Access protocols to support different service classes in an optical burst switching ring *

Vishwas S Puttasubbappa and Harry G Perros {vsputtas,hp}@csc.ncsu.edu

> North Carolina State University Computer Science Department Raleigh NC 27695-7534

Abstract. Several access protocols are proposed to support different service classes in an optical burst switched ring. Their performance is evaluated through simulation. Various performance metrics such as throughput, utilization, burst loss rate, end-to-end delay and fairness are used to analyze the behaviour of each protocol.

Keywords - Optical burst switching, MAN, access protocols, service classes

1 Introduction

Optical Burst Switching (OBS) is a novel method currently under study that can be used to transport data over a Wavelength Division Multiplexing (WDM) optical network. Battestelli and Perros [1] provide a detailed survey on OBS and its variations. There is not much work done in the field of OBS over metropolitanarea rings. Xu et al [6] investigated access protocols for OBS rings based on the Just Enough Time (JET) scheme and a new scheme called the Only Destination Delay (ODD). Jong [4] proposed several access protocols for multicasting in such an environment. A new architecture called the LightRing has been proposed by Fumagalli and Krishnamoorthy [3] with multi-token protocol to prevent contention among bursts. Bouabdallah et al [2] proposed a collision avoidance MAC protocol for a metropolitan bus-based optical access network. Analytical models were developed to calculate the mean access delay of each node in such a shared-medium system. Fairness issues were also investigated.

The work done so far on OBS rings considered traffic to be best effort except in Fumagalli and Krishnamoorthy [3], where real-time and best-effort were considered. The aim of this paper is to investigate how an OBS ring can support different classes of traffic. In this study, we extend the OBS ring architecture proposed in Xu et al [6] in order to consider the following three different classes of traffic. The first class of traffic (Class 1) is a variable bit rate traffic with stringent end-to-end delay constraints, the second class (Class 2) is variable bit

 $^{^{\}ast}$ This work was supported by ARDA under contracts MDA904-00-C-2133 and MDA904-02-C-0482

rate with no delay constraints, and the third class (Class 3) is non-real time variable bit rate best effort traffic. Several access protocols are proposed and their performance evaluated through simulation.

The paper is organized as follows: Section 2 presents the system architecture including the structure of the metro ring and the OBS nodes. The proposed protocols are presented in section 3, and in section 4 we describe the arrival processes used in the simulation model. The results of the simulation are discussed in 5. Finally, section 6 gives the conclusions from the study.

2 System Architecture

The metro ring uses the WDM architecture and it consists of several nodes which serve as concentration points to incoming and outgoing traffic from several access networks. The ring is divided into two co-existing rings as follows: Some of the wavelengths are used to host SONET/SDH rings, and the remaining wavelengths are used for optical burst switching. The metro ring has to carry all types of traffic, such as circuit switched traffic, ATM traffic and IP traffic. SONET/SDH rings can cater to circuit switched traffic and data traffic can be transported through the OBS ring. In this paper, we only investigate the OBS part of the metro ring.

The wavelengths allocated for optical burst switching are divided into S sets of N+1 wavelengths. Within a set of wavelengths, each of the N wavelengths is allocated to a different node. This wavelength is referred as the *home* wavelength of the node. The (N+1)th wavelength of the set is the control channel for the wavelengths in the set and it carries control frames. The control frames implement the signalling necessary for OBS. Since there are S sets of wavelengths, each node is allocated to S home wavelengths. A node can only transmit bursts on its home wavelengths.

Each OBS node in the ring has S transmitters each fixed-tuned to one of the S home wavelengths, and S tunable receivers one per wavelength set. These S pairs of transceivers are used for transmitting and receiving bursts. A node can transmit a burst on any free home wavelength. A free receiver can tune to receive a burst arriving on any wavelength in its corresponding wavelength set. Each node is additionally equipped with S transceivers, one set per control wavelength. The OBS node is equipped with a control module which performs its functions based on the information each control frame carries around the ring. Each control wavelength carries back to back control frames. The structure of a control frame is as shown in figure 1. Each node has its own slot into which it can write information during transmission. The control frames on the S control wavelengths travel around the ring in a synchronous manner. That is, the control frames in the (i+1)st control wavelength lag behind their corresponding ones in the ith control wavelength by the time the control module requires to process them. This arrangement ensures efficient usage of control frames for burst transmission. For instance, if a node cannot transmit on the first control channel, it has an opportunity to transmit immediately using the control frame



Fig. 1. Control Frame structure

in the second control channel without having to wait even for a small amount of time.

Each node serves a number of access networks. The incoming data from these networks is queued in the transmission queues of a node. Specifically, each node maintains N-1 transmission queues of each class, where N is the total number of nodes in the ring. Since we consider three classes, each node maintains 3(N-1) queues. In this paper we assume that class i has non-preemptive priority over class i+1, i=1,2. That is, the transmitter will always transmit a burst from a class 1 queue. If there is no class 1 traffic, it will transmit a burst from class 3. The N-1 class 1 transmission queues are assumed to form a single logical queue. Bursts in these N-1 class 1 queue. The N-1 class 2 transmission queues are served in the order in which they arrive at this logical queue. The N-1 class 2 transmission queues are served in a roundrobin fashion. Likewise, round robin is used to serve the N-1 class 3 transmission queues.

3 The Access Protocols

In this paper, we assume that the class 1 traffic consists of multiple HDTV streams. Each HDTV frame constitutes a single burst. In the case of class 2 and class 3 traffic, a burst is comprised of several data packets which may be IP packets, ATM cells etc. A class 2 and class 3 transmission queue is eligible to be served if there are enough packets in the queue whose aggregate size exceeds a minimum burst size.

We have defined and analyzed the following five protocols. *Destination*-*Reservation Free* which provides no guaranteed delivery to any traffic class, *Ack* and *Token* which provide guaranteed delivery to class 1 traffic and *Token*-*Token* and *Ack-Ack* which provide guaranteed service to both class 1 and class 2 traffic. All these protocols transmit class 3 bursts when bandwidth is available. These access protocols fall into two categories: *collision-free* and *collision* protocols. Collision-free protocols reserve resources at the destination and hence no burst loss at the destination and in the ring. None of these five protocols are collision-free for class 3 traffic. Token-Token and Ack-Ack are class 1 and class 2 collision-free. Token-Token uses tokens and Ack-Ack makes use of acknowledgements. Ack and Token are class 1 collision-free protocols. Each of the protocols is explained in detail in the following sub-sections.

3.1 Destination-Reservation Free Protocol (Dest-Resv-Free)

Nodes transmit bursts without making any reservations at the receiver node. This may result in multiple bursts arriving at the destination at the same time and hence collisions. This is the Tell-and-Go protocol currently used in OBS mesh networks.

Single transceiver case: Upon arrival of a control frame, the bursts are transmitted with the priority scheme as stated in section 2 in case the transmitter is not busy. On multiple bursts arriving at the same time, priority for reception is given to class 1 bursts. If multiple class 1 bursts arrive at the same time, one of them is randomly selected. There is pre-emption of class 2 and class 3 bursts on arrival of a class 1 burst. Class 2 bursts are given the next priority and in case many of them arrive at the same time, one of them is randomly chosen. Class 3 bursts are given the least priority and if many of them arrive at the same time, one of them is randomly selected.

Multiple transceivers case: In the case of multiple home wavelengths, the control module can choose any free home wavelength to transmit. The reception mechanism is identical in each of the different sets of wavelengths.

3.2 Token Protocol

The Token protocol uses the concept of tokens to resolve receiver collisions. Tokens are used only for class 1 bursts. Class 2 and class 3 is serviced through best-effort. Every node has a token circulating around the ring. If a source wants to transmit a class 1 burst to a particular destination, it has to have the token for that destination. All the nodes maintain a queue to hold tokens. The token is released after the transmission is completed. Since only the node that has possession of the token can transmit a burst to the appropriate destination, the destination can only receive a single burst at a time, and therefore the token protocol is a collision-free protocol, but only for class 1 traffic.

Single transceiver case: Each node monitors the control frames for tokens. If a control frame is carrying a token for destination k, then the flag field of slot k has the value 1. The node takes the token out of the control frame and queues it in its FIFO queue, provided that there is traffic for that destination node. The token has to be queued because the node may not be able to transmit the burst immediately because of busy transmitter. The token is released after the transmission is completed. This guarantees no collision at the destination between class 1 bursts. Class 2 and class 3 bursts do not have a collision-free reception mechanism. Bursts are simply transmitted following the Dest-Resv-Free protocol. On the receiver's side, class 1 bursts are given the highest priority. Since this is a class 1 collision-free protocol, there are no collisions among class 1 bursts. Priority of reception is similar to the Dest-Resv-Free protocol.

Multiple transceivers case: Each node maintains a separate token queue for each home wavelength. A node cannot use more than one home wavelength simultaneously to transmit bursts to a particular destination. This is achieved by making sure that only one token to a particular destination can be held in any of the multiple token queues. A class 1 burst can be transmitted on any home wavelength as long as the corresponding token queue has the token for the appropriate destination. Reception in any one set of wavelengths is independent and identical to the other sets.

3.3 Ack Protocol

This protocol also ensures guaranteed reception of class 1 bursts. It is different to the Token protocol and is based on the Tell-And-Wait (TAW) protocol proposed in OBS networks. The protocol uses a *Request* and *Acknowledgement* mechanism. A minimum of a round-trip delay is required for the node to transmit a burst after it can be formed. The acknowledgement mechanism is available only for class 1 traffic. Class 2 and class 3 bursts are transmitted as in the case of the Dest-Resv-Free protocol.

Single transceiver case: As soon as a HDTV frame arrives at any of the class 1 queues of the node, a request is sent out to the destination node requesting it to return an acknowledgement in which it indicates the earliest time it is free to receive this burst. The source is not allowed to send out a request to any destination as long as it has an outstanding request. This makes sure that there are no transmitter conflicts. The flag field of the control frame is used to indicate a *Request*. When the destination node sends an acknowledgement, the offset field in the destination's slot will be the earliest time by which the source can start transmitting. This is to make sure that by the time the source starts transmitting the burst, the destination is free to receive it. The transmission of class 2 and class 3 bursts is without any acknowledgement mechanism and is similar to the Dest-Resv-Free protocol. A destination node cannot send out an acknowledgement for a request that it has received until it starts receiving the burst from the previous acknowledgement that it sent. This is necessary because the source node to which the last acknowledgement was sent may not be able to transmit the burst immediately because of a busy transmitter. Reception is similar to Dest-Resv-Free protocol. The receiver has a queue to hold requests and the request with the earliest time-stamp will be served first. Class 2 bursts are given higher priority than class 3 bursts.

Multiple transceivers case: While sending a request on any of the control channels, the source node makes sure that no other home wavelength is being used for the same destination. If this is the case, other transmission queues are served which send out requests to different destinations. Reception is similar to the Token protocol.

3.4 Token-Token Protocol

This is a collision free protocol for class 1 and class 2 bursts. Nodes use the token mechanism.

Single transceiver case: The token for a particular destination is captured by the node only if it has bursts of class 1 or class 2 to be sent to that particular destination node. Priority for transmission is given for class 1 traffic. Unlike the token protocol, class 2 bursts require a token for transmission. Class 3 queues are served whenever the transmitter is free and the bursts can be formed. Reception is simpler in this case since there is no collision between class 1 and class 2 bursts and hence no priority between them. Class 3 bursts are subject to pre-emption in case either of the other classes' burst arrives.

Multiple transceivers case: The operation is similar to that of the Token protocol except that tokens are required for classes 1 and 2.

3.5 Ack-Ack Protocol

Guaranteed reception both for class 1 and class 2 bursts is provided by extending the acknowledgement scheme to cover both traffic classes. Class 3 is served through best effort.

Single transceivers case: The operation of the transmitter is similar to the Ack protocol, except that class 2 bursts need an acknowledgement before transmission. A request to transmit a class 2 burst is sent only if all the class 1 transmission queues are empty. Thus, this is a collision-free protocol for class 1 and class 2. Class 3 bursts are transmitted whenever the transmitter is free. The receiver exercises its decision to send out acknowledgements not in the FCFS manner, but based on priority. Once the receiver starts receiving a burst for the acknowledgement it last sent, it scans all the requests in its queue and sends out the next acknowledgement according to the following rules: (1). Requests which are for class 1 traffic are given priority. If there are none, then one of the class 2 requests is arbitrarily picked and served (2). If there are multiple class 1 requests, then the request that has the earliest time-stamp will be served first. Rule 2 makes sure that requests are served not in their order of arrival (which may be biased towards nodes close by) but by the earliest time-wise arrival of frames at their respective source nodes. Reception is similar to Token-Token protocol.

Multiple transceivers case: The operation is similar to that of the Ack protocol except that acknowledgements are required for classes 1 and 2.

4 The Simulation Model

An event-based simulation model was developed with a view to analyzing the performance of the proposed five protocols. For details, see Puttasubbappa and Perros [7].

For each node i, i=1,2,...,N, a number of HDTV streams are setup at the beginning of the simulation. Each of these streams originate at node i and terminate at destination node j. In each stream, frames are generated at a rate of 60 frames per second giving an inter-frame arrival time of 16.667 milliseconds. We assume that the frames follow the MPEG 2 Group Of Pictures (GOP) structure of IBB PBB PBB PBB. The size of each frame is generated using the auto-regressive model, see Bragg [9], $S(t) - S(t-12) = e(t) - 0.69748 \times e(t-3)$, where S(t) is the size of frame t, and $e(t) \sim N(0,\sigma^2)$ with σ^2 =4849.5.

A class 2 source in our simulation is a variable bit rate source with no end-toend time constraints. In our simulation experiments, we assume that the packets are generated from a storage area network (SAN), with the following packet-size distribution: 44 % of 64Kbytes, 18 % of 56K, 21 % of 40K, 4 % of 32K, 4 % of 24K and 6 % of 8Kbytes, see Trevitt [8]. The arrival process consists of packets arriving in succession with an exponentially distributed inter-packet delay. The time it takes for each packet to arrive is taken into account.

Finally the class 3 traffic arrival process is best effort traffic and it is modelled as in Xu et al [6] by a modified Interrupted Poisson Process. The ON and OFF periods are exponentially distributed. Packets arrive back to back during the ON period at the rate of 2.5 Gbps. The last packet that arrives when the ON period ends is truncated. During the OFF period, no packets are generated. The mean packet size is 500 bytes and any packet size above 5000 bytes is truncated to 5000 bytes. To calculate the ON and OFF periods, we use the coefficient of variation c^2 , defined as the ratio of the variance of the packet inter-arrival time divided by the squared mean of the packet inter-arrival time. c^2 indicates the burstiness of the arrival process.

$$c_{IPP}^2 = 1 + 2\lambda\mu_1/(\mu_1 + \mu_2)^2 \tag{1}$$

where $\frac{1}{\lambda} = (500 \text{bytes})/(2.5 \text{Gbps}) = 1.6 \mu \text{s}$, and $\frac{1}{\mu_1}$ and $\frac{1}{\mu_2}$ are the mean times of the ON and OFF periods. The arrival process of class 3 traffic is completed by the following equation:

Average Arrival Rate = 2.5 Gbps $\times \mu_1/(\mu_1 + \mu_2)$ (2)

5 Discussion of Simulation Results

We simulated a ring consisting of 10 nodes and each node is separated by a distance of 5 km. Each wavelength was assumed to have a bandwidth of 2.5 Gbps and the control wavelength works at a rate of 622 Mbps. For each class i, i=1,2,3, the transmission queue in a node was assumed to have a buffer size of 1 MB. In the single transceiver case, 11 wavelengths are required for the OBS network. In the multiple transceivers case, the number of wavelengths used is an integral multiple of 11. The simulation results are plotted with 95% confidence interval estimated by the method of batch means, see Perros [5]. Each batch is completed when each node generates 10,000 bursts. The confidence intervals are very tight and are not discernible in the graphs.

The simulation model was used to evaluate the performance of each of the protocols discussed in section 3. For all the results obtained, the class 2 average arrival rate at each node was fixed to 0.8 Gbps and the average arrival rate of class 3 traffic to 0.5 Gbps. The x axis is always the number of HDTV streams originating at each node. Specifically, in each simulation experiment, the same number of HDTV streams originate at each node, and the destination node of each stream is randomly selected. Each stream contributes an average of 20 Mbps of the total traffic. The total average arrival rate is the sum of the average

arrival rates of the 3 traffic classes. In all the experiments, the overall traffic a node transmits is less than the bandwidth of the home wavelength(s).

Most of the results are self-explanatory. For a detailed discussion, please refer our full paper Puttasubbappa and Perros [7]. Figure 2(a) plots the *Mean node overall throughput* versus the number of HDTV streams per node. The *mean node overall throughput* is defined as the average number of bits received (class 1, class 2 and class 3) by all the nodes in a unit time divided by the number of nodes in the ring. The Token-Token protocol has the highest mean node overall throughput followed by the Token protocol. Ack-Ack protocol provides less service to class 2 traffic as can be verified with figure 2(e). Token-Token is not as biased towards class 1 traffic as Ack-Ack. This can be confirmed from figure 2(c) where we note that even though the class 1 arrival rate increases, Token-Token cannot increase the throughput provided to class 1 traffic. The mean node overall throughput for the three transceivers case is plotted in figure 2(b). A notable feature is a better performance by Ack-Ack. It can be seen that Token-Token scales well.

Figure 2(c) plots the *Mean node class 1 throughput* versus the number of HDTV streams per node. The *mean node class 1 throughput* is defined as the average number of class 1 bits received by all the nodes in a unit time divided by the number of nodes in the ring. Ack-Ack performs very closely to Token and Ack despite the fact that it provides acknowledgement services to both class 1 and class 2 traffic. Token-Token does not perform as well as its counterpart. Due to more frequently available tokens, Token and Token-Token scale well for the mean node class 1 throughput when there are three transceivers per node as can be seen in the figure 2(d). The propagation delay limits the scalability of acknowledgement based protocols.

The % Bandwidth utilization is defined as the amount of time a home wavelength is busy transmitting bursts. Two graphs for the utilization of the home wavelength for node 1 are given: one depicting % bandwidth utilization for class 2 traffic and the other for class 1 traffic. Figure 2(e) plots the % bandwidth utilization for class 2 traffic for the 1 transceiver case. Token-Token supports class 2 traffic better than Ack-Ack. For the multiple transceivers case (figure not shown), Ack-Ack has higher utilization for class 2 traffic because of higher bandwidth available in the 3 transceivers case. Figure 2(f) plots the % bandwidth utilization of transmit wavelength of node 1 for class 1 traffic for the 1 transceiver case. From these plots, it can be seen that although Dest-Resv-Free protocol has a high % bandwidth utilization, it has a low class 1 throughput because of collisions. For the 3 transceivers case, Token and Token-Token scale up better than Ack and Ack-Ack (figure not shown).

The Overall burst loss rate is defined as the total number of bursts (class 1, class 2 and class 3) lost because of receiver collisions divided by the total number of bursts transmitted by all the nodes. Figure 2(g) plots the overall burst loss rate with varying number of HDTV streams per node.

Figure 2(h) plots the *hit ratio* for the five protocols. This is an important performance metric that describes how good a protocol is to support class 1



transmit wavelength for class 2 traffic (Single transceiver)

(e) % Bandwidth utilization of node 1(f) % Bandwidth utilization of node 1 transmit wavelength for class 1 traffic (Single transceiver)



Fig. 2. Simulation results

traffic. Hit ratio is defined as the total number of class 1 bursts (i.e. HDTV frames) received in time by all the nodes, divided by the total number of frames sent by all the nodes. A frame is received in time if it arrives within 17 milliseconds of receiving the previous frame from the same stream, otherwise it is a miss. A higher hit ratio for a larger number of streams per node is the desired performance criterion. Token and Token-Token perform remarkably well when more bandwidth is available as can be seen in figure 2(i) compared to Ack and Ack-Ack.

The *fairness* of a protocol is an important criterion in ring networks, since it shows whether the positioning of a node in the ring has any effect on the protocol's performance metrics. Two types of fairness indices were calculated. The *throughput fairness index* of a protocol which determines how fair a protocol is with respect to the throughput of individual nodes, and the *delay fairness index* of a protocol which determines how fair a protocol is with respect to the delay in a node. The definition of the two indices metrics are the same as those used in Xu et al[6]. The throughput fairness index of a node i is defined as the c^2 of the throughput from node i to all other nodes.

Throughput Fairness Index of Node i =
$$\left(\sum_{j=1, j\neq i}^{10} (H_{ij} - \overline{H_i})^2\right) \times \frac{1}{\overline{H_i}^2}$$
 (3)

where H_{ij} is the throughput from node i to node j, and $\overline{H_i} = (\sum_{j=1, j \neq i}^{10} H_{ij})/9$. The throughput fairness index of the protocol is defined as the average of the throughput fairness indices of all the nodes. The throughput fairness index of the protocol was computed only for the class 1 traffic. We note that the number of HDTV streams from node i to the other N-1 nodes may not be the same, since the destination of each stream is randomly selected. In view of this, the term H_{ij} is normalized by dividing it by the number of HDTV streams between node i and node j.

Figure 2(k) plots the throughput fairness index of the protocols (considering only class 1 traffic) versus the number of HDTV streams per node. Since all the protocols have a value very close to zero, all of them are throughput fair. If the throughput fairness index is calculated considering all the three classes, the protocols would still be fair because the protocols themselves do not distinguish between closer and farther nodes.

The delay fairness index of a node is similar to (3) except that W_{ij} (the mean frame delay from node i to node j) and $\overline{W_i} = (\sum_{j=1, j \neq i}^{10} W_{ij})/9$ are used instead of H_{ij} and $\overline{H_i}$ respectively. Here, the mean frame delay counts only the queueing delay that all frames in queue j of node i experience. The delay fairness does not include the propagation delay. The delay fairness index of the protocol is defined as the average of the delay fairness indices of all the nodes.

Figure 2(1) plots the delay fairness index of the protocols (considering only class 1 traffic) versus the number of HDTV streams per node. It shows that none of the protocols are delay fair. This is not due to the physical positioning

of the nodes around the ring, because as mentioned earlier, the protocols do not distinguish between closer and farther nodes. Due to the asymmetric traffic pattern of class 1, there is difference in delays experienced by class 1 bursts, translating to a higher delay fairness index.

6 Conclusion

Five different protocols for the support of different service classes on an optical burst switched ring network were proposed and their performance was evaluated through simulation. Dest-Resv-Free protocol was based on the best-effort type of service. Ack and Token provided guaranteed delivery to class 1 bursts but do not ensure zero burst loss for classes 2 and 3. These two protocols can support more HDTV streams per node than other protocols because of guaranteed delivery only to class 1 bursts. Ack-Ack and Token-Token provided zero burst loss delivery for both classes 1 and 2 of traffic and they ensured that class 1 bursts are received in time and without too much jitter. Class 3 received best-effort service. For the single home wavelength case, Ack-Ack provides better service to class 1 traffic than Token-Token. But, the Token-Token and Token protocols perform better than their acknowledgement-based counterparts when multiple home wavelengths are available for transmission. Additional simulations experiments performed indicate that token based protocols are more scalable with ring size and number of nodes in the ring.

References

- T. Battestilli and H. G. Perros. An introduction to optical burst switching. *IEEE communications optical magazine*, 41:510–515, 2003.
- N. Bouabdallah, A-L. Beylot and G. Pujolle. Collision avoidance and fairness issues in metropolitan bus-based optical access networks. Technical report, Paris 6 university, 2003.
- A. Fumagalli and P. Krishnamoorthy. A low-latency and bandwidth-efficient distributed optical burst switching architecture for metro ring. In Proceedings of IEEE ICC, 2:1340–1344, 2003.
- 4. James Jong. Multicast access protocols in an optical burst switched WDM ring network. Master's thesis, NC State University, 2002.
- 5. H. G. Perros. *Computer Simulation Techniques: The definitive introduction*. Available for free download from http://www.csc.ncsu.edu/faculty/perros/hp.html.
- L. Xu, H. G. Perros and G. N. Rouskas. A simulation study of optical burst switching access protocols for WDM ring networks. *Computer Networks*, 41(2):143–160, January 2003.
- 7. V.S. Puttasubbappa and H.G. Perros. Access protocols to support different service classes in an optical burst switching ring, 2003. Available from World Wide Web: http://www.csc.ncsu.edu/faculty/perros/puttasubbappa-perros.pdf.
- S. Trevitt. Traffic patterns in fiber channel fabrics, 2002. Available from World Wide Web: http://www.mcdata.com/downloads/mkt/wpaper/TrafficPatterns.pdf.
- 9. Bragg, A. W. MPEG-4 Analysis. private e-mail communication, March 2003.