Journal of the Korean
OR/MS Society
Vol. 22, No. 3, September 1997

Packet Delay Analysis in the DQDB Network with a Saturated Station

Seung J. Noh*

Abstract

This paper presents an analytical model for estimating packet waiting times at stations in the DQDB network, where the most upstream station is saturated. This model is useful in comparing the extreme unfairness which downstream stations experience due to their geographical locations in accessing the medium. Each station is modeled as an M/G/1, where the service time is defined to be the time a packet spends in the transmission buffer. The service time is decomposed into five components, and in turn, the first and second moment of each component are derived in three different modes of operation. Simulation experiments are presented for model validation and results are discussed.

1. Introduction

The DQDB network has been adopted by the IEEE Standard Committee as the subnetwork for the IEEE 802.6 MAN. Performance of the network is crucial in the network design phase since it allows us to predict the performance of various design alternatives. Only a limited amount of work has been reported on the performance evaluation of the network. The main difficulty for the analysis of the DQDB network resides in the high degree of complexity and interdependence of the various processes that describe the operation of the network.

In this paper, we consider a system where the most upstream station is saturated and all other stations are stable. This system is probably unrealistic. However, the modeling of this system enables us to gain some insights into the complex behavior of the DQDB protocol, such as the effect of interactions of the stations and the unfairness feature among the stations. Each station is modeled as

* School of Business Administration, Ajou University

an M/G/1 queueing system. Since the most upstream station is saturated, the service time for a packet has a special structure, which enables us to decompose it into serveral independent components. The mean and variance of each component in the service time are computed in three different modes of operations.

There is a number of work on the DQDB protocol [2,3,4,10,13], and a comprehensive review of the work is provided in Mukheriee and Bisdikian [8]. Most of the work, however, has been conducted using simulation. There is relatively little work on analytical modeling of performance. Moreover, analytical work on the unfairness feature of the network has not been reported. This motivated our study.

The organization of the paper is as follows. Section 2 briefly introduces the DQDB Protocol. Sections 3 and 4 present the model and experimental results, respectively. Section 5 concludes the paper with a summary.

2. Overview of the DQDB Protocol

The DQDB is a medium access control protocol for forming a global First-Come-First-Served(FCFS) transmission queue among stations. It controls access to the shared medium for the stations that wish to transmit packets. It is based on a dual bus architecture as shown in Figure 1.

In the DQDB network, the upstream heads of the two buses (HOBs) continuously generate slots of fixed duration(the same in both buses) that travel along their respective buses. The duration of a

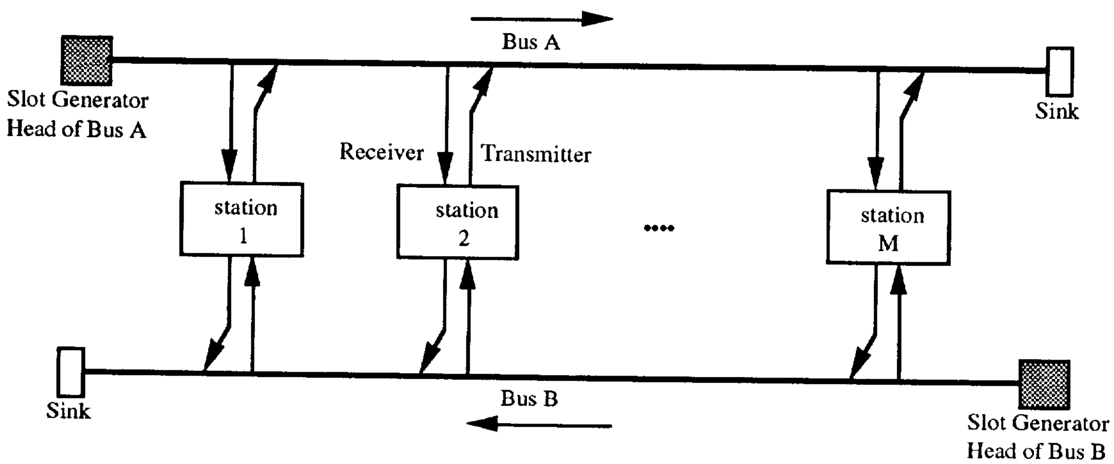

Figure 2.1 Dual Bus Topology

slot is equal to the size of a data packet. Each slot contains in its header, a *busy* bit and three *request* bits. The busy bit indicates whether or not the slot is occupied by a packet, while each request bit is used for sending a request for future packet transmission for each priority level.

Every station maintains two buffers for each bus: the Transmission Buffer(TB) for the "distributed-queue" and a local buffer. For ease of presentation, we will only discuss transmissions on bus A. The local buffer stores the arriving packets according to the FCFS discipline. Only the packet at the head of the local buffer joins the TB. Packets in the individual TBs compete for the transmission on the bus. The Distributed Queue is thus a protocol for forming a global distributed queue among ready stations.

Every station has a request counter(req-counter), a countdown counter (cd-counter), and a local request queue counter (q-counter) for packet transmission on a bus. These counters actually govern the DQDB protocol. The req-counter maintains the number of requests from downstream stations. It is increased by one whenever a request bit set passes on bus B and decreased by one whenever an empty slot passes on bus A until it becomes 0. The cd-counter maintains the number of empty slots that a station should pass on before it transmits its packet. As soon as a packet in a station enters the TB, the station copies the value of req-counter into cd-counter and sets the value of req-counter to zero. The station will then continuously decrease the cd-counter by one for every empty slot passing by bus A until it becomes zero. It will then set the busy bit in the next empty slot and transmit its packet. The q-counter keeps the number of local packets which arrived at a station and not yet sent a request on bus B. It is increased by one whenever a packet enters the TB and decreased by one whenever the station sets a request bit on bus B until it becomes 0.

3. The model

In this section, we consider the system in which station 1 is assumed to be saturated, i.e., it always has a packet ready for transmission at any instant in time.

We focus on the transmission of packets on bus A, since the protocol operates independently and in an indentical manner for packets transmitted on bus B. To transmit a packet, a station reserves a slot on bus A by marking the request bit on a slot on bus B. When we refer to a "marked slot," or the act of "marking a slot," we are referring to a slot on bus B. Similarly, When we refer to a "reserved slot," we are referring to a slot on bus A released in response to a request.

We normalize the time scale with respect to one slot transmission time. In other words, we set the time unit to be one slot time. We assume that packets arrive at station m according to a Poisson process at a rate λ_m, $m=2, \cdots, M-1$, where M is the number of stations in the network. Note that $\lambda_M=0$, since station M never transmits its packets on bus A. Also note that λ_1 is undefined since station 1 is assumed to be saturated. If all the other stations operate in a stable mode, we must have $\Sigma_{k=2}^{M-1}\lambda_k<1$. We can then estimate the probability, β_m, that the slot passing by station m is marked, as

$$\beta_m = \lim_{t\to\infty}\left[\frac{\#\text{ marked slots passing by station m on bus B in time t}}{\#\text{ slots passing by station m on bus B in time t}}\right]$$

$$= \lim_{t\to\infty}\frac{(\lambda_{m+1}+\lambda_{m+2}+\cdots+\lambda_{M-1})t}{t} = \sum_{k=m+1}^{M-1}\lambda_k, \quad m=1,\cdots,M-1. \tag{3.1}$$

Note, thereby, that we assume λ_m is the long run probabability that station m marks a slot. We adopt the convention that an empty sum=0, and empty product=1.

Since station 1 is assumed to be saturated, it consumes all the available slots on bus A that are not reserved by downstream stations. Thus, as far as the downstream stations are concerned, station 1 acts as a controller which releases slots in response to requests received by it.

Consider an arbitrary packet arriving at the Transmission Buffer(TB) of station m. We will refer to this packet as the "tagged packet" [5]. The tagged packet enters the TB in one of two ways: either a) station m has no other waiting packets at the packet arrival instant, in which case the packet immediately joins the TB, or b) the tagged packet waits in the queue till the packet immediately ahead of it is transmitted.

Definition The term "Slot_1" is used to refer to the first slot that passes by station m on bus B, following the arrival of the tagged packet at the TB. ∎

Following its arrival at the TB, the tagged packet attempts to mark Slot_1. Even if Slot_1 is already marked by a downstream station, j, the tagged packet behaves as if it marks this slot, in the sense that it is transmitted ahead of the packet at station j that marked Slot_1. Therefore, the tagged packet has an unfair advantage, as far as the packet at station j is concerned. Of course, it is quite likely that a station upstream of station m attempts to mark Slot_1. in this case the upstream station has an unfair advantage as far as station m is concerned.

This unfairness feature, inherent in the DQDB protocol, complicates the analysis of the protocol for the mean waiting times. This is due to the fact that although the tagged packet marks a slot on bus B, it is not necessarily transmitted on the slot reserved by this request. We will denote, by

Slot_m, the(marked) slot on bus B which ultimately obtains a slot on bus A for the transmission of the tagged packet.

Let S_m denote the time spent by the tagged packet in the TB. Let. $E[S_m]$, $E[S_m^2]$, and $Var[S_m]$ denote the mean, second moment, and variance of S_m, and let $\rho_m = \lambda_m E[S_m]$. If we can estimate $E[S_m]$ and $E[S_m^2]$, then we can obtain $E[W_m]$, the expected time the tagged packet spends in the local queue of station m, using the Pollaczek-Khinchin formula [5] as

$$E[W_m] = \frac{\lambda_m E[S_m^2]}{2(1-\rho_m)} \cdot \qquad (3.2)$$

We will refer to S_m as the service time at station m. We can express S_m as the sun of five components as follow:

$$S_m = U + V_m + x_{m1} + T_m + x_{1m}, \qquad (3.3)$$

where a) U is the time interval from the instant the packet arrives at the TB until Slot_1 passes station m, b) V_m is the time interval from the instant Slot_1 passes station m until Slot_m passes station m, c) x_{m1} is the time taken to propagate Slot_m to station 1 and registers its request there, d) T_m is the time from the instant Slot_m passes station 1 until the slot reserved by this request is released by station 1, and e) x_{1m} is the time taken to propagate this slot from station 1 to station m. When we say that a slot "passes" a station we are referring to the instant at which the head of the slot passes by the station.

We assume that the propagation speed is the same on both buses, namely that $x_{1m} = x_{m1}$. We refer to V_m as the "marking delay," and to T_m as the "reservation delay" at station 1. In general, there may also be a constant "phase delay" y, which represents the time between the instant a slot on bus B reaches station 1 and the instant a slot on bus A just passes station 1. This term does not affect the computation of the unknown components of the service time. Hence, for ease of discussion we ignore the effect of y on the service time for the moment.

The mean of S_m can be computed if we obtain the means for the terms on the right hand side of equation (3.3). Under the assumption that the terms in equation (3.3) are independent, the variance of S_m is obtained as the sum of the variance of the individual terms. We compute the first and second moment of V_m and T_m by considering four exclusive modes of operation, based on the following possible outcomes of the attempts to mark Slot_1.

i) Slot_1 is not marked by any downstream station, the tagged packet marks the slot, and no upstream station attempts to mark the slot. The tagged packet is thus transmittted in the slot actually reserved for it. We term this the "normal" mode of operation. Let $\pi_m(N)$ denote the long-run probability that the system operates in this mode, and let $S_m(N)$ denote the service time in this mode.

ii) Slot_1 is marked by a downstream station, the tagged packet attempts to mark the slot, and no upstream station attempts to mark the slot. The tagged packet is transmitted in the slot reserved by a downstream station, j. We term this the "gain" mode of operation. Let $\pi_m(j,G)$ denote the long-run probability that system operates in this mode, and let $S_m(j,G)$ denote the service time in this mode.

iii) Slot_1 is not marked by any downstream station, the tagged packet marks the slot, and at least one upstream station attempts mark the slot.

iv) Slot_1 is marked by a downstream station, the tagged packet attempts to mark the slot, and at least one upstream station also attempts to mark the slot.

When the system is operating in either mode iii) or iv), we say it is operating in the "loss" mode. Let $\pi_m(j,L_a)$ ($\pi_m(j,L_b)$) denote the long-run probability that the system operates in mode iii) (mode iv), where j is the station most upstream of station m that attempts to mark Slot_1. Let $S_m(j,L_a)$ ($S_m(j,L_b)$) denote the service time in this mode.

The terms "gain," and "loss" have an obvious meaning if we compare the operation of the DQDB protocol in which a station can only transmit its packet on a slot reserved for it. Although the tagged packet takes unfair advantage over packets at downstream stations in mode iv), we stil classify it as a loss mode since one or more packets at upstream stations take unfair advantage over the tagged packet. Note that Slot_ m is the same as Slot_1 in both the normal mode and the gain mode.

We can estimate the long run probabilities quite easily. First. $\pi_m(N)$ is the joint probability that Slot_1 is not marked by any downstream station and no station upstream of station m attempts to mark this slot. If we assume each station attemps to mark slots independently, then we can estimate $\pi_m(N)$ as

$$\pi_m(N) = (1-\beta_m) \prod_{k=2}^{m-1} (1-\lambda_k). \tag{3.4a}$$

Following a similar reasoning, $\pi_m(j,G)$ is the probability that Slot_1 is marked by a downstream station, j (with probability λ_j), and no station upstream of station m attempts to mark Slot_1. Note that station m finds Slot_1 marked with probability β_m, in which case the probability that it is marked by station j is λ_j/β_m. Hence,

$$\pi_m(j,G)= \lambda_j \prod_{k=2}^{m-1} (1-\lambda_k), \quad j=m+1,\cdots,M-1. \tag{3.4b}$$

In the loss mode, there may be more than one upstream station that attempts to mark Slot_1. Each one of these stations will transmit their packet before the tagged packet. Let station j be the station most upstream of station m which attempts to mark Slot_1. The probabilities $\pi_m(j, L_a)$ and $\pi_m(j, L_b)$ are obtained as

$$\pi_m(j,L_a)=\lambda_j(1-\beta_m) \prod_{k=2}^{j-1} (1-\lambda_k), \quad \pi_m(j,L_b)=\lambda_j\beta_m \prod_{k=2}^{j-1} (1-\lambda_k), \quad j=2,\cdots,m-1, \tag{3.4c}$$

The service times in the different modes of operation are (implicitly, in the gain mode the index j ranges from m+1 through M-1, and in the loss mode the index j ranges from 2 through m-1):

$$S_m(N)=U+V_m(N)+T_m(N)+2x_{1m}, \qquad S_m(j,G)=U+V_m(j,G)+T_m(j,G)+2x_{1m}, \tag{3.5a}$$

$$S_m(j,L_a)=U+V_m(j,L_a)+T_m(j,L_a)+2x_{1m},\ S_m(j,L_b)=U+V_m(j,L_b)+T_m(j,L_b)+2x_{1m}, \tag{3.5b}$$

If we can compute $S_m(\cdot)$ under the four modes of operation, then we obtain the first and second moments of S_m(and, thereby, $Var[S_m]$) as:

$$E[S_m]=\pi_m(N)E[S_m(N)]+\sum_{j=m+1}^{M-1}\pi_m(j,G)E[S_m(j,G)]+\sum_{i\in\{a,b\}}\sum_{j=2}^{m-1}\pi_m(j,L_i)E[S_m(j,L_i)]. \tag{3.6a}$$

$$E[S_m^2]=\pi_m(N)E[S_m^2(N)]+\sum_{j=m+1}^{M-1}\pi_m(j,G)E[S_m^2(j,G)]+\sum_{i\in\{a,b\}}\sum_{j=2}^{m-1}\pi_m(j,L_i)E[S_m^2(j,L_i)]. \tag{3.6b}$$

Since the arrivals are Poisson, it can be shown that U is accurately described by a uniform random variable distributed over the range [0,1]. Therefore, the mean and variance of U are $E[U]=1/2$, and $Var[U]=1/12$. The propagation time x_{1m} is a constant which is computed from given physical data on the location of station m with respect to station 1, the propagation speed of the medium, and the speed of the bus. Thus, we only need to estimate the mean and variance of $V_m(\cdot)$ and $T_m(\cdot)$ in the different modes of operation. We now describe how these terms are computed.

Consider an arbitrary request marked by some station, and let J denote the time from the instant the request reaches station 1 until the instant the slot for this request is released on bus A by station 1. It is convenient to refer to this request as the "tagged request." We first obtain the distribution of J, and use it to compute the mean and variance of T_m under the different modes of operation. (Recall that T_m is the time from the instant Slot_m reaches station 1 until the slot for this request is released on bus A by station 1.) We note that J depends on the values of the req-counter and the cd-counter at station 1, at the instant the tagged request reaches station 1. Figure 3.1 depicts the situation at this instant. In Figure 3.1, the tagged request is identified by a *; this Figure also depicts the phase delay, y.

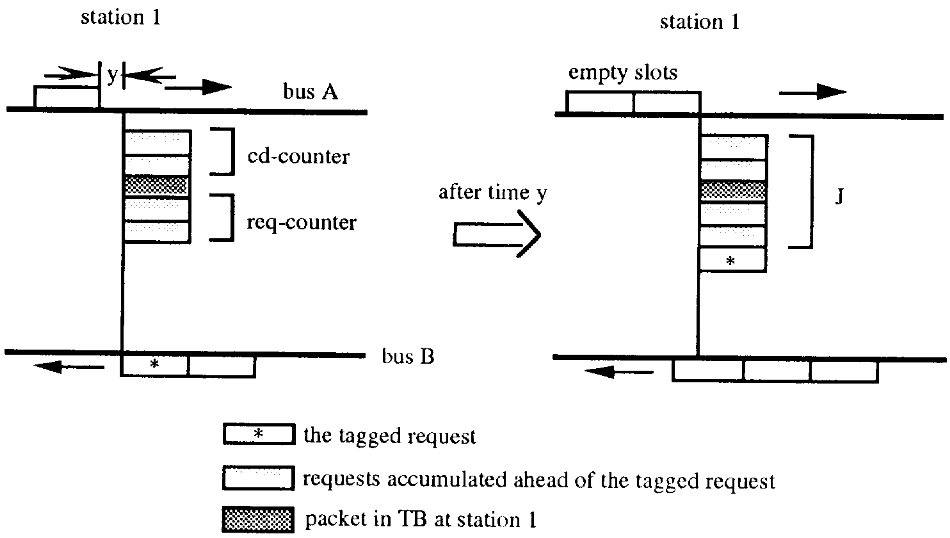

Figure 3.1 Pictorial representation of J

In general, slots passing station 1 on bus B are marked with probability β_1, and each marked slot increases the req-counter by 1. Since station 1 always has a packet ready for transmission, every time its cd-counter becomes zero station 1 immediately transfers the contents of its req-counter to its cd-counter and transmits a packet on the next slot on bus A. Note that the value of the req-counter transferred to the cd-counter represents the number of requests registered at station 1 between successive transmission of packets from station 1. If we view station 1 as a server serving the requests(releasing slots) registered at the station. and the time taken by station 1 to transmit a packet as a vacation, we have the following analogy.

Observation From the viewpoint of the tagged request, station 1 behaves like *a B/G/1 queueing* *system with vacations*, where the server closed the gate when it begins a vacation. Requests arrive at this system according to a Bernoulli process with rate β_1. The service time for a request and the vacation time are both equal to 1 slot time. ∎

Based on Observation, we obtain the probability distribution of J by analyzing the B/G/1 gated queueing system with vacations.

3.1 The Discrete Time B/G/1 Gated Queueing System with Vacations

The gated service mechanism in our model is different from the gated service mechanism usually considered in the literature in which the server closes the gate when it returns from a vacation and services the customers that arrived during the previous cycle. In our model, the server closes the gate *when it begins a vacation*. That is, if we number the sequence of vacations and service periods as (v_n, t_n), $n=1,\cdots$, then the customers arriving during vacation v_n and service period t_n are served during t_{n+1}. This system has been analyzed by Noh [9].

Let G(z) denote the probability generating function (p.g.f) of the number of customers that a random departing customer leaves behind. Using the well known BASTA property (Bernoulli Arrivals See Time Averages), the distribution seen by a departing customer is the same as the distribution of the number of customers at a random point in time. The M/G/1 decomposition property presented by Fuhrmann and Cooper [6] applies to our model as well. This results in the following expression for G(z):

$$G(z) = F(z)\frac{1-A(z)}{A'(1)(1-z)}D(z). \tag{3.7}$$

In equation (3.7), $F(\cdot)$, $A(\cdot)$, and $D(\cdot)$ are, respectively, the p.g.f.s for the stationary distribution of the number of customers already present when a vacation begins, the number of customers that arrive during a vacation, and the number of customers that a random departing customer leaves behind in the corresponding standard B/G/1 queueing system where the server is always available.

The model we consider has the service time and vacation time both equal to one slot time. However, we obtain $A(\cdot)$, $D(\cdot)$, and $F(\cdot)$ for arbitrarily distributed service and vacation times, and subsequently specialize the result to our model. Let B(z) and E(z) denote, respectively, the p.g.f of the service time and vacation time distributions. Let C(z) denote the p.g.f. for the distribution of the number of customers arriving during a service time. Since arrivals follow a Bernoulli process,

$$C(z) = B(1-\beta_1+\beta_1 z), \text{ and } \quad A(z) = E(1-\beta_1+\beta_1 z). \tag{3.8}$$

Let $\rho = \beta_1 B'(1)$. We can write

$$D(z) = \frac{(1-\rho)(z-1)C(z)}{z - C(z)} = \frac{(1-\rho)(z-1)B(1-\beta_1+\beta_1 z)}{z - B(1-\beta_1+\beta_1 z)} \quad (3.9)$$

Denote by $\Theta\{j \mid H + B^{*^{(k)}}\}$ the probability of j arrivals during a time interval that consists of a vacation of duration H followed by k services of duration $B^{*^{(k)}}$. Note that $B^{*^{(k)}}$ is the convolution of k independent services having p.g.f. $(B(z))^k$. Let N denote the number of customers present in the system when the vacation begins. Since the gate closes when the servier begins a vacation, it can be easily shown that

$$P\{N=j\} = \sum_{k=0}^{\infty} P\{N=k\}\, \Theta\{j \mid H + B^{*^{(k)}}\},$$

and so

$$F(z) = \sum_{j=0}^{\infty} P\{N=j\}\, z^j = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} P\{N=k\} \Theta\{j \mid H + B^{*^{(k)}}\}\, z^j.$$

Since H and $B^{*^{(k)}}$ are indepent, using equation (3.8), the above equation is rewritten as

$$\begin{aligned} F(z) &= \sum_{k=0}^{\infty} P\{N=k\} (B(1-\beta_1+\beta_1 z))^k E(1-\beta_1+\beta_1 z) \\ &= F(B(1-\beta_1+\beta_1 z))\, E(1-\beta_1+\beta_1 z). \end{aligned} \quad (3.10)$$

Equation (3.10) presents a recursive expression for F(z), which is used to compute the desired moments of $F(\cdot)$. Thus we obtain the moments for the distribution of the number of customers in the system, in a straightforward manner, from equations (3.7), (3.8), (3.9) and (3.10).

Let R_v denete the residence time (queueing time + service time) for a customer in the vacation system and let $R_v(\cdot)$ denote the p.g.f of its distribution. Similarly, let R_B and $R_B(\cdot)$ denote the residence time and the p.g.f of the residence time distribution in the corresponding standard B/G/1 queueing system (the system without vacations). Note that $J = R_V - 1$ since the service time is equal to one slot time. To determine $R_V(\cdot)$, we first observe that $G(\cdot)$ and $D(\cdot)$ are also the p.g.f for the steady state distribution of the number of customers present in a vacation system and in the corresponding standard system at a customer arrival epoch. Next, we observe that, under the FCFS service discipline, the number of customers left behind by a departing customer are precisely those customers arriving during the residence time of the departing customer. Hence,

$$D(z) = R_B(1-\beta_1+\beta_1 z) \quad \text{and} \quad G(z) = R_V(1-\beta_1+\beta_1 z). \tag{3.11}$$

From equations (3.7) and (3.11), we obtain

$$R_V(z)=F\left(\frac{z-(1-\beta_1)}{\beta_1}\right)\frac{1-A\left(\frac{z-(1-\beta_1)}{\beta_1}\right)}{A'(1)\left(1-\frac{z-(1-\beta_1)}{\beta_1}\right)}R_B(Z). \tag{3.12}$$

Since the first and second moment of the residence time in the standard B/G/1 queueing system are obtained from equations (3.9) and (3.11), we can easily compute the first and second moments of $R_V(z)$ from equation (3.12).

We now specialize the above result to our model where the length of a service and a vacation are both equal to one slot time. Hence, $B(z) = E(z) = z$, and $A(z) = 1-\beta_1+\beta_1 z$. From equation (3.9), we observe that $D(z) = 1-\beta_1+\beta_1 z$, and so from equation (3.11), $R_B(z)=z$. Hence, equation (3.12) simplifies to

$$R_V(z)=F\left(\frac{z-(1-\beta_1)}{\beta_1}\right)z. \tag{3.13}$$

From equation (3.10), we obtain the following recursive expression for $F(z)$,

$$F(Z)=(1-\beta_1+\beta_1 z)\ F(1-\beta_1+\beta_1 z), \tag{3.14}$$

and so the first and second factorial moment of $F(z)$ are

$$F'(1) = \frac{\beta_1}{1-\beta_1}, \quad \text{and} \quad F''(1) = \frac{2\beta_1^{\,3}}{(1-\beta_1)(1-\beta_1^{\,2})} \tag{3.15}$$

Hence, from equations (3.13) and (3.15),

$$R_V'(1) = 1+\frac{1}{1-\beta_1}, \quad \text{and} \quad R_V''(1) = \frac{2}{(1-\beta_1)}\left(\frac{\beta_1}{1-\beta_1^{\,2}}+1\right).$$

Finally, since $J=R_V-1$, we obtain

$$E[J] = \frac{1}{1-\beta_1}, \quad \text{and} \quad Var[J] = \frac{\beta_1}{1-\beta_1^2} \tag{3.16}$$

3.2 Computing $S_m(N)$

Since Slot_1 is the same as Slot_m in the normal mode of operation, $V_m(j, N)=0$. We use the result of Section 3.1(equation 3.16) to compute $T_m(N)$. However, equation (3.16) uses the parameter β_1 which includes λ_m, the probability that slot is marked by station m. We assume that once station m makes a request, it does not make further requests until the packet at this station begins transmission. Therefore, if we consider the stream of slots preceding a request made by station m, these slots would be marked with probability $\beta_1 - \lambda_m$ when they reach station. 1. Let

$$\gamma_m = \beta_1 - \lambda_m. \tag{3.17}$$

We compute $E[T_m(N)]$ and $Var[T_m(N)]$ using equation (3.16) with γ_m in place of β_1. Hence,

$$E[V_m(N)] = 0, \qquad Var[V_m(N)] = 0, \tag{3.18a}$$

$$E[T_m(N)] = \frac{1}{1-\gamma_m}, \qquad Var[T_m(N)] \frac{\gamma_m}{1-\gamma_m^2}. \tag{3.18b}$$

3.3 Computing $S_m(j, G)$

Note that Slot_1 is the same as Slot_m in the gain mode of operation as well, and so $V_m(j, G)=0$. In this mode of operation, Slot_m is marked by the downstream station, j, and so the mean and variance of $T_m(j, G)$ are obtained by using equation (3.16) with γ_j in place of β_1. Hence,

$$E[V_m(j,G)] = 0, \qquad Var[V_m(j,G)] = 0, \qquad j=m+1,\cdots,M-1, \tag{3.19a}$$

$$E[T_m(j,G)] = \frac{1}{1-\gamma_j}, \qquad Var[T_m(j,G)] = \frac{\gamma_j}{1-\gamma_j^2}, \qquad j=m+1,\cdots,M-1. \tag{3.19b}$$

3.4 Computing $S_m(j, L_a)$

This is the service time for the tagged packet when it is transmitted in mode iii), namely, the case where Slot_1 is marked by station m, one or more upstream stations attempt to mark Slot_1, and station j is the station most upstream of station m that attempts to mark Slot_1.

Figure 3.2 depicts the loss delay in mode iii) for the case where the tagged packet, which is at station 4, is able to mark Slot_1(the slot marked 4 in the Figure) and station 2 and 3 also attempt to mark Slot_1. For the case considered in the Figure, it is assumed that the two slots following Slot_1 are unmarked when they pass station 4, so that station 3 marks the slot following Slot_1 , and station 2 marks the subsequent slot 2 will therefore transmit on the slot reserved for station 4, and station 4 will transmit on the slot reserved for station 2.

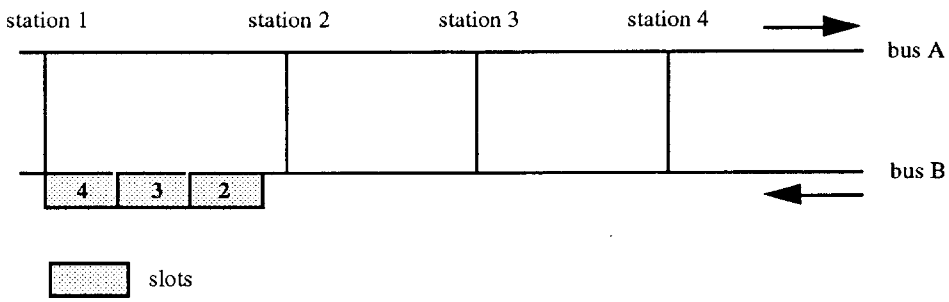

Figure 3.2 Illustration of loss delay in mode iii)

3.4.1 Estimating $E[Vm(j, L_a)]$ and $Var[V_m(j, L_a)]$

In general, if each station $k, j < k < m$, attempts to mark Slot_1 independently with probability λ_k, we can show that $E[V_m(j, L_a)] = 1 + \Lambda_{jm}$, where

$$\Lambda_{jm} = \sum_{k=j+1}^{m-1} \lambda_k. \tag{3.20}$$

It is, no doubt, possible to estimate $Var[V_m(j, L_a)]$ accurately, by considering all the possibilities with regard to the attempts made by station j+1 through m-1 to mark Slot_1. However, this combinatiorial approach does not result in a simple closed form expression for $Var[V_m(j, L_a)]$. Therefore, as a first order approximation, we aggregate all upstream stations lying between stations m and j into an aggregate station with an arrival rate equal to the sum of the arrival rates at the individual stations. Note that simplifying assumption results in exactly the same value for $E[V_m(j, L_a)]$ although it would underestimate $Var[V_m(j, L_a)]$. However, since the sum of arrival rates is less 1, the degree of underestimation is unlikely to be significant. Thus, we estimate $E[V_m(j, L_a)]$ and $Var[V_m(j, L_a)]$ as

$$E[V_m(j, L_a)] = 1 + \Lambda_{jm}. \qquad Var[V_m(j, L_a)] = \Lambda_{jm}(1 - \Lambda_{jm}), \quad j = 2, \cdots, m-1. \tag{3.21}$$

3.4.2 Estimating $E[T_m(j, L_a)]$ and $Var[T_m(j, L_a)]$

Consider next, the mean and variance of $T_m(j, L_a)$, which is the reservation delay for Slot_m. Note that Slot_m is not necessarily marked by station j (although this is the case in Figure 3.2). For example, consider the situation where only one upstream station, say j, attempts to mark Slot_1(which is already marked by station m). If the slot following Slot_1(call it Slot_2 is marked by a downstream station, say k, then station j can only mark a subsequent slot. Station m, however, will transmit on the slot reserved by Slot_2 (station j will transmit on the slot reserved by Slot_1). Therefore, if we follow the same reasoning as before, we must consider equation (3.16) with γ_k in place of β_1. It is, of course, possible to evaluate $E[T_m(j, L_a)]$ and $Var[T_m(j, L_a)]$ by considering all the realizations of marking attempts, on the slots following Slot_1, made by the downstream stations. However, once again this combinatorial approach does not lead to simple expressions for $E[T_m(j, L_a)]$ and $Var[T_m(j, L_a)]$. Hence we obtain simple, approximate expressions for $E[T_m(j, L_a)]$ and $Var[T_m(j, L_a)]$ assuming that Slot_m is always marked by station j.

With this simplifying assumption, we can consider using equation (3.16) with γ_j in place of β_1 to estimate $E[T_m(j, L_a)]$ and $Var[T_m(j, L_a)]$. However, $E(J)$ and $Var(J)$ in equation (3.16) are estimated based on a probabilistic analysis in which it is assumed that the slots ahead of the tagged request are each independently marked with probability β_1. Based on a simplistic reasoning, each slot ahead of the tagged slot thus probilistically contributes to J a factor β_1. We are now considering the case where each slot ahead of Slot_m is marked probability γ_j with the exception of Slot_1 which is marked with probability 1. Hence we use equation (3.16) with γ_j in place of β_1, but we add a factor (1- γ_j) to account for the marked slot when estimating $E[T_m(j, L_a)]$. We ignore the effect of Slot_1 on $Var[T_m(j, L_a)]$. The presence of a marked slot ahead of

Slot_m also increases the likelihood of a packet transmission by station 1 before Slot_m reserves a slot on bus A. Let ε_i denote the probability that a packet transmitted by station 1 leaves behind i outstanding requests. The probability that station 1 transmits a packet between the slots released for two adjacent marked slots is approximated as $(1-\varepsilon_0)$. Hence, we estimate $E[T_m(j, L_a)]$ and $Var[T_m(j, L_a)]$ as

$$E[T_m(j, L_a)] = \frac{1}{1-\gamma_j} + (1-\gamma_j) + (1-\varepsilon_0), \quad Var[T_m(j, L_a)] = \frac{\gamma_j}{1-\gamma_j{}^2}, \quad j=2,\cdots,m-1. \tag{3.22}$$

The term ε_0 is obtained from the analysis for the B/G/1 gated system with vacations where the arrival rate is γ_j. Recall that $F(\cdot)$ is the p.g.f. for the distribution of the number of customers present when a vacation begins(when a customer from station begins transmission). Although the expression for F(z) given by equation (3.10) is recursive in nature, we can numerically evaluate ε_0 quite accurately and efficiently.

3.5 Computing $S_m(j, L_b)$

This is the service time for the tagged packet when it transmitted in mode iv), namely, the case where Slot_1 is marked by a downstream station, and one or more upstream stations attempt to mark Slot_1. Station j is the station most upstream of station m that attempts to mark Slot_1. Following a similar line of reasoning as used to estimate $E[S_m(j, L_a)]$ and $Var[S_m(j, L_a)]$, we obtain

$$E[T_m(j, L_b)] = \frac{1}{1-\gamma_m} + (1-\beta_m) + (1-\varepsilon_0), \quad Var[T_m(j, L_b)] = \frac{\gamma_m}{1-\gamma_m{}^2}, \quad j=2,\cdots,m-1. \tag{3.23}$$

4. Numerical Evaluation of the Model

We conducted two simulation experiments for revealing the accuracy of the analytical model and for illustrating the degree of unfairness downstream stations suffer from in accessing the medium.

For each experiment, we consider an $M=7$ station DQDB network with stations 1 and 7 playing the role of the head of the buses. Stations 2 throught 6 are designed to have the same traffic rate of $\lambda=0.05$(packets per one slot time). Note that these stations are expected to have the same delay if there does not geographical unfairness among stations.

In example 1 the interstation spacing(the distance between two adjacent stations) is set to 1 slot, whereas in example 2 it is set to 0.5 slot. Simulation runs were conducted using the SIMAN simulation language [11]. For each example, 20 independent replications were made. The elapsed time for each run is 6,000 slot times. The packet arrival rate at each station was normalized to packets/slot-time. All performance measures are scaled to one-slot-time. In each example, we assume that the stations are equally spaced, i.e., the distance between any two adjacent stations is the same We refer to this distance as 'interstation spacing'. The distance is measured in slot transmission time units. We set the phase delay y to $\frac{1}{2}$ for all cases. For all examples, we provide the mean and variance of the service time, and the mean packet waiting time in the local FCFS queue. The mean packet access time is omitted since it is simply the sum of $E[W_m]$. We also provide the 95% confidence intervals for $E[S_m]$, $Var[S_m]$, and $E[W_m]$.

As can be seen in the following results, our model estimates the mean packet waiting times with an acceptable accuracy. Furthermore, it is observed that the mean service time and the mean waiting time for a packet are absolutely influenced by the relative position of stations. It is also observed that the distance between stations has an impact on the unfairness feature. That is, the far the stations are apart from each other, the higher the degree of unfairness in the downstream stations appear.

Example 1

$\lambda_2 = 0.05$, $\lambda_3 = 0.05$, $\lambda_4 = 0.05$, $\lambda_5 = 0.05$, $\lambda_6 = 0.05$, and $\lambda_7 = 0$,
$y = 0.5$, interstation spacing = 1

station #	$E[S_m]$		$Var[S_m]$		$E[W_m]$	
	Analysis	Simulation	Analysis	Simulation	Analysis	Simulation
1	1.333	1.333±0.004	-	-	-	-
2	4.250	4.263±0.012	0.274	0.267±0.007	0.582	0.549±0.043
3	6.356	6.353±0.017	0.478	0.455±0.024	1.498	1.424±0.124
4	8.459	8.459±0.025	0.665	0.644±0.024	3.129	3.319±0.293
5	10.560	10.571±0.020	0.831	0.813±0.033	5.950	5.693±0.651
6	12.656	12.632±0.025	0.977	0.990±0.043	10.972	10.472±1.132

Example 2

$\lambda_2 = 0.05$, $\lambda_3 = 0.05$, $\lambda_4 = 0.05$, $\lambda_5 = 0.05$, $\lambda_6 = 0.05$, and $\lambda_7 = 0$.

y = 0.5, interstation spacing = 0.5

station #	$E[S_m]$		$Var[S_m]$		$E[W_m]$	
	Analysis	Simulation	Analysis	Simulation	Analysis	Simulation
1	1.333	1.333±0.007	-	-	-	-
2	3.250	3.318±0.022	0.278	0.246±0.025	0.324	0.302±0.045
3	4.356	4.359±0.026	0.487	0.453±0.026	0.622	0.547±0.061
4	5.459	5.474±0.030	0.677	0.648±0.045	1.048	1.127±0.168
5	6.560	6.538±0.029	0.848	0.834±0.043	1.632	1.598±0.184
6	7.656	7.690±0.037	0.998	0.965±0.062	2.415	2.410±0.386

5. Summary and Conclusions

We consider a special system in which the most upstream station is saturated in the DQDB network. This model is used to estimate the mean access time that data packets experience at each station in the network. The mean number of packets queued at each station can also be obtained using Little's law.

Each station is modeled as an M/G/1 queueing system, in which the service time for a packet is defined as the time spent by the paket at the Transmission Buffer (TB) of the station. It is observed that, since station 1 consumes all available slots, stations usually use slots which are reserved for themselves. This special structure enables us to decompose the service time into a number of independent components including the reservation delay, the delay that a request experiences until it reserves an empty slot at station 1. The reservation delay is modeled by employing B/G/1 gated queueing system. The mean and variance of unknown terms in the decomposed service time are computed for three different mode of operations:normal, gain, and loss. The mean waiting time for packets are obtained using the Pollaczek-Khinchin formula. Simulation experiments show that our model accurately estmates the mean waiting time for packets in a moderately loaded network. This model also enables us to reveal the degree of unfairness which downstream stations undergo in an extreme situation where the most upstream station is saturated.

References

[1] IEEE Standard 802.6-1990, ***IEEE Standards for Local and Metropolitan Area Networks: Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network (MAN)***, New York: IEEE, 1991

[2] Bisdikian, C., "Waiting Time Analysis in a Single Buffer DQDB(802.6) Network," ***IEEE INFOCOM*** 90 (1990), pp. 610-616.

[3] Bisdikian, C., "A Queueing Model for a Data Station within the IEEE 802.6 MAN," IBM Research Division, RC 15587, 1990.

[4] Bisdikian, C., "A Note on the Conservation Law for Queues with Batch Arrivals," IBM Research Division, 1990.

[5] Cooper, R. B., ***Introduction to Queueing Theory***, North Holland, New York, 1981.

[6] Fuhrmann, S. W. and Cooper, R. B., "Stochastic Decompositions in the M/G/1 Queue with Generalized Vacations," ***Operations Research***, Vol. 33, No. 5 (1985), pp. 1117-1129.

[7] Mukherjee, B. and Banerjee, S., "Alternative Strategies for Improving the Fairness in and an Analytical Model of DQDB Networks," Proc., ***IEEE INFOCOM*** '91 (1991), Bal Harbour, FL, pp. 879-888.

[8] Mukherjee, B., and Bisdikian, C., "A Journey Through the DQDB Network Literature," ***Performance Evaluation*** (1992).

[9] Noh, Seung J. "A note on the B/G/1 Gated Queueing System with Vacations.," ***Korean Journal of Management Science***, Vol. 19, No. 3 (1994).

[10] Noh, Seung J. "Two Queue Single Sever Model for the DQDB MAM," ***Korean Journal of Management Science***, Vol. 22, No. 2 (1997).

[11] Pegden, C. D., ***Introduction to SIMAN***, System Modeling Corporation, State College, PA, 1987.

[12] Potter, P. G. and Zukerman, M., "Analysis of Discrete Multipriority Queueing System Involving Central Shared Processor Serving Many Local Queues," ***IEEE Jour. on Selec. Areas in Comm.***, Vol. 9, No. 2 (1991), pp. 194-202.

[13] Tran-Gia, P. and Stock, T., "Approximate performance analysis of the DQDB access protocol," ***Proc., Intl. Teletraffic Congress (ITC)***, Adelaide, Australia, Sep. 1989. (Also, ***Computer Nets, and ISDN Syst.***, Vol. 20 (1990), pp. 231-240.)