The Effect of the Gateway Location on the Performance of the DQDB Network

Hitoshi Nagano and Shuji Tasaka
Department of Electrical and Computer Engineering,
Nagoya Institute of Technology, Nagoya 466, Japan
e-mail: nagano@lap.elcom.nitech.ac.jp

Abstract

This paper presents a study of the DQDB network when one of its stations has a role of a gateway. Under this context, some of the packets generated by each station are destined to another network. These packets are switched in the gateway, which is a DQDB station. In particular we study the effect of the physical location of the gateway on the performance. We present the traffic workload and a positioning policy for the gateway. Our results show us that bad positioning leads to unfairness conditions.

1 Introduction

The distributed queue dual bus (DQDB) medium access control (MAC) protocol is a potential candidate to be an international standard for Metropolitan Area Networks (MANs) in the near future. At present, it has been adopted by the IEEE 802.6 standards committee [1]. There is a necessity for investigating the performance of the DQDB.

We propose here a problem concerning the DQDB network when one of its stations has a role of a gateway. In particular we study the effect of the position of the gateway on the network performance. For the type of traffic utilized (assymmetric load), we will find that bad positioning of the gateway can lead to high unfairness.

In most of the work done until now, the performance of the DQDB in terms of bus access delay and throughput was analyzed with respect to only one of the two buses. This analysis is satisfactory if we consider that the workload type follows a symmetric policy[3]. In the present paper, we discuss an assymetric workload, and therefore the analysis of both buses becomes necessary.

Much research on the performance of the DQDB network has been done up to now. In [3] network simulations are performed under various conditions. It was pointed out that an increase in the offered load to the bus can lead to unfairness behavior, which means that the performance of a station would depend on its physical location. In this paper, we use this result to develop our gateway positioning policy. We analytically determine the range of permissible positions for the gateway in order to keep the load on both buses in low levels. The relationship between the unfairness and the offered load can be obtained by analytical models such as those given in [4] or [5]. In order to discuss the degree of unfairness, we define an unfairness measure in the present work.

There are other ways to attack the unfairness problem, such as the Bandwidth Balancing Mechanism (BWB) [2]. This approach is effective when the network is overloaded[3]. By adding an extra counter to the station, fairness (in terms of equally distributed bandwidth) is achieved. The BWB has been adopted by the IEEE standards commitee.

In the IEEE standard, three priority levels are used, and two types of service, namely, isochronous and asynchronous, are offered. In this paper, only the asynchronous type of service with one priority level is considered. Also, we do not use the BWB mechanism.

Section 2 describes the DQDB protocol. Section 3 presents the gateway positioning problem and its solution. Section 4 provides numerical results. Finally, in Section 5 some conclusions are given.

2 DQDB Protocol Description

The network consists of two high-speed unidirectional buses, which we call bus A and bus B in this paper. Fig. 1 shows the network topology. These two buses carry data in opposite directions. The stations are placed along the buses. As the speed is high and as the distance between stations is considerable, the propagation time becomes an important issue. It is reasonable that between two adjacent stations, a certain number of slots can be placed. For example, the simulations done in [3] consider the interstation gap as being a distance that corresponds to three slots.

Each bus has two components called head and end. The head continuously generates slots of fixed duration, which travel along the bus. The slot is composed of two parts. In the single priority case, the part that carries the protocol control information uses only two bits: the busy bit and request bit. The busy bit indicates whether or not the slot is occupied by a packet. It is set when the slot is not empty. The request bit is set whenever a user wants to make a request. The data part follows this protocol control information part.

Once a station writes into a slot, the slot passes through all the downstream stations, and the data part is read by the destination station(s). It is not permitted for any downstream station to overwrite into a slot. A given user (say user #r) uses an empty slot on bus A to transmit information to user #s ($s > r$), which we call a downstream user with respect to bus A. The same user #r also uses bus B to inform user #t ($t < r$) about a packet it wants to transmit on bus A. In other words, a slot reservation for one bus is done on the other bus. As the operation of both buses are the same, we call bus A the forward bus and bus B the reverse bus.

Each user is either idle, when there is nothing to transmit, or otherwise, active. When the user is active, it may be either counting down or transmitting a packet. When the user is idle, it keeps count of the number of outstanding requests. When a queued packet enters the MAC layer, the countdown process begins. The countdown process finishes when the user enters a state in which it is capable to transmit the packet in the next empty slot. After transmission the station returns to the idle state.

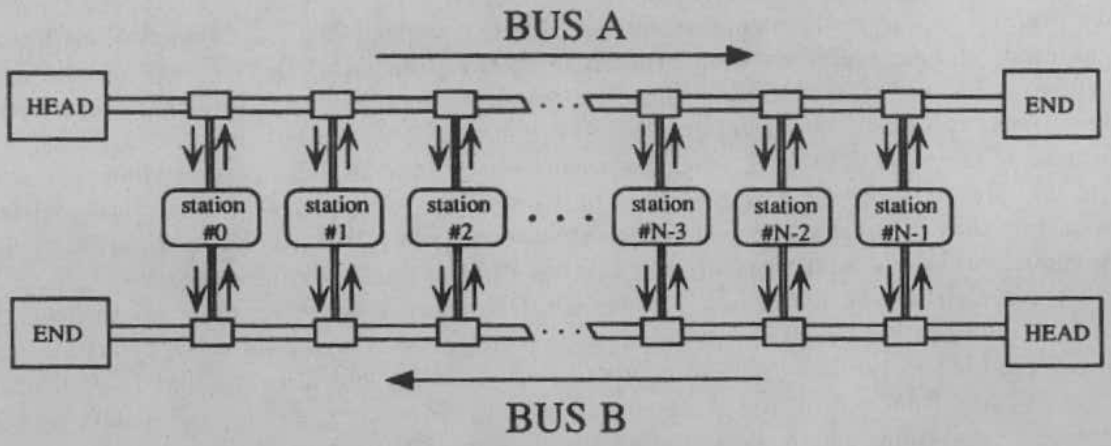

Figure 1: DQDB network topology

In practice, each user has a counter named request counter (RQ_CTR). In the idle state, it increases by one for each slot (on the reverse bus) containing a request. It also decreases by one for each slot (on the forward bus) that does not contain a packet, until it reaches zero. When a user becomes active, that is a packet has entered the MAC layer, it transfers the contents of the RQ_CTR to a second counter named countdown counter (CD_CTR) and resets the RQ_CTR. The user now begins the countdown process. The user also sends on the reverse bus a request in the first slot with request bit not set. The CD_CTR decreases by one for each empty slot (busy bit not set) passing on the forward bus. In the meantime, RQ_CTR increases by one for each slot with the request bit set on the reverse bus. When the CD_CTR reaches zero, the user is permitted to transmit the packet in the first empty slot that comes. After transmitting, the user returns to the idle state if it has no queued packet. If there is any packet waiting to enter the MAC layer, the

user begins the countdown process again.

We can see that both counters register the position of a self-request in the distributed queue. It is also clear that each user sees a different distributed queue, because all users are not informed of the reservations at the same time. This characteristic leads to the unfair and unpredictable behavior at high user loads [3],[2]. In order to solve this problem, many mechanisms have been proposed. The bandwidth balancing machine (BWB)[2] is one proposal that fixes the unpredictability issue, but with loss of a small fraction of the usable bandwidth.

In the present model we do not consider the BWB. Instead, we try to find conditions that minimize the offered load on both buses, in order to avoid unfairness.

3 The DQDB network with one gateway

In the standard [1] the network presented in Fig. 1 can be connected to others via a gateway. In Fig. 2 we present this case. It is clear that the gateway influences the workload type.

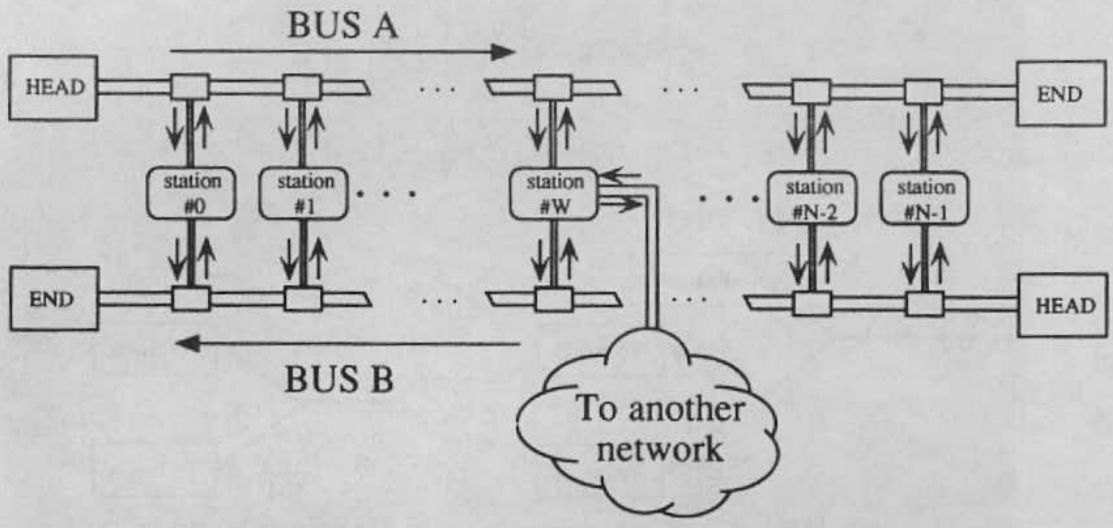

Figure 2: DQDB network when one of its stations is a gateway

3.1 Workload type

In the following discussion we consider that the traffic is measured by means of the packet generation probability per slot. In the symmetric policy for the workload type, each station generates a packet with a probability proportional to the number of downstream stations. In other words, for a station #r in Fig. 1, a generated packet is routed to bus A with probability $(N-1-r)/(N-1)$, while it is routed to bus B with probability $r/(N-1)$. We now consider that one of the station works exclusively as a gateway. We take the following assumptions:

A1): The network has N stations as in Fig. 2.

A2): The station index of the gateway is W.

A3): Each station except for the gateway generates a packet with probability σ in a slot

A4): A generated packet has its destination in another network (i.e. it goes to the gateway) with probability γ.

A5): The packets that are destined to the present network follow the symmetric policy.

A6): The gateway generates a packet with probability σ_w. This traffic is also routed according to the symmetric policy.

In Fig.3 we present the routing probabilities of a given station $\#k$, for the three cases: $0 \leq k \leq (W-1)$; $k = W$; $(W+1) \leq k \leq (N-1)$. We can easily derive the packet generation probability

per slot for each station corresponding to each one of the two buses. In Equations (1)-(6) let the superscript denote the bus name and the subscript the station index.

For $0 \leq k \leq W-1$,

$$\sigma_k^A = \sigma[\gamma + (1-\gamma)\frac{N-2-k}{N-2}] \tag{1}$$

$$\sigma_k^B = \sigma[(1-\gamma)\frac{k}{N-2}] \tag{2}$$

When the station is the gateway we have,

$$\sigma_W^A = \sigma_W \frac{N-1-W}{N-1} \tag{3}$$

$$\sigma_W^B = \sigma_W \frac{W}{N-1} \tag{4}$$

When $W+1 \leq k \leq N-1$ we obtain,

$$\sigma_k^A = \sigma[(1-\gamma)\frac{N-1-k}{N-2}] \tag{5}$$

$$\sigma_k^B = \sigma[\gamma + (1-\gamma)\frac{k-1}{N-2}] \tag{6}$$

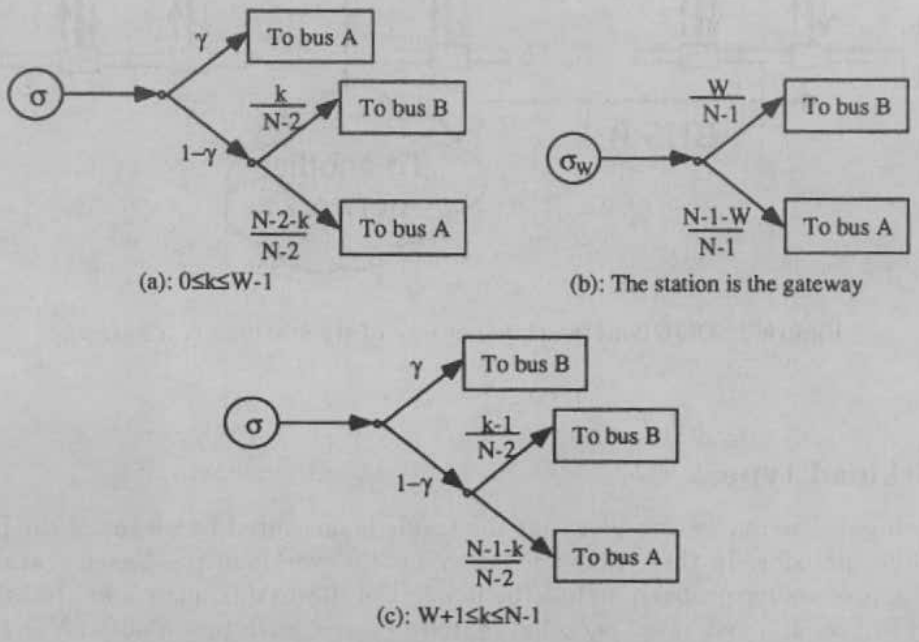

Figure 3: Routing probability

Let G_A and G_B denote the the offered load to buses A and B, respectively. We can calculate G_A and G_B and find

$$G_A = \sum_{k=0}^{N-1} \sigma_k^A = \frac{\sigma(1-\gamma)(N-1)}{2} + W\gamma\sigma + \sigma_W(\frac{N-1-W}{N-1}) \tag{7}$$

$$G_B = \sum_{k=0}^{N-1} \sigma_k^B = \frac{\sigma(1-\gamma)(N-1)}{2} + (N-1-W)\gamma\sigma + \sigma_W(\frac{W}{N-1}) \tag{8}$$

We can see that in both (7) and (8), the second term corresponds to the traffic to the gateway (outernet traffic), while the third term to the traffic from the gateway (gateway traffic). The first term stands for the total traffic closed to the present network (intranet traffic).

3.2 Unfairness measure

We now define a simple measure for the unfairness as the difference ΔD in the average delay between the most downstream station and the most upstream one. For example, for bus A in Fig. 1, this measure is expressed as

$$\Delta D_A = D_{N-1} - D_0 \tag{9}$$

where D_k represents the average packet delay of station #k.

We define the packet delay as the time (measured in slots), from the moment of a packet generation until the completion of its transmission. In [3], by means of simulations, this measure was shown to be an increasing function of the total offered load. We use this result for the next subsection.

3.3 Positioning a gateway

We now discuss the following problem. Given a constant traffic workload condition $(\sigma, \gamma, \sigma_W)$ is there any suitable position W for the gateway? By "suitable", we mean that unfairness is not problematic. From the previous subsection, we observe that imposing a restriction on the total offered load G_A or G_B would be a good criterion for keeping low unfairness.

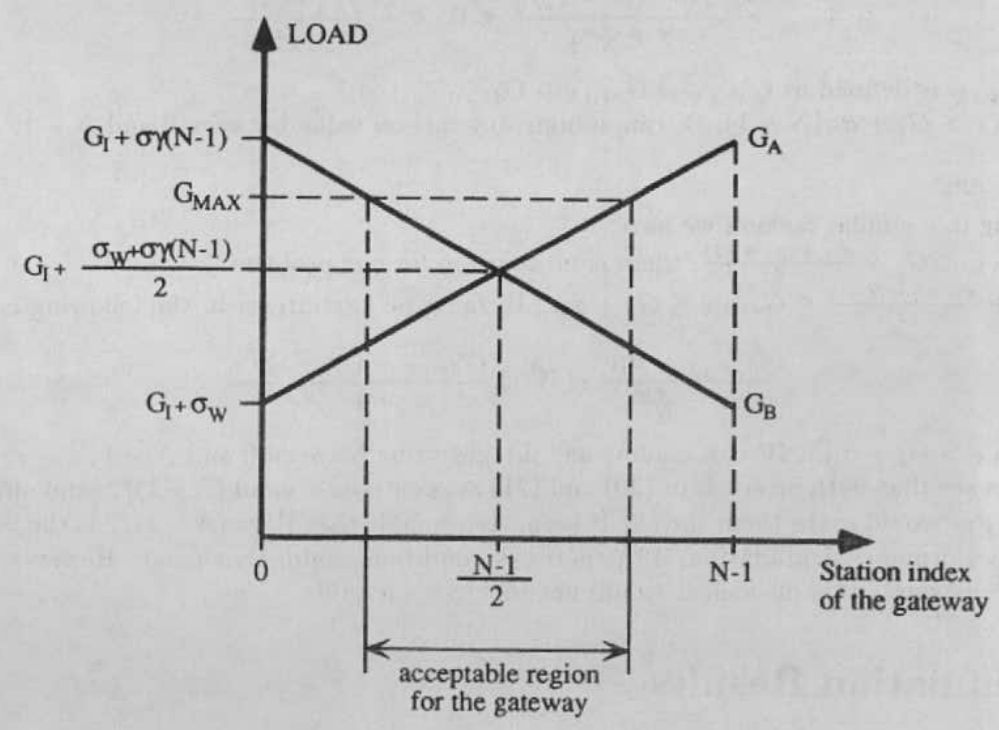

Figure 4: Buses offered load G_A and G_B as a function of W

We estabilish the maximum allowable value of the total offered load as being G_{MAX}. The choice of this limit depends on which kind of application the network is submitted to. In other words, applications that have tight delay constraints would call for low values of G_{MAX}. In any case, the relationship between ΔD and G_{MAX} can be obtained by using analytical models such as in [4] or [5]. Then we have to determine the permissible values of W such that

$$\begin{aligned} G_A &< G_{MAX} \\ G_B &< G_{MAX} \end{aligned} \tag{10}$$

Taking the first derivative of Equations (7) and (8), we have,

$$\frac{dG_A}{dW} = \sigma\gamma - \frac{\sigma_W}{N-1} \tag{11}$$

$$\frac{dG_B}{dW} = -\sigma\gamma + \frac{\sigma_W}{N-1} \tag{12}$$

and we observe that

$$\frac{dG_A}{dW} = -\frac{dG_B}{dW} \tag{13}$$

We denote G_I as the intranet traffic, which is constant to be $\frac{\sigma(1-\gamma)(N-1)}{2}$ in the present discussion. Let us divide our discussion into two separate cases:

i): $\sigma\gamma > \frac{\sigma_W}{N-1}$
We plot both G_A and G_B as a function of W in Fig. 4. The intersection of these two lines is easily determined as $(\frac{N-1}{2}, G_I + \frac{\sigma_W+\gamma\sigma(N-1)}{2})$. We observe that,
a) If $G_{MAX} < G_I + \frac{\sigma_W+\gamma\sigma(N-1)}{2}$, there is no solution for our problem.
b) If $G_I + \frac{\sigma_W+\gamma\sigma(N-1)}{2} \leq G_{MAX} \leq GI + \sigma\gamma(N-1)$, W must be any integer in the following interval:

$$\frac{G'_{MAX} - (N-1)\sigma\gamma}{-\gamma\sigma + \frac{\sigma_W}{N-1}} < W < \frac{G'_{MAX} - \sigma_W}{\gamma\sigma - \frac{\sigma_W}{N-1}} \tag{14}$$

where G'_{MAX} is defined as $G'_{MAX} = G_{MAX} - G_I$
c) If $G_{MAX} > G_I + \sigma\gamma(N-1)$, W can assume any integer value between 0 and $N-1$.

ii): $\sigma\gamma < \frac{\sigma_W}{N-1}$
Proceeding in a similar fashion we have,
a) If $G_{MAX} < G_I + \frac{\sigma_W+\gamma\sigma(N-1)}{2}$, there is no solution for our problem.
b) If $G_I + \frac{\sigma_W+\gamma\sigma(N-1)}{2} \leq G_{MAX} \leq G_I + \sigma_W$, W must be any integer in the following interval:

$$\frac{G'_{MAX} - \sigma_W}{\gamma\sigma - \frac{\sigma_W}{N-1}} < W < \frac{G'_{MAX} - (N-1)\sigma\gamma}{-\gamma\sigma + \frac{\sigma_W}{N-1}} \tag{15}$$

c) If $G_{MAX} > G_I + \sigma_W$, W can assume any integer value between 0 and $N-1$.

We can see that both intervals in (20) and (21) are centered around $(N-1)/2$, and any increase in σ, γ or σ_W would make them shrink. It seems reasonable that $W = (N-1)/2$ is the best choice to avoid performance degradation, if the network conditions suddenly change. However, owing to physical constraints this placement would not be always feasible.

4 Simulation Results

We adopt the following common parameters for the network. Some of them resemble the simulations present in [3].

TABLE I : NETWORK PARAMETERS VALUES

Network speed = 150 Mb/s
Slot size = 53 octects
Interstation gap = 3 slots
Number of nodes(N) = 50
Buffer size = 5 packets
Intranet traffic(G_I) = 0.6

Concerning the gateway, we present the simulation results for two cases:
(a) Low gateway utilization
$\gamma = 0.1$
$\sigma_W = 0.05$

(b) High gateway utilization
$\gamma = 0.2$
$\sigma_W = 0.10$

In Fig. 5 we plot the average packet delay as a function of the station index in case (a), for three positions for the gateway, namely $W = 10$, $W = 20$ and $W = 25$. In Fig. 6 we also plot case (b) for three positions for the gateway, namely $W = 10$, $W = 20$ and $W = 30$. The MAC protocol treats packets in a uniform way, independent of its destination. Therefore, the average packet delay considers all packets from outernet, gateway and intranet traffics. Tables II and III present the values of ΔD corresponding to cases (a) and (b), respectively.

The choice of G_{MAX} turns to be a difficult decision. In this paper, we are not specifying which kind of application we are dealing with. For the present discussion let us choose $G_{MAX} = 0.85$, for instance. It is shown by simulations done by the authors and by many researchers as well that an offered load of 0.85 corresponds approximately to ΔD of 10 slots time.

In case (a), we have from the previous section that any position for the gateway would keep the network fair. We can see from Fig. 5 that the average delay in the two buses is not too much influenced by the change in the position of the gateway. However, in case (b) we have from Eq. (14) that W must be in the interval: $12.25 < W < 36.75$. We see from Fig. 6 that when $W = 10$ stations with low index number experience high delay on bus B. This case can be easily seen in Fig. 4, which shows that for stations with low index number, the delay on bus B tends to get high, while the delay on bus A is not so problematic.

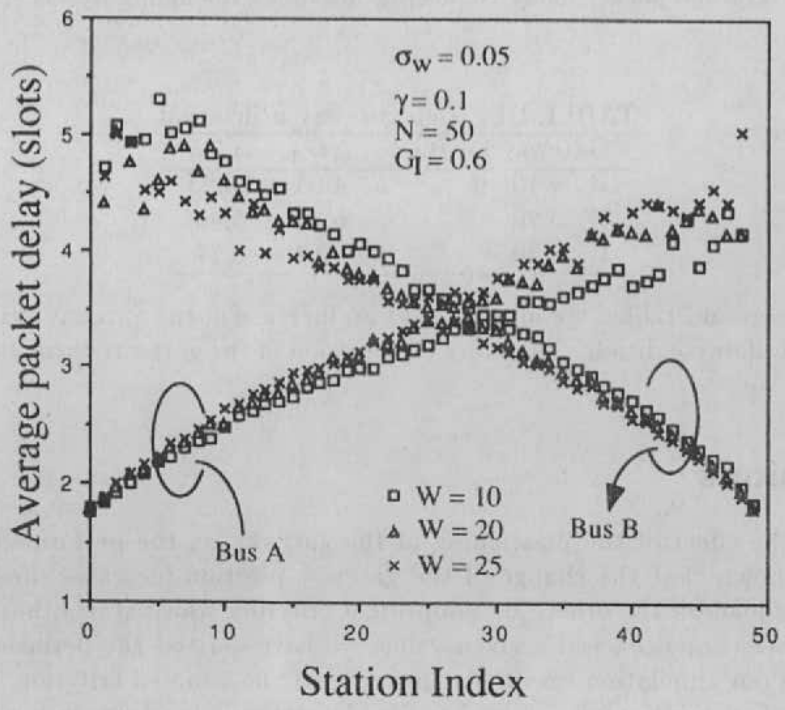

Figure 5: Average packet delay versus station index for low gateway utilization

TABLE II : Low gateway utilization

Gateway location	ΔD_A	ΔD_B
$W = 10$	2.40	2.87
$W = 20$	2.36	2.59
$W = 25$	3.23	2.83

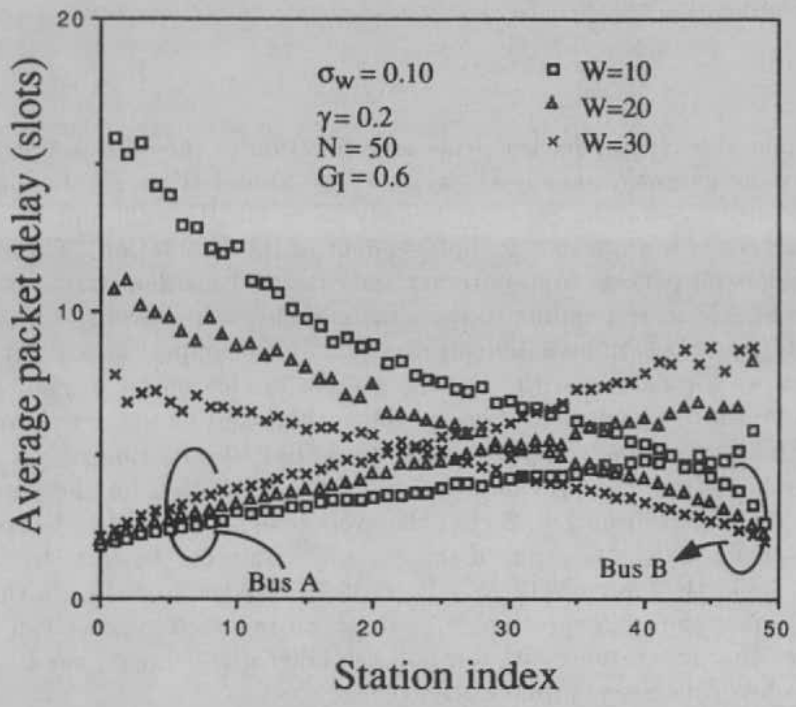

Figure 6: Average packet delay versus station index for high gateway utilization

TABLE III : High gateway utilization

Gateway location	ΔD_A	ΔD_B
$W = 10$	4.03	13.23
$W = 20$	6.30	8.40
$W = 30$	6.59	5.77

From these figures and tables, we observe that an increase in the gateway utilization can drive the network to an unfair condition. Therefore the position of the gateway turns to be an important issue.

5 Conclusions

We have studied the effect of the positioning of the gateway on the performance of the DQDB network. It was shown that the change of the gateway position increases the load on one bus, while decreases the load on the other. By adopting a criterion which states that the load on each one of the two buses cannot exceed a given value, we have derived the permissible positions for the gateway. From our simulation results we proved that the adopted criterion is reasonable. We can also state that bad positioning can lead to visible performance degradation in one of the two buses.

Although we have studied the DQDB network in the presence of a gateway, this work can be extended to a case in which one station of the DQDB network concetrates traffic. For example, the present analysis is also valid when one of the stations is a file server.

References

[1] IEEE Std 802.6-1990, Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network (MAN), IEEE, 1991.

[2] E. L. Hahne, A. K. Choudhury and N. F. Maxemchuk, "Improving the fairness of Distributed-Queue, Dual-Bus Networks," in *Proc. IEEE INFOCOM' 90* , San Franscisco, pp.175–184, June 1990.

[3] M. Conti, E. Gregori and L. Lenzini, "A Methodological Approach to an Extensive Analysis of DQDB Performance and Fairness," *IEEE J. Select. Areas Commun.*, vol. 9, no. 1, pp.76–87, January 1990.

[4] C. Bisdikian,"Waiting time analysis in a single buffer DQDB (802.6) network", *IEEE J. Select. Areas Commun.*, vol.8, no.8, pp.1565–1573, October 1990.

[5] H. Nagano and S. Tasaka, "An analytical model of a Finite Buffered DQDB station", *Technical Report of IEICE*, IN92-67, October 1992.