COLLISION AVOIDANCE IN METROPOLITAN OPTICAL ACCESS NETWORKS

Nizar Bouabdallah [1, 3], Andre-Luc Beylot [2] and Guy Pujolle [1]

[1] LIP6, University of Paris 6, 8 rue du Capitaine Scott, F-75015 Paris, France; [2] ENSEEIHT - IRIT/TeSA Lab., 2, rue C. Camichel, BP7122, F-31071 Toulouse; [3] Alcatel Research & Innovation, Route de Nozay, F-91460 Marcoussis, France

Abstract: Packet-based optical ring becomes the standard access medium in metropolitan networks. Its performance depends mainly on how optical resource sharing, among different competing access nodes, takes place. This network architecture has mostly been explored in regard to synchronous transmission. However, in the present paper, we focus on the performance of asynchronous transmission-based metropolitan networks with variable packet sizes. Analytical models are presented in an attempt to provide explicit formulas that express the mean access delay of each node of the bus-based optical access network. In addition, we prove that in such a network, fairness problems are likely to arise between upstream and downstream nodes sharing a common data channel. Furthermore, we show that sharing the available bandwidth fairly and arbitrarily between access nodes, as in slotted WDM rings, does not resolve the fairness problem in asynchronous system. A basic rule, in order to achieve fairness, consists in avoiding random division of the available bandwidth caused by the arbitrary transmission of the upstream nodes.

Key words: Bus-based optical access network, Medium Access Control (MAC) Protocol, Fairness control, Access delay evaluation.

1. INTRODUCTION

In next-generation metropolitan networks, internet traffic is deemed to be stamped by three important characteristics. In fact, packet-based data traffic of bursty nature will become prevalent 1. Moreover, it is believed that traffic will fluctuate heavily and on a random basis. Finally, internet traffic will keep on growing in the next few years up to, and eventually beyond, 1 Tbit/s. The architecture of next-generation metro networks must consequently evolve enabling to tackle the new challenges, which are set by the aforementioned characteristics. In this regard, three major enabling factors can be identified as crucial for the evolution process: optics, packet switching and protocol convergence.

In the metropolitan segment, infrastructures are generally organized over a ring topology. We have proposed a new architecture named DBORN (Dual Bus Optical Ring Network), which satisfies all the requirements of next-generation metro networks. The DBORN architecture will be described in this paper. For more detailed information about this architecture the reader is invited to refer to 2. Nonetheless, the work presented in this study, is more pertaining to the design of the media-access-control (MAC) protocol planned for the bus-based optical access networks such as DBORN. This protocol is designed for efficient transport of variable-sized IP packets, whereas it does not address the inherent fairness control issue, characteristic of shared medium networks.

Generally, in order to avoid collisions on the individual WDM channels of such networks and arbitrate the bandwidth access, MAC protocols are needed. In the mean time, several access protocols for all-optical slotted WDM rings have been proposed in the literature 3, 4, 5. Most of them consider as many wavelength channels as nodes in the network, resulting in serious scalability issues, especially for MANs (Metropolitan Area Networks). Moreover, some proposals require transmitter/receiver arrays at each node leading to high equipment costs and control complexity 5. In order to deal with the aforementioned limitations, we proposed a novel access protocol for a packet-based optical metropolitan network supporting much more ring nodes than the available wavelengths in the network. The proposed MAC protocol addresses the case of non slotted WDM rings.

Since several source nodes share a common channel, one upstream node can grab all the available bandwidth, and possibly starve downstream nodes competing to access the same channel. Protocols at various levels (such as MAC or CAC – Call Admission Control) must be introduced to ensure good utilization of transmission resources and alleviate fairness problems. In general, fairness control mechanisms limit the transmission of upstream nodes in an attempt to leave enough bandwidth for downstream stations 6,7. These schemes may be efficient in the case of slotted WDM rings (i.e. synchronous transmission). However, they do not perform well in the case of asynchronous transmission based architectures like DBORN. We present here analytical models that aim to illustrate this issue. Despite its importance and up to now, the analytical study of asynchronous transmission in bus-based optical access networks has not been tackled.

The key behavior metric in such networks is the access delay at each node competing to access the shared data medium. By presenting a specific two-nodes bus as a first case study, we examine the average access delay of each node thanks to an exact analytical model. Afterwards, approximate models handling the general case, with much more nodes, are developed. The fairness issues are also dealt with in the proposed models. Simulation results show that the analytical models remain highly accurate under various traffic loads.

The remaining parts of this article are organized as follows. Section II focuses on the MAC context including a description of the network and node architectures along with the main features. Analytical models for evaluating the access delay performance of each ring node are developed in Section III. Then, section IV validates the accuracy of the models by comparing the analytical results with that obtained by means of simulations, and it discusses the effects of unfair access to the data channel. Finally, some conclusions are drawn in section V.

2. NETWORK ARCHITECTURE AND MAC DESIGN

This section describes the DBORN architecture and the proposed MAC protocol. DBORN can be described as a unidirectional fiber split into downstream and upstream channels spectrally disjoint (i.e. on different wavelengths) as shown in figure 1. The downstream bus, initiated at the hub node, is a medium shared in reading, while the upstream bus, initiated in the ring nodes, is a multiple access-writing medium.

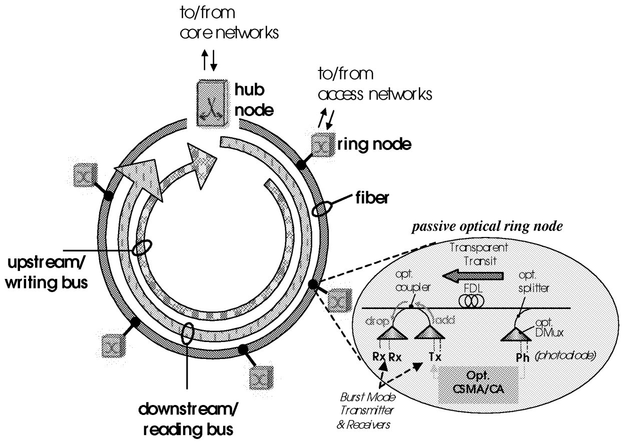

Figure 1. Overview of DBORN network and node architecture

In terms of logical performance, the main issue is related to the collision-free packet insertion on a shared writing bus. Since the transit path remains transparent and passive, no packet is dropped once transmitted on the ring (optical memory is still in the research stage). Hence, traffic control mechanisms are required at the electronic edge of the ring nodes to regulate data emission. In this regard, each DBORN ring node is equipped with void/null-detection mechanism in its upstream operating plane. This mechanism tends to retain the upstream traffic flow within the optical layer while monitoring the medium activity.

In a fixed-slotted ring system with fixed-packet size, void (i.e. slot) filling can be carried out immediately upon its detection, since the void duration is either one or multiple series of fixed-packet size duration. The detected void is therefore guaranteed to have a minimum duration of one fixed-packet length. However in non slotted ring systems with variable packet length and arbitrary void duration, it is very likely for a collision to occur if a packet is immediately transmitted upon detecting the edge of a void.

To avoid the abovementioned problem, a very simple collision avoidance system is implemented through photodiode power detection on each locally accessible upstream wavelength (figure 2). So, ring nodes first use an optical coupler to separate an incoming signal into two identical signals: the main transit signal and its copy used for control. A Fiber Delay Line (FDL) creates on the transit path a fixed delay between the control unit and the add function realized through a 2:1 coupler. With regard to the control part, as in 8, low bit rate photodiodes (ph) –typically 155 MHz- are used to monitor the activity on upstream wavelengths.

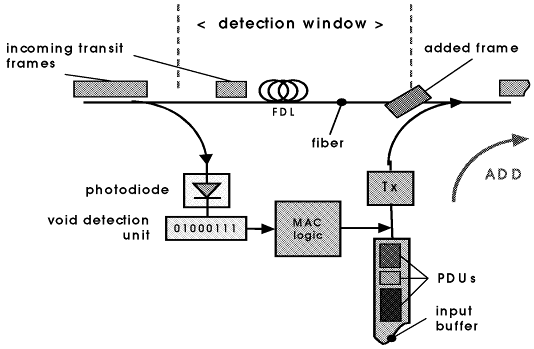

Figure 2. Schema of the CSMA/CA based MAC of DBORN

This way, voids are detected and a fixed length FDL – slightly larger than the MTU (Maximum Transmission Unit) size allowed on the network – ensures collision free packet insertion on the upstream bus from the add port. The introduction of a FDL delays the upstream flow by one maximum frame duration plus the information processing time, so that the MAC unit will have sufficient time to listen and measure the medium occupancy. The ring node will begin injecting a packet to fill the void only if the null period is large enough (i.e. at least equal to the size of the packet to be inserted). Undelivered data will remain buffered in the electronic memory of the ring node until a sufficient void space is detected.

However, considering only this basic mechanism, HOL (Head Of the Line) blocking and fairness issues arise. A direct resulting effect is performance degradation for ring nodes that are close to the hub node on the upstream bus. Additional flow control mechanisms have thus to be considered, both at the MAC layer and in upper layers at edge nodes.

3. ANALYTICAL MODELS

3.1 Framework

In this section, we will analyze the performance of the network in term of access delay. The proposed MAC protocol, which is based on CSMA/CA principle, avoids collision between local and transient packets competing to access the shared medium. As described earlier, the MAC protocol detects a gap between two packets on the optical channel, then it tries to insert a local packet into the perceived gap. However, in such an environment, fairness issues could arise.

In this study, the network is composed of N ring nodes sharing a common medium (e.g. one wavelength) used to contact the hub. Packets arrive to each node according to a Poisson process with an arrival rate λ (the analysis presented in this paper can easily be extended to unbalanced traffic conditions). We assume that the transmission time of the packets S forms a sequence of iid random variables, distributed according to some common distribution function f_S with a mean $E[S]$, a second moment $E[S^2]$ and a Laplace transform B^*. Moreover, we assume that the length of the packets emitted by the different nodes has the same distribution. The input load ρ_i of a node i ($i = 1,..., N$) is consequently equal to:

$$\rho_i = \rho = \lambda E[S] \tag{1}$$

The aim of this study is to determine the mean waiting time (or the access delay) of the different nodes $E[W_i]$, defined as the time spent by a packet in the queue i until successfully starting its transmission. Once a packet is emitted, it will not be blocked anymore and will only experience constant delays up to the hub.

We will first study the performance of the first two nodes. An exact model is presented. Approximate analytical methods are then proposed to extend the results to the following nodes, giving upper and lower bounds of the waiting time.

3.2 Analysis of the first two nodes

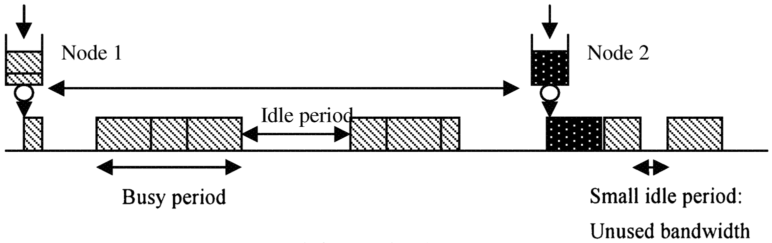

Figure 3. Activity on the data channel

In order to simplify the analysis, let us primarily consider the first two nodes. The traffic of the first node has a higher priority to access the medium. The head-of-line packet of the second queue can only access the channel if the medium is free for a sufficient time period, larger than its transmission time (figure 3). So, the emission process of the second node depends on the activity of the first one. The first queue can be simply modeled by an M/G/1 queue. Hence, the waiting time of the first node is given by:

$$E[W_1] = \frac{\lambda E[S^2]}{2(1-\rho)} \tag{2}$$

So in the remainder study, we will focus on the second queue analysis. This method will be iterated to determine the performance of the other nodes. In this paper, the "link state" refers to the state of the link when a packet, from a ring node, attempts to access the data channel. The wavelength channel can be in one of two states: free (idle) or occupied (busy). It is obvious that for packets from the upstream node, i.e. node 1, the channel is always idle. However, when packets from the downstream node, i.e. node 2, try to access the channel, the latter can be either free or occupied by upstream traffic. It is important to note in this regard that the state of the medium, as seen by node 2, alternates continuously between an idle and a busy period.

Let $\{A_i(t), t \in \mathbf{R}\}$ denote the arrival process of packets to the queue i. As the delays due to the propagation from node 1 to node 2 (Δ_1) and to the FDL (Δ_2) are constant, the whole system can be analyzed as a priority queue with Preemptive Repeat Identical (PRI) discipline 9 (i.e. if a packet can not be sent because the idle period is not long enough, the packet size will not change). The arrival process of packets is defined as, $A(t) = A_1(t - \Delta) + A_2(t)$, where traffic 1 has the higher priority and $\Delta = \Delta_1 + \Delta_2$. The workload for the queue consists of two classes of jobs. The objective is to determine the average waiting time for jobs of each class in the queue. Note that the waiting time of the higher priority class, $E[W_1]$, is simply the waiting time in an ordinary M/G/1 queue as described in (2). Below, we will focus on the waiting time of the class i customers where $i \geq 2$.

Under a preemptive repeat policy, service is interrupted whenever an arriving customer has higher priority than the one in service. The new arrived customer begins service at once. A preempted job will restart service from the beginning as soon as there are no higher priority jobs remaining in the queue. In other words, the preemptive repeat strategy stipulates that the work already done on an interrupted job is lost. In this case, the transmission time of the interrupted packet may be re-sampled according to the service time distribution after every preemption (preemptive repeat different discipline) or it may be the same as in the first service attempt (preemptive repeat identical discipline). In this study, we adopt the PRI discipline since it coincides with the real behavior of the network.

We can consequently apply the results presented in 9 based on 10. Let C_i denote the completion time of a class i customer (i.e. the time between starting and finishing service, including the preemption time). Let S_2 be the transmission time of the packet of class 2 that is chosen first. Suppose that $\tilde{n}$ preemptions occur because of the arrival of packets of class 1. Let $I(n)$ be the service time futilely expended due to the nth preemption, and $B(n)$ be the duration of the nth preemption. Note that $I(n)$ is the nth unusable idle period encountered by the packet while trying to access the data channel and $B(n)$ is the nth busy period of packets of class 1. The completion time C_2 for a packet of class 2 can be written as:

$$C_2 = S_2 + \sum_{n=1}^{\tilde{n}} I(n) + \sum_{n=1}^{\tilde{n}} B(n) \tag{3}$$

Which leads to:

$$E[C_2] = \frac{\overline{X}_{1,1}}{1-\rho}, \text{ with } \overline{X}_{1,1} = \frac{B^*(-\lambda)-1}{\lambda} \tag{4}$$

The mean waiting time may be derived as follows. In the book referenced above, the mean waiting time $E[Z_2]$ is the time spent by a class 2 packet from its arrival until service begins. It does not include the completion time. The mean response time $E[R_2]$ is consequently equal to:

$$E[R_2] = E[Z_2] + E[C_2] \tag{5}$$

As explained before, we refer in this paper to the waiting time, as the time spent by a packet in the queue until its transmission successfully begins. The mean waiting time $E[W_2]$ can be written as:

$$E[W_2] = E[R_2] - E[S] \tag{6}$$

In the case of *p* traffic classes, we have:

$$E[Z_p] = \frac{\sum_{k=1}^{p} \lambda \left(\overline{X}_{k,2}\left(1-\rho_{k-1}^{+}\right) + \frac{2\left(B^*(-2(k-1)\lambda) - 2B^*(-(k-1)\lambda) + 1\right)\rho_{k-1}^{+}}{((k-1)\lambda)^2} \right)}{2\left(1-\rho_{p-1}^{+}\right)\left(1-\rho_{p}^{+}\right)} \tag{7}$$

with

$$\rho_k^+ = \sum_{i=1}^{k} \lambda \overline{X}_{k,1} \tag{8}$$

$$\overline{X}_{k,1} = \frac{B^*(-(k-1)\lambda) - 1}{(k-1)\lambda}, \quad \overline{X}_{1,1} = E[S] \tag{9}$$

$$\overline{X}_{k,2} = \frac{2\left\{B^*(-2(k-1)\lambda) - B^*(-(k-1)\lambda) + (k-1)\lambda B'^*(-(k-1)\lambda)\right\}}{(k-1)\lambda} \tag{10}$$

$$\overline{X}_{1,2} = E\left[S^2\right] \tag{11}$$

Solving (7) for *p*=2, we can determine the mean waiting time of the second node queue, which is given by:

$$E[W_2] = E[Z_2] + E[C_2] - E[S] \tag{12}$$

3.3 Extension to N nodes

3.3.1 An upper bound for the mean waiting time

Unfortunately, the previous method can not be applied to the following nodes. Indeed, in the single priority queue with PRI discipline, the emission time already elapsed on an interrupted job is lost and can not be used anymore by lower priority jobs ($i+1,...,N$). However, in reality, if the idle period is not long enough to support the queue i head-of-line packet, the medium remains free and this idle period can be used by downstream nodes.

Using this method, it can be shown that the analysis of the system with a single priority queue will lead to an upper bound of the mean waiting time for the downstream node k where $k > 2$:

$$E[W_k] \le E[W_k^+] = E[Z_k] + E[C_k] - E[S] \tag{13}$$

Where $E[Z_k]$ is derived using (7) and

$$E[C_k] = \frac{\overline{X}_{k-1,1}}{1-\rho_{k-1}} \tag{14}$$

3.3.2 A lower bound for the mean waiting time

Conversely, the following method leads to a lower bound for the waiting time. In each node, the upstream traffic has a higher priority than the local traffic. So, the emission process of the local queue depends only on the activity of the upstream nodes and the profile of busy and idle periods generated by upstream flows. The method consists on aggregating all the upstream traffics in a single flow. The packets of the aggregated flow arrive according to a Poisson process. Then, we analyze each node as a single queue with two traffic classes under PRI priority discipline where the local traffic has the lower priority.

This approximate analysis leads to an underestimation of the mean response time because it may cause longer busy period duration and consequently longer idle period duration as well. This is true as long as the distribution of the packet length is the same in the different nodes (on average, if a packet of a previous node can not be sent in the medium because it is too large, a packet of the local node will not pass either). In reality, the free bandwidth seen by a downstream node is much more fragmented than the one generated by the aggregated upstream flow. One then obtains the following results by applying the method of paragraph 3.2. to each node *k* with two flow classes (i.e. upstream and local traffic) with respective arrival rates:

$$\lambda_{k-1}^- = (k-1)\lambda, \quad \lambda_k = \lambda \tag{15}$$

It corresponds to "equivalent loads":

$$\rho_{k-1}^- = \lambda_{k-1}^- E[S], \quad \rho_k^- = \lambda_{k-1}^- E[S] + \frac{B^*\left(-\lambda_{k-1}^-\right) - 1}{(k-1)} \tag{16}$$

The lower bound of the waiting time is given by:

$$E[W_k^-] = E[Z_k^-] + E[C_k^-] - E[S] \tag{17}$$

Where $E[Z_k^-]$ is derived using (7) and

$$E\left[C_k^-\right]=\frac{B^*\left(-\lambda_{k-1}^-\right)-1}{\lambda_{k-1}^-\left(1-\rho_{k-1}^-\right)} \tag{18}$$

3.4 Example

Different packet length distributions can be considered. In the present paper, we consider packets of variable length (50, 500 and 1500 bytes) more or less representative of the peaks in packet size distribution in Ethernet.

Let p_i be the probability of the different packet sizes and d_i the corresponding emission time.

The mean waiting time of the first queue (2), of the second queue (12) and the bounds on the waiting times for the following nodes (13) (17) can be derived using the following parameters:

$$E[S^k]=\sum\nolimits_i p_i d_i^k\,,\; B^*(s)=\sum\nolimits_i p_i e^{-sd_i} \tag{19}$$

4. NUMERICAL RESULTS

To evaluate the accuracy of the proposed analytical models, we compare their results with those obtained from a simulation conducted on, network simulator 2. In the following, only a subset and a synthesis of the results are presented. In all our simulations, unless otherwise specified, we assume that (1) all the ring nodes share a common upstream wavelength modulated at 1 Gbit/s ; (2) the packets arrive according to a Poisson process; (3) the arrival rate of the packets to each node is the same in order to highlight the fairness issues; and (4) all the ring nodes transmit only to the hub. In all the figures depicting the simulation results, the traffic load on x-axis denotes the average traffic load ρ sourced from every node to the hub.

The analysis results of access delay for the first two nodes are presented in figure 4, revealing a perfect match with the simulation results: analytical results practically coincide with the simulation results. We consider packets of variable length (50, 500 and 1500 bytes) more or less representative of the peaks in packet size distribution in Ethernet. The total traffic volume comprises 50% of 1500 Bytes, 40% of 500 Bytes and 10% of 50 Bytes packets size. We observe that:

- Under light traffic load, the access delay of the downstream node is more important than upstream node one. As a result, the Fairness issue is pronounced even under light traffic load.

- Under high traffic load, the difference between the performance of upstream and downstream nodes sharing the optical channel increases. The main reason is that upstream node grabs more bandwidth thus leaving less capacity to the downstream node.

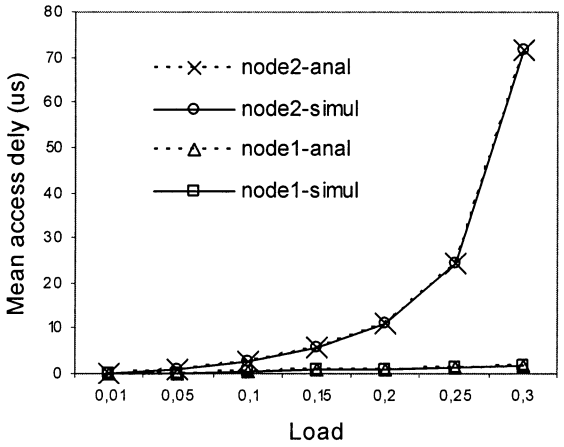

Figure 4. Mean access delay of the first two nodes with variable-packet size traffic

The analysis results in this special scenario are so significant. We observe that even when the upstream node uses a small part of the available bandwidth, the downstream nodes performance is strongly affected. The fairness issue is always present in shared medium networks. This is mainly due to the lack of organization of the emission process in the network and the absence of control mechanisms. In fact, the mismatch, between the idle period distribution resulting from upstream node utilisation and the packets' size distribution of the downstream node, often leads to bandwidth waste as well as fairness problems with regard to resource access.

Once a packet of maximum size is at the head of the insertion buffer, it blocks the emission process until finding an adequate void: this is the well-known HOL blocking problem. Thus, sharing the bandwidth fairly and arbitrarily between nodes is not sufficient to ensure satisfying results. The sharing process must thus be done smartly in order to preserve a maximum of useful bandwidth for downstream nodes. In general, fairness control mechanisms limit the transmission of upstream nodes to keep enough bandwidth for downstream stations. These schemes may be efficient in the case of slotted WDM rings. However, they don't perform well in the case of asynchronous transmission based architectures like DBORN.

Hence, we suggest preserving bandwidth (represented by idle periods) by upstream nodes in order to satisfy downstream nodes requirements in an organized way. A basic rule consists in avoiding random division of the resource, which would lead to inadequacy between idle periods length and the layer 2 PDUs (Protocol Data Units) size. Therefore the control mechanism has to prevent greedy upstream stations from taking more than their fair share by forcing them to keep idle periods of sufficient size.

The analysis results for the general case of six-node bus, depicted in the figure 5, emphasize the abovementioned results. The traffic load ρ sourced by each node is 0,05. The access delay of each node is found to increase monotonically when progressing towards the hub. Indeed, the closest nodes to the hub encounter relatively large delays, incompatible with performances expected in metropolitan networks. We insist that the performance degradation of downstream nodes is not due to the medium saturation since the medium occupation is not beyond 30%. The upper and lower bound curves are very close to the simulation result curve. So, the approximate analytical models can achieve high accuracy. But, we make the observation that the bounds become less accurate for the closest nodes to the hub, especially the upper one.

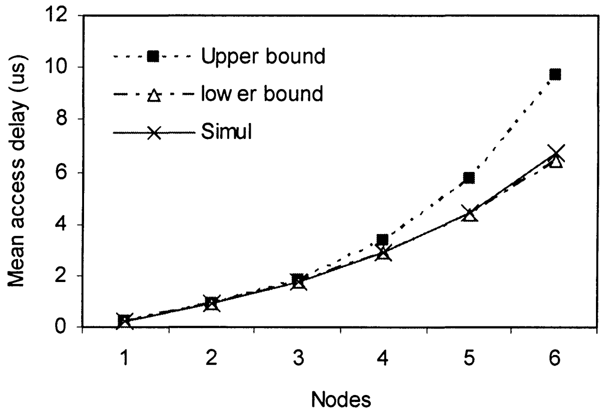

Figure 5. Mean access delay of the six-node bus with variable-packet size traffic

5. CONCLUSION

This paper, to the author's knowledge, provides the first analysis of shared bus network behavior with asynchronous transmission. We analyzed the system performance in terms of access delay required by each node to inject a packet on the shared medium. The analysis results showed that fairness issues are likely to arise between upstream and downstream nodes even under light loads. We observed that sharing the available bandwidth fairly and arbitrarily between nodes does not resolve the fairness problem. Consequently, additional flow control mechanism has to be considered, not only to limit the transmission of the upstream nodes but also to organize their emission process. Simulations results showed that the proposed analytical models are extremely accurate under various traffic loads.

REFERENCES

1. M. J. O'Mahony, D. Simeonidou, D. K. Hunter, A. Tzanakak, "The Application of Optical Packet Switching in Future Communication networks", IEEE Commun. Mag., pp. 128-135, March 2001.
2. N. Le Sauze et al., "A novel, low cost optical packet metropolitan ring architecture", Proc. Of ECOC '01, Amsterdam, Netherlands, Vol. 4, pp. 66-67, October 2001.
3. M. A. Marsan, A. Bianco, E. Leonardi, M. Meo, and F. Neri, "MAC protocols and fairness control in WDM multirings with tunable transmitters and fixed receivers", IEEE/OSA J. Ligh. Tech., vol. 14, pp. 1230-1244, June 1996.
4. J. Fransson, M. Johansson, M. Roughan, L. Andrew, and M. A. Summerfield, "Design of a medium access control protocol for a WDMA/TDMA photonic ring network", Proc. of GLOBECOM '98, Sydney, Australia, Vol. 1, pp. 307-312, November 1998.
5. A. Fumagalli, M. Johansson, and M. Roughan, "A token-based protocol for integrated packet and circuit switching in WDM rings", Proc. of GLOBECOM '98, Sydney, Australia, pp. 2339-2344, November 1998.
6. M. A. Marsan et al., "Metaring Fairness Control Schemes in All-Optical WDM Rings", Proc. of INFOCOM '97, Kobe, Japan, vol. 2, pp. 752-760, April 1997.
7. J. S. Yih, C. S. Li, D. D. Kundlur, and M. S. Yang, "Network access fairness control for concurrent traffic in gigabit LANs", Proc. of INFOCOM '93, San Francisco, California, vol. 2, pp. 497-504, March 1993.
8. R. Gaudino et al., "RINGO: a WDM Ring Optical Packet Network Demonstrator", Proc. of ECOC '01, Amsterdam, Netherlands, Vol. 4, pp. 620-621, September 2001.
9. H. Takagi, Queueing Analysis Vol I: Vacation and Priority Systems Part I, North Holland, 1991.
10. N.K. Jaiswal, Priority Queues, Academic Press, 1968.