Occupancy Distribution for a DQDB Station Based on a Queueing System with Markov-Structured Service Requirements

Ioannis Stavrakakis   Sophia Tsakiridou
Department of Electrical Engineering and Computer Science
University of Vermont
Burlington, VT 05405
Tel: (802) 656-8506   FAX: (802) 656-0696

Abstract

In this paper, the queue occupancy distribution of the DQDB station is derived through the formulation and exact analysis of a queueing model with the following characteristic: the packet service time requirement depends on the state of an underlying Markov-structured process at the time when the packet is forwarded to the server (head of the queue). Numerical results are presented and compared with simulations to evaluate the accuracy of the involved approximations.

1 Introduction

The Distributed Queue Dual Bus (DQDB) medium access protocol has been adopted as a Metropolitan Area Network (MAN) Standard (IEEE 802.6), [1]. The DQDB MAN is a high speed integrated services network designed to accommodate traffic generated by a variety of information sources, as well as provide for the interconnection of LAN's. A detailed description of the DQDB protocol may be found in [1]. Various design and performance analysis issues are investigated in [2] - [7] and the references cited there. Bandwidth balancing mechanisms or priorities [1, 3, 6] are not considered in the present DQDB study.

The queueing behavior of the DQDB station is modeled in this paper in terms of a single server queueing system. The service discipline - which reflects the availability of network capacity to the station - is described in terms of an appropriately defined process (balance process), which captures the station's perception of the state of the global queue at time instants of interest. A general queueing system of potentially wide applicability is formulated, analyzed and applied for the analysis of the developed DQDB station model. The effective packet service time of the system, defined as the waiting time of a packet at the head of queue until its service completion (departure), is determined by the state of an underlying Markov process at the time when the packet is forwarded to the head of the queue.

2 A queueing model for the DQDB station

In this section the queueing behavior of the DQDB station is modeled in terms of a centralized queue operating under a service policy which reflects the distributed queueing protocol. The basic operation of the protocol is described only to the extent that is relevant to the DQDB station model developed in this paper.

The DQDB network consists of two high speed unidirectional buses carrying information in opposite directions. Each network user is capable of transmitting and receiving information on both buses. Information segments (packets) are transported in fixed size slots which are continuously generated by the head of each bus. Since the media access protocol is identical for both network buses, packet transmissions only along one of the buses (forward bus) will be considered. A station uses the forward bus to transmit packets to its downstream stations and the reverse bus to register requests to its upstream stations as explained below.

At any given time a DQDB station may be either busy or idle; it is busy if a packet is queued in its buffer (to be transmitted on the forward bus), and it is idle otherwise. A busy station must notify all upstream stations of its intent to transmit by registering a request. This is implemented by setting the request bit of the first slot passing by on the reverse bus,

Research supported by the National Science Foundation under grant NCR-9011962.

email : stavraka@uvm.edu , tsakirid@uvm.edu

whose request bit is reset. In turn, an upstream station that detects a request must satisfy it by allowing an empty slot to pass by on the forward bus. Stations keep track of the pending requests which need to be satisfied before they are allowed to transmit, by using a request counter (REQ-CNTR). The REQ-CNTR of an idle station is incremented by one each time a request is detected on the reverse bus; it is decremented by one each time an empty slot is detected on the forward bus.

When a station enters the busy state, the value of the REQ-CNTR is downloaded into a count-down counter (CD-CNTR) and the REQ-CNTR is reset to zero. A request is registered on the reverse bus as explained above. The REQ-CNTR of the busy station is incremented by one each time a set request bit is detected; no decrement of this counter occurs in the busy state. The CD-CNTR, which is operational only while the station is busy, is decremented by one each time an empty slot is detected on the forward bus. When the value of the CD-CNTR reaches zero, the station is permitted to transmit into the next empty slot on the forward bus. Following the packet transmission, the station enters the idle state. If another packet has been waiting in the buffer, the station enters the busy state instantaneously.

The implementation of the distributed queueing protocol requires that both buses be continuously monitored by each station. The empty-slot process, $\{e_j\}_{j\geq 0}$,and the request process, $\{r_j\}_{j\geq 0}$, observed by the tagged station under study on the forward and reverse bus, respectively, are defined as follows

$$e_j = \begin{cases} 1 & \text{if the } j^{th} \text{ slot on the forward bus } \underline{\text{leaving}} \\ & \quad \text{the tagged station is empty} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

$$r_j = \begin{cases} 1 & \text{if the request bit of the } j^{th} \text{ slot on the reverse} \\ & \quad \text{bus } \underline{\text{arriving}} \text{ at the tagged station is set} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Time instant (or slot) j denotes the end of the j^{th} slot; all events of interest are assumed to occur at the end of slots.

The request and the empty-slot processes determine the value of the REQ-CNTR of the tagged station, denoted by $\{(REQ\text{-}CNTR)_j\}_{j\geq 0}$. By monitoring the two processes, the station obtains the information regarding the state of the distributed global queue based on which it can position its service request (packet) in a conceptual centralized queue. Due to the propagation delay, the state information provided to the station may be outdated. The "perception" of the station about the state of the distributed global queue is determined by the balance process $\{b_j\}_{j\geq 0}$, defined by

$$b_0 = 0 \ , \ b_j = [b_{j-1} + r_j - e_j]^+ \quad , j \geq 1 \tag{3}$$

where $[x]^+ = \max\{x, 0\}$. Notice that b_j is equal to the number of unsatisfied requests from the downstream stations at slot j. The state space of the balance process is denoted by $S_b = \{0, 1, 2, \cdots, N_b\}$.

The time sequences defined below will be involved in the study of the DQDB station.

- $\{t_k^d\}_{k\geq 0}$: Defined to be the sequence of slots at which a packet departure occurs ($t_0^d = 0$).
- $\{t_k^{adv}\}_{k\geq 1}$: Defined to be the sequence of slots at which a packet advances to the Head of the Queue (HoQ) of the station.
- $\{\hat{t}_k\}_{k\geq 0}$: Defined to be the sequence of slots at which a change occurs at the HoQ. A change at the HoQ occurs if a packet advances to or leaves from the HoQ; $\{\hat{t}_k\}_{k\geq 0} = \{t_k^d\}_{k\geq 0} \cup \{t_k^{adv}\}_{k\geq 1}$.

The following proposition is important for the formulation of the proposed DQDB station model.

Proposition 1:

The states of processes $\{(REQ\text{-}CNTR)_j\}_{j\geq 0}$ and $\{b_j\}_{j\geq 0}$ are identical in slots at which a change occurs at the HoQ of the station. That is,

$$(REQ\text{-}CNTR)_{\hat{t}} = b_{\hat{t}} \quad , \ \forall\, \hat{t} \in \{\hat{t}_k\}_{k\geq 0}$$

The proof may be found in [8]. □

Since $\{t_k^{adv}\}_{k\geq 1} \subset \{\hat{t}_k\}_{k\geq 0}$ the above proposition implies that whenever a packet advances to the HoQ, the value (state) of the balance process is identical to the value of the REQ-CNTR which is downloaded to the CD-CNTR. Thus, the value of the balance process at these time instants provides the information about the state of the distributed queue that is utilized by the tagged station of the DQDB network.

In this paper, the packet arrival process to the DQDB station is assumed to be an independent and identically distributed (i.i.d.) discrete-time process; the station's queueing capacity is assumed to be infinite. Based on Proposition 1 and the DQDB protocol operation, the queueing behavior of the DQDB station can be modeled in terms of a single server queueing system. The packet service (transmission) time is deterministic and equal to one time slot. The service policy that captures exactly the availability of the server (network capacity) to the queue - as determined by the station's perception about the state of the distributed global queue - is described below, in terms of server vacations or the effective packet service time.

According to the DQDB protocol the server will be unavailable (on vacation) for a period w_k, following the time instant t_k^{adv} at which the k^{th} packet advances to the HoQ of the station. Over this period, the pending requests $(REQ\text{-}CNTR)_{t_k^{adv}}$ from the downstream stations will be satisfied and some service will be provided to the upstream users. w_k depends on $(REQ\text{-}CNTR)_{t_k^{adv}}$, which in view of Proposition 1 is identical to $b_{t_k^{adv}}$, and the evolution of the empty-slot process after t_k^{adv}, $\{e_j\}_{j=t_k^{adv}+1}^{\infty}$.

It may be alternatively assumed that, instead of taking a vacation of length w_k, the server remains at the queue but the service time of the k^{th} packet is increased to the effective value $s_k = w_k + 1$. Note that s_k depends on $b_{t_k^{adv}}$ and $\{e_j\}_{j=t_k^{adv}+1}^{\infty}$ and it is, in general, different for different packets. The determination of the sequence of effective service times $\{s_k\}_{k\geq 1}$ would require knowledge of the sequences $\{e_j\}_{j\geq 0}$ and $\{r_j\}_{j\geq 0}$ so that $b_{t_k^{adv}}$ and $\{e_j\}_{j=t_k^{adv}+1}^{\infty}$ be determined for any t_k^{adv}, $k \geq 1$. Such a requirement would result in an analytically intractable queueing model. In section 4, processes $\{e_j\}_{j\geq 0}$ and $\{r_j\}_{j\geq 0}$ are approximated in a way such that the effective service time s_k is determined based on $b_{t_k^{adv}}$ only, for all $k \geq 1$.

3 Queueing analysis under a Markov service policy

In this section the general queueing system is formulated and analyzed. Consider a discrete-time, single server, infinite capacity queueing system with i.i.d. packet arrivals and effective packet service time dependent on the state of an underlying Markov chain $\{c_j\}_{j\geq 0}$ at the time when the packet advances to the HoQ; $\{c_j\}_{j\geq 0}$ is assumed to be independent of the packet arrival process. Let $S_c = \{0, 1, 2, \ldots, N_c\}$ denote its state space and let $p_c(i,j)$ and $\pi_c(i)$ denote its transition and steady state probabilities, $i, j \in S_c$. Given the state i of $\{c_j\}_{j\geq 0}$ when a packet advances to the HoQ, its effective service time $s(i)$ depends only on i and it is given by a general distribution $p_{s_c}(i,k)$, $i \in S_c$,$k \geq 1$. Consider the following definitions in addition to those presented in section 2 :

x_j : Denotes the queue occupancy at slot j.

a_j : Denotes the number of packet arrivals at slot j.

a^l : Denotes the generic random variable for the number of packet arrivals over l slots, $l \geq 1$; let $p_a^l(n)$ denote the probability that n packets arrive over l slots, $n \geq 0$; $p_a^l(n)$ can be determined as the l-fold convolution of $p_a^1(n)$.

τ : Denotes the generic random variable for the number of slots until the first packet arrival to the queue that was left empty by the last packet departure; its probability distribution is easily seen to be given by $p_\tau(n) = \left[p_a^1(0)\right]^{n-1}\left[1 - p_a^1(0)\right]$, $n \geq 1$.

τ_k : Denotes the length of time that the HoQ remains empty after the $(k-1)^{st}$ packet departure; $\tau_k = t_k^{adv} - t_{k-1}^d$ (Fig. 1). Note that τ_k depends on $x_{t_{k-1}^d}$. If $x_{t_{k-1}^d} > 0$, that is, if the queue is left non-empty upon departure of the $(k-1)^{st}$ packet, the k^{th} packet is instantaneously forwarded to the HoQ and $\tau_k = 0$. If $x_{t_{k-1}^d} = 0$, $\tau_k \neq 0$, $k \geq 1$, is distributed identically to τ.

s_k : Denotes the effective service time of the k^{th} packet, defined as the length of time that the packet stays in the HoQ; $s_k = t_k^d - t_k^{adv}$. According to the packet service policy, s_k depends on $c_{t_k^{adv}}$. Given $c_{t_{k-1}^d}$, $c_{t_k^{adv}}$ is determined by the evolution of $\{c_j\}_{j\geq 0}$ in τ_k slots. Therefore, given $\{x_{t_{k-1}^d}, c_{t_{k-1}^d}\}$, $c_{t_k^{adv}}$ and consequently s_k, are determined in a probabilistic sense.

d_k : Denotes the time interval between the $(k-1)^{st}$ and the k^{th} packet departures; $d_k = t_k^d - t_{k-1}^d = \tau_k + s_k$ (Fig. 1). By definition of τ_k and s_k it is easily established that d_k is probabilistically determined by $\{x_{t_{k-1}^d}, c_{t_{k-1}^d}\}$.

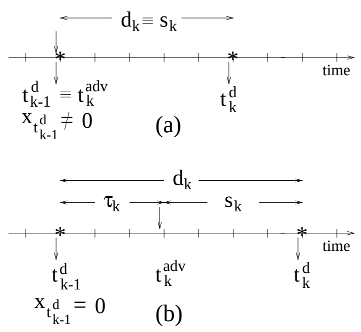

Fig. 1: Illustration of the definitions of τ_k and d_k: (a) $\tau_k = 0$ (b) $\tau_k > 0$. An arrow pointing to (away from) the time axis indicates the movement of a packet to (away from) the HoQ.

The distribution of the queue occupancy upon packet departures, is derived by considering the process $\{x_{t_k^d}, c_{t_k^d}\}_{k\geq 0}$ embedded on packet departure time instants, $\{t_k^d\}_{k\geq 0}$. Note that $\{x_{t_k^d}\}_{k\geq 0}$ evolves as follows

$$x_{t_k^d} = x_{t_{k-1}^d} + a^{d_k} - 1$$

where a^{d_k} denotes the number of packet arrivals over d_k slots. By definition of d_k, the state of $\{x_{t_k^d}\}_{k\geq 0}$ at t_k^d is probabilistically determined from $\{x_{t_{k-1}^d}, c_{t_{k-1}^d}\}$. Similarly for $\{c_{t_k^d}\}_{k\geq 0}$, since $c_{t_k^d}$ is determined by the evolution of $\{c_j\}_{j\geq 0}$ in d_k slots. From the above arguments it is established that $\{x_{t_k^d}, c_{t_k^d}\}_{k\geq 0}$ is a Markov chain. Let $S_{xc} = Z_0^+ \times S_c$ denote its state space, where Z_0^+ denotes the set of nonnegative integers; let $\pi_{xc}(i,j)$, $p_{xc}(i_1,j_1,i_2,j_2)$ and $\tilde{P}_{xc}$ denote its steady state probabilities, transition probabilities and transition matrix, respectively. By identifying $x_{t_k^d}$ and $c_{t_k^d}$ as the level and the phase, respectively, of Markov chain $\{x_{t_k^d}, c_{t_k^d}\}_{k\geq 0}$, its transition matrix, $\tilde{P}_{xc}$, may be written in the form

$$\tilde{P}_{xc} = \begin{bmatrix} \tilde{B}_0 & \tilde{B}_1 & \tilde{B}_2 & \tilde{B}_3 & \cdots \\ \tilde{A}_0 & \tilde{A}_1 & \tilde{A}_2 & \tilde{A}_3 & \cdots \\ 0 & \tilde{A}_0 & \tilde{A}_1 & \tilde{A}_2 & \cdots \\ 0 & 0 & \tilde{A}_0 & \tilde{A}_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{4}$$

$\tilde{A}_n$ and $\tilde{B}_n$, $n \geq 0$, are matrices of order (N_c+1) with elements $A_n(j_1,j_2)$ and $B_n(j_1,j_2)$, $j_1, j_2 \in S_c$,

$$A_n(j_1,j_2) = \Pr\left\{x_{t_k^d} = i_1 - 1 + n, c_{t_k^d} = j_2 / x_{t_{k-1}^d} = i_1, c_{t_{k-1}^d} = j_1\right\}, \ i_1 \geq 1$$

$$B_n(j_1,j_2) = \Pr\left\{x_{t_k^d} = n, c_{t_k^d} = j_2 / x_{t_{k-1}^d} = 0, c_{t_{k-1}^d} = j_1\right\}$$

Let $\tilde{A}_n(l)$ ($\tilde{B}_n(l)$) be defined similarly to $\tilde{A}_n$ ($\tilde{B}_n$) but imposing the additional requirement that the interdeparture time $d_k = \tau_k + s_k$ is equal to l, $l \geq 1$. That is,

$$A_n(j_1,j_2,l) = \Pr\left\{x_{t_k^d} = i_1 - 1 + n, c_{t_k^d} = j_2, s_k = l / x_{t_{k-1}^d} = i_1, c_{t_{k-1}^d} = j_1\right\}, \ i_1 \geq 1 \tag{5}$$

$$B_n(j_1,j_2,l) = \Pr\left\{x_{t_k^d} = n, c_{t_k^d} = j_2, \tau_k + s_k = l / x_{t_{k-1}^d} = 0, c_{t_{k-1}^d} = j_1\right\} \tag{6}$$

To facilitate the application of the formulae that follow to the DQDB study in the next section, let $\tilde{P}_{c/>0}$ ($\tilde{P}_{c/0}$) denote the one-step transition probability matrix of $\{c_j\}_{j\geq 0}$ over the periods when the queue is non-empty (empty); let $\tilde{P}_{c/>0}(l)$ and $\tilde{P}_{c/0}(l)$ denote the corresponding l-step transition probability matrices. Since $\{c_j\}_{j\geq 0}$ evolves independently of the queue, $\tilde{P}_{c/>0}(l) = \tilde{P}_{c/0}(l) = [\tilde{P}_c]^l$. Let $\tilde{P}_{s_c}(l) = \mathrm{diag}\,[p_{s_c}(0,l), p_{s_c}(1,l), \ldots, p_{s_c}(N_c,l)]$ denote the diagonal matrix of the effective packet service time probabilities associated with the states of $\{c_j\}_{j\geq 0}$. From (5) and (6) it is easily established that matrices $\tilde{A}_n(l)$ and $\tilde{B}_n(l)$, $l \geq 1$, are given by

$$\tilde{A}_n(l) = p_a^l(n)\tilde{P}_{s_c}(l)\tilde{P}_{c/>0}(l) \tag{7}$$

$$\tilde{B}_n(l) = \sum_{i=1}^{l-1} \left[p_a^1(0)\right]^{i-1} \sum_{j=1}^{n+1} p_a^1(j)\tilde{P}_{c/0}(i)\tilde{A}_{n-j+1}(l-i) \tag{8}$$

By applying total probability arguments, $\tilde{A}_n$ and $\tilde{B}_n$ can be written as follows

$$\tilde{A}_n = \sum_{l=1}^{\infty} \tilde{A}_n(l) = \sum_{l=1}^{\infty} p_a^l(n)\tilde{P}_{s_c}(l)\tilde{P}_{c/>0}(l) \tag{9}$$

$$\tilde{B}_n = \sum_{l=1}^{\infty} \tilde{B}_n(l) = \tilde{P}_{c/0}\left(\tilde{I} - p_a^1(0)\tilde{P}_{c/0}\right)^{-1} \sum_{j=1}^{n+1} p_a^1(j)\tilde{A}_{n-j+1} \tag{10}$$

where $\tilde{I}$ is the identity matrix of order (N_c+1).

The queue occupancy distribution upon departures, $\{q_x(i)\}_{i=0}^{\infty}$, - as determined by the stochastic matrix $\tilde{P}_{xc}$ - is derived based on the matrix analytic approach, as outlined in [8]; details on this approach may be found in [9, 10] and the references cited therein. The queue occupancy distribution at arbitrary time instants, $\{\psi_x(i)\}_{i=0}^{\infty}$, is derived in terms of $\{q_x(i)\}_{i=0}^{\infty}$ by invoking Markov renewal theory arguments; it is given by, [8],

$$\psi_x(0) = \lambda_a\left(1 - p_a^1(0)\right)^{-1} q_x(0)$$

$$\psi_x(i) = \left(1 - p_a^1(0)\right)^{-1} \cdot \left\{\sum_{j=1}^{i} p_a^1(j)\psi_x(i-j) + \lambda_a\left[q_x(i) - q_x(i-1)\right]\right\}, i \geq 1$$

where λ_a denotes the packet arrival rate.

4 The DQDB station queueing analysis

The analysis of the queueing behavior of the DQDB station is carried out in this section as follows. At first approximations on $\{e_j\}_{j\geq 0}$ and $\{r_j\}_{j\geq 0}$ are presented. Then the approximate DQDB station model is related to the general queueing system by establishing weaker sufficient conditions for the application of the analysis developed in section 3.

In section 2, the queueing model for the DQDB station was described in terms of the balance process. Equation (3) may be written as follows:

$$b_0 = 0, \qquad b_j = [b_{j-1} + r_j - \hat{e}_j + 1_j^d]^+ \qquad (11)$$

where $1_j^d = 1$ if a packet departure occurs at time j and it is zero otherwise. $\{\hat{e}_j\}_{j\geq 0}$ denotes the process of empty slots arriving to the station:

$$\hat{e}_j = \begin{cases} 1 & \text{if the } j^{th} \text{ slot on the forward bus } \underline{\text{arriving}} \\ & \quad \text{at the tagged station is empty} \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, $\hat{e}_j = e_j + 1_j^d$, where $\{e_j\}_{j\geq 0}$ is the empty slot process leaving the tagged station (1). An exact characterization of processes $\{\hat{e}_j\}_{j\geq 0}$ and $\{r_j\}_{j\geq 0}$ is difficult, since they reflect the cumulative activities of the upstream and the downstream stations, respectively, as modulated by the DQDB protocol and the stations' perception about the state of the system. In order to develop a tractable model for the balance process, processes $\{\hat{e}_j\}_{j\geq 0}$ and $\{r_j\}_{j\geq 0}$ are assumed to be Bernoulli with intensities $p_{\hat{e}}$ and p_r equal to $1-\lambda_{up}$ and λ_{down}, respectively; λ_{up} (λ_{down}) denotes the packet traffic rate generated by the upstream (downstream) stations. The following proposition will be useful for the study of the DQDB station.

Proposition 2: Under the Bernoulli assumption for $\{\hat{e}_j\}_{j\geq 0}$ and $\{r_j\}_{j\geq 0}$ the balance process in (11) is not Markov, but satisfies the following:

(a) $b_{t_k^{adv}}$ is probabilistically determined by $x_{t_{k-1}^d}$ and $b_{t_{k-1}^d}$,

(b) s_k is probabilistically determined by $b_{t_k^{adv}}$, and,

(c) $b_{t_k^d}$ is probabilistically determined by $b_{t_k^{adv}}$.

Conditions (a)-(c) are sufficient for $\left\{x_{t_k^d}, b_{t_k^d}\right\}_{k\geq 0}$ to be an embedded Markov chain on $\{t_k^d\}_{k\geq 0}$.

Proof: (a) If $x_{t_{k-1}^d} > 0$ then $t_k^{adv} \equiv t_{k-1}^d$ and, thus, $b_{t_k^{adv}} = b_{t_{k-1}^d}$. If $x_{t_{k-1}^d} = 0$ then $\tau_k = t_k^{adv} - t_{k-1}^d \neq 0$. From (11) it is easily established that the evolution of the balance process over a non-zero interval $(t_{k-1}^d, t_k^{adv}]$, $k \geq 1$, is described by

$$b_j = [b_{j-1} + r_j - \hat{e}_j]^+ \,,\; j \in (t_{k-1}^d, t_k^{adv}] \qquad (12)$$

Under the Bernoulli assumption for $\{r_j\}_{j\geq 0}$ and $\{\hat{e}_j\}_{j\geq 0}$, (12) implies that over the non-zero intervals $(t_{k-1}^d, t_k^{adv}]$, $k\geq 1$, process $\{b_j\}_{j\geq 0}$ is Markovian with transition probability matrix $\tilde{P}_b$,

$$\tilde{P}_b = \begin{bmatrix} p_0+p_1 & p_2 & 0 & 0 & \cdots \\ p_0 & p_1 & p_2 & 0 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \qquad (13)$$

where $p_0 = p_{\hat{e}}(1-p_r)$, $p_1 = p_{\hat{e}}p_r + (1-p_{\hat{e}})(1-p_r)$ and $p_2 = (1-p_{\hat{e}})p_r$. Thus, $b_{t_k^{adv}}$ is determined as the τ_k-step transition of the Markov process with transition matrix $\tilde{P}_b$.

(b) Let $T_{t_k^{adv}}(i)$ be defined to be the time interval until the occurrence of the stopping time of process $\{\hat{e}_j\}_{j\geq 0}$, defined as the time instant at which the i^{th} empty slot arrives at the tagged station following time instant t_k^{adv}. By definition of the effective packet service time, s_b, and the DQDB protocol it is easily established that if $b_{t_k^{adv}} = i, i \in S_b$, then $s_b(i) = T_{t_k^{adv}}(i+1)$ and thus, under the Bernoulli assumption for $\{\hat{e}_j\}_{j\geq 0}$,

$$p_{s_b}(i,l) = \Pr\left\{T_{t_k^{adv}}(i+1) = l\right\} = \binom{l-1}{i} p_{\hat{e}}^{i+1}(1-p_{\hat{e}})^{l-1-i} \qquad (14)$$

where $l > i$ and $i \in S_b$. The above prove part (b).

(c) It was shown in (b) that s_k is probabilistically determined by $b_{t_k^{adv}}$. Given that $s_k(b_{t_k^{adv}}) = l$, $l \geq 1$, $b_{t_k^d}$ is determined completely by the evolution of $\{r_j\}_{j\geq 0}$ over l slots. This is easily explained in terms of the DQDB protocol description. The content of the CD-CNTR decreases from $b_{t_k^{adv}}$ to zero during the effective service time interval $(t_k^{adv}, t_k^d]$ of length l and the state $b_{t_k^d}$ of the balance process at t_k^d reflects the new requests registered over the same interval. In view of this observation, the l-step probabilities of $\{b_j\}_{j\geq 0}$ over a busy sub-period s_k of length l, are easily determined in terms of the parameters of $\{r_j\}_{j\geq 0}$:

$$p_{b/>0}(m_1, m_2, l) = \binom{l}{m_2} p_r^{m_2}(1-p_r)^{l-m_2} \qquad (15)$$

for $m_1, m_2 \in S_b$ and $l \geq m_2$. It is concluded from the above that $b_{t_k^d}$ is determined by $b_{t_k^{adv}}$. Note that given the length of the effective service time l, $b_{t_k^d}$ is independent of $b_{t_k^{adv}}$ as illustrated in (15) where the right hand side is independent from m_1.

By applying arguments similar to those used in the case of the general queueing system it can be easily shown that $\left\{x_{t_k^d}, b_{t_k^d}\right\}_{k\geq 0}$ is a Markov chain. $\square$

A careful consideration of the study of the general queueing system presented in section 3 reveals that the analysis is valid for a similar queueing system whose underlying process $\{b_j\}_{j\geq 0}$ satisfies (a) and (c) and whose effective packet service time satisfies (b), in Proposition 2. Equations (13), (14) and (15) determine the elements of matrices $\tilde{P}_{b/0}$, $\tilde{P}_{s_b}(l)$ and $\tilde{P}_{b/>0}(l)$ which - together with $\tilde{P}_{b/0}(l) = [\tilde{P}_b]^l$ - replace the corresponding matrices for process $\{c_j\}_{j\geq 0}$ in (7)-(10) and, finally, determine transition matrix $\tilde{P}_{xb}$ (as in (4)) of the DQDB station model.

Note that conditions (a),(c) of Proposition 2 are also satisfied by a Markov process which evolves according to different transition matrices over idle and busy periods. The balance process is different from the above in the sense that it is not Markovian over busy periods.

The numerical complexity of the matrix analytic approach requires that the cardinality $N_b + 1$ of the state space S_b be finite and of small to moderate size. This requirement imposes a restriction on the maximum number of pending requests, N_b, as captured by the state of the balance process $\{b_j\}_{j\geq 0}$. The larger the network size in number of stations and physical length, the larger the maximum number of pending requests. A maximum value N_{max} on the number of pending requests is considered for the derivation of the numerical results on the DQDB station queue occupancy distribution presented in the next section. In particular, the evolution of the balance process over the idle periods $(t^d_{k-1}, t^{adv}_k]$, $k \geq 1$, given by (12), is truncated and it is determined by

$$b_j = \min\{[b_{j-1}+r_j-\hat{e}_j]^+, N_{max}\}\ ,\ j \in (t^d_{k-1}, t^{adv}_k]$$

Thus, the transitions of the balance process over the idle periods are determined by the stochastic matrix

$$\tilde{P}_{b,\max} = \begin{bmatrix} p_0+p_1 & p_2 & 0 & \dots & 0 & 0 \\ p_0 & p_1 & p_2 & \dots & 0 & 0 \\ 0 & p_0 & p_1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & p_0 & p_1+p_2 \end{bmatrix} \quad (16)$$

of order $(N_{max} + 1)$. Note that the stationary probabilities of (16) and (13) are given, respectively, by

$$\pi_{b,max}(m) = \frac{1-p}{1-p^{N_{max}+1}} p^m\ ,\ m \in S_{max}$$
$$\pi_b(m) = (1-p)p^m\ ,\ m \in S_b$$

where $p = p_2/p_0$, $S_{max} = \{0, 1, \dots, N_{max}\}$, and $S_b = \{0, 1, 2, \dots, \infty\}$. For sufficiently large N_{max}, $p^{N_{max}+1} < \delta$ and $\pi_{b,max}(m) - \pi_b(m) < \epsilon(\delta, m)$ for $m \in S_{max}$. For $\delta = 10^{-8}$ it was observed that an increment of N_{max} beyond the minimum value specified by this value of δ, did not result in any significant change of the values of the computed queue occupancy distributions. This value of N_{max} was also used for the truncation of the state space containing the allowed transitions of $\{b_j\}_{j\geq 0}$ over the busy sub-periods $(t^{adv}_k, t^d_k]$, $k \geq 1$, determined by (15).

5 Numerical results and conclusions

Results for the queue occupancy distribution of DQDB stations are presented in this section, under various values of network parameters.

The traffic load of a station is assumed to be linearly dependent on its position on the bus (number of downstream stations). That is, the traffic load of station i is given by

$$\lambda_i = (M-i)\lambda_0\ ,\ \lambda_0 = \frac{2\lambda}{M(M-1)} \quad (17)$$

where M denotes the number of stations in the DQDB network and λ is equal to the total traffic load on the bus. Various models have been considered for the packet arrival process to the DQDB station: Bernoulli, (truncated) Poisson, i.i.d with fixed size batches of lengths 3 (Batch-3), 4 (Batch-4) and 8 (Batch-8).

Figs. 2, 3 present the results obtained for stations No. 5 and 9, respectively, for $M = 10$, $\lambda = .7$ and packet arrival processes modeled as Bernoulli, Poisson and Batch-4. Simulation results are also presented for a DQDB network with identical parameters and interstation distance equal to one time slot. The analytically derived results are seen to be in good agreement with the simulations. Note that for both stations considered, the buffer occupancy is the largest under the Batch-4 arrival process and the smallest under the Bernoulli one. Note that the results obtained for station No. 9 under Bernoulli and Poisson arrivals almost coincide due to the fact that these processes almost coincide for small rates, which is the case for λ_9.

Fig. 4 (Fig. 5) presents analytical and simulation results for station No. 3 obtained for $M = 5$ and Bernoulli (Poisson) arrivals under traffic loads λ = .7, .8 and .9. As expected, the queue occupancy increases as λ increases.

Fig. 6 presents the analytical results obtained for station No. 2 obtained for $M = 5$, $\lambda = .7$ and under Bernoulli, Poisson, Batch-3 Batch-4 and Batch-8 packet arrival processes. Note that the queue occupancy increases with the batch size.

Fig. 7 presents the analytical results obtained for stations No. 2, 5 and 9 for $M = 5$, $\lambda = .7$ and under Poisson and Batch-4 packet arrivals. Notice that (for the traffic load distribution as in (17)) the larger the station number (more downstream) the smaller the queue occupancy and, thus, the smaller the required buffer size for certain packet loss probability.

Finally, Fig. 8 presents results obtained for station No. 3, $M = 5$, $\lambda = .7$ and under Poisson and Batch-3 arrivals. Simulation results are shown under interstation distances of 1 and 4 slots. The analytical results cannot capture the interstation distance. Notice that the buffer occupancy increases as the interstation distance increases under both arrival processes. Nevertheless, the analytical results seem to be in satisfactory agreement with the simulations, especially under Poisson arrivals.

The contribution of this paper may be summarized as follows. At first a queueing system of potentially wide interest has been formulated and studied, assuming that service time requirements depend on the state of an underlying Markov process at the time when the packet advances to the HoQ. Furthermore, it has been shown that the developed analysis is applicable under more relaxed conditions regarding the structure of the underlying process.

The second contribution of this work is the development of the queueing model for the DQDB station by properly identifying the process which captures the station's perception of the global distributed queue (balance process). An analytically tractable model for $\{b_j\}_{j\geq 0}$ is derived by assuming Bernoulli approximations to the key processes $\{\hat{e}_j\}_{j\geq 0}$ and $\{r_j\}_{j\geq 0}$. Although the correlated nature of these processes is not taken into consideration under this model, the analytical and simulation results for the queue occupancy distribution of the station were shown to be in satisfactory agreement. The effect of the truncation of the state space S_b of the approximate balance process on the derived results was seen to be insignificant. The model does not capture the effect of interstation distance.

The DQDB queueing model can be easily extended to include an approximation to the Bandwidth Balancing Mechanism by assigning the proper probability distributions to the effective service times.

References

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

[2] C. Bisdikian, "Waiting Time Analysis in a Single Buffer DQDB (802.6) Network," IEEE J. Select. Areas Commun., vol. SAC-8, pp. 1565-1573, Oct.1990.

[3] B. Mukherjee and C. Bisdikian, "A Journey Through the DQDB Network Literature," IBM Report RC 170166 (75478), 7/11/91, T.J. Watson Research Center, Yorktown Heights.

[4] M. Conti, E. Gregori and L. Lenzini, "On the Approximation of the Slot Occupancy Pattern in a DQDB Network," Proc. IEEE Infocom'92, Florence, Italy, pp. 518-526, May 1992.

[5] W. Jing and M. Paterakis, "Message Delay Analysis of the DQDB (IEEE 802.6) Network," Proc. IEEE Infocom'92, Florence, Italy, pp. 527-535, May 1992.

[6] J. Liebeherr, I. Akyildiz and A. Tantawi, " An Effective Scheme for Pre-Emptive Priorities in Dual Bus Metropolitan Area Networks," Proc. ACM Sigcomm'92, Baltimore.

[7] I. Stavrakakis and R. Landry, "Delay Analysis of the DQDB Man Based on a Simple Model," Proc. ICC'92, Chicago, IL, pp. 154-158, June 1992.

[8] I. Stavrakakis and S. Tsakiridou, "A Queueing System with Markov-Structured Service Availabilities," Research Report CSEE/92/06-03, CSEE Dept., University of Vermont.

[9] M.F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Marcel-Dekker, New York, 1989.

[10] J. Daigle, *Queueing Theory for Telecommunications*, Addison Wesley, 1992.

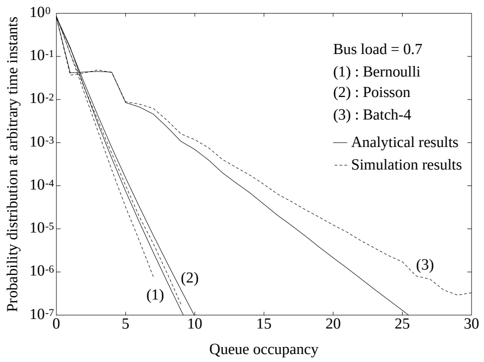

Fig. 2: Results for station No. 5 of a DQDB bus with 10 stations under various arrival processes.

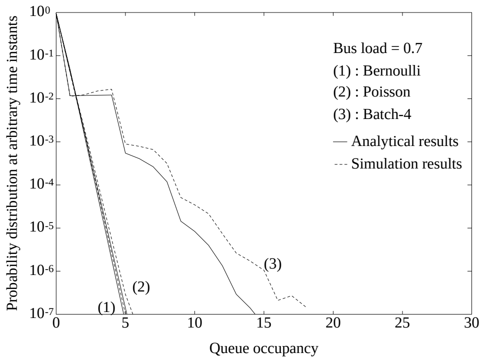

Fig. 3: Results for station No. 9 of a DQDB bus with 10 stations under various arrival processes.

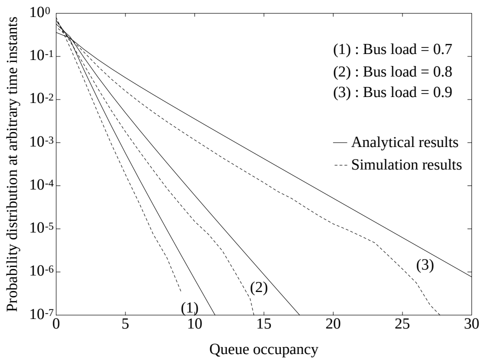

Fig. 4: Results for station No. 3 of a DQDB bus with 5 stations under Bernoulli arrivals and various bus loads.

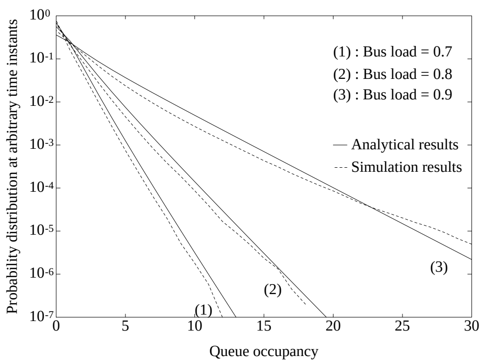

Fig. 5: Results for station No. 3 of a DQDB bus with 5 stations under Poisson arrivals and various bus loads.

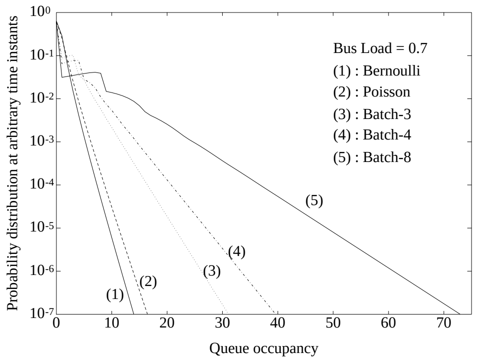

Fig. 6: Analytical results for station No. 2 of a DQDB bus with 5 stations under various arrival processes.

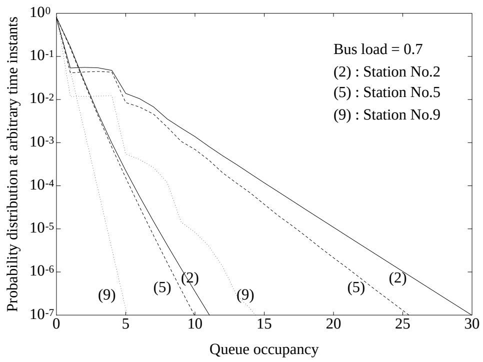

Fig. 7: Analytical results for stations No. 2,5,9 of a DQDB bus with 10 stations under Poisson and Batch-4 arrivals.

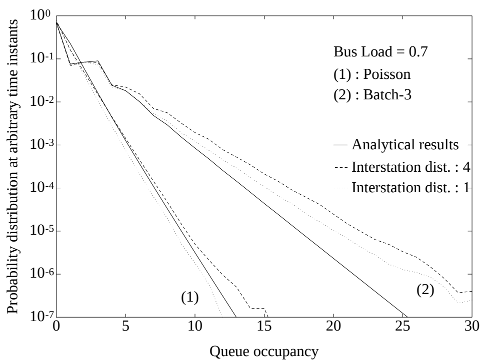

Fig. 8: Results for station No. 3 of a DQDB bus with 5 stations, under Poisson and Batch-3 arrivals and different interstation distance.