A Distributed Bandwidth Allocation Algorithm for Gbit/s LANs

T. Antonakopoulos, S. Koutroubinas, J. Koutsonikos and V. Makios

Euromicro Journal of Microprocessing and Microprogramming

VOL. 38 NO. 1-5, SEPTEMBER 1993, pp. 301-308

Microprocessing and Microprogramming 38 (1993) 301-308
North-Holland

A Distributed Bandwidth Allocation Algorithm for Gbit/s LANs

T. Antonakopoulos, S. Koutroubinas, J. Koutsonikos and V. Makios

Laboratory of Electromagnetics, University of Patras, 26500 Patras, Greece

In this paper, a medium access protocol for high speed Local Area Networks is described. The access protocol uses fixed length cells in slotted operation and supports both synchronous and asynchronous types of traffic. The network bandwidth is allocated to the various stations using a cell reservation scheme during the connection set-up of synchronous traffic. The synchronous traffic of each station is composed of various independent information streams of constant or variable bit rate, has higher priority than the asynchronous traffic and is served using the nonpreemptive discipline. The access method ensures the transmission of synchronous traffic by fulfiling its Quality of Service (QOS) requirements in terms of transmission delay variations and cell rejection rate. Various performance results are presented and discussed in order to highlight the ability of the access method to support various types of traffic in the Gbit/sec LAN area.

1. INTRODUCTION

Current research in the area of Local and Metropolitan Area Networks (LANs/MANs) is performed in order to provide transmission speeds beyond 1 Gbit/sec and cover large geographical areas [1], [2]. The medium access schemes used in these networks must operate efficiently and be independent of the inter-station distances and the transmission speed. Due to the high operational speed, the access protocol must be as simple as possible in order to be implemented in hardware cost-effectively.

The basic requirement of a high-speed LAN is to support different types of traffic which usually have contradictory requirements. The synchronous traffic requires small transmission delay, bounded transfer delay jitter and can afford a small undetectable bit error rate, while for the asynchronous traffic which has no timing requirements, no undetectable bit error rate is acceptable. In the LAN/MAN environment, new access methods with these characteristics have already been proposed and developed [3], [4], since the typical LAN medium access protocols are inadequate for the high-speed LAN/MAN environment [1], [5]. Even the FDDI access method which operates at 100Mbit/sec and covers distances of tens of kilometers, cannot operate effectively in the MAN environment, since its throughput drops when the network speed or the size of the covered area increases.

In this paper, a new access method for usa in the HSLAN/MAN environment is described. This access method is applicable in ring topologies (single or dual) for supporting various types of traffic, like data, voice, video etc. The access method uses fixed length packets, called cells, in a slotted operation. The method incorporates a distributed algorithm for allocating the bandwidth to the various stations by comparing the total requested bandwidth for synchronous traffic to the actually measured traffic and adjusting the offered load accordingly. The basic idea of the proposed access technique is the following: when a new synchronous request is applied, a bandwidth allocation procedure is performed by requesting a number of idle cells per time unit. If the request has been accepted, the service of the synchronous traffic begins, otherwise no transmission is allowed. Each station transmits its own cell if there are no data in its ring buffer, otherwise the station defers its access until the ring buffer is empty. In cell-based transmission systems, the buffer is considered empty when it contains an idle cell. The idle cells are used to insert new data in the cell stream or for decoupling the differences in the transmission speed between adjacent nodes. The cells are removed from the network by the destination station achieving an increased network throughput, which approaches 400% in dual ring structures. No asynchronous cells are transmitted in the network from a station if there are synchronous cells in its buffers. The proposed access method has been based on the load-control method proposed by J. Limb in [6], but it has been extended to be used in a network with various types of traffic.

As it has been shown, the management of the network load using traffic measurements can be used to improve the network performance and probably to satisfy the requirements of various traffic classes. The Load-Controlled Scheduling of Traffic method measures the traffic flow on the transmission medium in each station independently and adjusts the station transmission rate in order to maintain the total traffic flow on a target value. In our method the load control is used not only for adjusting the offered load in different stations but also to implement the bandwidth reservation scheme for the synchronous traffic and to distribute the allocated bandwidth within the different types of traffic.

Section II gives a brief description on the implications improved by the introduction of the cell concept in the LAN environment and its relation with the B-ISDN development, emphasizing the main points which will be used in section III. In Section III, the basic medium access method is described along with various reliability improvement procedures. Finally in section IV, the proposed access method is identified as an appropriate method for Gbit/s LANs and MANs.

2. THE CELL CONCEPT IN THE LAN ENVIRONMENT

The evolution in the services provided by telecommunication systems has a great impact to the techniques used in the various levels of communication networks, e.g. in the transport protocol. This evolution has mainly affected the used access method since it must be flexible enough to be adapted easily to the requirements of the forthcoming services. The most promising medium access technique is the Asynchronous Transfer Mode (ATM) which is a special case of the Fast Packet Switching concept [2]. Its basic characteristics are the use of small, constant length cells as the primary transmission unit, the use of statistical multiplexing in combining different services into the same transmission medium and that it is service independent. The ATM uses no error protection or flow control on a link-by-link basis, operates in a connection-oriented mode and requires minimum header functionality. These characteristics result to a better time transparency performance, (especially for real-time services) and lower overall bit error rate.

Although the ATM concept was initialy developed for the Broadband ISDN, it has also affected seriously the HS LAN/MAN environment. The access methods of DQDB, CRMA etc [1], [4], [5] use small cells, like the cells used in ATM, and incorporate various control mechanisms in order to mix delay sensitive services with delay tolerant data.

The use of the ATM concept in the LAN/MAN area must take into account that the transmission medium is shared between the various stations using distributed switching. This is one of the differences between MANs and B-ISDN, where a centralized connection-oriented approach is used. When a new access method is examined if it is appropriate for use in the LAN/MAN area, various functional parameters are verified, such as the independence of the protocol performance from the number of stations and from the distance of the covered area, as well as its ability to perform bounded transfer delays for real time traffic.

In the integrated services LAN environment three classes of traffic are usually considered, the isonchronous, the synchronous and the asynchronous class [7]. The isochronous traffic, which has the most strict requirements, requires constant delay while the synchronous traffic requires bounded transfer delays. The asynchronous traffic has no delay limit and usually is supported by the network bandwidth not used by the other two classes of traffic. In our discussion, the isochronous traffic will be considered as a special case of the synchronous traffic, where the delay bounds approach the mean delay value by using a type of elastic buffer at the receiver side.

3. THE BANDWIDTH ALLOCATION ACCESS PROTOCOL

The network under consideration operates in 1 Gbit/sec and uses small and fixed length cells with a destination release method. Although the destination release scheme increases the MAC complexity, it provides improvement in delay and throughput levels, allowing spatial reuse of the available bandwidth. The use of constant length packets allows easy implementation of the access mechanism and for high

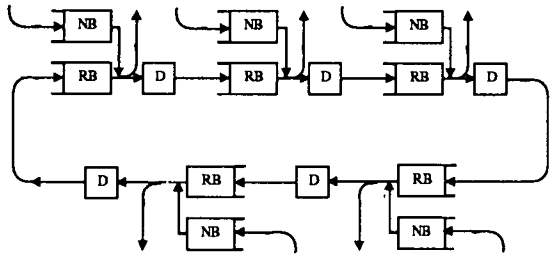

Fig. 1: The Medium Access Method model.

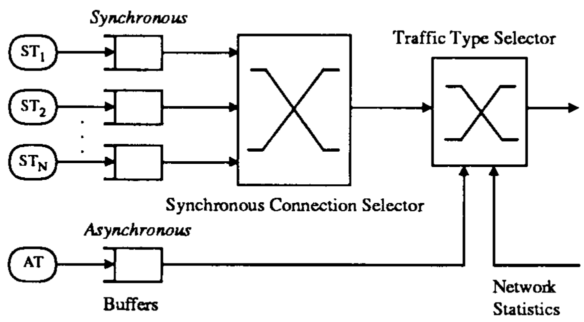

Fig. 2: Node Buffer Structure.

speed networks this can be used to achieve constant delay during the transmission of a cell in a specific connection.

The network topology is a single ring but the description that follows applies also on dual ring networks where the second ring is used to double the network throughput or for fault-tolerance purposes. The Medium Access Control sublayer is composed of three types of buffers. The ring buffer stores each cell which passes from the node, irrespective of its destination address and is one cell long. For supporting the synchronous traffic, various synchronous traffic buffers are used but they are managed by the access protocol as a single, high priority buffer. The multiplexing of the various synchronous buffers is independent of the access protocol and can be based either on their mean cell generation rate or on the relation of their instantaneous offered load. The asynchronous buffer is used for storing the cells generated by connectionless frames transmitted through the network. The data frames are segmented in the ATM Adaptation Layer (AAL) and the generated cells are handled by the medium access layer with no time relation. The operation and the characteristics of the ATM Adaptation are exactly the same as they have been described in [2]. In Figure 1 the model of the proposed Medium Access method is given. Each node is represented by the ring buffer and the node buffer. The node transmits a cell (irrespective of the cell's type) when the ring buffer contains an idle cell or a cell destined to the specific node. The node buffer structure is given in Figure 2. The Traffic Type Selector collects the network statistics and determines if a cell will be transmitted from the node buffers or the node will defer. The algorithm used for that decision will be described later in this section. The Synchronous Connection Selector implements the synchronous buffer multiplexing and determines which buffer will transmit in the next access round.

From the above description it is obvious that, although the access protocol is very simple, it can not guarantee the time restrictions of a real-time traffic and a control mechanism must be incorporated. The MAC inserts a constant delay to the passing-through cells irrespective of their destination. This delay is added in each passing-through cell, so a cell transmitted from a source node will arrive at the destination node with constant delay, while the instantaneous traffic conditions do not affect its transmission delay. The total cell delay is composed of two parts, the 'queueing' delay and the 'passing-through' delay. The 'queueing' delay is determined as the time elapsed from the reception of a cell in the output buffer up to the transmission of the cell into the network. The 'passing-through' delay is determined as the time a cell needs to pass through all the intermediate stations from the source to the destination node. The 'passing-through' delay is constant and depends only on the cell length, the transmission speed and the number of stations the cell has to pass through. In long inter-station connections, the delay imposed by the MAC layer is small compared to the delay caused by the point-to-point connections. Due to the high transmission speed and small cell length, an inter-station link of 10km acts like a buffer of 78 cells long, when the transmission speed is 1 Gbit/sec and the cell is 53 bytes long.

The problem which arises in order to support synchronous traffic is due to the queueing delay variations caused by the upstream traffic inside the network. Since the 'passing-through' delay is constant for a given connection, the jitter of the 'queueing' delay must be regulated for supporting synchronous traffic. In order to minimize the effect of the network traffic to the delay variations of the same synchronous stream, a traffic monitoring method is used to control the type and the amount of the offered load.

The offered-load control algorithm used in this access method is based on traffic monitoring and is fully distributed because each station decides on its offered load independently from the others and by using only its traffic monitoring statistics. This control algorithm uses the basic concept of LOCOST [6] which states that the appropriate scheduling of traffic can improve the network performance and if it is properly managed, it can change the characteristics of the access method.

Before starting the description of the bandwidth allocation algorithm, we will examine the way the synchronous cells are generated. The synchronous traffic generates cells in small bursts during a connection and the number of cells contained in each burst depends on the type of the synchronous traffic (Constant Bit Rate (CBR) or Variable Bit Rate (VBR)), the cell length, and on some service dependent characteristics (e.g. the interframe period of a video source). Each synchronous traffic is characterized by its mean cell generation rate (the average number of cells generated during a time unit) and its maximum cell generation rate, which is measured in VBR services when the primary information complexity is maximized.

The basic parameter of the access method is the Total Traffic Window, W_T, which gives the portion of the traffic used by the algorithm to collect its statistics. As it will be shown in the performance analysis, the value of W_T will determine the protocol efficiency.

The 'Total Traffic Window' which is an indirect expression of the available bandwidth, is divided in two parts: the first part is devoted to the synchronous traffic only and the second part is used to support both classes of traffic. The portion of the bandwidth devoted to the synchronous traffic is determined by two parameters: the Connections' Window, W_C, and the instantaneous synchronous traffic measurement, S. The Connections' Window is the sum of the number of cells generated during the Total Traffic Window by each synchronous connection. The W_C is updated each time a new connection is established and when a synchronous connection is released. The instantaneous synchronous traffic measurement, S, indicates the number of synchronous cells measured in the last W_T cells passed through each node.

When a new synchronous connection must be established, the node estimates the mean number of cells which will be generated by the specific service during W_T and if this number is less than the difference of W_T and W_C, then the connection is accepted and a Connections' Window Update procedure begins. Otherwise, the connection is rejected because there is no available bandwidth to support the specific service effectively. During the Connections' Window Update procedure, the node transmits a control type cell with destination address its own address. This cell passes through all the network nodes, so each node is informed about the new status of the synchronous traffic connections. The same procedure is used when a connection is released and the Connections' Window Update procedure informs all nodes to decrease their W_C value by the same number of cells. Using this mechanism all stations use the same protocol parameters and the protocol fairness is guaranteed.

Figure 3 shows the synchronous traffic access probability function. Each node knows the mean and the maximum number of cells, w_{ij}, reserved for its own usage during W_T. If a smaller number of cells, than its mean value, has been transmitted by this station, then when an idle cell passes, the node removes it and transmits its own cell. Otherwise, if the node has transmitted more cells than its mean value, the access probability depends also on the parameter S. When a small percentage of the allocated bandwidth is used, the node is allowed to transmit up to its maximum offered load. The allowed synchronous offered load decreases as the parameter S increases. The access probability function of the asynchronous traffic is shown in Figure 4. When the node has no synchronous cells to transmit and detects an idle cell in its ring buffer, the transmission of an asynchronous cell depends on the relation of W_C/W_T (the bandwidth percentage allocated exclusively for synchronous traffic) and on the number of cells not used for synchronous traffic during the Total Traffic Window. The smaller the bandwidth portion allocated to synchronous traffic, the bigger the probability of

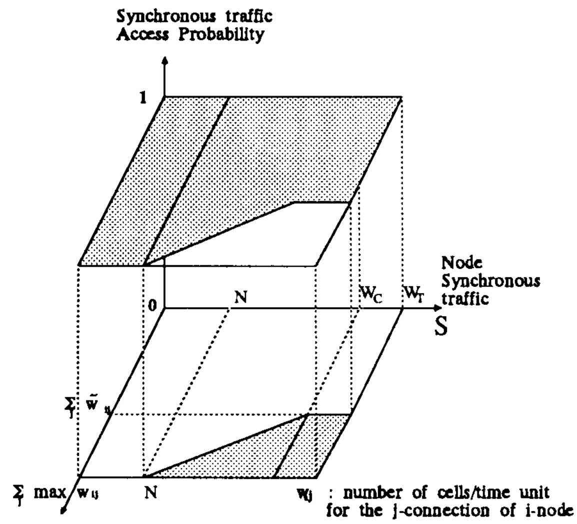

Fig. 3 : The access permission probability function for the synchronous traffic.

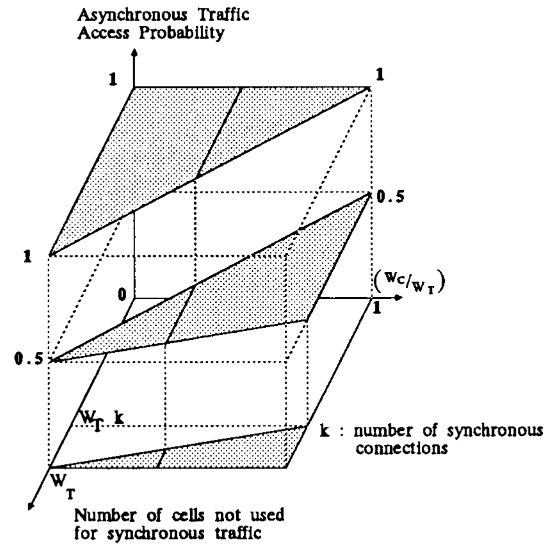

Fig. 4 : The access permission probability function for the asynchronous traffic.

asynchronous transmission. The 0.5 value of the probability can be easily implemented in hardware by using a simple flip-flop which toggles each time an idle cell is detected and the access protocol statistics fill in the 0.5 area of Figure 4.

The access method gives the requested bandwidth to the synchronous traffic by varying the asynchronous offered load. The synchronous traffic throughput of a node varies around its mean value, for a given number of connections, depending on the total statistical multiplexing of the synchronous traffic. When the portion of the transmitted synchronous cells to the total traffic decreases, the algorithm assumes that more bandwidth is available for asynchronous traffic, and the asynchronous offered load increases.

Up to now, the basic access protocol has been described. In real networks there are also some other important topics which must be addressed. These are: how the network handles the occurence of a transmission error in a control cell, and how a new node participates in the evolution of the bandwidth allocation algorithm. How a node is removed from the network without introducing a permanent error in the algorithm parameters? In the rest of this section an answer to the above mentioned problems will be given.

When a new node wants to participate in the algorithm evolution, first it must be informed about the value of the W_C parameter. This is achieved by sending a control cell down to the network. The node which will receive this cell, will change its information field by introducing the value of W_C and will transmit it again into the network. The cell will arrive at the generated node, and the value of W_C will be known.

When a transmission error occurs in a control cell, the originating node will not receive it back and a network reset procedure will begin. The node which did not receive its own control cell, will generate a reset cell by indicating to all the other nodes to clear the value of W_C. When a station receives a W_C-clear command, it sets W_C to zero and reinforms all the other nodes about its current synchronous connections by using the previously described Connections' Window Update procedure. After a small number of cells, the bandwidth allocation algorithm stabilizes again.

In normal circumstances, when a node has to be shut down, uses the Connections' Window Update procedure for informing the other nodes about the release of its own connections. In abnormal conditions, a node suddenly goes down and its preallocated bandwidth for synchronous traffic remains unchanged in the other nodes. In order to resolve these undesirable situations, the use of periodic reset procedures is proposed to be used. The reset procedure has negligible impact on the access protocol evolution and long inter-reset times can be used due to the high reliability of today's networks.

4. THE PROTOCOL PERFORMANCE ANALYSIS

In this section the simulation parameters are given and the respective results are presented. Initially, we examine the network performance for various number of network nodes and inter-node distances using the throughput offered load characteristic curves. Then the influence of the Total Traffic Window length on the performance of the medium access method is determined. Finally, it is indicated how the inter-arrival times of a CBR service at a receiving node are distributed. The ability of the proposed access method to support synchronous traffic is also verified.

Our simulation model considers a Metropolitan Area Network, based on optical fiber with transmission speed at 1Gbit/sec. The cell length is equal to 53 bytes, which consists of 48 bytes of information and 5 bytes of the cell header.

Although usually two traffic scenarios, the light and the heavy scenario, are considered in various simulation models, in this analysis we have considered only the 'heavy' scenario which gives 80% of the total node offered load to the synchronous traffic while the rest of the node's load is composed of asynchronous data. The 'heavy' scenario has been selected to be presented because it represents the worst case traffic conditions for the protocol under consideration.

In each station the asynchronous traffic consists of variable length file transfers, where the mean file length is 10 kbytes and the length distribution is exponential. The whole file is considered to be available for transmission immediately after its generation time in the station. The file generation procedure is a Poisson process, with mean value equal to 90 files/sec. The synchronous traffic consists of connections established in time instants according to a Poisson process, while their duration is uniformly distributed between 7 and 9 secs. The synchronous data stream is either of constant bit rate or of variable bit rate. The CBR services generate 6 Mbits/sec, while the VBR services generate a mean of 9 Mbits/sec and their upper limit is 12 Mbits/sec. Each node can support up to four synchronous services in the same time. The synchronous traffic model has been based on the analysis presented in [8]. The synchronous traffic is generated in bursts of cells and the number of cells in each burst depends on the traffic characteristics. When a new burst arrives at a synchronous buffer, the cells in the occupied buffer are dropped.

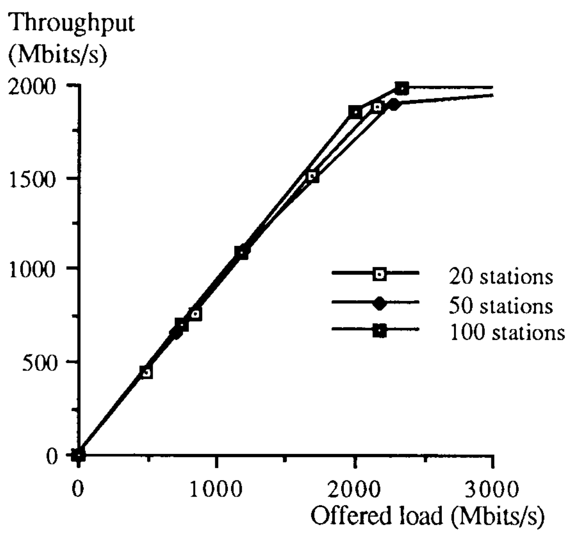

Fig. 5: Network throughput versus offered load for various number of network stations.

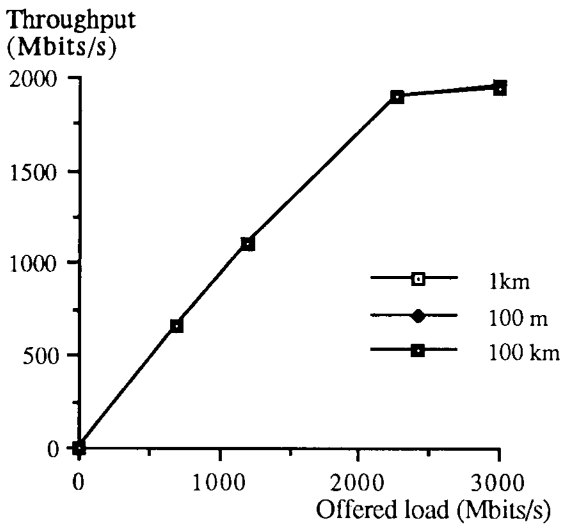

Fig. 6: Network throughput versus offered load for various inter-station distances.

Figure 5 gives the total network throughput versus the total offered load for various numbers of network nodes. The network performance is independent from the number of its nodes and the throughput remains in its maximum value even when the offered load increases remarkably. The offered load for the synchronous traffic, which is 80% of the total traffic, is

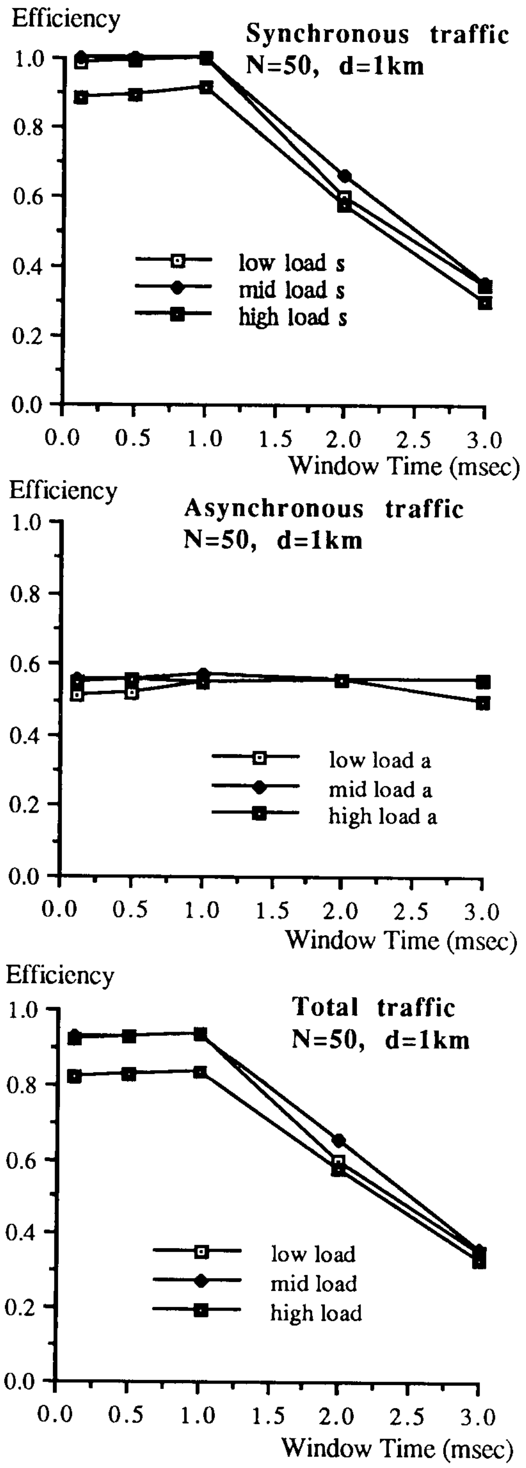

Fig. 7: The access protocol efficiency versus the Total traffic for various values of the offered load.

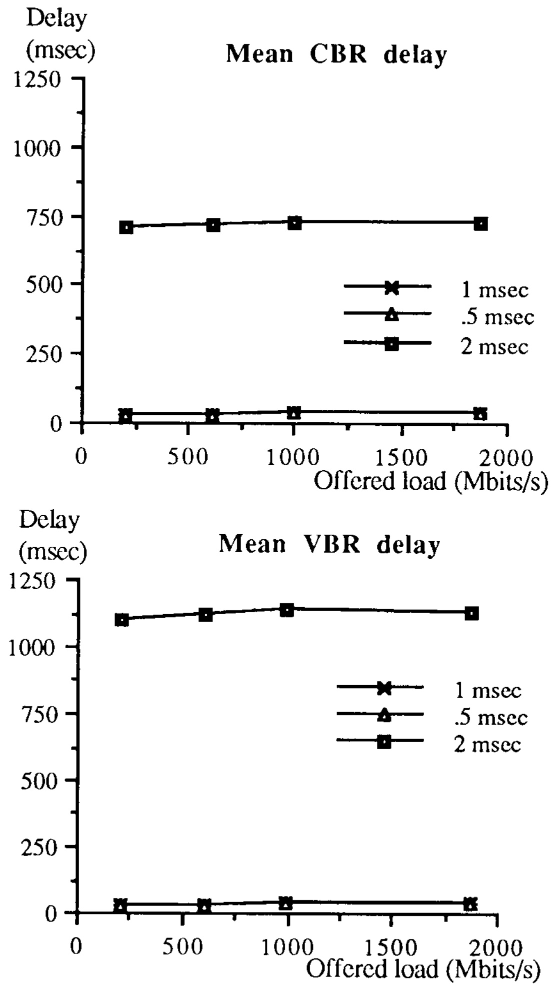

Fig. 8: The mean delay for synchronous traffic cells versus the network throughput for (a) CBR and (b) VBR services.

the requested offered load, but due to the used control mechanism, no connection is accepted if its QOS parameters cannot be guaranteed. As it will be shown in the delay analysis, this control mechanism results to small delay variations in the cells of the accepted connenctions. As it is shown in Figure 6, the network performance is also independent from the inter-node distances. From these results it is obvious that the proposed access method fulfils two basic requirements in order to be used in the HSLAN/MAN area.

The influence of the Total Traffic Window to the network efficiency is highlighted in Figure 7. As network efficiency is determined the ratio of the network throughput to the offered load. The synchronous traffic efficiency decreases as the W_T value increases, because the adaptability of the control algorithm decreases. The efficiency of the asynchronous traffic remains almost constant because its access probability does not depend on the value of W_T but on the ratio W_C/W_T, which depends only on the percentage of the synchronous traffic to the total traffic.

The mean delay for synchronous cells for CBR and VBR traffic is shown in Figure 8. Due to the used control mechanism, which does not allow new synchronous connections to participate in the bandwidth allocation procedure if their QOS parameters are not satisfied, and due to the high priority of the synchronous traffic, the mean delay is not seriously affected by the network traffic. When W_T increases the synchronous delay increases since the control algorithm slows down.

5. CONCLUSION

In this paper, a new Medium Access protocol appropriate to support integrated services in the Gbit/sec LAN environment was presented. The access protocol uses a distributed algorithm for implementing a cell reservation scheme for the synchronous types of traffic, and a traffic measurement technique is used for deciding the access probability for both types of traffic. The use of constant length cells, in combination with the pipeline structure of the ring buffer, guarantees a constant transmission delay to the cells of the same connection and the end-to-end delay is affected mainly by the variable queueing at the source station due to the upstream traffic. These delay variations are bounded due to the used control mechanism.

As the performance analysis has shown, the proposed access method is appropriate for use in the LAN/MAN environment because its performance is independent from the number of the network nodes and from the distances of adjacent nodes. The proposed access method can guarantee bounded transfer delays for the synchronous traffic and is appropriate for integrated services networks.

REFERENCES

[1] G.C. Kessler and D.A. Train: "Metropolitan Area Networks", McGraw-Hill, 1991.

[2] Martin de Prycker: "Asynchronous Transfer Mode: Solution for broadband ISDN", Ellis Horwood Ltd., 1991.

[3] J. Y. Chao, W.T. Lee and L.Y. Kung: "A New Buffer Insertion Ring with Time Variant Priority Scheme to Facilitate Real-Time Image Transmission on a High Speed Integrated Local Area Network", Fourteenth Conference on Local Computer Network, Minneapolis, USA, October 1989, pp. 212-218.

[4] M. Mehdi Nassehi: "CRMA: An Access Scheme for High-Speed LANs and MANs", Supercomm/ICC'90, Atlanta, USA, April 1990.

[5] H. Ohnishi, N. Morita, S. Suzuki: "ATM Ring Protocol and Performance", IEEE International Conference on Communications, Boston, MA, June 1981, pp. 394-398.

[6] John O. Limb: "Load-Controlled Scheduling of Traffic on High-Speed Metropolitan Area Networks", IEEE Trans. on Communications, Vol. 37, No 11, November 1989, pp. 1144-1150.

[7] D.B. Hehmann, M. G. Salmony and H.J. Stüttgen: "High-Speed Transport Systems for Multimedia Applications", IFIP Workshop on Protocols for High-Speed Networks, North Holland, May 1989, pp. 303-321.

[8] D. P. Heyman, A. Tabatabai and T. V. Lakshman: "Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks", IEEE Trans. on Circuits and Systems for Video Applications, Vol. 2, No 1, March 1992, pp. 49-58.