Performance Evaluation of the DBORN MAC Layer*

Guoqiang Hu, Christoph M. Gauger, Sascha Junghans
University of Stuttgart, Institute of Communication Networks and Computer Engineering (IKR)
Pfaffenwaldring 47, 70569 Stuttgart, Germany
e-mail: {hu, gauger, junghans}@ikr.uni-stuttgart.de

Abstract

DBORN (Dual Bus Optical Ring Network) is a new network architecture for wavelength division multiplexing metropolitan area networks. In DBORN, metro clients like enterprise networks are connected to the optical metro-ring by edge nodes which share the wavelength channels on the ring. In order to allow for a high-speed optical interface on the upstream wavelengths while avoiding active optical switching components, a carrier sense multiple access/collision avoidance (CSMA/CA) medium access control (MAC) protocol is realized. In this paper, the MAC protocol and the closely related transmitter interface are described first. Then, the performance of the MAC protocol is evaluated with respect to important system parameters like slotted/unslotted working mode, burst size distribution and node position in terms of the mean waiting time as well as the node-to-hub delay. Finally, buffer dimensioning for the transmitter and the performance impact of a bottleneck interconnection link in the local edge node are analyzed.

1 Introduction

Due to new broadband access technologies and the increasing number of Internet users as well as due to the trend of enterprise networking, demands for higher capacity metropolitan area networks (MAN) are rising. After the fast growing networking capacity of the last years, internet providers today ask for equipment with higher bandwidth and lower costs. The demand for high bandwidth leads to optical solutions with wavelength division multiplexing (WDM) technology. As active optical switching elements are still a high cost factor, it is desired to provide an optical network architecture without active optical switching elements. The DBORN (Dual Bus Optical Ring Network) architecture satisfies these demands with a new MAN technology. It connects metro edge nodes, which do not employ any optical active switching elements, to the core network via a hub node. As this cost effective solution leads to some constraints in the area of medium access, a new medium access control (MAC) protocol and interface card design is required. In this paper, we present a detailed performance evaluation of the MAC protocol of this new network architecture.

In Section 2 we will introduce the network architecture, its MAC protocol and the transmitter interface in edge nodes. Section 3 presents the results of the performance evaluation of the MAC protocol and Section 4 inspects the buffer dimensioning for the transmitter as well as the performance impact of the bottleneck interconnection link in the local edge node. In Section 5 conclusions are drawn and further work is outlined.

2 Network architecture and MAC protocol

DBORN is a high speed network solution for metropolitan areas [7]. On the basis of advances in the optical transmitter and receiver technology [3], the carrier sense multiple access with collision avoidance (CSMA/CA) is realized in DBORN.

2.1 Network architecture

DBORN is an optical metro ring architecture connecting several edge nodes, e. g., metro clients like enterprise, campus or local area networks (LAN), to a regional or core network. The ring consists of two parallel fibers called working and protection fiber in order to provide resilience in case of single link failures. Each ring employs WDM and carries a set of wavelengths which are further classified into downstream and upstream wavelength channels (**Fig. 1**). While downstream wavelength channels start from the transmitters in the hub, upstream wavelength channels are terminated by the receivers in the hub. In our studies, the bandwidth of the metro client interfaces is set to 1Gbps. The metro ring itself is studied for 2.5Gbps and for 10Gbps.

Several edge nodes share upstream and downstream channels respectively in asynchronous time division multiplexing. For load balancing purposes, an edge node can be attached to more than one upstream or downstream channel. In order to keep the edge node interface cards as simple as possible, all traffic—external and intra-ring—has to pass the hub. Specifically, no edge node receives or even

* This work was funded within MultiTeraNet project BurstNet by the German Bundesministerium für Bildung und Forschung.

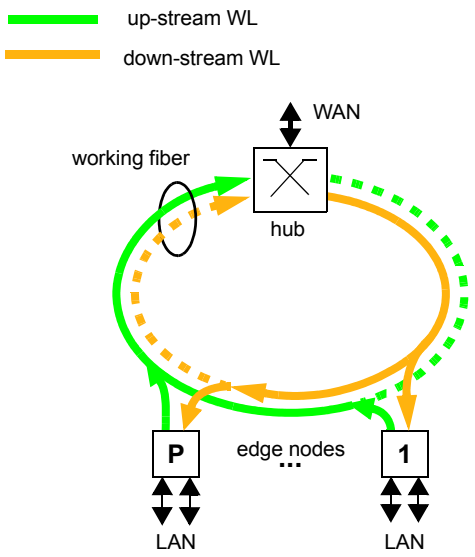

Fig. 1 DBORN architecture

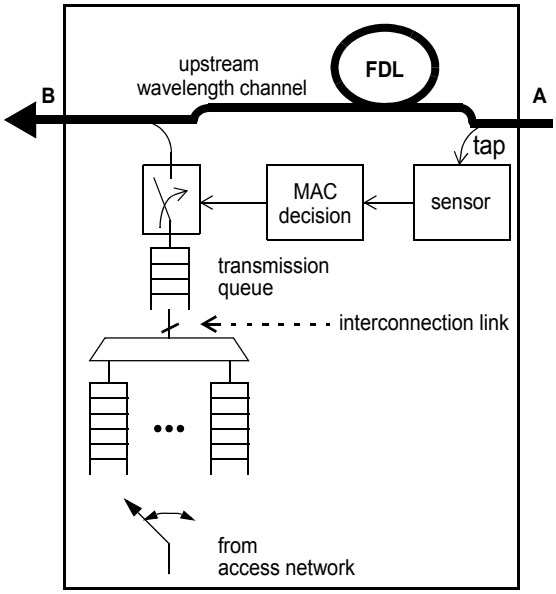

Fig. 2 Functional model of transmitter interface in edge nodes

removes traffic on upstream channels or inserts traffic on downstream channels. Thus, both upstream and downstream channels can be modelled as shared unidirectional buses.

As the hub node exclusively transmits on the downstream channel, traditional scheduling mechanisms can be applied here. However, medium access of edge nodes has to be controlled on the upstream channel which will be inspected in depth in Section 3.

2.2 Burst size and burst assembly

In order to provide for safe transmitting and receiving on the ring a guard time has to be inserted between consecutive optical transmission units. A typical value of the guard time with current technologies is $50\,ns$ [3], which on a 10Gbps channel corresponds to the transmission time of about 63 bytes.

DBORN targets transaction data and Internet traffic which is commonly transported over Ethernet, i. e. client layer packet sizes are in the range of 40 to 1500 bytes [9] bounded by the Ethernet maximum transmission unit (MTU). As transmission of individual client layer packets/frames would lead to a significant overhead due to guard times, all client layer traffic is assembled into larger units called *bursts* for transmission on the optical ring. A considerable amount of literature on burst assembly is available in the context of optical burst switching (e.g. [4][5][8]).

In the current version of the DBORN prototype the optical ring employs the Ethernet frame as burst format. Thus, the small MTU value only allows a limited degree of assembly gain. In future versions, this could be improved by segmentation of client layer packets [1] or by selection of a different optical layer burst format, e. g. ITU-T's G.709 frame format with a size of about 16K bytes [6]. As this paper focuses on MAC performance, we do not consider the effects of (suboptimal) burst assembly and use a maximum burst size of 16K bytes in our studies.

2.3 MAC protocol

As DBORN targets a cost efficient optical ring solution no active optical components, e. g. switches, are used on the interface cards and transmitting and receiving part are strictly separated. **Fig. 2** depicts a functional model of the transmitter interface, which was designed to allow a collision-free medium access.

Between the input (point A) and the output (point B) of the edge node a fiber delay line (FDL) is inserted into the ring. The length of the FDL should correspond to a delay equal to or greater than the transmission time of the maximum burst size. At the input (point A) of the edge node, a simple sensor taps the upstream channel and constantly monitors the channel status—busy or idle. On the other side of the FDL a laser is coupled into the same channel and controlled by the decision unit to send bursts safely. Due to the delay introduced by the FDL, the edge node can determine the duration of voids on the channel up to the FDL delay before they pass the coupling point of the laser and thus decide on the medium access avoiding collisions.

There are two possible operation modes for DBORN: slotted and unslotted. In the slotted mode, the channel is divided into constant duration slots and the transmission is

allowed if the edge node finds an idle slot on the upstream channel. On the one hand, this requires some basic synchronization between network nodes, on the other hand edge nodes only have to check whether a slot is idle or used.

In the unslotted mode, no synchronization is required and bursts can have an arbitrary transmission time up to the FDL delay. By comparing the duration of an available void on the channel and the transmission time of the first burst in the transmission queue the edge node can decide when to transmit a burst.

2.4 Architecture of edge node

The edge nodes provide the functionalities of data assembly, scheduling and medium access in upstream direction and reading medium access and data disassembly for the downstream path. For the assembly mechanism, data to different destinations and data of different QoS classes must be buffered in a memory until a complete burst can be sent out. For this purpose, a large buffer is needed in the edge node.

As the bandwidth of the memory must be at least the bandwidth of the fastest link, a large high speed memory device would lead to high costs. As the link rate to the DBORN ring is significantly higher than the rate of the client networks, it would be preferable, to use a small but very fast transmission buffer for serving the high link rate and a large but slower and cheaper memory for the assembly queues. (**Fig. 2**).

The minimum size of the fast transmission buffer is the maximum length of a single burst. If a gap on the ring is detected, this burst in the transmission buffer can be sent out. For this minimum size buffer, the system must wait after transmission until the transmission buffer is filled up again with data from the assembly queues. During this time no transmission is possible, even when larger gaps appear on the ring. Therefore, it would be desirable to use longer busy times on the ring for filling more than one burst in the transmission buffer and sending out these bursts in larger idle times.

In Section 4, we will discuss the ideal size of the transmission buffer from the point of view of system performance.

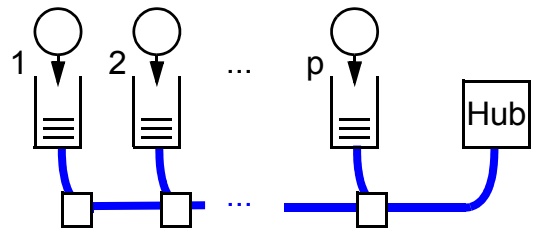

Fig. 3 System model of MAC protocol

3 Performance evaluation of the MAC protocol

In this section, performance evaluation will be concentrated on the MAC protocol on the upstream channel. Slotted and unslotted mode will be compared and the impact of the burst size distribution will be evaluated. Finally, node-to-hub delay performance of DBORN are assessed.

The system model for the upstream channel is illustrated in **Fig. 3**. Two scenarios are considered: a system with 10 edge nodes attached to a single 10 Gbps upstream channel and a system with 5 edge nodes on a 2.5 Gbps channel. Since the results are quite similar in these two cases, only the results from the former scenario are presented in Section 3. Homogenous traffic is uniformly distributed over all edge nodes and bursts arrive according to a Poisson process. For unslotted mode both fixed burst size and variable burst size are considered while for slotted mode only the fixed size is treated. As motivated in Section 2.2 we set the burst size to 16K bytes for the fixed size case. In the variable size case, we use independent discrete uniform distributions in order to systematically cover a broad spectrum of burst size variability in the presence of a fixed upper bound of 16K bytes. For illustration, a 16K byte burst has a transmission duration of 12.8 μs. The term load always refers to the ratio of average traffic bitrate and channel capacity. In all graphs, mean waiting time is normalized by the mean burst transmission time.

3.1 Comparison of slotted and unslotted mode

Mean waiting times for slotted and unslotted mode with fixed burst size are compared in **Fig. 4**. It can be observed that in both operation modes the mean waiting time is in the order of only 1 to 20 mean burst transmission times, i. e., less than 0.25 ms. Also, downstream nodes experience a larger delay due to the intrinsic priority property of the DBORN MAC protocol, i. e. , an edge node can only make use of bandwidth (voids) on the channel which was left over by other nodes located further upstream. However, at small and medium load levels, the unfairness between the edge nodes is not really prominent. Depending on load the unslotted mode yields lower waiting times for upstream nodes up to a certain ring location. This can be explained by the residual slot lifetime at arrival. In case of high load this effect diminishes and the slotted mode outperforms the unslotted mode. Because in unslotted mode the edge nodes send asynchronously, there may be voids becoming too small to be filled by any of the bursts, so called channel fragmentation. This leads to significant performance degradation at high load level.

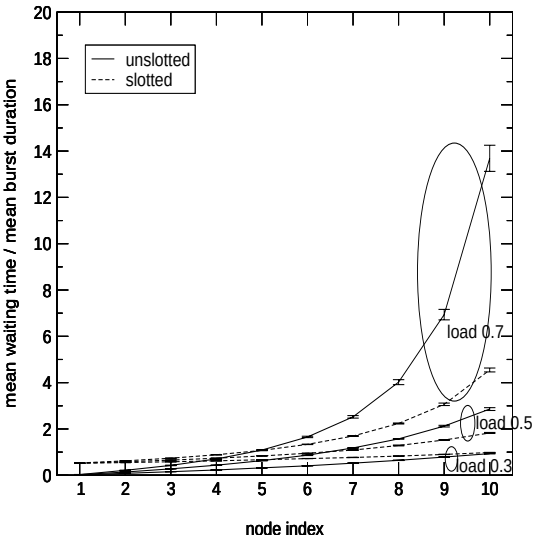

Fig. 4 Mean waiting time wrt. node position

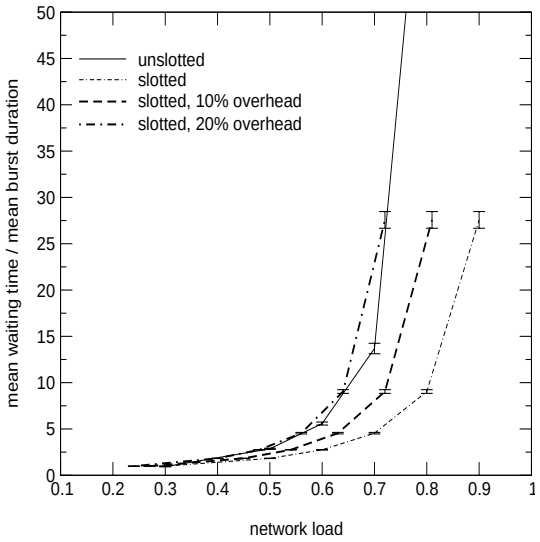

Fig. 5 Mean waiting time in node #10 wrt. network load

In **Fig. 5**, the mean waiting time of the 10th edge node, which has the worst waiting time performance, is observed regarding different network loads. The performance gap between slotted and unslotted opens increasingly with the network load. The high sensitivity of the unslotted mode to the high load is closely related to its non-work-conserving property, i. e., the channel bandwidth is not fully utilized due to channel fragmentation.

However, in practice slotted operation implies that not all bursts are perfectly filled (c.f. Section 2.2) which introduces an overhead to the slotted case not considered so far. To make this comparison more accurate curves with a 80 % and 90 % filling efficiency are included in **Fig. 5**— the load is increased respectively. It can be clearly seen that the difference between unslotted operation and slotted operation with even 80 % filling efficiency is marginal. Consequently, burst assembly for fixed size bursts has to yield very high filling degrees.

3.2 Impact of burst size distribution for unslotted operation

The mean waiting time for the unslotted mode with variable burst size is drawn in **Fig. 6**. Three discrete uniform distributions are applied with ranges for the sample values of [11276, 16000], [5058, 16000] and [1150, 16000] respectively. They are selected to systematically analyse the impact of increasing the coefficient of variation c moving from the fixed size case ($c = 0$) to $c \approx 0.1, 0.3, 0.5$. Note that as burst size is limited, so is its variability.

It can be seen from the graph that a higher variability in the burst size results in an increased waiting time. However, this impact is small compared to the impact of load or node position on the ring. Thus, we restrict our following evaluations to the case with fixed burst size.

3.3 Node-to-hub delay between edge node and hub

While the MAC protocol introduces a clear unfairness with respect to mean waiting time for downstream nodes, these nodes have the advantage of a small propagation delay towards the hub. In order to consider both effects we evaluate the node-to-hub delay in the ring. It comprises the waiting time for transmission in the edge nodes as well as the propagation delay to the hub. The scenario is a reference metro ring with a total length of 120 km to which 11 equidistant nodes (10 edge nodes and 1 hub) are attached.

Fig. 7 depicts the node-to-hub delay for the most upstream and most downstream node and for both slotted and unslotted mode with fixed burst size versus the load. It can be observed that the delay of edge node 1 is insensitive to the load and equals the constant propagation delay. In contrast, the node-to-hub delay of edge node 10 is dominated by the waiting time and thus the load. Node 1 and 10 have the same node-to-hub delay at a load greater than 0.9 for slotted mode and around 0.75 for unslotted mode. This indicates that DBORN can operate even at high load without worrying about the fairness regarding mean node-to-hub delay.

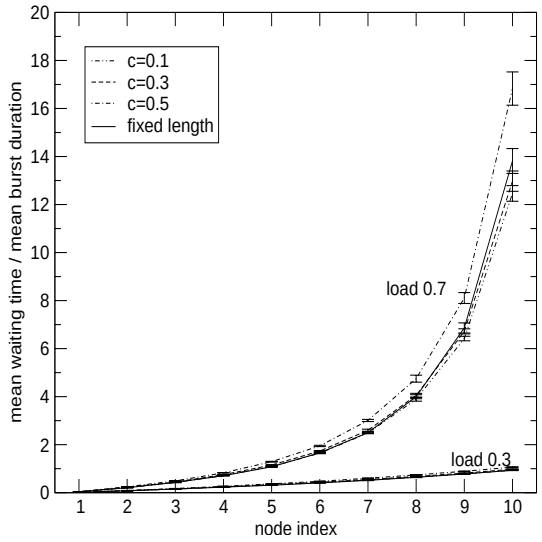

Fig. 6 Impact of variable burst sizep

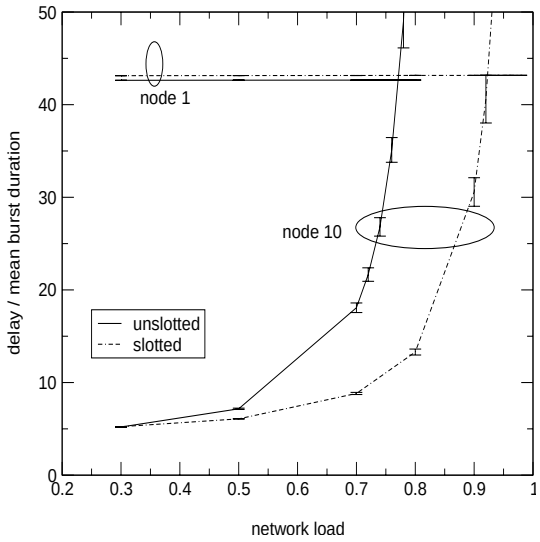

Fig. 7 Node-to-hub delay

4 Performance issues regarding the transmission buffer

In this section the observed system model is extended to include the transmission buffer in each edge node. The corresponding system model is shown in **Fig. 8**. The traffic from the LAN is sent to the transmission buffer (buffer B) through a 1Gbps interconnection link. To avoid the burst loss due to buffer overflow, buffer B is able to signal buffer A to block the transmission in case buffer B is full, which is the so called backpressure mechanism. When later the content of the buffer B decreases below a threshold and B is ready to accept more bursts, it sets another signal to activate the transmission again.

As mentioned in Section 2.4, the transmission buffer dimensioning is critical in the system development because of the high cost of the high-speed buffer. This issue will be checked in Section 4.1 with respect to the mean transition time (**Fig. 8**). In Section 4.2 a special performance degradation caused by the bit rate mismatch between the 1Gbps interconnection link and upstream channel is shown for the system with a 2.5Gbps upstream wavelength.

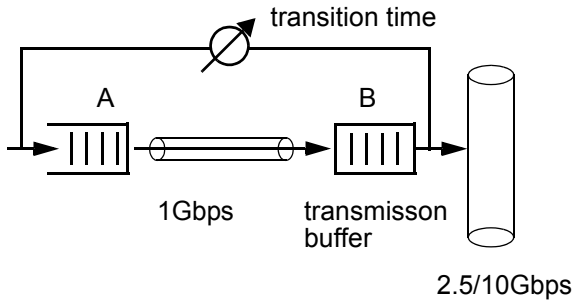

Fig. 8 Singe edge node in extended model

The traffic model applied in the simulation is the same as that of Section 3 except that here only the fixed burst size is looked at. The buffer A is set to be unbounded. Interesting system parameters include the effective size of buffer B and the threshold of buffer B at which it unblocks the transmission of buffer A. The threshold is quantified by the ratio of the queueing length to the buffer size.

4.1 Transmission buffer dimensioning

As long as the transmission buffer is larger than the maximal size of one burst it does not affect the performance of the upstream nodes very much since there is sufficient free bandwidth left on the channel and the transmission buffer is seldom filled due to the bottleneck of 1Gbps link. On the other hand, the downstream nodes face the problem of only few available channel bandwidth. Under heavy network load, it is important for them to take advantage of each suitable void on the channel to achieve an acceptable performance. This means they are supposed to keep a number of bursts in the transmission buffer in order to fill the coming voids, instead of spending the time in fetching bursts from buffer A and missing the valuable voids. Therefore, downstream nodes should have a larger demand of transmission buffer than the upstream nodes. So in the following attention will be drawn on the last node on the upstream channel.

In **Fig. 9** the normalized mean burst transition time in this last node is plotted with respect to the transmission buffer size for different threshold levels (0.5, 0.75, 1). Again, we look at the 10 node 10 Gbps scenario. Note that the buffer

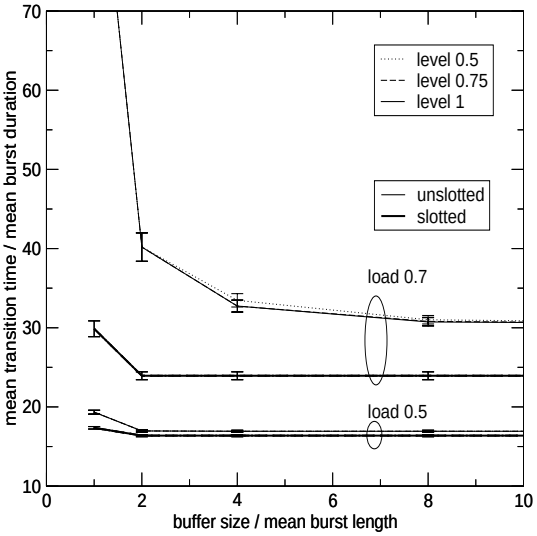

Fig. 9 Transition time in node #10 wrt. transmission buffer size

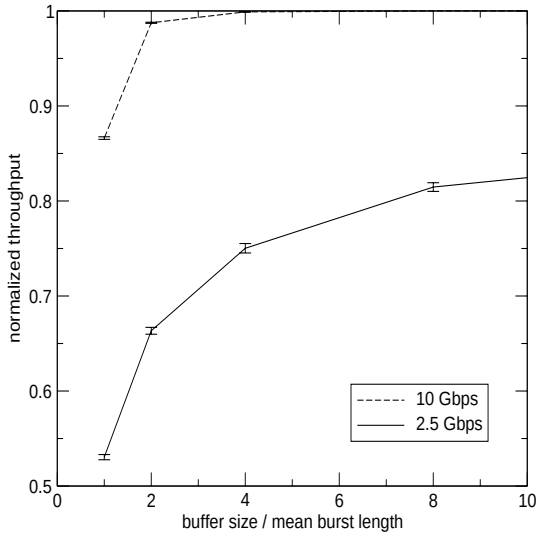

Fig. 10 Throughput of the most downstream node wrt. transmission buffer size

size here is normalized by the mean burst size. It can be seen that the performance is not sensitive to the buffer size when the load is not too high, since there is enough free bandwidth on the channel. The slotted mode benefits less from large buffer size in comparison to the unslotted mode. In the high load situation (0.7) and the unslotted mode, large performance gains are achieved through increasing the buffer size. However, even in this case increasing buffer size is not effective any more when the buffer size reaches 4 times of the mean burst size. In general, it can be figured out that a buffer size of 4~8 times of the mean burst size is enough for the transmission buffer.

Another significant observation from **Fig. 9** is that the threshold level of the backpressure mechanism does not have obvious influence on the mean burst transition time. Note that for a very small buffer size, e. g. one burst, the different threshold levels are equivalent so their curves overlap completely..

4.2 Impact on the throughput

In this section we show that the application of 2.5 Gbps upstream channel can cause some special performance degradation. For comparison both scenarios are taken into account: 10 edge nodes on one 10 Gbps channel and 5 edge nodes on one 2.5Gbps channel. In the former case the total input traffic intensity to the 9 upstream nodes are set to 0.5, i. e., 50% of the total channel bandwidth (5 Gbps) is available for node 10 not considering the bandwidth fragmentation. The node 10 is fed with a greedy source. Similarly, in the latter case the 4 upstream nodes have also a total traffic intensity of 0.5 so that they leave 1.25 Gbps for node 5 which has a greedy traffic source. Due to the limitation of the 1 Gbps interconnection link, however, the maximal throughput of the last downstream node in either case can not exceed 1 Gbps. The buffer threshold level is set to 0.5 for both cases.

In **Fig. 10**, the achievable throughput normalized by the ideal throughput (1 Gbps) is plotted for node 10 of the 10 Gbps scenario and node 5 of the 2.5 Gbps scenario. It can be seen that in the 10Gbps case the ideal throughput is comfortably obtained as long as the normalized transmission buffer size is greater than 4, whereas in the 2.5 Gbps case the ideal throughput cannot be reached even with large transmission buffer.

This observation is explained by observing node 1 of the 2.5 Gbps scenario as an example. Suppose there are more than two bursts in buffer A and the transmission buffer is large. Because no one competes with node 1 for channel access, a burst that arrives at the transmission buffer is immediately sent onto the channel and has a duration of L equal to the transmission time, as illustrated in **Fig. 11**. However, due to the 1 Gbps interconnection link, the next burst will not be completely received by the transmission buffer until $2.5L$ after the previous burst. As a result, the inter-departure time of the two bursts is $2.5L$ and leaves a void of $1.5L$ of which $0.5L$ is definitely not usable for the

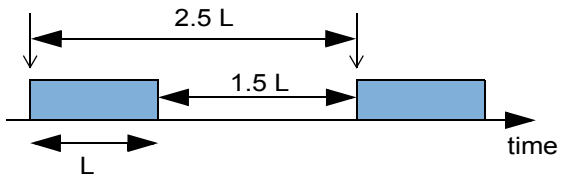

Fig. 11 Burst interval on the 2.5Gbps channel

downstream nodes. Therefore, in 2.5 Gbps scenario with unslotted operation and fixed burst length there is additional loss of usable bandwidth, which does not occur in the 10 Gbps scenario. In the system development, special measures must be taken to cope with it.

5 Conclusion and outlook

This paper analyzes the performance issues of the CSMA-CA MAC protocol in DBORN. The impact of key design parameters on the performance is studied. Slotted mode outperforms unslotted mode in high load situations. However, this advantages can be reduced by inefficient filling of fixed size bursts. For unslotted mode, variability in burst size shows only little influence on the mean waiting time, which is another credit for introducing variable size assembly. We also discuss the node-to-hub delay and show that the different propagation delays of edge nodes at different positions can balance the unfairness in the mean waiting time to a large degree. Furthermore, the buffer dimensioning for the transmitter of the edge node is inspected, which is a critical factor in system design and closely coupled with the MAC protocol. We show that only small transmission buffers are required to assure the performance. At last, the problem of bitrate mismatch between the wavelength channel and interconnection link is disclosed in the case of 2.5 Gbps channel adopted.

Our future work will be focused on the design and evaluation of fairness mechanisms for heavy load and overload situation with reference to the available work [2]. Also, we will study an extended system scenario including the traffic assembler, multiple channels and scheduling in edge nodes and the hub node.

6 References

[1] T. Atmaca, D. Popa: “Logical performance of the optical packet metropolitan ring architecture.” *Proceedings of 17th International Symposium on Computer and Information Sciences (ISCIS)*, Orlando, Oct. 28-30, 2002, pp. 98-102

[2] N. Bouabdallah, L. Ciavaglia, E. Dotaro, N.L. Sauze: “Matching fairness and performance by preventive traffic control in optical multiple access networks.” *Optical Networking and Communications (OptiComm 2003)*, Dallas, Oct. 13-17, 2003, pp. 424-429.

[3] G. Eilenberger, L. Dembeck, W. Lautenschläger, J. Wolde: “Optische Burst-Techniken in Metronetzen.” *ITG-Fachtagung Photonische Netze 2002*, pp. 13-18

[4] C.M. Gauger: “Trends in Optical Burst Switching.” *Proceedings of the SPIE ITCom Conference (ITCom 2003)*, Orlando/FL, September 2003

[5] G. Hu, K. Dolzer, C.M. Gauger: “Does burst assembly really reduce the self-similarity.” *Optical Fiber Communication Conference (OFC 2003)*, Atlanta, Mar. 2003, pp. 124-126.

[6] ITU-T Recommendations on the OTN Transport Plane, G.709

[7] N.L. Sauze, A. Dupas, E. Dotaro, L. Ciavaglia, M.H.M Nizam, A. Ge, L. Dembeck: “A novel, low cost optical packet metropolitan ring architecture.” *European Conference on Optical Communication (ECOC 2001)*, Amsterdam, Oct. 2001.

[8] X. Yu, Y. Chen, C. Qiao: “A study of traffic statistics of assembled burst traffic in optical burst switched networks.” *Opticomm 2002*, pp. 149-159

[9] http://www.caida.org/analysis/AIX/plen_hist/