some of the advantages of the dynamic instantiation such as the support of non-uniform traffic flows, or better resilience capabilities in front of network failures. The dynamic instantiation may be done using the Multi-Spanning Tree Wavelength Assignment (MSTWA) protocol proposed in [8] or with some type of GMPLS reservation mechanism. MSTWA ensures a low burst blocking probability inside the network by restricting the use of certain wavelengths to each destination. The light-branches defined with MSTWA in the optical domain are very similar to a bus in the electrical domain. Branches are created between two endpoints, being the head and tail of the bus the branch root node and the branch destination node, respectively. In this document we may indistinctly use the terms “root” of the light-branch or “head” of the bus (HoB).

In the proposed architecture, a bus instance holds two counters $-RQ_C$ and CD_C for every priority level, a one-position buffer for the scheduled burst to transmit, and a buffer of bursts waiting to be served.

Figure 1 depicts the relationship between these elements. The Bus Waiting Queue (BWQ) receives bursts from the Virtual Output Queues (VOQ) as soon as the wavelength assignment module has successfully processed them. Later, according to the processing of the control and request packets, the scheduler moves the first burst of the BWQ to the Bus Ready Queue (BRQ). The burst is transmitted from the BRQ as soon as the offset time elapses after successfully occupying a free reservation slot. The total waiting time of a burst in the node’s queue system is sum of the channel access delay and offset time:

$$T_{delay} = T_{access} + T_{offset} \quad (1)$$

The offset time is fixed for each node that belongs to the bus, but decremented at every hop on the downstream channel. This means that nodes closer to the ToB have a lower offset time. Nevertheless, the access delay depends on the values of the protocol’s counters, which at the same time change their value according to the bus load. In the results section we evaluate how this delay depends on the bus utilization.

A VOQ is created for each set of bursts destined to the same end node, e.g. if a node transmits bursts to two different destinations, it puts them into their respective queues.

The distributed queue policy does not permit a node to transmit a new reservation request unless the BRQ is again empty. As a result, nodes can only transmit bursts one after another, and therefore, the BRQ can only be occupied by one element.

C. Wavelength assignment

The wavelength assignment is responsible of assigning bursts to a specific bus. Every wavelength can only be part of a unique bus, hence assignment either to a bus or a wavelength means the same.

We run an algorithm to carry out the wavelength assignment according to the values of the counters RQ_C and CD_C .

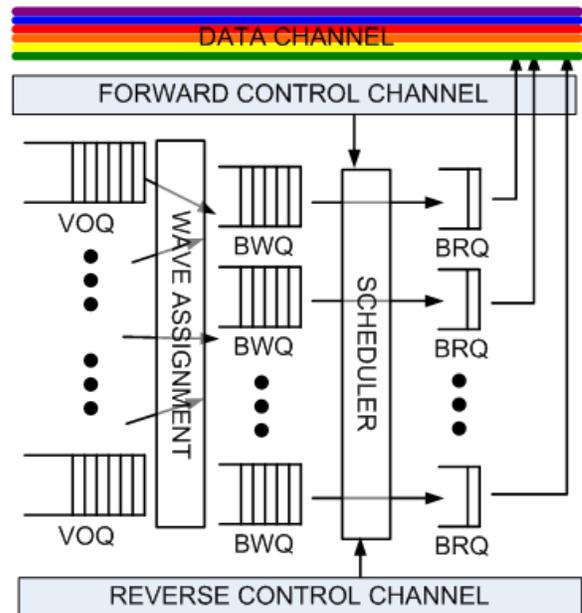


Fig. 1. Bus instantiation and queue scheduler.

As we have previously introduced, the value of these counters determines the position of the node in the distributed FIFO. Depending on these values, the node will take more or less time to transmit a burst on that specific bus, thus incrementing or decreasing the channel access delay, and consequently, the end-to-end delay of the burst. What we do here is to select the bus that will give less delay, and therefore, the bus that by adding the two counter values gives the lower value.

D. Reservation scheduler and MAC protocol

Initially, the upstream node, the head of bus (HoB), is allowed to fill all the control packets with burst transmissions (i.e. setting always the BUSY bit) when it does not see any request packet on the reverse bus with the REQ bit set. Downstream nodes examine the value of the BUSY bit and if it is set to 1, they process the control information to reserve resources for the upcoming burst switching.

The MAC scheduler is composed mainly of three algorithms. The first one is specific of the HoB, which is responsible of generating and transmitting the control packets. As we have noted previously, the node is allowed to fill all the control packets with transmissions when it does not see any request from any downstream node. If the RQ_C is not zero, meaning that any of the downstream nodes have requested a transmission slot and the HoB has a burst to transmit, then it starts the count-down counter and transmits a control packet with the BUSY bit unset. In this process, the HoB moves the first burst from the BWQ to the corresponding BRQ. Later, at the next slot, if no more requests have to be attended, the HoB will have the chance to occupy that slot to transmit the burst from the BRQ.

The second algorithm is run by the downstream or forward nodes along the bus and it is very similar to the first one. If

the BUSY bit of the control packet is set to 1, the node simply performs the reservation of resources to switch the upcoming burst after an offset time, in the same way control packets are used in an OBSN. Otherwise, the algorithm run by the current node forwards or initiates a new reservation according to the counters and the input buffers. To decrease the channel access delay on the downstream nodes, these nodes are allowed to occupy a free slot, thus initiating a new burst transmission reservation, if the node has its RQ_C equal to 0.

Under normal circumstances, downstream nodes need to run the third algorithm to request free slots. It should be noted that a node can only set the REQ bit if this has not been previously set by a downstream node. When this bit is unset, and the node has a burst to transmit, it changes the bit value to 1, starts the count-down counter and moves the first burst from the BWQ to the corresponding BRQ. In any case, a new request cannot be initiated if there is an existing burst waiting in the ready queue. Finally, when the request packet reaches the HoB, on the next time slot, it will transmit an unset BUSY bit control packet for satisfying this request.

V. RESULTS

The performance of the protocol has been evaluated by simulation. The simulator uses the DESMO-J simulation framework [10] made by the University of Hamburg.

In the simulations we consider the use of a ring network with a fixed number of network nodes. In all the network scenarios we assume a wavelength capacity of 1 Gbps, a control packet processing time of 10 μ s (including the clock recovery time), a switch fabric configuration time of 5 μ s and a first burst query processing of 3 μ s. The simulations presented here assume no wavelength conversion capability in the network.

To provide fixed burst lengths, the burst assemblers in every node follow a volume-based approach. In order to provide a more accurate burst generator, and to increase the efficiency of the simulator by decreasing the number of events to treat, we implemented the volume-based algorithm presented in [11]. This approach works under the assumption that the process of packet arrival to the assembly buffer follows a Poisson distribution. The authors mention a speedup gain as high as 40 when their algorithm is used.

The values given in this section have been averaged from 30 simulations. Some graphs show the confidence intervals and others do not in order to improve the readability.

In this section we evaluate the properties of the DAOBSR protocol regarding the following features: 1) burst blocking probability, 2) throughput, 3) end-to-end delay, 4) queue performance and 5) channel access delay.

In the graphs, the traffic load is defined as [12]:

$$\rho = \frac{NR}{C\lambda} \quad (2)$$

where N is the number of nodes, R denotes the data rate of the arrival traffics added into the ring, C is the capacity per WDM channel, and λ corresponds to the number of

wavelengths around the ring. As pointed in section IV-B, in this protocol evaluation we have used a static bus instantiation scheme to provide the best performance when traffics are distributed uniformly around the ring and loads are similar among the set of links.

A. Comparison of protocols

In this section we evaluate the burst blocking probability and other parameters of the DAOBSR protocol versus other protocols that use different ways of wavelength assignment. The access control in these other protocols is not distributed, hence when a burst is scheduled to use a certain wavelength is transmitted immediately after its offset time has elapsed. These protocols and their respective terms are: a random wavelength assignment scheme (RANDOM); the First-Fit with Traffic Engineering (FFTE) [13], in which the start wavelength is different for each node; and the MSTWA, which assigns the wavelengths on the basis of a set of spanning trees of lambdas. In each of these scenarios, but DAOBSR, the nodes use Fiber Delay Lines (FDL) to resolve some of the contentions. In comparison to the rest, DAOBSR does not need any type of burst contention resolution technique to achieve a good network performance.

The simulations in this test are of a network composed of 16 nodes and 8 wavelengths. The minimum burst length is 19,000 bytes, very close to the maximum burst-slot length value established in this case at 20,000 bytes –corresponding to a slot time of 0.16 ms. The results of these simulations are drawn in the set of graphs in Figure 2.

Figure 2(a) depicts the total burst blocking probability in each scenario. As it can be seen, the plot of the DAOBSR is not present in this graph, which means that we do not have any burst contention in the network. In a different way, the rest of protocols experiment some burst blockings. The one that offers the poorest performance is the FFTE protocol. At low loads, MSTWA is worse than RANDOM, whereas at loads close to 100%, the performance of these two are similar. As long as the network load is below the capacity limit, DAOBSR always provides a way to transmit and switch bursts without interferences between nodes. There will be a counterpart to satisfy such behavior, as we will see. Figure 2(b) complements the results given above showing the average effective network throughput per node. It can be observed that the receiver's throughput slightly decreases for all the protocols but DAOBSR. Due to the perfect performance of DAOBSR in terms of burst blocking, the throughput relation is very close to 1.0.

Figure 2(c) shows the average end-to-end delay as a function of the network traffic load. The delay calculated in this section is composed of the following delays: burst assembly, channel access, offset time, burst transmission, propagation and reception. Therefore, the value gives an upper bound of the end-to-end delay, that is, the delay that the first packet of a burst may experience. It can be observed that the DAOBSR protocol introduces an extra delay in comparison to the rest. This delay is primarily produced by the extra time needed

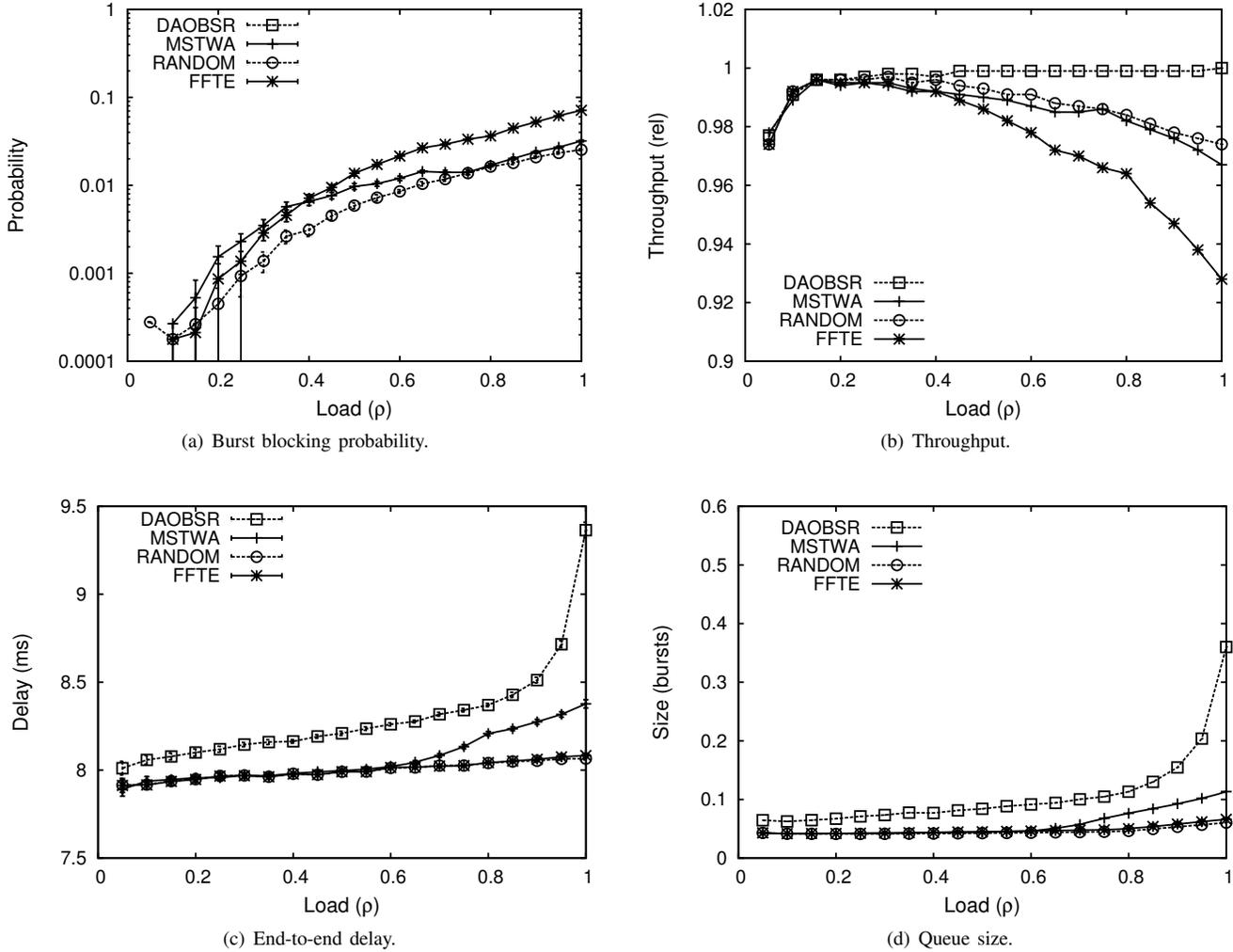


Fig. 2. Comparison between protocols (8 waves, 16 nodes, burst of 19,000 bytes and slots of 20,000 bytes).

to access the channel –the counterpart we have mentioned above. This delay has something to do with the number of requests that have to be processed in the globally distributed FIFO before the present node, i.e. the node that is processing the current burst, gets access to the channel. Remember that this value is given by the request counter RQ_C . For almost any load, the delay introduced by our approach is around 5% greater than the rest. The issue comes up at high loads where the delay tends to increase exponentially.

Figure 2(d) depicts the average size of the burst input queue as a function of the traffic load in the network. The values in this graph are closely related to the previous ones. At higher loads, the number of bursts queued in the input buffers are greater as a result of increasing the channel access delay due to the saturation of the buses.

The thing is that the protocol DAOBSR is hard to compare to the rest of the protocols. On one way, the protocol offers an excellent performance according to the number of burst blockings. On the other way, the delay is higher at any traffic load. However, this added delay is not so high, just around

5% for most of the load range (except at high loads). Next subsection deals about this issue.

B. Access delay and network size

As it has already been mentioned, the access delay of the DAOBSR is its main drawback. If a node needs to transmit a burst, it must wait its turn according to its position in the virtual distributed FIFO queue. In this section we quantify this access delay on the same assumptions and network conditions established in the previous subsection, plus two extra scenarios with 8 nodes and 4 lambdas, and 4 nodes and 2 lambdas, respectively.

In figure 3(a) we show these results. We can appreciate that the access delay can not be neglected, specially under heavy load scenarios. Another interesting issue is the behavior in each network scenario: a greater number of nodes does not seem to change extremely the access delay in comparison to an scenario with fewer nodes.

Finally, for the same group of network scenarios, a comparison of the burst average delay in queue is shown in figure 3(b).

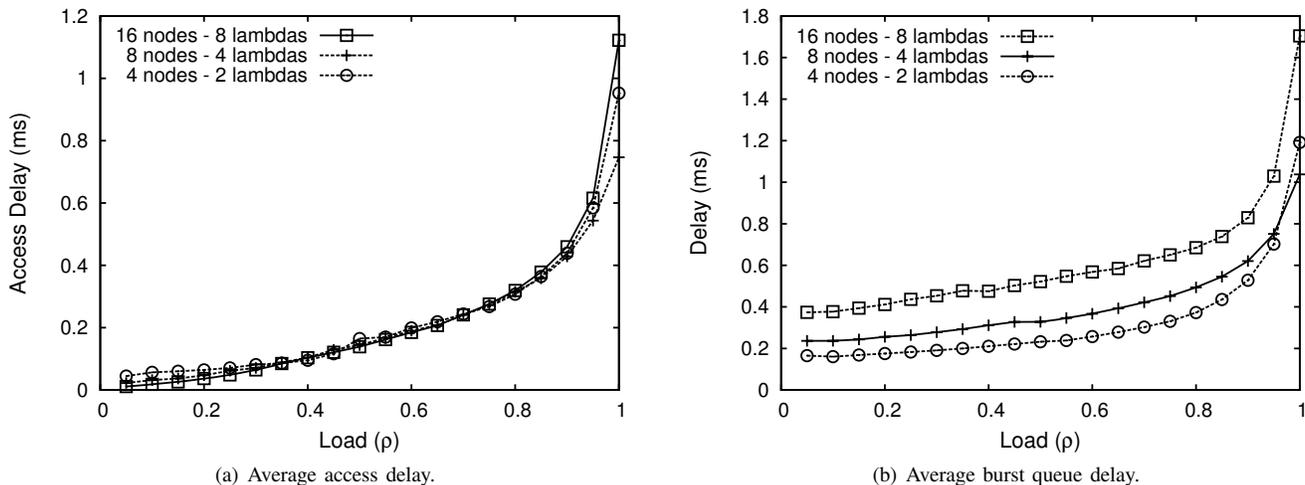


Fig. 3. Comparison of network size (burst of 19,000 bytes and slots of 20,000 bytes).

This delay, see Eq. 1, is a sum of the access delay analyzed in the previous figure and the average offset time. Enlarging the network implies that the average number of hops to reach any destination is also increased. A longer path needs a greater offset time, and as a result, the average waiting delay increases.

VI. CONCLUSIONS

In this paper we have presented a novel medium access protocol for Optical Burst-Switched ring networks that uses a distributed knowledge of the network state to control the way nodes access to the channel, i.e. transmit data bursts. The protocol minimizes the number of burst contentions, making into reality the objective of a loss free network, at the expense of introducing some extra average channel access delay.

The protocol has been compared to some common wavelength selection –channel assignment– protocols. The results have shown a very good impact in the burst blocking performance. At the same time, the delay and queue sizes in the traffic input nodes are worse in comparison to the rest of the protocols, but not so worse than it may be expected.

A further analysis shows that the access delay does not have a great dependency of the network’s size. However, a larger network means increasing the average network distance, so that a greater offset time is necessary for establishing the burst switching path.

As future research lines, an efficient way of managing the channel access priorities should be analyzed. This, as well, would offer us a method to provide fairness between nodes.

ACKNOWLEDGMENT

Work supported by the Spanish “Ministerio de Educación y Ciencia” and FEDER under project TSI2006-12507-C03-03, the i2CAT Foundation’s project TRILOGY, the EURO-NF Network of Excellence and by the Department of Education and University of the Government of Catalonia and the European Social Funding with a pre-doctoral grant.

REFERENCES

- [1] M. Herzog, M. Maier, M. Reisslein, “Metropolitan area packet-switched WDM networks: a survey on ring systems”, *IEEE Communications Surveys*, vol. 6, no. 2, pp. 2-20, 2004.
- [2] Y. Chen, C. Qiao and X. Yu, “Optical Burst Switching: A New Area in Optical Networking Research”, *IEEE Network*, vol. 18, pp. 16-23, May 2004.
- [3] Lisong Xu, Harry G. Perros, George N. Rouskas, “A Simulation Study of Optical Burst Switching and Access Protocols for WDM Ring Networks”, *Computer Networks*, vol. 41, no. 2, pp. 143-60, Feb. 2003.
- [4] A. Fumagalli, P. Krishnamoorthy, “A Low-Latency and Bandwidth-Efficient Distributed Optical Burst Switching Architecture for Metro Ring”, *Proceedings of ICC 2003*, vol. 2, pp. 1340-44, May 2003, Alaska (USA).
- [5] Hui-Tang Lin, Wang-Rong Chang, “CORNet: A Scalable and Bandwidth-Efficient Optical Burst Switching Ring Architecture for Metro Area Networks”, *Proceedings of 2006 International Conference on Networking and Services*, Jul. 2006, California (USA).
- [6] M. Nord, S. Bjornstad, M.L. Nielsen, “Distributed medium-access-control protocol for an optical-packet-switched ring network supporting variable-length packets”, *Journal of Optical Networking*, vol. 4, no. 4, pp. 213-25, Ed. Optical Society of America, April 2005.
- [7] IEEE Standards Board, “Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Network (MAN) - 802.6”, Dec. 1990.
- [8] Joan Triay and Cristina Cervelló-Pastor, “QoS Multi-Spanning Tree for Optical Burst Switching”, *Proceedings of the 2008 4th International Telecommunication Networking Workshop on QoS in Multiservice IP Networks (IT-NEWS)*, Feb. 2008, Venice (Italy).
- [9] E. Kehayas, K. Vyrosokinos, L. uStampoulidis, K. Christodouloupolos, K. Vlachos, H. Avramopoulos, “ARTEMIS: 40Gb/s All-optical Self-Routing Node and Network Architecture Employing Asynchronous Bit and Packet-Level Optical Signal Processing”, *Journal of Lightwave Technology*, vol. 24, no. 8, pp. 2967-77, Aug. 2006.
- [10] University of Hamburg, Department of Computer Science, “DESMO-J (Java): A Framework for Discrete-Event Modelling and Simulation”, Available online at <http://www.desmoj.de>.
- [11] Ahmad Rostami and Adam Wolisz, “Modelling and Synthesis of Traffic in Optical Burst-Switched Networks”, *Journal of Lightwave Technology*, vol. 25, no. 10, pp. 2942-52, Oct. 2007.
- [12] Y.-L. Hsueh, J. Kim, C.-F. Su, R. Rabbat, T. Hamada, C. Tian and L.G. Kazovsky, “Traffic Grooming on WDM Rings Using Optical Burst Transport”, *Journal of Lightwave Technology*, vol. 24, no. 1, pp. 44-53, Jan. 2006.
- [13] J. Teng and G.N. Rouskas, “Wavelength Selection in OBS Networks using Traffic Engineering and Priority-based Concepts”, *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 8, pp. 1658-69, Aug. 2005.