Efficient Modeling of Merging and Splitting Processes in Large Networking Structures

Ioannis Stavrakakis, *Member, IEEE*

Abstract—**The packet traffic generated by a source of information undergoes transformation at both the access points and within a large networking structure, due to merging and splitting operations. Such points of traffic transformation are present in almost all networking systems, such as high-speed switching systems, and systems of interconnected LAN's and B-ISDN's. In this paper, simple models are developed for the description of a first-order Markovian (bursty) process modulated by merging and/or (independent or correlated) splitting operations. The developed models can be adopted for the packet traffic description in large networking structures, supporting multimedia packet traffic. This is due to the fact that the complexity of these models does not change as the number of points of transformation increases. On the other hand, the bursty traffic model is capable of describing a variety of packet sources. The induced packet delay at the merging points is evaluated. A queueing system is studied for this purpose under a general class of arrival processes, which include the adopted packet traffic models. Numerical results are obtained which are compared with simulations whenever approximations are involved.**

I. Introduction

PACKET networks have been evolved as the most efficient supporting structure for the communication among geographically separated information systems. Efficient utilization of the network resources requires that they be shared by a number of potential users. Accessing to the common resources (links) is regulated by efficient protocols. The performance of these protocols is usually evaluated in terms of the maximum amount of information that can be delivered (with finite delay) and the induced packet delay [1].

In today's complex networking structures, the packetized information may be generated by sources with quite different characteristics. In addition, the process of the packets generated by a particular source may be modulated by a number of network resource access mechanisms before it reaches its destination. The diversity of the packet generation mechanism of the sources, and the various multiplexing phases the packets have to go through, result in complex packet processes. The lack of an accurate description of the packet processes arriving at a network resource presents the major difficulty in evaluating its performance. The fundamental problem in this case is that of the characterization of the packet output processes generated by the network elements (nodes) supported by the networking structure. The latter is a major issue in most of today's networking structures such as those in a multihop environment, systems of interconnected networks, metropolitan area networks (MAN's), and switching networks [2]–[7], [14].

The analytical tractability inherent to the memoryless models has tempted many researchers to adopt such models for the description of complex network traffic processes. Such processes are, for instance, the internetwork traffic in systems of interconnected networks [2]–[4], [8], [9] and the internode traffic in interconnection networks [14]. The adoption of such models, however, may lead to erroneous identification of the bottlenecks of the system and to erroneous packet delay calculations [5]. Accurate models for the description of the dependent packet processes in systems of interconnected networks are usually either difficult to develop or lead to system models which are not analytically tractable [6], [10].

The discussion of this paper is confined to packetized slotted communication networks. Discrete time queueing models are developed for the analysis of large systems of interconnected nodes, where packet processes are described by discrete time point processes. In this paper, the exogenous packet traffic is assumed to be generated by bursty sources and it is modeled as a first-order discrete-time Markov process. This traffic is served by the networking structure through sequential transmission over fixed-speed lines. Unlike a memoryless model, the first-order Markov model captures some of the dependencies introducted by the source packet generating mechanism. At the same time, this model is simple enough to lead to a tractable analysis of the system performance. It has been shown in [13] that the first-order Markov model is a well-performing one for the description of the packet traffic modulated by mixing and splitting operations in a large network. This model has been shown to outperform the Bernoulli one in describing the output process generated by a random-access multiuser communication network [6]. A procedure for the calculation of the parameters of the model has been developed based on the true statistics of the relevant events in the output process. The results in [5] show that the output process of a statistical multiplexer (buffer) can be modeled by a first-order Markov process quite adequately. It has also been observed that this model can capture, to some extent, the strong corre-

Manuscript received January 30, 1991; revised May 6, 1991. This paper has been supported by the National Science Foundation Grant NCR-9011962.

The author is with the Department of Electrical Engineering and Computer Science, University of Vermont, Burlington, VA 05405.

IEEE Log Number 9101960.

0733-8716/91/1000-1336$01.00 © 1991 IEEE

lation between packets over consecutive cells of an ATM network [11]. Finally, the first-order Markov model has been adopted for the description of packetized voice and video traffic.

The exogenous packet traffic undergoes transformation as its travels through a large networking structure, due to merging and splitting operations. Traffic merging is necessary for the efficient utilization of common resources, due to the probabilistic nature of the network packet traffic. Traffic splitting is also unavoidable due to network routing decisions. Appropriate models for the description of the packet processes generated by traffic merging and splitting operations (points of transformation) are developed in this paper. These models take into consideration the structure and the values of the parameters of the packet processes before the point of transformation. An important characteristic of these models is that the complexity in the calculation of their parameters does not change as the number of points of transformation (stages) increases. The developed models are exact in the case of a splitting point and they are approximate in the case of a merging point. Both independent and correlated splitting operations are considered. The queueing systems formulated at the nodes of the networking structure are studied, and numerical results are obtained. Simulation results are provided for the cases in which the packet traffic processes are approximately described.

Approximate models for the traffic description in large networks have been considered in [16]–[18]. These models describe traffic as modulated by splitting and merging operations, whose complexity does not increase with the number of points of transformation. The systems considered there are continuous time, and the traffic description is based on the rate and variance. No Markovian structure is assumed, and only the case of independent splitting is considered.

The rest of the paper is organized as follows. In Section II, the bursty traffic model is described and the models for the characterization of the bursty traffic, as modulated by splitting and/or merging operations, are presented. In Section III, a relevant queueing system is studied under arrival processes described according to the traffic modeling introduced in Section II. In Section IV, the traffic modeling and queueing results developed in Sections II and III are incorporated in the performance evaluation of a simple interconnecting topology. Numerical results, both analytical and from simulations, are presented, and conclusions are drawn regarding the effects of splitting and merging operations on the performance of networking structures. Finally, the conclusions of this work are presented in the last section.

II. The Packet Traffic Models

In this section, the various packet traffic models adopted in this work are presented. For the reasons explained in the introduction, the sources of packetized information are considered to be bursty. These sources generate an exogenous packet traffic which is modeled as a first-order Markov process with state space $\{0, 1\}$ and parameters λ and γ; λ denotes the packet traffic rate and γ denotes the traffic burstiness coefficient defined by

$$\gamma = p(1, 1) - p(0, 1). \tag{1}$$

$p(i, j)$ denotes the transition probability from state i to state j, $i, j \in \{0, 1\}$. Throughout this paper, the packet traffic generated by the exogenous sources is assumed to be delivered to the networking structure through fixed-speed slotted transmission lines. A time slot is assumed to be equal to the packet transmission time. State 0 (idle state) generates no packet with probability one (empty slot); state 1 (active state) generates one packet with probability one (busy slot). The state transition diagram of the bursty traffic model is shown in Fig. 1(a). A bursty traffic source, which generates packets according to the Markov model shown in Fig. 1(a) with parameters λ and γ, will be denoted by $B(\lambda, \gamma)$. This traffic process undergoes transformation at the nodes of the network. Models for the resulting process are developed in the sequel. The complexity of these models is independent of the number of transformations.

At certain points of the interconnecting structure, routing decisions need to be taken. A packet is assumed to be routed along the tagged direction with probability p; it is routed along any other direction with probability $1 - p$. This routing policy is defined as the independent splitting operation and it is depicted in Fig. 2(a). Since the independent splitting destroys some of the memory present in the original bursty traffic $B(\lambda, \gamma)$, a Bernoulli approximation $\tilde{B}(\lambda')$ of the resulting process would appear to be meaningful. Although the independent splitting generates a memoryless process within the duration of a burst (active period), it fails to completely destroy the memory present in the original bursty traffic process. The independent splitting of the original bursty traffic results in a process which generates packets according to a Bernoulli process over a geometrically distributed length, followed by periods of inactivity of geometrically distributed length. This process clearly has memory and it is not a first-order Markov process. One way to capture this memory, to some extent, would be by approximating the resulting process by a bursty model $\tilde{B}(\lambda', \gamma')$, whose parameter λ' and γ' are related to those of the original process $B(\lambda, \gamma)$. A meaningful procedure for the calculation of the parameters of the approximate model would be to equate its probabilities to the probabilities that certain corresponding events occur in the true resulting process. This is done in the following proposition.

Proposition 1: Let $\tilde{B}(\lambda', \gamma')$ be a bursty traffic (first-order Markov) model approximating the packet traffic process generated by applying independent splitting with parameter p [Fig. 2(a)] on a bursty traffic process $B(\lambda, \gamma)$. The values of the parameters of $\tilde{B}(\lambda', \gamma')$, which determine a process that generates the true probabilities of occurrence of any specific slot or pair of consecutive slots, are given by:

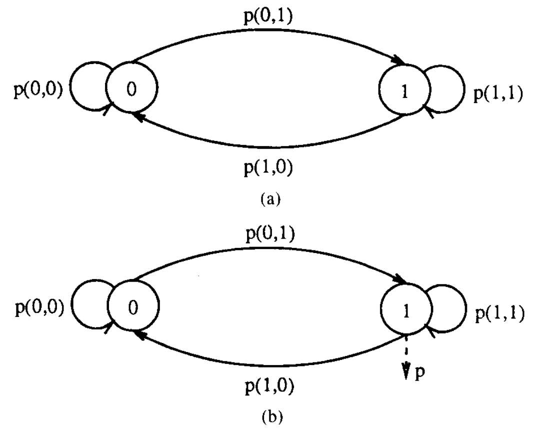

Fig. 1. (a) The Markov model for the bursty traffic $B(\lambda, \gamma)$. (b) The Markov model for the bursty traffic $B(\lambda', \gamma', p)$ after independent splitting.

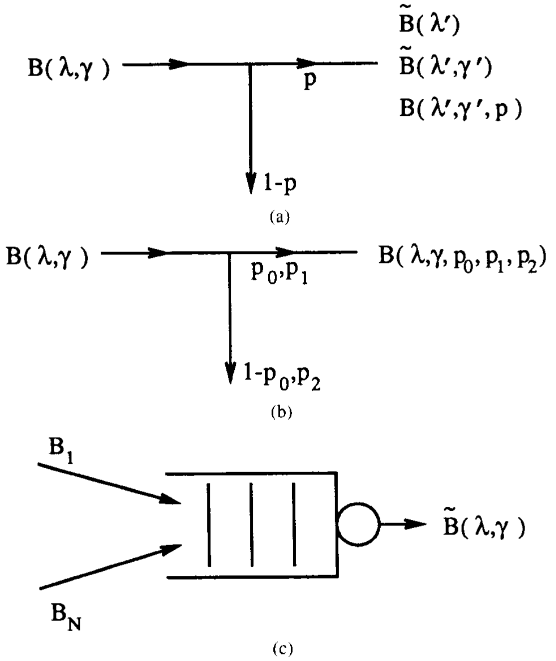

Fig. 2. (a) The independent splitting operation. (b) The correlated splitting operation. (c) The merging operation.

$$\lambda' = p\lambda \qquad \gamma' = \frac{1-\lambda}{1-\lambda p} p\gamma. \tag{2}$$

The proof of this proposition is given in Appendix A. □

Although the bursty traffic approximation $\tilde{B}(\lambda', \gamma')$ is expected to outperform the Bernoulli one $\tilde{B}(\lambda')$ (with parameter $\lambda' = p\lambda$), it is expected that this approximation will be inadequate for the description of the resulting process. Although the parameter calculation approach presented in Proposition 1 seems to be meaningful, it fails to "look" a the process beyond the previous slot and get more information as to whether the original process is in the active state or not. As a result, the equivalent burstiness coefficient γ' is expected to determine an approximating process $\tilde{B}(\lambda', \gamma')$, which causes less severe queueing problems than the true process. The latter is expected due to the inadequate amount of memory captured by $\tilde{B}(\lambda', \gamma')$, as reflected by the value of γ' calculated by (2).

The objective in the above discussion and the development of the approximating models $\tilde{B}(\lambda')$ and $\tilde{B}(\lambda', \gamma')$ is to help get insight into the true process and illustrate its considerably different structure from those of processes $\tilde{B}(\lambda')$ and $\tilde{B}(\lambda', \gamma')$. The performance of the approximations $\tilde{B}(\lambda')$ and $\tilde{B}(\lambda', \gamma')$ will be evaluated in terms of the accuracy of the packet delay results induced by a multiplexer fed by such approximating processes. These queueing results, as well as those under exact modeling of the resulting process, will be obtained in the next section. An exact modeling of the process generated by applying independent splitting on bursty traffic is described in the sequel.

The true packet traffic process generated by applying independent splitting on the bursty process $B(\lambda, \gamma)$ [Fig. 2(a)] is exactly described in terms of a Markov modulated Bernoulli process (MMBP). Its state transition diagram is shown in Fig. 1(b). The underlying first-order Markov model is identical to that in the original traffic $B(\lambda, \gamma)$. A packet is assumed to be generated when the underlying Markov chain is in state 1 (active) with probability p; no packet is generated from state 0 (idle). This MMBP will be denoted by $B(\lambda', \gamma', p)$, where $\lambda' = \lambda/p$, $\gamma = \gamma'$ and p is determined by the splitting probability.

The analysis of certain queueing systems is, in general, less complex under packet arrival processes described by $B(\lambda)$ (memoryless process) compared to that under processes described by $B(\lambda, \gamma)$ (process with memory). Since process $\tilde{B}(\lambda', \gamma')$ is a special case of the process $B(\lambda, \gamma, p)$, the analysis of certain queueing system is, in general, less complex under packet arrival processes described by $\tilde{B}(\lambda', \gamma')$ compared to that under processes described by $B(\lambda', \gamma', p)$. For instance, important boundary probabilities are trivially computed under $B(\lambda, \gamma)$ arrival processes (Section III). This is not the case under $B(\lambda, \gamma, p)$ arrival processes; in the latter case, the development of a complicated approach is required. In the next section, a simple queueing system is studied under the exact modeling of split bursty traffic. Exact queueing results are derived to be compared with those obtained under the simple approximation models. Based on this comparison, conclusions regarding the performance of the approximate models will be drawn and the expectations for good accuracy of these models in more complex queueing situations will be better shaped.

In this sequel, the case of correlated splitting of the bursty traffic $B(\lambda, \gamma)$ is considered. The resulting (after the splitting) traffic along the tagged direction is denoted by $B(\lambda, \gamma, p_0, p_1, p_2)$, where p_0, p_1, and p_2 are the parameters of the splitting process [Fig. 2(b)]. According to the correlated splitting policy adopted in this paper, the first packet of an active period in process $B(\lambda, \gamma)$ appears in process $B(\lambda, \gamma, p_0, p_1, p_2)$ with probability p_0. Given that a packet appeared in $B(\lambda, \gamma, p_0, p_1, p_2)$ in the previous slot, a packet appears in $B(\lambda, \gamma, p_0, p_1, p_2)$ in the current slot with probability p_1, provided that process $B(\lambda, \gamma)$ is still active. Given that a packet did not appear in $B(\lambda, \gamma, p_0, p_1, p_2)$ in the previous slot although process $B(\lambda, \gamma)$ was active, a packet appears in the current slot with probability $1 - p_2$, provided that $B(\lambda, \gamma)$ is still active. The values of the parameter p_1 and p_2 determine the degree of the correlation in the splitting process; the larger the value of p_1 the larger the probability that consecutive packets will be routed along the tagged direction. Unlike the independent splitting policy, which generates a memoryless process within an active period of $B(\lambda, \gamma)$, the correlated splitting policy results in a process which has memory within an active period of $B(\lambda, \gamma)$. This splitting policy becomes an independent one if $p_0 = p_1 = 1 - p_2$. The correlated splitting operation captures the realistic situation in which consecutive packets delivered by a source (or node) are highly likely to have the same destination and, thus, follow the same network path. The latter might be the case when consecutive packets are originated from the same network user or when they are part of a long message.

The correlated splitting policy applied on process $B(\lambda, \gamma)$ is easily seen to generate a process $B(\lambda, \gamma, p_0, p_1, p_2)$ that is exactly described in terms of a three-state Markov modulated Bernoulli model with the state transition diagram shown in Fig. 3. State 0 corresponds to the idle state of process $B(\lambda, \gamma)$. State 1_1 describes the state in which process $B(\lambda, \gamma)$ is active and the packet is routed along the tagged direction. State 1_0 describes the state in which process $B(\lambda, \gamma)$ is active and the packet is not routed along the tagged direction. The transition probabilities are described in terms of those of the original process $B(\lambda, \gamma)$ and the splitting probabilities p_0, p_1, and p_2. State 1_1 generates a packet with probability 1. States 1_0 and 0 generate no packets with probability 1.

In the case of the correlated splitting, it is important that a measure of the correlation be defined. The correlation coefficient, defined by:

$$c = p_1 + p_2 - 1 \tag{3}$$

possesses a number of useful properties. When $p_0 = p_{eq}$, where p_{eq} is equal to the portion of packets forwarded along the tagged direction, then $c = 0$ corresponds to the case of independent splitting. In this case, $p_0 = p_1 = 1 - p_2 = p_{eq}$, which implies that a packet of the original bursty process $B(\lambda, \gamma)$ is forwarded along the tagged direction with probability p_{eq} independently of the decision taken in the previous slot. The maximum value of c, $c = 1$, is achieved when $p_1 = p_2 = 1$. In this case, states 1_1 and 1_0 do not communicate directly and splitting is performed on a burst-by-burst basis. That is, a (whole) burst is forwarded along the tagged direction with probability p_0 (bursty splitting). If $c = 1$ and $p_0 = 1$, then no splitting

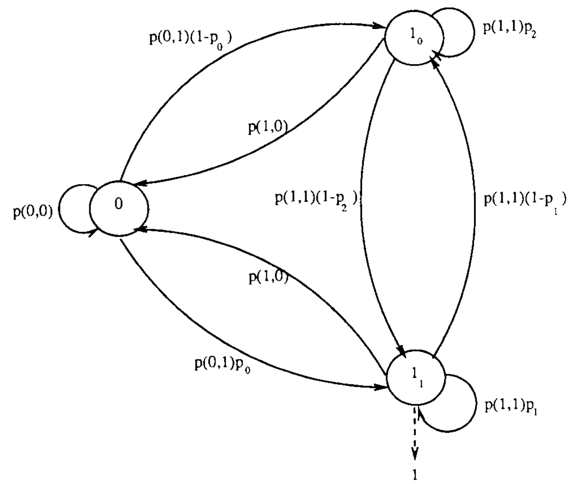

Fig. 3. The three-state Markov modulated Bernoulli model for the bursty traffic $B(\lambda, \gamma, p_0, p_1, p_2)$.

is in effect and the resulting process is identical to the original process $B(\lambda, \gamma)$.

Unlike the case of independent splitting, which reduces significantly the amount of memory in the resulting packet process, it is expected that correlated splitting will preserve (to a greater extent) the memory present in the original process $B(\lambda, \gamma)$. Thus, it appears to be meaningful to construct an equivalent bursty traffic model $\tilde{B}(\lambda', \gamma')$ to approximately describe the resulting process $B(\lambda, \gamma, p_0, p_1, p_2)$. This is done by simply merging states 0 and 1_0 of the exact three-state Markov model $B(\lambda, \gamma, p_0, p_1, p_2)$. The parameters of $\tilde{B}(\lambda', \gamma')$ are given in the next proposition. Its proof is similar to that of Proposition 1.

Proposition 2: Let $\tilde{B}(\lambda', \gamma')$ be a bursty traffic model approximating the packet traffic process generated by applying correlated splitting, with parameters p_0, p_1, p_2, on a bursty traffic process $B(\lambda, \gamma)$. The values of the parameters of $\tilde{B}(\lambda', \gamma')$, which determine a process that generates the true probabilities of occurrence of any specific slot or pair of consecutive slots, are given by:

$$\lambda' = \pi(1_1),$$

$$\gamma' = \frac{p_1 p(1, 1) - \pi(1_1)}{1 - \pi(1_1)} = \frac{p_1 p(1, 1) - 1}{1 - \pi(1_1)} + 1 \tag{4}$$

where $\pi(1_1)$ is the steady-state probability that the three-state Markov chain, shown in Fig. 3, is in state 1_1. $p(1, 1)$ is the transition probability from state 1 to state 1 in the original bursty traffic $B(\lambda, \gamma)$. □

In the previous paragraphs, exact models were developed for the description of the packet traffic generated by independent or correlated splitting, applied on a bursty packet traffic process. When traffic is merged, the resulting packet process is more difficult to be exactly described. This is due to the complexity of the merging (queueing) process, due to the fact that packets may have to be temporarily stored in a buffer before they are transmitted [Fig. 2(c)]. Unlike the splitting process, which tends to destroy the original burstiness and make bursty modeling less accurate, the merging process results in more regular packet traffic. This traffic generates one packet when the buffer is nonempty and zero packets when the buffer is empty. In this paper, it is assumed that the buffer output line has the same speed with the feeding lines and, thus, the slots at the input and output have the same length. The potential packet departure and arrival points (slot boundaries at the input and the output) are assumed to coincide.

The bursty traffic (first-order Markov) model is adopted for the characterization of the output of a merging process. State 0 corresponds to the empty buffer state; state 1 corresponds to the nonempty buffer state. It has been observed in previous studies [5] that this model performs quite satisfactorily. The parameters of the approximate model $\tilde{B}(\lambda', \gamma')$ for the buffer output process are computed in terms of the true probabilities that any specific slot (empty or busy) or pair of consecutive slots appear in the buffer output process. These parameters are derived in the next proposition.

Proposition 3: Let $\tilde{B}(\lambda', \gamma')$ be a bursty traffic (first-order Markov) model describing the packet output process of a buffer fed by N bursty or split bursty traffic processes B_k, $1 \leq k \leq N$, where $B_k \in \{B(\lambda), B(\lambda, \gamma), B(\lambda, \gamma, p), B(\lambda, \gamma, p_0, p_1, p_2)\}$. The values of the parameters of $\tilde{B}(\lambda', \gamma')$, which determine a process that generates the true probabilities of occurrence of any specific slot or pair of consecutive slots, are given by:

$$\lambda' = 1 - \rho_0, \qquad \gamma' = p_{00}\frac{1 + \rho_0}{\rho_0} \tag{5}$$

where ρ_0 denotes the probability that the buffer is empty and p_{00} denotes the joint probability of occurrence of a pair of consecutive empty slots in the output process. These probabilities are given by:

$$\rho_0 = 1 - \sum_{k=1}^{N} \sum_{x \in S^k} \pi^k(x) q_1^k(x) \tag{6}$$

$$p_{00} = \sum_{\bar{x} \in \bar{S}} \tau(0; \bar{x}) q_0(\bar{x}) \tag{7}$$

where S^k denotes the state space of the Markov chain which describes the kth input process; $S^k = \{0\}$ if $B_k = B(\lambda)$, $S^k = \{0, 1\}$ if $B_k = B(\lambda, \gamma)$ or $B_k = B(\lambda, \gamma, p)$ and $S^k = \{0, 1_1, 1_0\}$ if $B_k = B(\lambda, \gamma, p_0, p_1, p_2)$; $\pi^k(x)$ denotes the steady-state probability that the kth input process is in state x, $x \in S^k$; $q_j^k(x)$ denotes the probability that state x of the kth input process generates j packets, $j = 0, 1$; $\bar{S} = S^1 \times S^2 \times \cdots \times S^N$ denotes the state space of the N-dimensional Markov chain formed by the N input Markov chains; $q_j(\bar{x})$ denotes the probability that state $\bar{x} \in \bar{S}$ generates j packets; $0 \leq j \leq N$; $\tau(0; \bar{x})$ denotes the (boundary) probability that the buffer is empty and the input Markov chain is in state $\bar{x}$ (packets generated by $\bar{x}$ are assumed to arrive at the end of the current slot). These boundary probabilities are computed in the next section for the case of arbitrary input processes B_k, $1 \leq k \leq N$. When the input processes B_k are such that $B_k \in \{B(\lambda), B(\lambda, \gamma)\}$ for all k, $1 \leq k \leq N$, then the boundary probabilities are easily shown to be given by:

$$\tau(0, \bar{x}) = \rho_0 \prod_{k=1}^{N} p^k(0, x^k), \qquad \bar{x} = (x^1, \cdots, x^N) \in \bar{S} \tag{8}$$

where $p^k(0, x^k) = 1$ if $B_k \in \{B(\lambda)\}$; $p^k(x^k, y^k)$ denotes the transition probability that the underlying Markov chain associated with the kth input process moves from state x^k to state y^k. The proof of this proposition is given in Appendix A. □

The computation of the boundary probabilities $\tau(0; \bar{x})$ receives special attention in the next section. A method for the derivation of these probabilities is developed and an important queueing system is studied.

III. The Queueing System

In this section, the queueing system depicted in Fig. 2(c) is considered. This system is adopted as a model for the study of the traffic merging points (nodes) of the networking structure. The input lines are assumed to be slotted and packet arrivals and service completions are assumed to be synchronized with the end of the slots. A slot is defined to be equal to the fixed service (transmission) time required by a packet. At most, one packet may be served over a slot. The first-in first-out (FIFO) service discipline is adopted. Packets arriving at the same slot are served in a randomly chosen order. The buffer capacity is assumed to be infinite. Each of the N (discrete time) input lines is assumed to deliver packetized information according to one of the processes $B(\lambda)$, $B(\lambda, \gamma)$, $B(\lambda, \gamma, p)$ or $B(\lambda, \gamma, p_0, p_1, p_2)$. Let B denote the collection of these processes.

The discrete-time queueing system described above has been studied in [12], under packet arrival processes from the class of the $m/MM/r/B$ processes. An $m/MM/r/B$ process is a discrete-time Markov modulated generalized Bernoulli process, which may deliver up to r packets at a time. This process is defined in terms of an underlying Markov chain and a set of probabilities associated with its states. The state space S of the Markov chain has cardinality M equal to m. Given the current state j of this Markov chain, k packets are generated with probability $q_k(j)$, $0 \leq k \leq r$, $j \in S$. The moments of the buffer occupancy process have been computed in terms of the solutions of $M^1 \times M^2 \times \cdots \times M^N$ linear equations, where M^k is the value of M associated with the kth input process. The mean packet delay has been computed by invoking Little's theorem. These equations are given in Appendix B.

Application of the queueing results derived in [12] require knowledge of the boundary probabilities that the buffer is empty and the state of the N-dimensional underlying Markov chain is $\bar{x}$, $\bar{x} \in \bar{S}$. This Markov chain is formulated by the underlying Markov chains associated with the N inputs. Its state space is denoted by $\bar{S} = S^1 \times$

$S^2 \times \cdots \times S^N$, where S^k is the state of the kth input Markov chain. As discussed in [12], when there exists only one state $\bar{x}_0$, $\bar{x}_0 \in \bar{S}$, such that $q_0(\bar{x}_0) > 0$ and $q_0(\bar{x}) = 0$ for $\bar{x} \in \bar{S} - \{\bar{x}_0\}$, then the boundary probabilities can be computed from the following equation.

$$\tau(0; \bar{x}) = \rho_0 \prod_{k=1}^{N} p^k(x_0^k, x^k) \tag{9}$$

where superscript k indicates quantities associated with the kth input process.

Clearly, B is a subclass of the class of the $m/MM/r/B$ processes. Process $B(\lambda, \gamma, p_0, p_1, p_2)$ is a $3/MM/1/B$ process based on an underlying Markov chain with two states (states 1_0 and 0) which generate no packets. Thus, since $q_0(0) = 1 \neq 0$ and $q_0(1_0) = 1 \neq 0$, (9) is not applicable. Similarly, process $B(\lambda, \gamma, p)$ is a $2/MM/1/B$ process based on an underlying Markov chain with two states (state 0 and 1) which generate no packets. Thus, since $q_0(0) = 1 \neq 0$ and $q_0(1) = p \neq 0$, (9) is again not applicable.

The above discussion implies that the analysis of the queueing system under study, when input processes from $\{B(\lambda, \gamma, p), B(\lambda, \gamma, p_0, p_1, p_2)\}$ are present, requires the development of an approach for the calculation of the boundary probabilities $\tau(0; \bar{x})$, $\bar{x} \in \bar{S}$. These probabilities are also required for the evaluation of the parameters of a bursty traffic approximation of the packet process generated by merging processes from $\{B(\lambda, \gamma, p), B(\lambda, \gamma, p_0, p_1, p_2)\}$, as described in Proposition 3. The boundary probabilities $\tau(0; \bar{x})$, $\bar{x} \in \bar{S}$ are derived in the next theorem.

Theorem 1: The boundary probabilities $\tau(0; k)$ that the buffer is empty and the state of the N-dimensional input Markov chain is k, $1 \leq k \leq \bar{M}$ ($\bar{M}$ is the cardinality of the N-dimensional input Markov chain), is given by:

$$\tau(0; k) = \pi_w(k)\rho_0, \qquad 1 \leq k \leq \bar{M} \tag{10}$$

where ρ_0 is the probability that the buffer is empty and $\pi_w(k)$, $1 \leq k \leq \bar{M}$, are the solutions of the matrix equation

$$[\Pi_w] = [\Pi_w][P_w] \tag{11a}$$

$$[\Pi_w]\bar{e} = 1 \tag{11b}$$

where $[\Pi_w] = [\pi_w(1), \cdots, \pi_w(\bar{M})]$, $\bar{e} = [1, \cdots, 1]^T$ is the $\bar{M}$-dimensional unit column vector and $[P_w]$ is an $\bar{M} \times \bar{M}$ matrix with elements p_{ij}, $1 \leq i, j \leq \bar{M}$, computed from the matrix equation

$$[P_w] = \sum_{k=0}^{N_0} [Q_k][P][P_w]^k \tag{11c}$$

where N_0 is the maximum number of packet arrivals in a slot; $[Q_k]$ is an $\bar{M} \times \bar{M}$ diagonal matrix with elements $q_i(k)$, $1 \leq k \leq \bar{M}$, $0 \leq i \leq N_0$, where $q_i(k)$ is defined in Proposition 3; $[P]$ is the transition matrix of the N-dimensional input Markov chain; $[P_w]^k$ is the kth power of $[P_w]$ ($[P_w]^0 = [I]$). □

Proof of Theorem 1: Let $J = \{0, 1, 2, \cdots\}$ be the index set of the instants (slot boundaries) when the buffer is empty. Let w_j, $w_j \in \bar{S}$ be the state of the N-dimensional input Markov chain at instant j, $j \in J$. Clearly, $\{w_j\}_{j\in J}$ is a Markov chain with state space $\bar{S}$; let $\pi_w(j)$ and $p_w(i, j)$, $i, j \in \bar{S}$, denote the steady-state and transition probabilities of $\{w_j\}_{j\in J}$, respectively. $p_w(i, j)$ is the probability that the input Markov chain moves from state i, at an instant when the buffer is empty, to state j at the first instant in the future when the buffer becomes empty again. The transition probabilities $p_w(i, j)$ and $i, j \in \bar{S}$ satisfy the following equations:

$$p_w(i, j) = p(i, j)q_0(i) + \sum_{k=1}^{N_0} q_k(i) \sum_{n\in\bar{S}} p(i, n)P\{E_{nj}^k\}, \quad i, j \in \bar{S} \tag{12}$$

where $q_k(i)$, $0 \leq k \leq N_0$, $i \in \bar{S}$, and $p(i, j)$, $i j \in \bar{S}$, are defined in Proposition 1. N_0 is the maximum number of packet arrivals in a single slot. E_{nj}^k denotes the event that the input Markov chain moves from state n, at an instant when the buffer content is equal to k, to state j at the first instant in the future when the buffer becomes empty. This event is equivalent to event R_{nj}^k; R_{nj}^k denotes the event that the input Markov chain moves from state n, at the beginning of a sequence of k consecutive R reductions, to state j at the end of the last of these R reductions. An R reduction with respect to time instant i_1 occurs at the first instant (following i_1), i_2, at which the buffer content is reduced by one with respect to its content at instant i_1; instants i_1 and i_2 are defined as the beginning and the end of the R reduction, respectively. In a sequence of consecutive R reductions, the end of the R reduction is assumed to coincide with the beginning of the one that follows. A realization of the buffer occupancy process $V(k)$, in which the event E_{nj}^5 occurs is shown in Fig. 4, where the beginning and the end of the R reductions of interest are marked; $V(k)$ is assumed to be right-continuous. The state of the input Markov chain at the time instants of interest is shown in parentheses. The equivalence of the events E_{nj}^5 and R_{nj}^5 is easily established in this realization.

Let $c(i, j)$, $i, j \in \bar{S}$ denote the probability that the input Markov chain moves from state i at the beginning of an R reduction to state j at the end of this R reduction; let $[C]$ denote the stochastic matrix with elements $c(i, j)$, $i, j \in \bar{S}$. Let $c^{(k)}(i, j)$ denote the probability that the input Markov chain moves from state i, at an instant when the buffer content is equal to k, to state j, at the first instant in the future when the buffer becomes empty. $c^{(k)}(i, j)$ is equal to the probability that the end of the last of a sequence of k consecutive R reductions finds the input Markov chain in state j, given that the input Markov chain is in state i at the beginning of the first of those R reductions. This probability is the (i, j)th element of the kth power of $[C]$. The previous discussion establishes that:

$$P\{E_{nj}^k\} = P\{R_{nj}^k\} = c^{(k)}(n, j)$$

and, thus, (12) becomes:

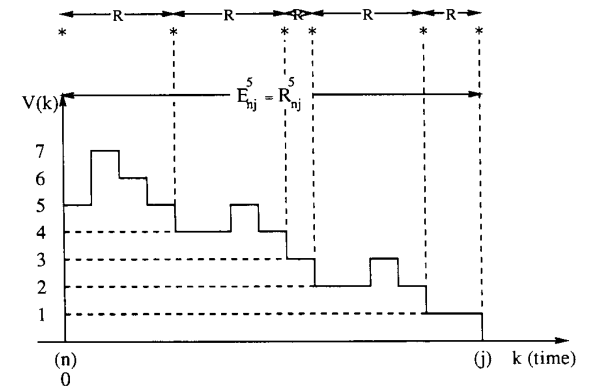

Fig. 4. A realization of the buffer occupancy process $V(k)$ where the events R^5_{nj} and E^5_{nj} are shown, as well as the five consecutive R reductions.

$$p_w(i, j) = p(i, j)q_0(i) + \sum_{k=1}^{N_0} q_k(i) \sum_{n \in \bar{S}} p(i, n)c^{(k)}(n, j). \tag{13}$$

To derive (11) and complete the proof of the theorem, the following lemma is invoked.

Lemma 1:

$$[P] = [C]. \tag{14}$$

Proof of Lemma 1: It is easy to establish the following equations for the transition probabilities $c(i, j)$, $i, j \in \bar{S}$:

$$c(i, j) = p(i, j)q_0(i) + \sum_{k=1}^{N_0} q_k(i) \sum_{n \in \bar{S}} p(i, n)c^{(k)}(n, j). \tag{15}$$

Equations (13) and (15) establish (14). The result in (14) can be also shown by identifying that the event of the transition of the buffer occupancy process from state 0 (empty buffer) to state 0 for the first time in the future, corresponds to an R reduction. The easiest way to see the previous is to consider a realization of an R reduction starting with arbitrary nonzero buffer content and the corresponding realization (i.e., using the same realization of the packet arrival process) of the buffer occupancy process between two consecutive instants in which the buffer is empty. It will be easily seen that both the R reduction and the empty to first-time empty processes have the same time duration; that is, the same number of transitions of the input Markov chain. Thus, $p_w(i, j) = c(i, j) = p^n(i, j)$ if n such transitions are required. Notice that the evolution of the R reduction process does not depend on the initial value of the buffer content. □

The proof of Lemma 1 completes the proof of theorem. □

As a final comment, it should be noted that the complexity of the analysis of the queueing system considered in this section increases as the number of input streams and/or the cardinality of the underlying Markov chains increase. This is due to the increased dimensionality of the system of linear equations (Appendix B) and the convergence time in (11c). The convergence issue of equations of the type of that in (11c) is briefly discussed in [15] and the references cited there. These are not critical issues for the system considered in this paper, due to the very small cardinality (at most 3) of the state space of the underlying Markov chains associated with the processes in B.

IV. Performance Evaluation of Interconnecting Structures

In this section, the packet traffic modeling and the queueing results presented before are incorporated in the performance evaluation of networking structures. The focus at this point is twofold. The first objective is to illustrate the relative performance of the various traffic models, identify the regions and the reasons of the observed inaccuracies and shape accordingly the expectations for the relative performance under other queueing conditions. The second objective is to propose a procedure for the performance evaluation of large networking structures, based on proper modeling of the packet traffic as modulated by merging and splitting operations. As an example, a fully connected topology of interconnected nodes (points of transformation) is considered. This topology can present most of the situations which may be encountered in an arbitrary topology.

Consider four nodes interconnected according to a fully connected topology. Let a, b, c, and d denote these nodes. let l_{kj}, $k, j \in \{a, b, c, d\}$ denote the link that connects node k with node j. Let U_{kj}, $k, j \in \{a, b, c, d\}$ denote the uplink buffer which feeds link l_{kj}. This buffer is fed by the packet traffic generated at node k and forwarded to node j. Let D_{kj}, $k \in \{a, b, c, d\}, j \in [1, 2, \cdots, N_k\}$ represent the downlink buffer of node k which receives the internode traffic that is destined to the jth local destination within node k; N_k denotes the maximum number of local destinations supported by node k. The uplink buffer U_{ab} and the downlink buffer D_{b1} are shown in Fig. 5. S^j_a, $j = 1, 2, 3$ denotes the jth traffic source supported by node a. The splitting arrows at the inputs of uplink buffer U_{ab} represent exogenous traffic generated at the corresponding source which is not forwarded to link l_{ab}. The splitting arrows at the inputs of downlink buffer D_{b1} represent traffic arriving through the corresponding link, which is forwarded to one of the other local destinations supported by node b.

For the reasons outlined earlier in this paper, the exogenous packet traffic is assumed to be bursty and is described by a first-order Markov model $B(\lambda, \gamma)$. Thus, the input traffic to uplink U_{ab} is a bursty traffic modulated by the splitting operation. The situation is different at the input of the downlink buffer D_{b1}. The traffic delivered by lines l_{ab}, l_{cb}, and l_{db} (before the splitting) is the output from the uplink buffers at the other end of the links.

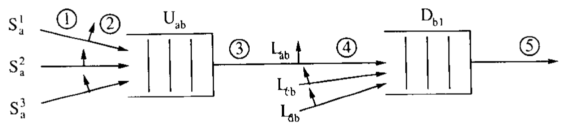

Fig. 5. Detailed description of the uplink buffer U_{ab} and the downlink buffer D_{b1}.

To evaluate the performance of the interconnection, the packet delay induced by both the uplink and downlink buffers needs to be calculated. The models and the results developed in the previous sections are used for this purpose. Let $B_i(\,)$, $i = 1, 2, 3, 4, 5$ denote the packet traffic at point i, as marked on Fig. 5. Without loss of generality, it is assumed that the system is symmetric; each of the exogenous traffic is modeled as $B(\lambda, \gamma)$.

Study of the Uplink Buffer U_{ab}: First, consider the case of independent traffic splitting at the input lines of the uplink buffer. The mean packet delay D_{ab} induced by the uplink buffer U_{ab}, is shown in Fig. 6 as a function of the splitting probability p. Results are presented under traffic modeling $\tilde{B}_2(\lambda')$, $\tilde{B}_2(\lambda', \gamma')$, and $B_2((\lambda', \gamma', p)$. The total packet input rate to the buffer, $\lambda_T = 3\lambda p$, is equal to 0.9 and the burstiness coefficient γ of each input traffic is equal to 0.9. The results indicate that the process B_2, generated by applying independent splitting on a bursty traffic $Bi(\lambda, \gamma)$, cannot be adequately described by a Bernoulli or bursty model. That is, the independent splitting process is not capable of significantly reducing the memory of the original process. On the other hand, it is capable of significantly hiding it from the mechanism (see Proposition 1) that constructs the equivalent bursty model. When no splitting is in effect ($p = 1$), the two models $\tilde{B}_2(\lambda', \gamma')$ and $B_2(\lambda', \gamma', p)$ coincide. When $p = 0.3$, in order for λ_T to be equal to 0.9 the packet rate of each input line has to be equal to $\lambda_T/(3p) = 1$; that is, $B_1(\lambda, \gamma)$ corresponds to a constant traffic stream process. By applying independent splitting on constant traffic, a Bernoulli traffic is generated. In this case, models $\tilde{B}_2(\lambda')$, $\tilde{B}_2(\lambda', \gamma')$, and $B_2(\lambda', \gamma', p)$ coincide, as illustrated in Fig. 6 for $p = 0.3$. In general, the smaller the value of p, the smaller the correlation between successive arrivals at the buffer. As p decreases, inactive intervals within a burst of the original traffic look probabilistically like those within an idle period of the original traffic. Thus, the gap between the delay results is decreased as p decreases. Similar delay results are shown in Fig. 7 as a function of the burstiness coefficient γ, for $\lambda_T = 0.9$ and splitting probability $p = 0.75$. As expected, the more bursty the original traffic is, the larger the gap between the three models. When $\gamma = 0$, the original process is a Bernoulli process and, thus, all three models coincide. It should be noted that model $B_2(\lambda', \gamma', p)$ is exact.

Fig. 8 presents some interesting results for the case of correlated splitting. The burstiness of the original bursty traffic $B_1(\lambda, \gamma)$ is equal to $\gamma = 0.7$ and the total packet arrival rate is equal to $\lambda_T = 0.75$. The results are derived under the exact and approximate (bursty) models. They

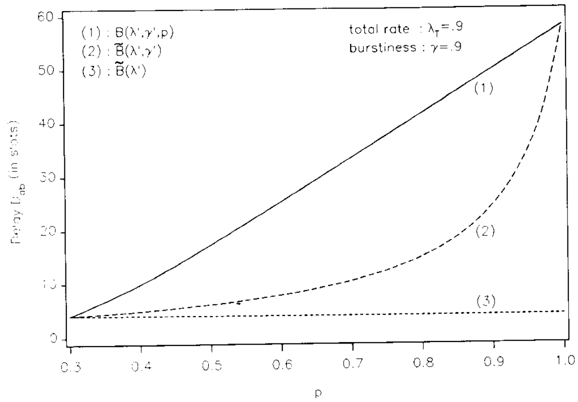

Fig. 6. Mean packet delay induced at the uplink buffer.

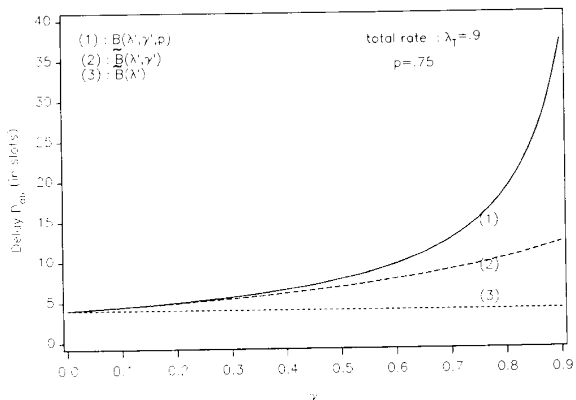

Fig. 7. Mean packet delay induced at the uplink buffer.

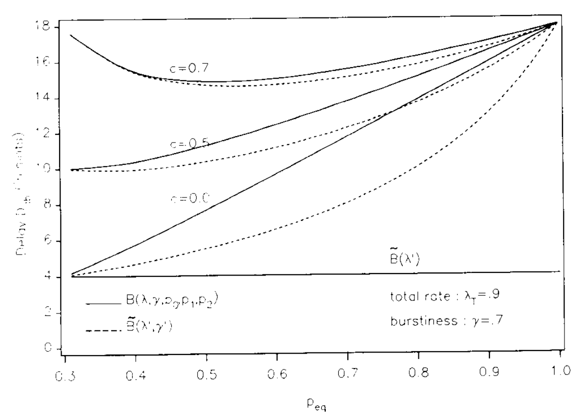

Fig. 8. Mean packet delay induced at the uplink buffer.

are plotted as a function of p_{eq}, which is the portion of the original traffic $B_1(\lambda, \gamma)$ which is forwarded along the tagged direction. Clearly, p_{eq} is equal to the splitting probability p when independent splitting is applied. Results are shown for $c = 0$, $c = 0.5$, and $c = 0.7$, where p_0 is set to be equal to p_{eq}. For given values of λ_T, p_{eq}, γ,

c, and p_0, the values of p_1 and p_2 are determined. When $p_0 = p_{eq}$, the case of $c = 0$ corresponds to independent splitting. Notice that the delay results under correlated splitting vary significantly with respect to those under the equivalent independent splitting. The larger the value of c, the greater the difference is. This observation implies that quite inaccurate results may be obtained under the independent splitting assumption, when the actual splitting operation is correlated.

The same plots as in Fig. 8 are shown in Fig. 9, with the addition of the case of $c = 1$. For $c = 1$ and $p_0 = p_{\text{eq}} = 1$, no splitting in the effect. In this case, $B_2(\lambda, \gamma, p_0, p_1, p_2) = B_1(\lambda, \gamma)$. Under a fixed $\gamma = 0.7$ for the original bursty process $B_1(\lambda, \gamma)$, and a fixed resulting packet rate of $\lambda_T = \lambda p_{\text{eq}}$ and $p_0 = p_{\text{eq}}$, λ increases as p_{eq} decreases. Given that $\gamma = p(1, 1) - p(0, 1)$ is fixed, increasing λ corresponds (under the above conditions) to increasing both $p(1, 1)$ and $p(0, 1)$. The increased value of $p(1, 1)$ implies that longer bursts are generated by the resulting original process $B_1(\lambda, \gamma)$. Since the whole burst is forwarded along the tagged direction when $c = 1$, the previous implies that increased queueing problems will appear as $p_{\text{eq}} = p_0$ decreases. This trend is clearly shown in Fig. 9. Notice also that the value of γ' in the approximate process $\tilde{B}_2(\lambda', \gamma')$ also increases as p_{eq} decreases (or $p(1, 1)$ increases) and the delay results obtained through this model increase monotonically as p_{eq} decreases [see (4)].

Study of the Downlink Buffer D_{b1}: Evaluation of the packet delay induced by the downlink buffer D_{b1} requires knowledge of the corresponding packet input processes. These processes are generated by applying merging and splitting operations on bursty traffic processes. Notice that models $B_2(\lambda', \gamma', p)$ and $B_2(\lambda, \gamma, p_0, p_1, p_2)$ exactly described the input processes to the uplink buffer U_{ab}. This is not the case with the input processes to the downlink buffer D_{b1}. These processes are described in terms of the process $\tilde{B}_3(\lambda', \gamma')$ which is only an approximation of the true packet traffic at the output of the buffer. It is this approximation whose accuracy is evaluated with simulation results of the delay induced by the downlink buffer D_{b1}.

The mean packet delay induced by buffer D_{b1} is shown in Figs. 10–13. This buffer is fed by three symmetric input lines, each of which carries traffic generated by the corresponding uplink buffer. p_1 and p_2 are the parameters of the independent splitting operation applied at the input and output of each of the involved uplink buffers. γ denotes the burstiness coefficient of the packet traffic generated by the sources feeding the uplink buffers. The packet rate λ of each source is selected to be such that the total input traffic λ_T to the downlink buffer takes a certain value from those considered in the plots (horizontal axis). Curve (1) corresponds to the simulation results; curve (2) corresponds to the results obtained if $B_4 = B(\lambda_3, \gamma_3, p_2)$, $B_3 = \tilde{B}(\lambda_3, \gamma_3)$, $B_2 = B(\lambda', \gamma', p_1)$ and $B_1 = B(\lambda, \gamma)$, where process B_3 is approximate and, thus, process B_4 is also approximate; curve (3) corresponds to the results when a bursty model is adopted after every splitting and merging; curve (4) corresponds to the results obtained under Bernoulli model for B_4, with parameter $\lambda_4 = \lambda_T/3$. The delay results have been derived as a function of λ_T and for $\gamma = 0.5$; four different pairs of values of (p_1, p_2) have been considered. Notice that the results obtained under the traffic modeling which incorporates the exact model for the splitting operation and the approximate

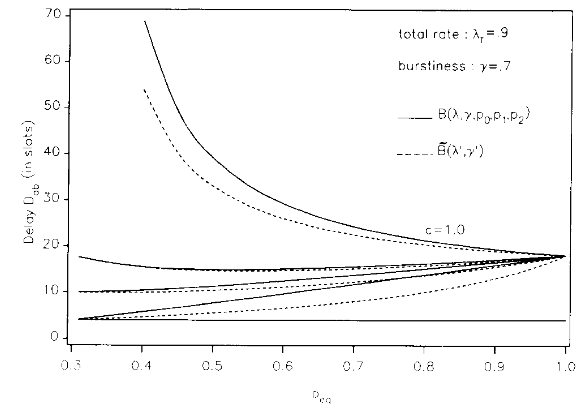

Fig. 9. Mean packet delay induced at the uplink buffer.

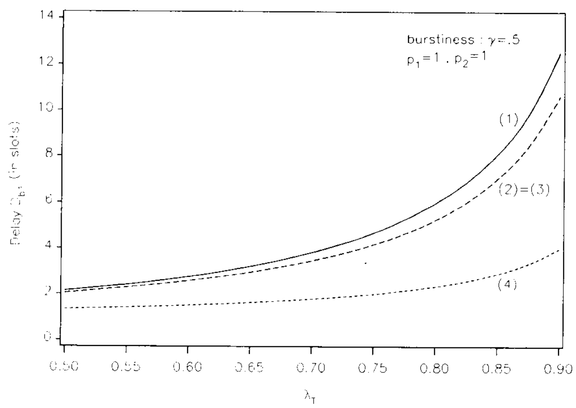

Fig. 10. Mean packet delay induced at the downlink buffer.

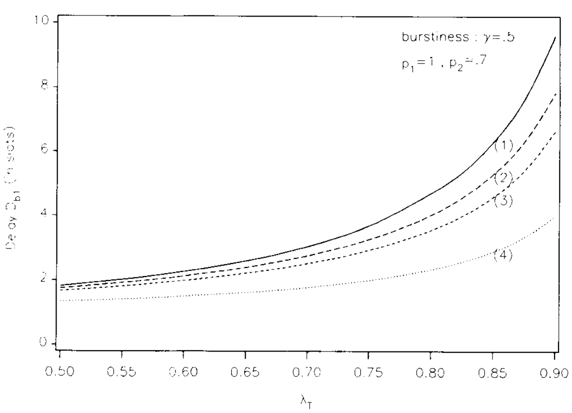

Fig. 11. Mean packet delay induced at the downlink buffer.

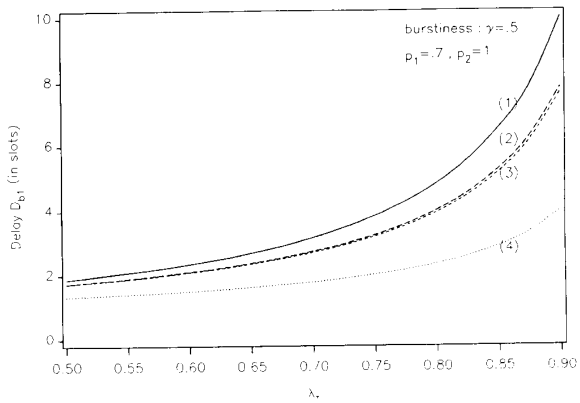

Fig. 12. Mean packet delay induced at the downlink buffer.

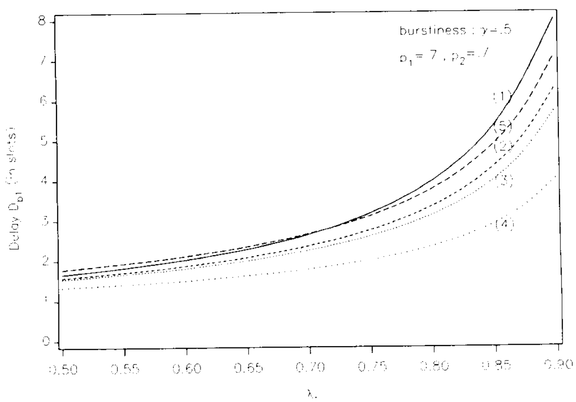

Fig. 13. Mean packet delay at the downlink buffer.

model for the merging operation [curve (2)] are close to the simulations and clearly outperfrom those under Bernoulli [curve (4)], and bursty [curve (3)] modeling. Notice that when $p_1 = p_2 = 1$ (Fig. 10), curves (2) and (3) coincide as expected.

When splitting occurs at the input of the uplink buffer, futher reduction in the memory captured by $\tilde{B}_3(\lambda_3, \gamma_3)$ is expected. As a result, the computed burstiness γ_3 is expected to underestimate the intensity of the queueing problems caused by the true process B_3. Indeed, it has been observed, that under symmetric input traffic and $p_1 < 1$, the value γ_3^*, computed as described in Proposition 3, is significantly smaller than γ_1. When $p_1 = 1$, it has been observed that the corresponding computed value, γ_3^{**}, is equal to or slightly larger than γ_1. To compensate for the superficially large reduction of the value of γ_3 when $p_1 < 1$, the following approximate value of γ_3 has been considered.

$$\tilde{\gamma}_3 = p_1\gamma_1 + (1 - p_1)\gamma_3^*.$$

Curve (5) in Fig. 13 corresponds to curve (2) when $B_3 = \tilde{B}_3(\lambda_3, \tilde{\gamma}_3)$. Notice that curve (5) overestimates D_{b1} when λ_T is below ≈ 0.7; it underestimates if for λ_T greater

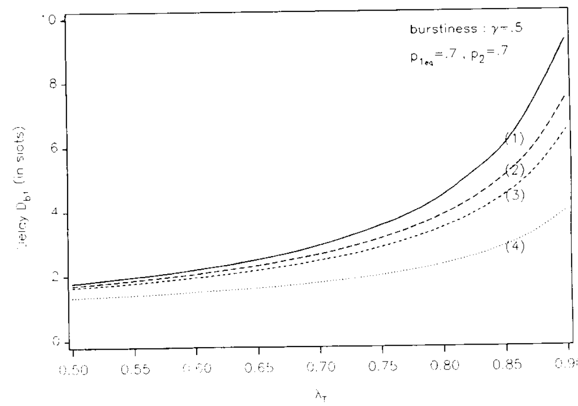

Fig. 14. Mean packet delay induced at the downlink buffer.

than ≈ 0.7. This behavior suggests that $\tilde{\gamma}_3$ be considered when the traffic load is significant.

Finally, when independent and correlated splittings are in effect at the inputs of the uplink and the downlink buffers, respectively, the delay results are shown in Fig. 14. Notice that, for the same value of p_{eq}, the corresponding results under independent splitting in Fig. 13 [curve (2)] are of significantly lower value, as expected.

V. Discussion and Conclusions

The objective in this paper has been the study of the effects of splitting and merging operations on bursty traffic $B(\lambda, \gamma)$. These operations are very common in networking structures, due to routing decisions and common sharing of the resources. Exact representation of the traffic generated by splitting traffic has been developed and a relevant queueing system has been studied. Approximate representation of the traffic at the output of a traffic merging point has also been proposed.

The effects of the independent splitting operation on bursty traffic $B(\lambda, \gamma)$ has been studied and a simple exact model for the resulting traffic has been developed. At the same time, the inadequacy of the Bernoulli or bursty traffic models for the description of the resulting traffic has been illustrated. The interesting case of correlated splitting, applied on bursty traffic, has also been considered. An exact model for the description of the resulting traffic has been developed. It has been shown that ignoring correlations in the splitting process may result in quite inaccurate performance calculations.

A simple exact model for the description of the packet process generated by merging packet traffic is difficult, due to the complexity of the queueing process at the merging point. A very accurate representation of this process is possible by adopting the approach in [5]. Unfortunately, such an approach would not be applicable beyond the first merging due to the increased complexity. In this paper, effort has been focused on the development of simple models which, on one hand, capture to some extent the dependencies in the true process and, on the other, present a complexity which does not change as the number of stages increases. The simplest model which possesses these characteristics is the Bernoulli one; which captures only the rate of the resulting process. The proposed bursty model, with parameters calculated as described in Proposition 3, captures not only the resulting packet rate but also dependencies between consecutive slots (taking into consideration the effect of possible splitting just before the merging point) as well. As a consequence, this model has been shown to clearly outperform the Bernoulli (i.i.d) one. At the same time, the complexity of its parameter calculation does not change as the number of stages increases.

Arbitrarily complex networking topologies can be studied by incorporating the splitting and merging models developed in this work. It is proposed in [13] that the internode traffic, modulated by merging and splitting operations, be approximated by a bursty model $B(\lambda, \gamma)$ with a rate-dependent value of γ. In particular, it is proposed that the value of γ be estimated by $\gamma = 1-(1 - e^{-\lambda})/\lambda$. Clearly, this bursty model is based on the resulting packet rate and fails to distinguish between a splitting or merging operation before the point of consideration. It also fails to capture the degree of burstiness that the traffic exhibits before the particular operation takes place. For $\lambda < 0.9$, the value of γ calculated in [13] cannot exceed 0.34 and, thus, it is not capable of representing more bursty traffic. In most of the examples considered here, the value of γ which is based on λ resulted in very inaccurate results. These results were better than those under Bernoulli modeling. Nevertheless, arbitration of the value of γ (based on λ) has been found to perform relatively well (and produce results close to those under the proposed models), when most of the traffic arriving at a buffer has been modulated by a large number of merging and independent splitting operations. When the packet traffic, generated by the sources supported by a node, is significant compared to the transit traffic through that node, then the arbitration of γ fails to produce accurate delay results. The same has been observed when the traffic of only a few sources dominates the arrival process to a particular node and/or highly correlated splitting occurs in the network.

When alternate routing in a fully connected network is possible or when the interconnecting topology is arbitrary, transit traffic appears in the uplink buffers. Since this traffic is modulated by a large number of splitting and merging and probably undergoes significant splitting at the particular node (since portion of this traffic is forwarded to other uplink buffers or to the destinations supported by the node), the memory present in this portion of the traffic arriving at the tagged uplink is not significant and it may be satisfactorily captured by a burstiness coefficient based on the (usually small) rate λ.

There are a number of contributions that could be attributed to this work. Although the idea of modeling packet traffic as a (first-order Markovian) bursty process is not a new one, its transformation under splitting and merging operations has not been studied in the past. First, a bursty model for the packet traffic after a transformation point (merging or splitting) has been developed; its parameter take into consideration the corresponding values before this point and they generate the true probabilities of single events and pairs of consecutive events (Proposition 1–3). It has been argued and shown that the bursty model is inappropriate for the characterization of the traffic right after a splitting point. Exact models have been developed for the resulting traffic in this case, under both independent and correlated splitting policies. The case of correlated splitting is more general and it may represent many realistic splitting situations, such as burst switching. A relevant queueing system has been considered under a general class of input processes, which contains the models developed in this paper. An approach for the calculation of important boundary probabilities has been developed (Theorem 1). This approach expands the applicability of the queueing system significantly. Queueing analysis has established that correlated splitting may cause significantly more intense queueing problems than the independent one. Thus, correlation in the splitting process should be properly taken into consideration. A common property to all models considered here is that their complexity, in both their description and their parameter calculation procedure, does not increase as the number of the points of transformation increases. Finally, it has been illustrated how the simple models can be applied for the evaluation of the performance of a networking structure. The objective at this point has been to illustrate the performance of the models rather than to analyze the performance of an important networking structure.

Appendix A

Proof of Proposition 1: The first part of (2) is obvious. Let $\pi_e(i)$, $p_e(i, j)$, $i, j \in \{0, 1\}$ denote the steady state and the state transition probabilities of the equivalent bursty traffic model. Let $p_e(i \cap j)$ denote the joint probability that the first and the second of a pair of consecutive slots along the tagged direction are in states i and j, respectively. It is easy to show that:

$$p_e(1, 1) = \frac{p_e(1 \cap 1)}{\pi_e(1)} = \frac{p(1 \cap 1)p^2}{\pi_e(1)} = \frac{p(1, 1)\pi(1)p^2}{p\pi(1)} = pp(1, 1)$$

where the notation without subscript e refers to the corresponding parameters of the bursty traffic model $B(\lambda, \gamma)$ before the splitting. The imposed Markovian structure on $\tilde{B}_2(\lambda', \gamma')$ implies that:

$$p_e(0, 1)\pi_e(0) + p_e(1, 1)\pi_e(1) = \pi_e(1) - \gamma' = \frac{p_e(1, 1) - \pi_e(1)}{\pi_e(0)}.$$

The above equations prove the second part of (2). □

Proof of Proposition 3: The first part of (5) is obvious since λ' is the total input rate. By using the definition of γ given in (1) and by imposing the Markovian structure on $\tilde{B}_2(\lambda', \gamma')$, the second part of (4) is proved. Two consecutive empty slots are observed at the output of the buffer whenever the buffer is empty (the first empty slot is generated) and the state of the input Markov chain results in no new arrival (the second empty slot is generated). This event can be expressed as the union (over all possible states $\bar{x}$) of the following events: {(buffer is empty) $\cap$ (state of input Markov chain is $\bar{x}$) $\cap$ (not packet generated from $\bar{x}$}. The latter implies (7).

The state space of the Markov chains associated with traffic models $B(\lambda)$ and $B(\lambda, \gamma)$ contains only one state (state 0) which may not generate a packet. As a result, whenever the buffer is empty the state of the input Markov chain in the previous slot is uniquely determined. Thus, the event that the buffer is empty and the input Markov chain is in the state $\bar{x}$ is equivalent to the event that the buffer is empty and a transition from the no packet generating state to state $\bar{x}$ takes place. Since state transitions of the input Markov chain do not depend on the buffer content, (8) is derived. □

Appendix B

In this Appendix, the equations for the calculation of the first moment of the buffer occupancy process in the queueing system described in Section III are presented. These results are taken from [12], where the moments of the buffer occupancy process are derived. By adopting the notation used in the main part of this paper, the mean buffer occupancy is obtained from the equation

$$Q = \overline{we}$$

where $\bar{e}$ is the $\overline{M}$-dimensional unit row vector and $\overline{w} = [w_1, \cdots, w_{\overline{M}}]$ with $\{w_i\}_{i=1}^{\overline{M}}$ being the solution of the following linear equations:

$$w_j = \sum_{i=1}^{\overline{M}} p(i, j)\,[w_i + \pi(i)(\mu_i - 1) + \tau(0; i), \quad 1 \le j \le \overline{M}$$

$$\sum_{i=1}^{\overline{M}} [2(\mu_i - 1)w_i + 2(\mu_i - 1)\tau(0; i) + (2 + \sigma_i - 3\mu_i)\pi(i)] = 0$$

where μ_i and σ_i denote the first and the second moments of the number of packets generated from input state i, $1 \le i \le \overline{M}$. The mean packet delay is computed from Little's theorem and it is given by $D = Q/\lambda$, where λ is the total packet input rate.

When all the input processes in the queue are described in terms of the $B(\lambda)$ or $B(\lambda, \gamma)$ modelings, then the induced mean packet delay may be calculated from the following closed form formula [13]

$$D = \left[1 + \frac{\sum_{n=1}^{N} \sum_{m>n}^{N} \lambda^n \lambda^m \left(1 + \frac{\gamma^n}{1-\gamma^n} + \frac{\gamma^m}{1-\gamma^m}\right)}{\left(1 - \sum_{n=1}^{N} \lambda^n\right) \sum_{n=1}^{N} \lambda^n} \right]$$

where N is the total number of input lines and λ_n, γ_n denote the rate and the burstiness of the nth input stream.

Acknowledgment

The contribution of S. Tsakiridou in the derivation of the numerical results is gratefully acknowledged.

References

[1] D. Bertsekas, and R. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prenctice Hall, 1987.
[2] F. A. Tobagi, "Analysis of a two hop centralized packet radio network-Part I: Slotted ALOHA," *IEEE Trans. Commun.*, vol. COM-28, no. 2, Feb. 1980.
[3] L. Merakos, G. Exley, and C. Bisdikian, "Interconnection of CSMA local area networks: The frequency division approach," *IEEE Trans. Commun.*, vol. 35, no. 7, July 1987.
[4] H. Takagi and L. Kleinrock, "Output processes in contention packet broadcasting systems," *IEEE Trans. Commun.*, vol. COM-33, no. 11, Nov. 1985.
[5] I. Stavrakakis, "An analysis approach to multi-level networking," presented at the Int. Conf. Commun., Atlanta, GA, 1990.
[6] I. Stavrakakis and D. Kazakos, "On the approximation of the output process of multi-user random access communication networks," *IEEE Trans. Commun.*, vol. 38, no. 2, Feb. 1990.
[7] I. Stavrakakis, "Analysis of a class of star interconnected networks," to appear in *IEEE Trans. Commun.*
[8] F. A. Tobagi, "Analysis of a two-hop packet radio network-Part II: Carrier sense multiple access," *IEEE Trans. Commun.*, vol. COM-28, no. 2, Feb. 1980.
[9] C. Ko, W. Wong, and K. Lye, "Performance of CSMA/CD networks connected by bridges," presented at the Int. Conf. Commun., Philadelphia, PA, June 1988.
[10] I. Stavrakakis and D. Kazakos, "Performance analysis of a star topology of interconnected networks under 2nd-order network output processes," *IEEE Trans. Commun.*, vol. 38, no. 10, Oct. 1990; also presented at the 22nd Ann. Conf. Inform. Sci. Syst., Princeton, NJ, Mar. 16-18, 1988.
[11] S. Q. Li, 'Performance of a non-blocking space-division packet switch with correlated input traffic," presented at the GLOBECOM'89 Conf., 1989.
[12] I. Stavrakakis, "Analysis of a statistical multiplexer under a general input traffic model," presented at the INFOCOM'90 Conf., San Francisco, CA, June 5-7, 1990.
[13] A. M. Viterbi, "Approximate analysis of time synchronous packet networks," *IEEE J. Select. Areas Commun.*, vol. SAC-4, no. 6, Sept. 1986.
[14] Y-C. Jenq, "Performance analysis of a packet switch based on single-buffered banyan network," *IEEE J. Select. Areas Commun.*, vol. SAC-1, no. 6, Dec. 1983.
[15] O. Yang and J. Mark, "Queueing analysis of an integrated services TDM system using matrix-analytic method," *IEEE J. Select. Areas Commun.*, vol. 9, no. 1, Jan. 1991.
[16] W. Whitt, "The queueing network analyzer," *Bell Syst. Tech. J.*, vol. 62, no. 9, Nov. 1983.
[17] ——, "Performance of the queueing network analyzer," *Bell Syst. Tech. J.*, vol. 62, no. 9, Nov. 1983.
[18] D. Mitra and R. Cieslak, "Randomized parallel communications on an extension of the omega network," *J. ACM*, vol. 34, no. 4, Oct. 1987.

Ioannis Stavrakakis (S'85–M'89) was born in Athens, Greece, in 1960. He received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1983, and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, in 1988.

In 1988, he joined the Department of Electrical Engineering and Computer Science, University of Vermont, Burlington, where he is presently an Assistant Professor. His research interests are in statistical communication theory, multiple access algorithms, computer communication networks, queueing systems, and system modeling and performance evaluation.

Dr. Stavrakakis is a member of Tau Beta Pi and the Technical Chamber of Greece.