
Eric Jonsson School of Engineering and Computer Science

9-1998

A Scalable and Adaptive Fair Access Protocol for Slotted WDM Bus Network

Andrea Fumagalli, *et al.*

For more information about this item go to: <http://libtreasures.utdallas.edu/xmlui/handle/10735.1/2589>

**A Scalable and Adaptive Fair
Access Protocol for Slotted WDM
Bus Network**

Andrea Fumagalli, Roberto Grasso

Department of Electrical Engineering

Technical Report UTD/EE-03-98

September 1998

A Scalable and Adaptive Fair Access Protocol for Slotted Wavelength Division Multiplexing Bus Network *

Andrea Fumagalli, Roberto Grasso

The University of Texas at Dallas

P.O. Box 830688 – MS EC33 – Richardson, TX 75083-0688

Tel: (972) 883-6853 – Fax: (972) 883-2710

E-mail: {andrea.f, r.grasso}@utdallas.edu

Abstract

The recent progress of optical technologies has made it possible to increase the number of wavelengths (channels) practically available in the fiber by almost one order of magnitude. With this rapid growth, access protocols originally designed for systems with a few wavelengths may not be adequate to fully utilize the optical bandwidth that is becoming available through this technological breakthrough. This is the case with access protocols designed for unidirectional fiber bus network.

The paper proposes a novel access protocol for multi-channel all-optical folded bus network whose performance scales well with the number of wavelengths. This result is achieved using a load balancing algorithm that runs independently at each node and evenly distributes the traffic among the wavelengths. With the proposed protocol, the packet average access delay at a given bandwidth utilization increases only marginally as the number of wavelengths (and proportionally the offered load) is grown in the system. In addition, at all source nodes fair access delay is guaranteed by the protocol under any stable load condition, including non uniform and bursty traffic.

1 Introduction

As traffic demand in data networks grows due to Web applications and emerging Internet telephony, the 30 THz bandwidth of the optical fiber becomes an attractive means to support such a growth without having to install additional cables. By using WDM (Wavelength Division Multiplexing) [2] the network throughput can be increased by using parallel channels in the same fiber, each operating at a transmission rate determined by a

*This research is sponsored by the NSF under contracts # NCR-9628189 and # NCR-9596242, and by CSELT. Corresponding author: R. Grasso

cost effective electro-optical interface, i.e., transmitters and receivers. This technology has been proven to be practical and potentially economical [1]. Commercially available WDM systems offer today 4-16 channels, with 40 and 80 channel systems being announced to be available in the near future. This is a quite fast growth, considering that systems with only 2-4 channels were available over the last few years. Consequently, existing access protocols designed for WDM networks may not necessarily cope with this rapid growth.

This paper investigates the scalability limitations of an existing access protocol proposed for all-optical¹ folded bus [5] and presents a novel protocol whose performance scales well with the number of channels. Scalability is assessed by evaluating the dependency of packet access delay on the number of channels, assuming that the bandwidth utilization remains constant in the network, i.e., the offered load is proportional to the number of channels. In this context, an ideally scalable solution yields access delay that remains constant as the number of channels grows.

In all-optical folded bus, e.g., [5], nodes are allowed to transmit packets on one bus (the transmission bus) and receive packets from the other bus (the reception bus). Transmitted packets are not converted back into the electronic domain until they reach the intended destination. This topology has the following properties:

- re-circulation of optical signal is not possible in the bus, thus preventing undesired effects caused by residual transmission due to non-ideal optical filtering [10]
- by monitoring the reception bus traffic every node can learn what is the current network load
- if time is divided into fixed slots, each containing one packet per wavelength, random transmission of (fixed length) packets is possible without generating collision by simply sensing the transmission bus at the beginning of each slot prior to the packets' transmission
- the folding point of the bus, i.e., where the transmission bus and the reception bus join, is the natural place for connecting a gateway towards other networks to yield system scalability [4].

However, as well know, uncontrolled transmission of packets in mono-directional bus originates unfairness among the source nodes: upstream nodes are more likely to sense empty slots than downstream nodes [3]. As a consequence, the access delay of packets transmitted by upstream nodes is on average smaller than the access delay of packets transmitted by downstream nodes.

This unfairness issue was originally addressed in the Distributed Queue Dual Bus (DQDB) standard [3] devised to control the sources' activities in a two mono-directional bus-single channel fiber MAN. According to the standard each node attempting transmission has to reserve an empty slot using a reservation bit that is broadcast to the upstream nodes with respect to the bus in which transmission is requested. With this reservation

¹The term all-optical is used in this paper to indicate that packets propagate from source to destination without opto-electronic and electro-optic conversion at the intermediate nodes.

mechanism each node in the network can build a virtual queue of packets that indicates the transmission order to achieve fairness in the system. In addition, since as discussed in [6, 7], the propagation delay of the reservation bits makes the reservation mechanism not totally fair, different bandwidth equalization techniques were introduced in the final draft of the DQDB standard [8] [9].

More recently, a stochastic access protocol called FairNet [5] was proposed to guaranteeing fair access in a multi-channel all-optical folded bus network. According to FairNet, each node chooses whether or not to transmit and the channel on which transmission will occur based on a set of probabilities. These probabilities depend on the network offered load and are calculated by a coordinator node every time there is a traffic change in the system. For each channel, the coordinator node collects the expected average offered load generated by every node, computes the transmission probabilities for each node under the assumption that traffic is non-bursty (by solving a Markov model [11]), and passes these probabilities on to the nodes.

Although fairness is achieved in the FairNet bus, at any given constant bandwidth utilization the average access delay grows linearly with the number of wavelengths. The main cause for this delay is twofold. Due to traffic fluctuations, at some time instant some channels may be heavily loaded while others are practically unused. Second, every time a node selects for transmission a channel for which the node does not have any packet awaiting transmission, packet transmission will not occur at that node even if there are packets awaiting transmission on other free channels. Overall, these two factors tend to increase the average access delay and their negative effect becomes more pronounced as the number of channels grows.

This paper proposes a novel access protocol for WDM folded bus, whose aim is to dynamically balance the traffic distribution throughout the channels. The protocol is based on a mixed approach that combines both a transmission probability associated with each node that determines whether or not the node is allowed to attempt transmission, and a balancing strategy that selects the channel for transmission. The node's transmission probability is calculated by the coordinator node considering the overall network offered load. The balancing strategy is local to the node and selects the channel for transmission according to two parameters: the occupancy of the transmitter queues and the instantaneous network traffic. The network traffic is measured by monitoring the reception bus using light sensors capable of detecting whether or not any wavelength is carrying a packet.

The proposed protocol is based on a self-adaptive control whose complexity grows only logarithmically with the number of channels and performance scales well with the number of channels. In addition, the proposed protocol offers satisfactory fair access delay at all nodes, unrespectively of their location in the bus.

2 The Folded Bus Architecture

The network under consideration is an optical folded bus in which a number (W) of parallel unidirectional channels, each corresponding to a distinct wavelength, are shared

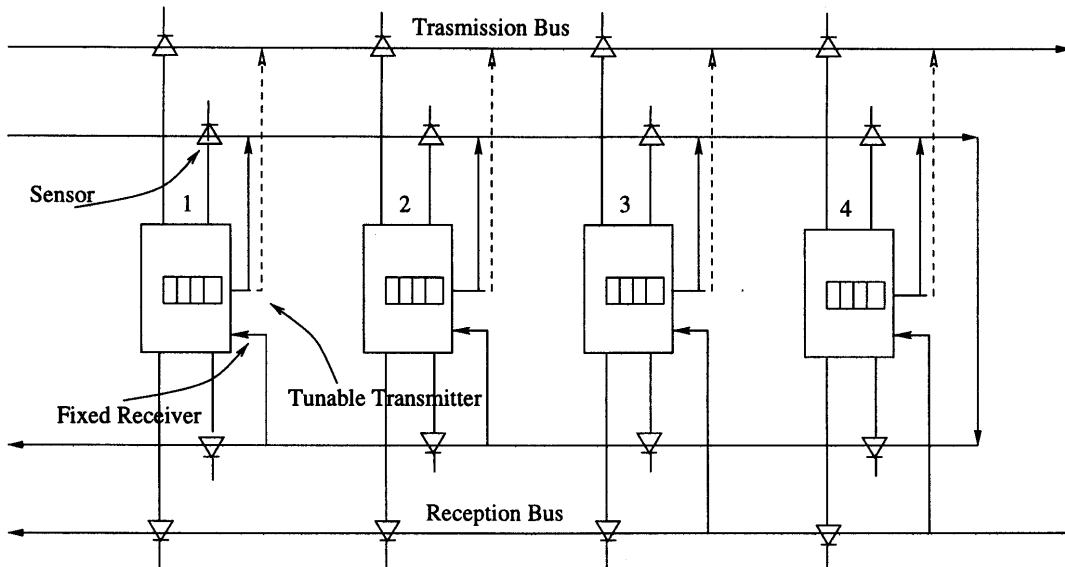


Figure 1: A two channel bus with four nodes.

by a number (N) of nodes. A 4 node-2 channel network is depicted in Fig.1. Nodes transmit packets using the transmitter bus and receive packets from the reception bus. Once transmitted, the optical signal circulates in the folded bus and it is broadcast to all nodes as it propagates within the reception bus. Node receives only packets intended to it. Time is slotted on each channel, and the slot length is equal to the packet transmission time plus the guard band necessary to take into account the finite tuning speed of the optical transceivers. Slots are aligned across the channels and aligned slots arrive at node simultaneously. Each node is equipped with three optical components:

- one tunable transmitter that can be tuned on any of the W channels,
- one fixed receiver which receives from a preassigned channel, e.g., receiver of node i is tuned on wavelength λ_j , $j = \lceil \frac{i \times W}{N} \rceil$,
- light sensors that detect the presence of packets in both transmission and reception busses.

Each node has W FIFO transmission queues, one for each channel.

3 Access Protocols

As discussed in section 1, if not controlled an upstream node can make use of the entire bandwidth of a channel at the expense of the downstream nodes that starve for empty slots. An appropriate access protocol is therefore necessary to assure that, independently from their location in the bus, nodes can fairly share the bus bandwidth. At the same time, the aim of the protocol is to minimize the packet average access delay and make

it dependent only on the bandwidth utilization, i.e., scalable with the number of wavelengths.

The protocol proposed in this paper achieves these goals by 1) dynamically assigning the node a set of transmission probabilities that determine how frequently transmission must be attempted and 2) using a self-adjusting control that determines on which channel transmission should be attempted. Once a channel is selected, the node transmits the head-of-the-line packet in the transmitter queue corresponding to that channel if

- the queue is not empty
- the passing slot is sensed empty on the selected channel.

Transmission Probabilities. The probability of attempting transmission is determined for every node by a coordinator node taking into account the overall offered load, considered known by the system. The transmission probabilities are calculated using the model proposed in [5] with the difference that for each node, j , only the total transmission probability is derived, pt_j , without differentiating on which channel transmission will be attempted. This probability is used by the node to determine whether or not transmission should be attempted in the passing slot.

Self-Adjusting Control for Traffic Balancing. If transmission must be attempted by the node, the channel for transmission, c , is selected according to a self-adjusting control whose goal is to dynamically balance the traffic across the channels. The channel selection is based on a set of probabilities $\{p_c\}$, that are functions of both the number of packets (or Protocol Data Units - PDUs) in the transmission queues and the instantaneous network throughput. The non-trivial objective here is to increase the probability of selecting a channel with *congested transmission queue* and a relatively *low throughput* without altering the *fair access* guaranteed by the set of probabilities $\{pt_j\}$.

Three observers are used by each node to determine the values of $\{p_c\}$.

The first observer is used to measure the throughput generated by the downstream nodes, separately on each channel. The downstream nodes' throughput can be measured comparing the state (free or busy) of wavelengths in the slot passing by in the transmission bus with the state of wavelengths in the same slot passing by in the reception bus after one round trip. The state of the slot passing by in the transmission bus is detected by an array of light sensors and it is recorded in an array of shift registers, one per wavelength, whose size is equal to the number of slots that are propagating in the fiber between the node's transmitter and the node's receiver. The state of wavelengths in the same slot passing by in the reception bus is detected by another array of light sensors and it is compared with the recorded values found in the shift register. Using the feedback provided by this observer the node can transmit more frequently on the channels that are not congested with traffic generated by downstream nodes. This is an important condition to maintain fairness between upstream and downstream nodes, i.e., make sure that upstream nodes do not impair downstream nodes' transmission capabilities.

The second observer measures the throughput on each channel. This is computed detecting the state (free or busy) of wavelengths using the array of light sensors of the reception bus. Using the feedback provided by this observer the node can transmit more

frequently on channels that are less congested, thus balancing the traffic across the channels.

The third observer counts the packets actually stored in the transmitter queues. This information is local and available to the node since each node monitors its own queues only. Using the feedback provided by this observer the node can balance the occupancy of the transmission queues by more frequently selecting packets from the most congested queues. Another advantage is the fact that contrary to the FairNet approach, empty queues are never selected.

To properly capture the evolution of the system, each of the three observers averages the observed values over a time window. This is obtained using the recursive equation

$$\omega(i+1) = a\omega(i) + (1-a)u(i) \quad (1)$$

where $\omega(i)$ is the value of the observer at time slot i , $u(i)$ is the value measured at time (i) (e.g., throughput in slot i), and a is a value in the 0 to 1 range that determines the practical length of the observation window. (A high value of a (0.94 ÷ 0.99) indicates that the observation window is quite large, implying that changes of the observer value are slow and capture only significant fluctuations of the network behavior.) It is useful to define a *time constant* for the observer defined as

$$\tau = \frac{1}{1-a} \quad (2)$$

This constant is proportional to the length of the observation window.

The three observers are updated at each node using Eq. 1 as follows.

We define as $\alpha_\lambda(i)$ the observer of the throughput on channel λ originated by downstream nodes. $u(i)$ is set to 1 when wavelength λ is occupied by a packet transmitted by a downstream nodes, 0 otherwise. As shown in A the time constant for this observer is quite critical and must be designed according to the bus length and the traffic burstiness. The value of this observer varies from node to node depending on the node's position.

We define as $\gamma_\lambda(i)$ the observer of the network throughput on channel λ . $u(i)$ is set to 1 when wavelength λ is detected busy in the receiver bus, 0 otherwise. To ensure that upstream nodes do not take advantage of sparing empty slots the time constant for this observer is chosen to be larger than the time constant of the previous observer. The values of this observer are the same for every node, although they may occur at different time slots due to the non-null propagation time of the signal in the bus.

We define as $\beta_\lambda(i)$ the observer of the occupancy of the transmission queue associated with channel λ . The time window for this observer is one slot, i.e., $a = 0$, and

$$\beta_\lambda(i) = \frac{\#packets\ in\ buffer\ c}{\sum_k \#packets\ in\ buffer\ k} \quad (3)$$

At each node, the probability to select channel c for attempting transmission at time i is given by

$$p_c(i) = \frac{z_c(i)}{\sum_{k=1}^W z_k(i)} \quad (4)$$

where weight $z_c(i)$ is

$$z_c(i) = (1 - \alpha_\lambda(i))\beta_\lambda(i)(1 - \gamma_\lambda(i)) \quad (5)$$

Weights $z_c(i)$ are computed taking into account three objectives. The first objective is to balance the packets stored in the transmission queues of the node: $z_c(i)$ is proportional to $\beta_\lambda(i)$, i.e., packets in longer queues are more likely to be selected for transmission. The second objective is to balance the traffic throughout the channels: $z_c(i)$ is inversely proportional to $\gamma_{c(i)}$, i.e., packets to be transmitted on less congested channels are more likely to be selected. The third objective is to maintain fair access throughout the nodes: $z_c(i)$ is inversely proportional to $\alpha_c(i)$, i.e., packets to be transmitted on channels less used by downstream nodes are more likely to be selected. This last objective is necessary to prevent that upstream nodes react to traffic changes more promptly than downstream nodes, and achieve a better-than-average access to the network.

3.1 Protocol Complexity

The high speed transmission rate of the considered network forces the node to make use of dedicated electronic circuitry to implement the proposed access protocol. Considering that at the node the access protocol must select the next packet to transmit within one slot time – as short as one microsecond – the time complexity of the protocol is critical and must be evaluated to determine its scalability in the number of wavelengths.

By assigning one distinct electronic circuitry to each channel, weights in Eq. 5 can be derived in parallel. The time complexity of this operation is therefore $O(1)$. The calculation of the probabilities in Eq. 4 is however the time consuming part of the packet selection. These probabilities are calculated after summing W terms. A multi-stage architecture of adders can perform this sum using $W - 1$ adders, with $W/2$ adders in the first stage summing pairs of terms, $W/4$ adders in the second stage summing pairs of output values from the adders in the first stage, etc. With such a multi-stage architecture the time complexity of the algorithm becomes $O(\log W)$.

4 Performance

The proposed access protocol needs to be analyzed in a variety of cases to establish its ability to guarantee fair access to all nodes and to scale in the number of wavelengths. The analysis must also include different traffic distributions, ranging from uniform to non uniform and bursty.

The results shown in this section are obtained using an object oriented C++ simulator specifically deployed to simulate a multi-channel slotted system with packet generation times not synchronized with the slots. The simulator allows to simulate a folded bus network with varying number of wavelengths, number of nodes and length of the bus. Results shown in the paper have 3% confidence interval at 99% confidence level.

The parameter of interest is the packet average access delay and its fairness throughout the nodes. Access delay is measured in time slots. The following values are used as time

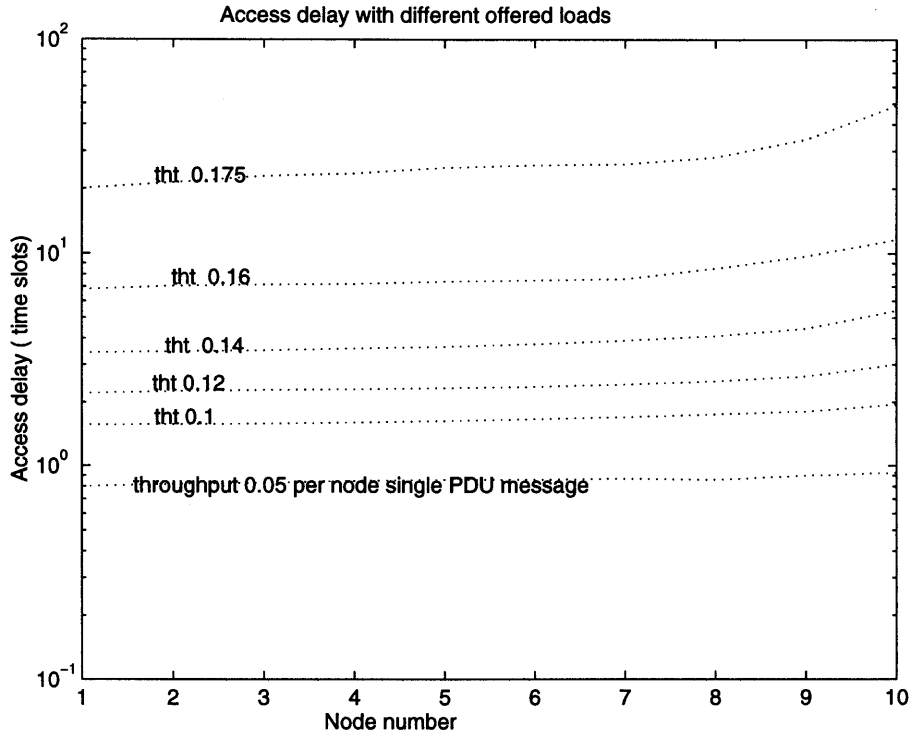


Figure 2: Access delay at different offered loads for a 10 node-2 wavelength bus when observers α and γ are *not* used by the node to select the channel for transmission.

constants for the observers: values in the $0.5 \div 0.7$ range are used for $\alpha_\lambda(i)$, values in the $0.95 \div 0.98$ range are used for $\gamma_\lambda(i)$, 0 is used for $\beta_\lambda(i)$. Offered traffic is modeled using a Poisson arrival process that generates messages whose length, measured in number of PDUs, is determined by a random integer variable uniformly distributed between one and the maximum number of PDUs, chosen to be PDUMAX. The length for the transmission queues is chosen to avoid buffer overflow and the generated traffic is such that network stability can be achieved. The analysis is carried out first assuming a uniform traffic distribution, i.e., nodes generate the same load and packets are evenly destined to any node except for the source node itself. Non-uniform traffic is analyzed in the second part of this section. To evaluate the influence of bursty traffic on system performance, varying values of PDUMAX are considered. Results obtained using the FairNet protocol are also shown to compare the two solutions.

4.1 Uniform Traffic

Before showing the results obtained using the proposed protocol, it is important to determine how relevant α and γ observers are in guaranting fairness in the system. Fig.2 shows the access delay drift from upstream nodes (left) to downstream nodes (right) if the protocol does not make use of α and γ observers, i.e., Eq. 4 is not used. We observe that the drift grows with the network throughput and becomes unacceptable (over 100% dif-

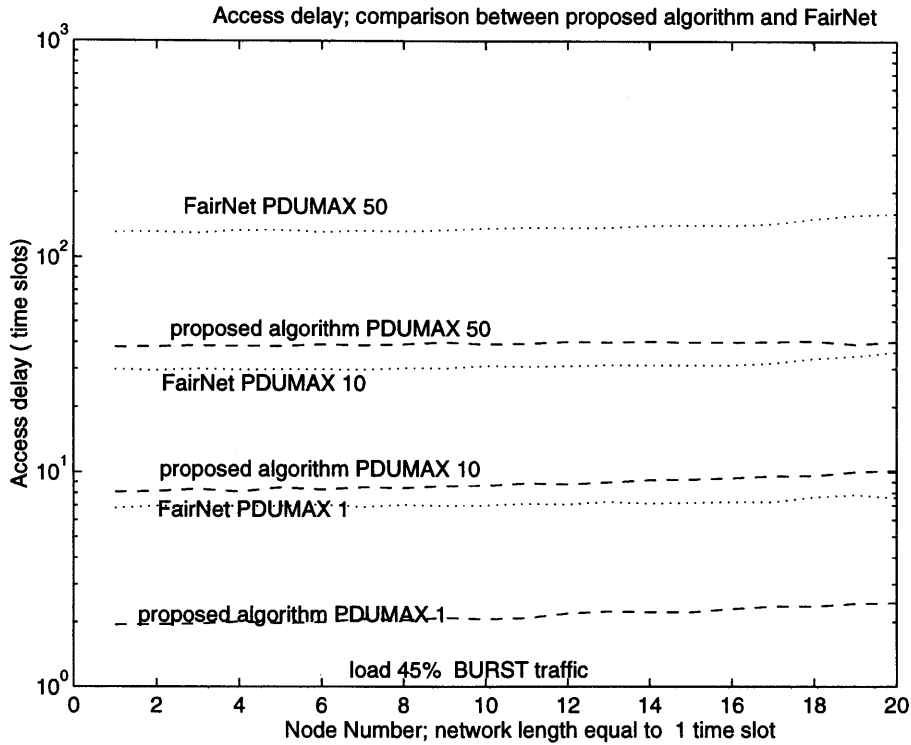


Figure 3: Access delay comparison between FairNet (dotted) and the proposed protocol (dashed), in a 20 node-4 wavelength bus network.

ference between the first (leftmost) node and the last (rightmost) node) when the offered load is above 80%.

Fig.3 shows access delay versus node's position in a 20 node-4 wavelength network for both the proposed access protocol and FairNet. Messages have varying maximum lengths, $PDUMAX = 1, 10, 50$, and throughput is 45%. The curves show that the proposed protocol reduces the average access delay by a factor of three when compared to FairNet. This reduction is a function of the number of wavelengths as shown in Table 1. Table 1 contains the average access delay measured in four networks with a 50% throughput. The access delay is obtained averaging the access delays measured at all node. As the number of wavelengths grows, the average access delay in FairNet grows linearly, whereas with the proposed protocol it remains almost unaffected. In the 16 wavelength bus the ratio between the two access delays is already 16.

Fig.4 shows access delay versus node's position when throughput is 80% and traffic is bursty, $PDUMAX = 1, 4$. The plots show that under bursty traffic the FairNet protocol originates a 50% delay drift between the first node and the last node in the bus, while the proposed protocol alleviates this unfairness to less than 15%. The reason for this result can be explained as follows. In presence of bursty traffic, detection of a busy slot in a channel may be the sign of a stream of PDUs that may follow in the same channel during the very next time slots. The observers used by the proposed protocol detect this

N	W	<i>FairNet</i>	<i>proposed</i>
10	2	4.22	1.71
20	4	9.176	1.83
40	8	19.15	1.9
80	16	39.14	2.45

Table 1: Comparison between average access delay obtained using the proposed control and the FairNet protocol with respect to number of wavelengths and nodes. Network load is 50%.

possibility and stochastically allocate the necessary bandwidth over the next time slots to support the transmission of the burst. Interestingly, the proposed protocol is able to self-adjust to bursty traffic in spite of the fact that the node’s transmission probabilities are derived under the assumption of non-bursty traffic.

Fig.4 reveals also a reduction of performance gap between the proposed protocol and FairNet to 30%. This is due to the fact that as throughput increases fewer empty slots are available in the system, making it more difficult to balance the transmission queues efficiently. With 90% throughput the system is practically unstable.

4.2 Non-Uniform Traffic

To investigate the protocol ability to adjust to non-uniform traffic, this section considers an extreme case of traffic distribution in which every node transmits to a subset of nodes that are all reachable using a single channel. Under this circumstance, each node has only one transmission queue that is non-empty, while different nodes may transmit on distinct channels. As described in section 3, in the proposed protocol the probability of transmission attempt at each node is calculated without distinguishing among the channels. In spite of this simplification, the access delay measured under non-uniform traffic distribution is still fair as shown in Fig. 5. These results are obtained using a 2 wavelength-10 node network with 50% offered load.

5 Conclusions

The paper proposed a novel self-adaptive protocol based on a probabilistic scheduling policy that yields fair and low access delay in WDM multi-channel folded bus, even in presence of large numbers of wavelengths and bursty traffic conditions. The protocol is based on a load balancing algorithm that runs independently at each node and evenly distributes the traffic among the wavelengths. As a result, traffic congestion on individual channels and overall packet access delay are minimized in the system.

Since it requires a computation time that is logarithmic in the number of wavelengths and yields a packet access delay that is only marginally sensitive to the number of wave-

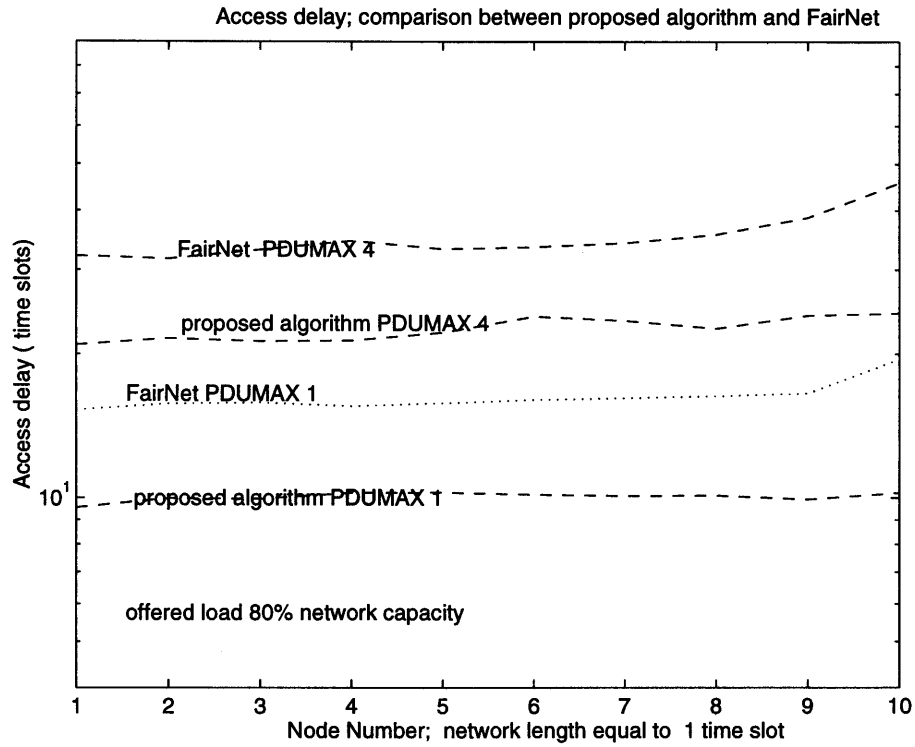


Figure 4: Comparison between access delay obtained by FairNet (dotted) and proposed algorithm (dashed) in a 10 node-2 wavelength bus network.

lengths, the proposed protocol is an ideal control solution to cope with the exponential growth of the number of wavelengths provided by today's WDM technologies.

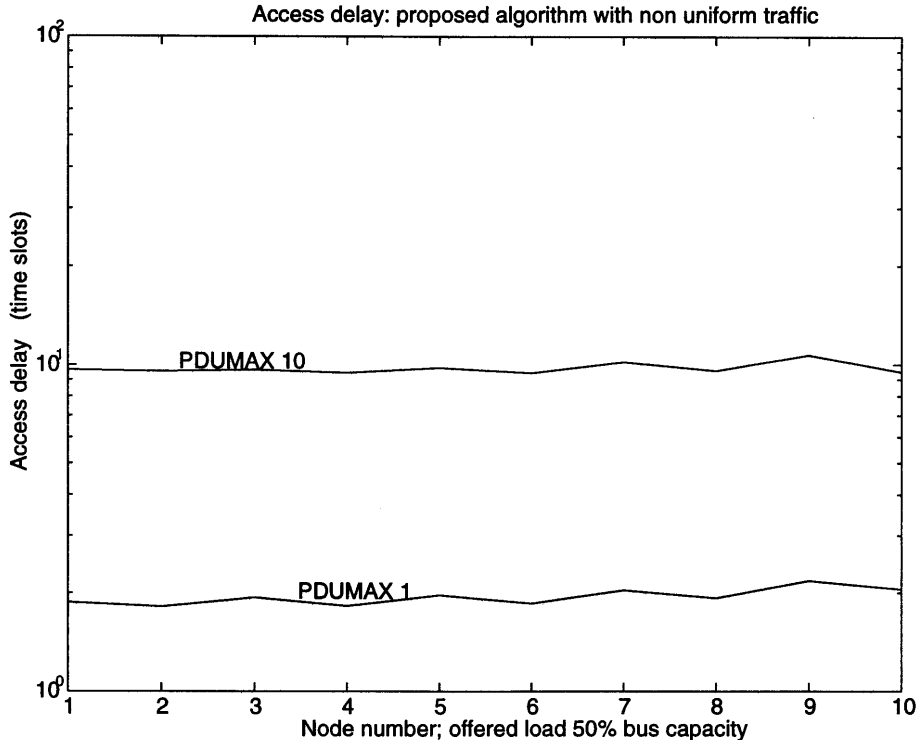


Figure 5: Access delay (dotted) obtained with non-uniform offered load for a 10 node-2 wavelength bus network.

A Design of the $\alpha_\lambda(i)$ Time Constant

How the length of the bus affects the protocol performance is another important issue that needs to be addressed to establish the stability of the proposed protocol. This appendix deals with this issue and presents some empiric rules to determine the time constant of one of the three observers used by the protocol. Since observer $\beta_\lambda(i)$ – monitoring the occupancy of transmission buffers – has time constant equal to 0, and observer $\gamma_\lambda(i)$ – monitoring the network traffic – requires a time constant quite large, the only critical time constant is the one for observer $\alpha_\lambda(i)$ – monitoring the traffic generated by downstream nodes.

The time constant of $\alpha_\lambda(i)$ that yields satisfactory fairness is a function of both the bus length and the maximum message length. Intuitively, the longer the message the slower the observer should be, the longer the bus the faster the observer should be to compensate for the network latency. These conjectures are supported by the simulation results shown in the following graphs.

Fig. 6 and Fig. 7 depict access delay versus node's position for varying time constant values ranging from 0.5 to 0.99. The bus length is 1 time slot in Fig. 6 and 50 time slot in Fig. 7. Maximum message length is equal to 50 PDUs. In these figures the y-axis is linear and each notch is equal to 10 time slot. Figures are self-explanatory. We note that fast time constants privilege downstream nodes, while slow time constants do not provide

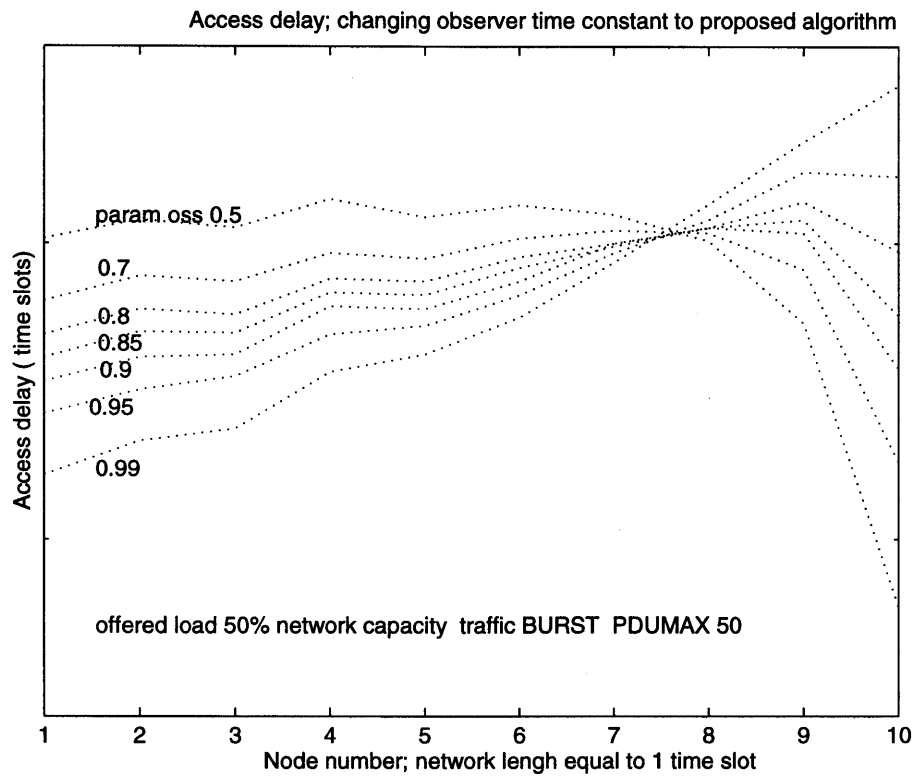


Figure 6: Access delay obtained by changing observer time constant (dotted) for a 10 node 2 wavelength 1 slot length bus network .

enough time for the upstream nodes to learn about the offered traffic of the downstream nodes. Longer bus networks tend to prefer faster observer to compensate for the higher network latency.

In conclusion, values in the $0.7 \div 0.85$ range are satisfactory for most of the network configurations.

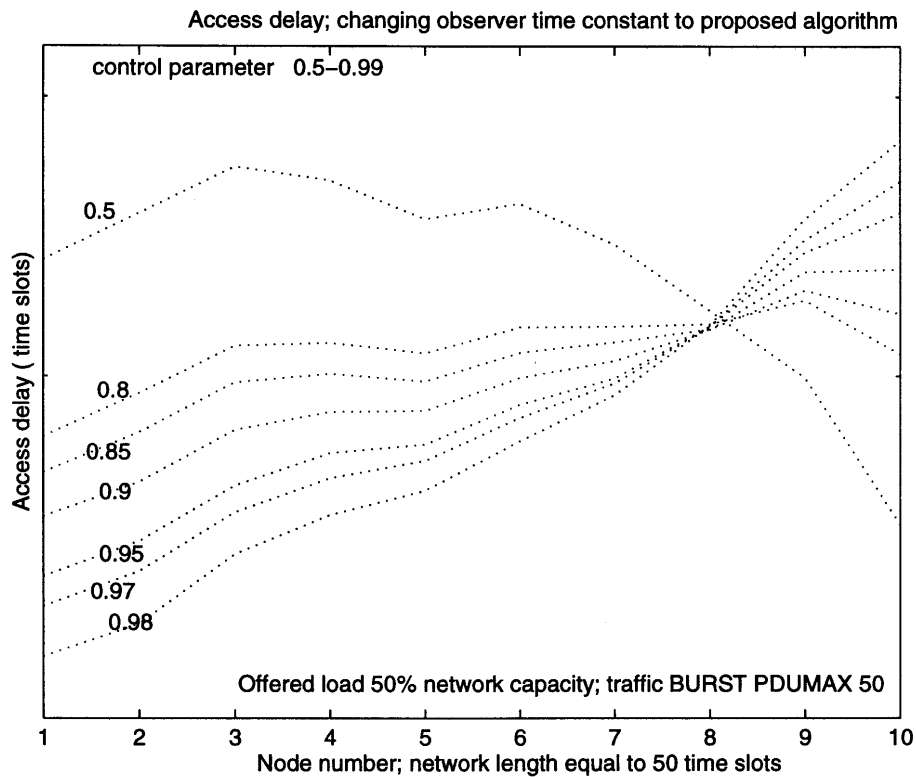


Figure 7: Access delay obtained by changing observer time constant (dotted) for a 10 node 2 wavelength 50 slot length bus network .

References

- [1] *IEEE Communications Magazine*, vol. 36, no. 2, February 1998.
- [2] C.A.Brackett, "Dense Wavelength Division Multiplexing Networks: Principles and Applications," *IEEE Journal on Selected Areas in Communications*, vol. 8, pp. 948-964, Aug. 1990.
- [3] IEEE Std 802.6, "Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network (MAN)", 1990
- [4] I. Chlamtac, V. Elek, A. Fumagalli, "A Fair Slot Routing Solution for Scalability in All-Optical Packet Switched Networks," *Journal of High Speed Networks*, vol. 6, no. 3, 1997.
- [5] S. Banerjee, B. Mukherjee, "FairNet: A WDM-based Multiple Channel Lightwave Network with Adaptive and Fair Scheduling Policy," *Journal of Lightwave Technology*, vol. 11, no. 5-6, pp. 1104-1111, May/June 1993.

- [6] M. Conti, E. Gregori, L. Lenzini "A Methodological Approach to an Extensive Analysis of DQDB Performance and Fairness", *IEEE Journal on Selected Areas in Communications*, Vol 9 (1991), No. 1, pp. 76-87.
- [7] M. Conti, E. Gregori, L. Lenzini "A Comprehensive Analysis of DQDB", *European Transactions on Telecommunications and Related Technologies*, Vol 9 (1991), No. 1, pp. 76-87.
- [8] E. Hahne, A.Choudhury, N. Maxemchuck, "Improving the Fairness of Distributed-Queue-Dual-Bus Networks", *Proc. IEEE INFOCOM90*, pp. 175-184.
- [9] E. Hahne, A.Choudhury, N. Maxemchuck, "DQDB Networks with and without Bandwidth Balancing", *IEEE Trans. on Communications*, Vol. 40 (1992), No. 7, pp. 1192-1204.
- [10] B. Mukherjee, "Optical Communication Networks" *McGraw-Hill*, New York, July 1997.
- [11] L. Kleinrock, "Queueing systems vol I e II," *Wiley Interscience*, New York, 1976.