INTERNATIONAL JOURNAL OF DIGITAL AND ANALOG COMMUNICATION SYSTEMS, VOL. 4, 249–260 (1991)

QUEUEING BEHAVIOUR OF TWO INTERCONNECTED BUFFERS OF A PACKET NETWORK WITH APPLICATION TO THE EVALUATION OF PACKET ROUTEING POLICIES

IOANNIS STAVRAKAKIS

Department of Electrical Engineering and Computer Science, University of Vermont, Burlington, Vermont 05405-0156, U.S.A.

SUMMARY

In this paper, the behaviour of two interconnected nodes of a packet communication network is analysed. The packet arrival process to the first node is described by an i.i.d. discrete time process. The packet arrival process to the second node consists of two mutually independent streams; one stream carries the packet output process from node 1; the other stream carries a packet traffic described by a dependent discrete time process. For this system, the following quantities are derived: (a) the first two moments of the buffer occupancy and the mean packet delay in the first node; (b) a description of the packet output process of the first node; (c) the first two moments of the buffer occupancy and the mean packet delay in the second node.

In the sequel, the behaviour of the buffers is investigated under some simple routeing policies in node 1. The objective at this point is to evaluate the queueing behaviour of both buffers under some routeing policies and it is not to seek for the optimal packet routeing policy in the first node. It is shown how the routeing policy in node 1 can be incorporated in the description of the packet output process of node 1 and, then, the behaviour of buffer 2 be accurately evaluated.

The numerical results show that the queueing behaviour of buffer 2 cannot be accurately evaluated by incorporating a Bernoulli approximation on the output process of node 1. Furthermore, it is shown that different routeing policies in the first node result in different intensities of the queueing problems in the second node, although the packet output rate from the first node is constant for all routeing policies considered. It turns out that a routeing policy which best randomizes the packet output process results in the least significant queueing problems in the second node. The latter implies that routeing decisions should be based not only on the desired packet traffic rates (which achieve optimality in the load distribution among the nodes), but also on the packet randomization affect of the routeing policy considered.

KEY WORDS Queueing systems Buffer behaviour Statistical multiplexing Routeing policies

1. INTRODUCTION

Packet communication networks have been widely adopted as an efficient means of transferring information. The information sources are assumed to generate a common quantity, the (fixed length) packet and they are (statistically) described by their packet generation process. Any component of a large packet communication network which outputs packets may be seen as a packet generating source. The description of the packet generation processes of the various components of a large packet communication network is essential to the accurate evaluation of the network performance.[1]

The discussion in this paper is confined to synchronous packetized communication networks. Discrete-time queueing models are incorporated in the analysis of such networks, where packet processes are described by discrete time point processes.[2,3] The performance of a large network of nodes (see, for instance, Figure 1) depends on a number of factors such as the packet traffics entering the system, and the routeing and service policies at the various nodes. For given packet input processes to a large network, different routeing procedures (which affect the intranetwork packet traffics) and service policies (for example priorities), result in different network performances.

In this paper, the queueing behaviour of the interconnected buffers shown in Figure 2 is investigated; R^{i}_{kj} denotes the packet traffic which enters node k and is to be forwarded to node j, through possibly more than one path, R^{o}_{kj} denotes the packet traffic departing from node k and forwarded directly to node j. This system of interconnected buffers may be found in the simple topology shown in Figure 3. At first, the behaviours of both buffers of the system shown in Figure 4 are studied, under a general independent discrete time packet arrival process R^{i}_{12} and a dependent discrete time packet arrival process R^{i}_{23}. Notice that no routeing is incorporated in node 1 and thus $\lambda^{i}_{12} = \lambda^{o}_{12}$, where λ^{i}_{kj} and λ^{o}_{kj} denote packet rates associated with the packet processes R^{i}_{kj} and R^{o}_{kj}, respectively. Three steps are followed in the analysis of this queueing system. At first, the queueing behaviour of buffer 1 is analysed. Then, the (dependent) packet output process R^{o}_{12} is described. Finally, buffer 2 is analysed by incorporating the (dependent) packet processes R^{o}_{12} and R^{i}_{23}.

1047–9627/91/040249–12$06.00

Received April 1991
Revised July 1991

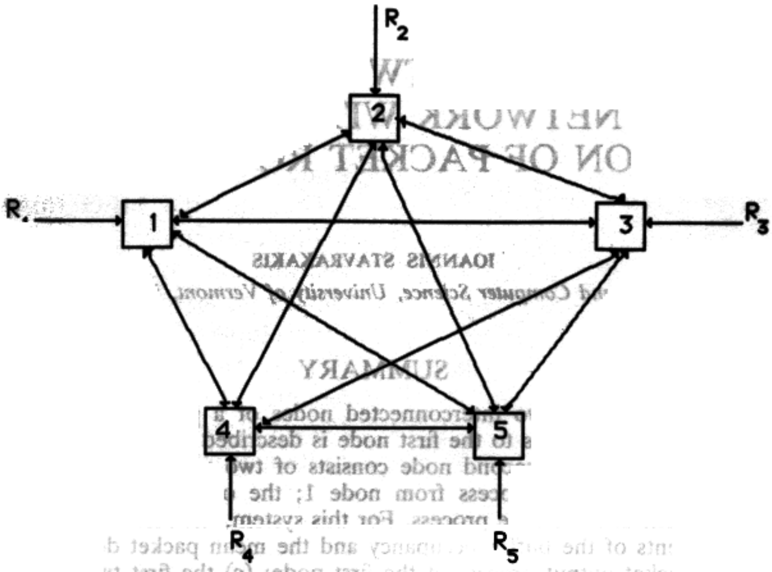


Figure 1. A topology of a packet communication network

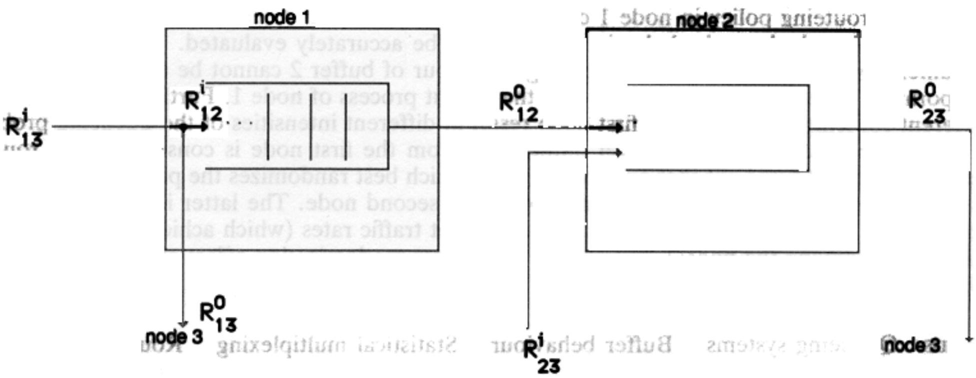


Figure 2. A system of two interconnected buffers with packet routeing

In the sequel, the behaviours of both buffers of the system in Figure 3 are studied under some simple routeing policies applied to node 1, by following the three steps described before. By maintaining a fixed packet output rate, λ^o_{12} (assuming that λ^i_{23} is constant), it is observed that the routeing policies in node 1 result in different queueing behaviours of buffer 2. This is due to the fact that, despite the equality of the intensity of the resulting packet rate, different routeing policies in node 1 generate statistically different packet output processes R^o_{12}. Although the objective of this paper is not the development of efficient routeing strategies, a unified description of the packet output process, which incorporates the routeing strategies, is developed and some interesting conclusions on the effect of the routeing policies on the queueing behaviour of the nodes are drawn. These conclusions suggest that a new performance measure could be incorporated in the evaluation of a packet routeing policy, which takes into consideration the queueing problems in the subsequent nodes of the network caused by the packet process generated by the routeing policy. This approach to routeing is quite different from the traditional ones which are either based only on desired packet rates to be generated by the routeing policy,[4,5] or try to minimize the queueing problems in the node where the routeing decision is taken.[6,7]

2. ANALYSIS OF TWO BUFFERS WITHOUT ROUTEING DECISIONS

In this section, the system shown in Figure 4 is analysed. The packet input process to the first node, R^i_{12}, is assumed to be a *generalized Bernoulli process* (GBP). That is, the number of packets arriving at node 1 at the potential arrival instants is an independent process and it follows a general distribution. This process is determined by a message (or multi-packet) arrival rate, r, and a general distribution, $g(j)$, $1 \leq j \leq N_B$, of the message (or multi-packet) size in packets; N_B is the maximum message length or number of packets which may arrive during a single slot. The packet input process to node 2 is assumed to be a compound process consisted of two independent packet streams. Input R^o_{12} represents the packet output process from the

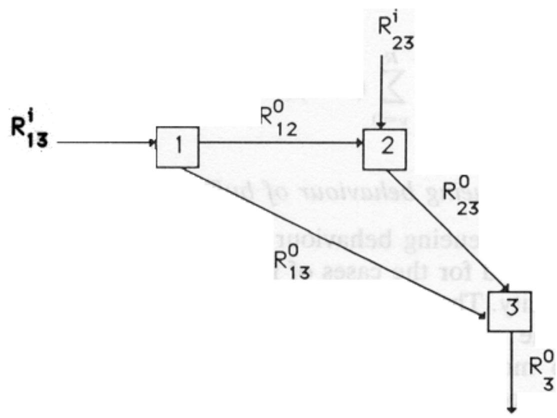


Figure 3. A three-node element of a topology of a packet communication network

to be the time distance between two consecutive potential packet arrival instants. The analysis of the system in Figure 4 requires the study of the queueing behaviour of buffer 1, the description of the packet output process R^o_{12} generated by buffer 1 and the study of the queueing behaviour of buffer 2.

Consider the FIFO (first-in–first-out) statistical multiplexer shown in Figure 5. The service rate is constant and equal to 1 time unit (slot). The buffer capacity is assumed to be infinite. Packets arrive at the multiplexer through N independent input streams (lines). The per stream packet arrival process is assumed to be the MMGBP described before. Packet arrivals are declared at the ends of the slots (potential packet arrival points).

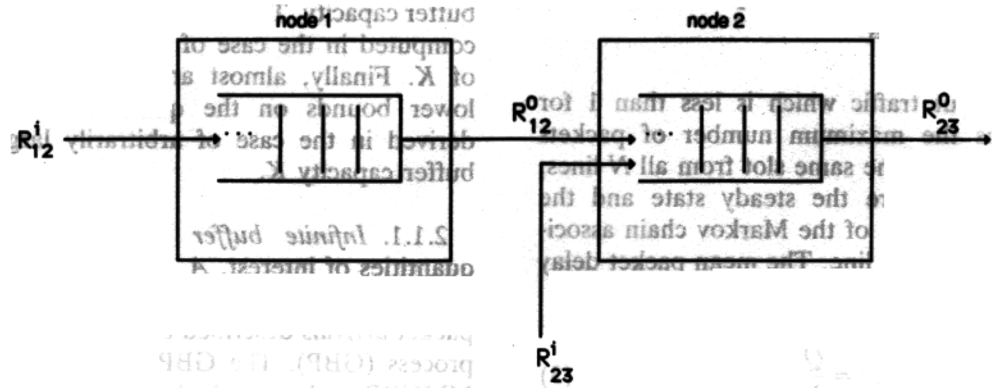


Figure 4. A system of two interconnected buffers without packet routeing

first node. Input R^i_{23} is assumed to be described by a *Markov modulated generalized Bernoulli process* (MMGBP) which is, in general, a dependent process. That is, R^i_{23} is described by a discrete time process $\{a_j\}_{j\geq 0}$ which depends on an underlying Markov chain $\{z_j\}_{j\geq 0}$. The detailed description of $\{a_j\}_{j\geq 0}$ is given below.

Let $\{z_j\}_{j\geq 0}$, be a finite state Markov chain imbedded at the potential packet arrival points; let $S = \{x_0, x_1, \ldots, x_{M-1}\}$, $M < \infty$, be the state space of $\{z_j\}_{j\geq 0}$. It is assumed that the state of the underlying Markov chain determines (probabilistically) the packet arrival process of the corresponding line. That is, if $a(z_j)$: $S\rightarrow Z_0$, is a probabilistic mapping from S into the non-negative finite integers, Z_0, then the probability that k packets arrive at the buffer at the end of the jth potential packet arrival instant is given by $\phi(z_j,k) = \Pr\{a(z_j) = k\}$. Furthermore, it is assumed that there is at most one state, x_0, such that $\phi(x_0,0) > 0$, and that the rest of the states of the underlying Markov chain result in at least one (but a finite number of) packet arrivals, i.e. $\phi(x_k,0) = 0$, for $1 \leq k \leq M$. To avoid instability of the buffer queue it is assumed that there is always one state x_0, such as described above.

The packet service rates at both nodes are constant and equal to one packet per slot; the slot is defined

The expected number of packets in the system is given by[8]

$$Q = \sum_{\bar{y}\in\bar{S}} W_1(\bar{y}) \tag{1}$$

where $\bar{S} = S^1 \times S^2 \times \ldots \times S^N$ and $W_1(\bar{y})$, $\bar{y}\in\bar{S}$, are the solutions of any $M^1 \times M^2 \times \ldots \times M^N-1$ linear equations (M^i is the cardinality of the state space of the Markov chain associated with the ith input line, S^i) given by

$$W_1(\bar{y}) = \sum_{\bar{x}\in\bar{S}} W_1(\bar{x})p(\bar{x},\bar{y}) + \sum_{\bar{x}\in\bar{S}} (\mu_{\bar{x}}-1)p(\bar{x},\bar{y})\pi(\bar{x}) + \sum_{\bar{x}\in\bar{S}} q_0(\bar{x})p(\bar{x},\bar{y}), \ \bar{y}\in\bar{S} \tag{2a}$$

and the linearly independent equation

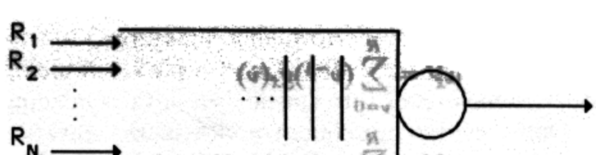


Figure 5. A FIFO statistical multiplexer with N inputs

$$\sum_{\bar{x}\in S}[2(\mu_{\bar{x}}-1)W_1(\bar{x})+2(\mu_{\bar{x}}-1)q_0(\bar{x}) +(2+\sigma_{\bar{x}}-3\mu_{\bar{x}})\pi(\bar{x})] = 0 \quad (2b)$$

where

$$\pi(\bar{x}) = \prod_{i=1}^{N}\pi^i(x^i),\ p(\bar{x},\bar{y}) = \prod_{i=1}^{N}p^i(x^i,y^i),$$
$$q_0(\bar{x}) = (1-\lambda)p(\bar{x}_0,\bar{x})$$
$$\mu_{\bar{x}} = \sum_{\nu=1}^{R}\nu g_{\bar{x}}(\nu),\ \sigma_{\bar{x}} = \sum_{\nu=1}^{R}\nu^2 g_{\bar{x}}(\nu),$$
$$g_{\bar{x}}(\nu) = \Pr\left\{\sum_{i=1}^{N}a^i(x^i) = \nu\right\}$$

and

$$\lambda = \sum_{\bar{x}\in S}\mu_{\bar{x}}\pi(\bar{x}) < 1$$

is the total input traffic which is less than 1 for stability. R is the maximum number of packets which may arrive at the same slot from all N lines; $\pi^i(x^i)$ and $p^i(x^i,y^i)$ are the steady state and the transition probabilities of the Markov chain associated with the ith input line. The mean packet delay is given by using Little's formula, i.e.

$$D = \frac{Q}{\lambda} \quad (3)$$

The variance of the number of packets in the system, V, is obtained from the expression

$$V = Q_2 - (Q_1)^2,\ Q_2 = \sum_{\bar{x}\in S}W_2(\bar{x}) + Q_1 \quad (4)$$

where Q_2 is the second moment of the buffer occupancy and $W_2(\bar{x})$, $\bar{x}\in\bar{S}$, are the solutions of the following system of linear equations:[8]

$$W_2(\bar{x}) = \sum_{\bar{x}\in S}p(\bar{x},\bar{y})W_2(\bar{x}) + \sum_{\bar{x}\in S}[\mu_{\bar{x}}^{2f}\pi(\bar{x}) +2\mu_{\bar{x}}^{1f}[W_1(\bar{x})+q_0(\bar{x})]]p(\bar{x},\bar{y}),\ \bar{y}\in\bar{S} \quad (5a)$$

and the linearly independent equation

$$\sum_{\bar{x}\in S}3[\mu_{\bar{x}}-1]W_2(\bar{x}) + \sum_{\bar{x}\in S}[\mu_{\bar{x}}^{3f}\pi(\bar{x})+3\mu_{\bar{x}}^{2f}[W_1(\bar{x})+q_0(\bar{x})]] = 0 \quad (5b)$$

where

$$\mu_{\bar{x}}^{1f} = \sum_{\nu=0}^{R}(\nu^{-1})g_{\bar{x}}(\nu)$$
$$\mu_{\bar{x}}^{2f} = \sum_{\nu=0}^{R}(\nu-1)(\nu-2)g_{\bar{x}}(\nu) \quad (5c)$$

and

$$\mu_{\bar{x}}^{3f} = \sum_{\nu=0}^{R}(\nu-1)(\nu-2)(\nu-3)g_{\bar{x}}(\nu)$$

2.1. *Queueing behaviour of buffer* 1

The queueing behaviour of the buffer in node 1 is studied for the cases of finite ($K<\infty$) and infinite capacity. The outcome of this study is the derivation of the expressions for the calculation of the first two moments, Q_1 and Q_2, respectively, and the variance, V, of the buffer occupancy and the mean packet delay, D, induced by buffer 1. We will refer to these queueing quantities by A, $A\in\{Q_1,Q_2,V,D\}$. Closed form expressions for the exact calculation of these quantities are derived in the case of infinite buffer capacity. The quantities of interest are exactly computed in the case of small or moderate values of K. Finally, almost arbitrarily tight upper and lower bounds on the quantities of interest are derived in the case of arbitrarily large but finite buffer capacity K.

2.1.1. *Infinite buffer capacity*. The queueing quantities of interest, A, are computed from equations (1)–(5) by considering a single input line with packet arrivals described by a generalized Bernoulli process (GBP). The GBP is a special case of the MMGBP under which the analysis in Reference 8 is valid. Since the GBP is an independent process, the state space, S, of the corresponding underlying Markov chain of the equivalent MMGBP consists of a single state, x'. By setting $\bar{x} = x'$ and since $\pi(x') = p(x',x') = 1$ and $\mu_{x'} = \lambda$, where λ is the packet input rate, the following equations are obtained directly from (1)–(3):

$$Q_1^I = W_1(x')$$
$$q_0(x') = 1-\lambda$$
$$2(\lambda-1)W_1(x')+2(\lambda-1)q_0(x')+2+\sigma_x'-3\lambda = 0$$

By manipulating the above equations, the following closed form expression for the mean buffer occupancy, Q_1^I, is obtained:

$$Q_1^I = \lambda + \frac{\sigma-\lambda}{2(1-\lambda)} \quad (6a)$$

where σ is the second moment of the packet arrival process. Equation (6a) is valid for any number of independent input streams, each of which is modelled by a GBP. In the case of N such streams,

$$\lambda = \sum_{i=1}^{N}\lambda^i,\ \sigma = \sum_{i=1}^{N}(\sigma^i-(\lambda^i)^2)+\lambda^2 \quad (6b)$$

where λ^i and σ^i are the first and second moments of the packet arrival process associated with the ith input line. By applying Little's theorem to (6a), the following closed form expression for the mean packet delay D^{I}, is obtained:

$$D^{\mathrm{I}} = 1 + \frac{\sigma-\lambda}{2\lambda(1-\lambda)} \tag{7}$$

In the case of Bernoulli per stream packet arrivals (i.e. at most one packet arrival per slot per line) (7) becomes

$$D^{\mathrm{B}} = 1 + \frac{\sum_{i=1}^{N}\sum_{j>i}^{N}\lambda_i\lambda_j}{\lambda(1-\lambda)}$$

which is a known result.[2]

The variance of the buffer occupancy is obtained by modifying equations (4) and (5) appropriately, and it is given by

$$V^{\mathrm{I}} = Q_2^{\mathrm{I}} - (Q_1^{\mathrm{I}})^2 \tag{8a}$$

where Q_1^{I} is given by (6a), Q_2^{I} is the second moment of the buffer occupancy given by

$$Q_2^{\mathrm{I}} = W_2(x') + Q_1^{\mathrm{I}} \tag{8b}$$

where

$$W_2(x') = \frac{1}{3(1-\lambda)}[\mu_{x'}^{3f}+3\mu_{x'}^{2f}(Q_1^{\mathrm{I}}+1-\lambda)] \tag{8c}$$

Equations (8) are valid for any number of independent input streams, each of which is modelled as a GBP. In this case, λ is given by (6b) and $\mu_{x'}^{2f}$ and $\mu_{x'}^{3f}$ can be computed from (5c) by setting $\bar{x} = x'$.

Notice that (6a), (7) and (8) imply that, under packet arrival processes described by a GBP, the queueing behaviour of the buffer depends on the moments of the packet input process and not on the detailed process, as described by the message rate, r, and the message length probability distribution $g(k)$, where k is the length of a message in packets. In particular, the first moment of the buffer occupancy (and, consequently, the mean packet delay, as well) depends on the first two moments of the packet arrival process; the second moment of the buffer occupancy depends on the first three moments of the packet arrival process. It can be shown that, in general, the kth moment of the buffer occupancy depends on the first $(k+1)$ moments of the packet arrival process. The previous claim is easy to establish by considering the general approach for the derivation of the equations (2) and (5), presented in Reference 8, and finally reducing the resulting equations to a single one with respect to $W_k(x')$, where $W_k(x')$ is the kth factorial moment of the buffer occupancy, as it is done for the derivation of W_1 and W_2 in this paper.

2.1.2. *Finite (moderate size) buffer capacity.* When the capacity K of the buffer in node 1 is finite and of small or moderate size, then the queueing quantities A induced by node 1 can be calculated from the following equations:

$$Q_1^{\mathrm{I}} = \sum_{i=0}^{K}\pi(i)i \tag{9a}$$

$$V^{\mathrm{I}} = Q_2^{\mathrm{I}}-(Q_1^{\mathrm{I}})^2\,,\; Q_2^{\mathrm{I}} = \sum_{i=0}^{K}\pi(i)i^2$$

$$D^{\mathrm{I}} = \frac{Q_1^{\mathrm{I}}}{\lambda}$$

where $\pi(i)$, $0\le i\le K$, are the steady-state probabilities of the Markov chain $\{d_j\}_{j\ge0}$, where d_j denotes the number of packets in the buffer of node 1 at the jth slot. Since

$$d_{j+1} = (d_j-1)^+ + a_j$$

where $(x)^+ = x$ if $x\ge0$ and is zero otherwise, and a_j denotes the packet arrivals during the jth slot, it is clear that $\{d_j\}_{j\ge0}$ is a Markov chain with transition probabilities given by

$$P(k,j) = \begin{cases} 1-r, \; j=0, \; 0\le k\le 1 \\ rg(j), \; 1\le j<K, \; k=0 \\ rg(j-k+1), \; 1\le j<K, k\ge1 \\ r\sum_{i=K}^{N_{\mathrm{B}}} g(i), \; j=K, \; k=0 \\ r\sum_{i=K-k+1}^{N_{\mathrm{B}}} g(i), \; j=K, \; k\ge0 \\ 0, \text{ otherwise} \end{cases}$$

The steady state probabilities $\pi(i)$, $0\le i\le K$, used in (9) are obtained from the equations

$$\sum_{i=1}^{k} \pi(i) = 1$$

$$\mathbf{\Pi P} = \mathbf{\Pi} \tag{11b}$$

where $\mathbf{\Pi}$ is the vector of the steady states probabilities and $\mathbf{P}$ is the matrix of the transition probabilities.

2.1.3. *Finite (arbitrarily large) buffer capacity.* When the buffer capacity K is finite but large, the accurate values of the quantities of interest are, as in case (b), obtained from equations (9). The steady-state probabilities $\pi(i)$, $0\le i\le K$, are obtained from the solution of a large number of equations given by (11). When the system operates outside its instability region (i.e. $\lambda<1-\epsilon$, $\epsilon>0$), the expected solutions $\pi(i)$, $0\le i\le K$, given by (9) become vanishingly small as i increases. These solutions may also be inaccurate particularly for $i>k_0$, for some $k_0<K$. To overcome this computational difficulty, bounds on the queueing quantities of interest are

derived, by introducing the concept of the dominant systems.

Consider the two nodes C^∞ and C^L which are identical to node 1, except for the capacities of the corresponding buffers. Node C^∞ has an infinite capacity buffer; node C^L has a buffer capacity of size $L<K$, where K is the capacity of node 1. Let A^j, denote the queueing quantity A associated with node C^j, $j=L,\infty$. Then,

$$Q_1^L \leq Q_1 \leq Q_1^\infty \tag{12a}$$

$$Q_2^L \leq Q_2 \leq Q_2^\infty \tag{12b}$$

$$D^L \leq D \leq D^\infty \tag{12c}$$

$$Q_2^L - (Q_1^\infty)^2 \leq V = Q_2 - (Q_1)^2 \leq Q_2^\infty - (Q_1^L)^2 \tag{12d}$$

where Q_1, Q_2, V and D are the queueing quantities, A, associated with node 1. The first two inequalities are justified by the fact that (assuming ergodicity)

$$d_j^L \leq d_j \leq d_j^\infty \tag{13}$$

where d_j^i denotes the number of packets at the jth slot in node C^i, $i=L$, ∞. The previous inequality is implied by the fact that, at any time instant, the occupancy of a buffer of capacity K_1 cannot be larger than that of a buffer of capacity K_2, where $1 \leq K_1 \leq K_2 \leq \infty$, assuming that the two buffers operate under identical conditions. The right part of inequality (12c) becomes obvious in view of the fact that packets in C^∞ which find more than K packets in the system are rejected by node 1 and, thus, these packets (of delay greater than K) do not contribute to the mean packet delay of node 1; their portion is substituted by packets of delay less than K in the calculation of the mean packet delay D. A similar argument justifies the left part of inequality (12c). Finally, (12d) becomes obvious in view of the inequalities (12a) and (12b).

Equations (12) establish upper and lower bounds on Q_1, Q_2, V and D. In most practical applications where buffer overflow is not desired, K should be sufficiently large so that $\pi(K)$ be extremely small. Under such condition $A \approx A^\infty$, $A \in \{Q_1, Q_2, D, V\}$. The smaller the value of $\pi(K)$ the larger the expected accuracy of the approximation of A by A^∞. On the other hand, if $\pi(L)$, i.e. the steady state probability that node C^L is in state L, is very small then $A \approx A^L$, $A \in \{Q_1, Q_2, D, V\}$. Again, the smaller the value of $\pi(L)$ the larger the expected accuracy of the approximation. If $\pi(L)$ is sufficiently small then $\pi(K)$, $\pi(K) \leq \pi(L)$, is even smaller, and thus $A \approx A^\infty$, as well. That is, if A^L is a good approximation of A and thus we can write $A \approx A^L$, then A^∞ is probably a good approximation of A as well, and we can write $A \approx A^\infty$. Under such conditions, the upper and the lower bounds in (12) are expected to almost coincide.

To analyse the queueing behaviour of buffer 2 an accurate description of the packet (output) process generated by node 1 is required.

2.2. *The packet output process of node* 1, R_{12}^o

Although the packet input process to node 1, described by a GBP, is an independent one, the process of the packets departing from this node is not, owing to the dependencies introduced by the operation of the node. This process is shown to be accurately described by an MMGBP.

Let $\{d_j\}_{j\geq 0}$ be the Markov chain defined in Section 2.1.2 which describes the number of packets in node 1 at the end of the jth slot. Let $S = \{0, 1, 2, \ldots, K\}$ be its state space, where K, $K \leq \infty$, is the buffer capacity of node 1. If K is finite and of small or moderate value, then $\{d_j\}_{j\geq 1}$ will serve as the underlying Markov chain of the MMGBP to be used for the description of R_{12}^o. If K is very large or infinite, then a new Markov chain, $\{d^L_j\}_{j\geq 0}$, with a state space of reduced cardinality L, $L<K$, will be incorporated in the description of R_{12}^o, to lead to tractable computations of the queueing quantities in node 2. Although the description of R_{12}^o based on $\{d^L_j\}_{j\geq 0}$ is an approximate one, it turns out that it results in very accurate calculation of the queueing quantities of interest in node 2. The parameters of $\{\bar{d}_j\}_{j\geq 0}$, $\bar{d}_j \in \{d_j, d^L_j\}$, are given by (10) and (11) if $\bar{d}_j = d_j$, and they are obtained from (11) if $\bar{d}_j = d^L_j$, where the following probabilities of the reduced state space Markov chain d^L_j are used:

$$\pi'(L) = \sum_{i=L}^{K} \pi(i)$$

$$\pi'(j) = \pi(j),\ 0 \leq j < L$$

$$p'(i,L) = \sum_{j=L}^{K} p(i,j),\ 0 \leq i < L$$

$$p'(L,L-1) = \frac{\pi(L)}{\pi'(L)} p(L,L-1)$$

$$p'(L,L) = 1 - p'(L,L-1)$$

$$p'(k,j) = p(k,j), \text{ otherwise}$$

The MMGBP which describes the packet output process R_{12}^o is completely defined by determining the probability distribution $\phi(\bar{d}_j,k)$. Clearly,

$$\phi(0,0) = 1,\ \phi(\bar{d}_j,1) = 1 \text{ for } \bar{d}_j > 0,\quad \bar{d}_j \in \{d_j, d_j^L\} \tag{14}$$

since node 1 outputs one packet when $\bar{d}_j > 0$ and zero packets otherwise.

Let $\{a_j\}_{j\geq 0}$ and $\{a^L_j\}_{j\geq 0}$ be the packet output processes determined by the probability distribution given by (14) and the underlying Markov chains $\{d_j\}_{j\geq 0}$ and $\{d^L_j\}_{j\geq 0}$, respectively. As mentioned before, the queueing quantities A associated with node 2 are accurately calculated under the approximation of the true packet output process $(a_j)_{j\geq0}$ by the process $\{a^L_j\}_{j\geq0}$, as long as node 1 operates outside its instability region. Under the latter conditions, states i, $i>k_0$, for some $k_0<L$, are almost never visited by the true Markov chain $(d_j)_{j\geq0}$. Thus, the Markov chain $\{d^L_j\}_{j\geq0}$, for some $L>k_0$, is a good approximation of $\{d_j\}_{j\geq0}$. In addition to that, the number of packets, a_j, generated by node 1 in the jth slot, is the same and independent of the state of the underlying Markov chain, as long as it is a non-zero state (see (14)). Owing to the latter observation, an additional refinement of the approximation of $\{d_j\}_{j\geq0}$ by $\{d^L_j\}_{j\geq0}$, as seen from the output process $\{a^L_j\}_{j\geq0}$, is introduced, as long as $L>0$. Finally, the queueing process in node 2 is a complex one and it may also introduce some smoothing on the differences between the processes $\{a_j\}_{j\geq0}$ and $\{a^L_j\}_{j\geq0}$, and consequently de-emphasize the difference between $\{d_j\}_{j\geq0}$ and $\{d^L_j\}_{j\geq0}$, as inferred from the values of the queueing quantities A evaluated at node 2. The previous arguments offer some non-rigorous explanations of the expected (and observed) accuracy of the approximation of $\{d_j\}_{j\geq0}$ by $\{d^L_j\}_{j\geq0}$ (and, consequently, of the approximation of $\{a_j\}_{j\geq0}$ by $\{a^L_j\}_{j\geq0}$), as measured by the accuracy of the calculation of the queueing quantities A at node 2.

2.3. *Queueing behaviour of buffer* 2

At this point, the queueing behaviour of node 2 is studied under the packet arrival processes R^o_{12} and R^i_{23}, each of which is modelled as an MMGBP; R^o_{12} is the packet output process from node 1 and is described in Section 2.2; R^i_{23} is an arbitrary MMGBP as defined at the beginning of Section 2. The queueing quantities A associated with node 2 are calculated when the cardinalities of the Markov chains associated with R^o_{12} and R^i_{23} are of small or moderate value.

When the capacity of buffer 2 is infinite, then the equivalent FIFO system, shown in Figure 5, with $N=2$ is considered. The queueing quantities of interest A, $A\in\{Q_1, Q_2, V, D\}$ are computed from equations (1)–(5). The obtained results are exact if the underlying Markov chain associated with R^o_{12} is the true one (i.e. if the capacity of buffer 1 is of small or of moderate size), and they are approximate otherwise.

When the capacity of buffer 2 is of small or moderate size or large but finite and the operation of node 2 is away from its instability region, then the queueing quantities of interest can be computed as described in Sections 2.1.2 and 2.1.3, where the Markov chain $\{d_j\}_{j\geq0}$ is replaced by the three-dimensional Markov chain $\{x^o_j, x^i_j, d_j\}_{j\geq0}$; x^o_j denotes the underlying Markov chain of the packet arrival process R^o_{12}; x^i_j denotes the underlying Markov chain of the packet arrival process R^i_{23}; d_j denotes the buffer occupancy of node 2.

3. ANALYSIS OF TWO BUFFERS UNDER SOME ROUTEING POLICIES

In this section, the queueing system shown in Figure 2 is studied. The only difference between this system and the one shown in Figure 4 (studied in section 2) is that routeing decisions divert some of the packet input traffic to node 1, R^i_{13}. The system shown in Figure 2 appears in network topologies such as the one shown in Figure 3, where the packet traffic which enters node 1 and is to be forwarded to node 3 has two alternate routes; a direct one from node 1 to node 3 and an indirect one through node 2. The routeing policies at node 1 may be adopted for the regulation of the rate of the traffic which is forwarded to node 2. As a result, overloading of the links between nodes 2 and 3, and between nodes 1 and 3, may be avoided. In this paper, four different routeing policies, P_1–P_4, are considered at node 1. The arrival process R^i_{13} is assumed to be the same, in probabilistic structure, under all routeing policies. Its intensity is modified accordingly to generate a fixed packet output rate under any of the routeing policies considered. Then, comparison of the queueing results at node 2 under the various routeing policies at node 1 is meaningful. The packet input process R^i_{13} determined by the parameters r_{13} and $g(k)$, $1\leq k\leq N_B$, where r_{13} is the message (or multi packet) arrival rate and $g(k)$ is the probability distribution of the message length (in packets). $g(k)$ is assumed to be the same under all policies considered, to maintain the same probabilistic structure of the packet arrival process in node 1; r_{13} is properly determined for each routeing policy, to generate the desired (fixed) packet output traffic to be forwarded to node 2. The following policies are considered; all packets to be forwarded to node 2 are stored in a single buffer, called buffer 12:

- P_1. Buffer 12 stores a new message (i.e. all packets arriving over a single slot) only if it is empty.
- P_2. Buffer 12 stores a new message (i.e. all packets arriving over a single slot) only if it has at most one packet stored.
- P_3. Buffer 12 stores all packets up to a maximum number $\Theta<\infty$.
- P_4. Buffer 12 stores half (or a portion) of the packets arriving over a slot (or half of them plus one in case of an odd number of packet arrivals), according to a deterministic splitting, up to a maximum Θ_4; Θ_4 can be infinite.

Clearly, the above routeing policies prevent all traffic R^i_{13} from being forwarded to node 2. It is assumed that the status of the buffer in node 2 is not available to node 1. Only the independent input traffic R^i_{23} is assumed to be known. The rate of this traffic determines the desired output traffic from node 1 (to be forwarded to node 2). Given the distribution of the packets arriving over a single slot at node 1 and a desired output rate from node 1, as determined by the desired total load in node 2, the best performing policy P_i, $P_i \in P = \{P_1, P_2, P_3, P_4\}$, may be identified. The queueing quantities A as induced by node 2 may serve as a measure of the performance of the routeing policies and the optimal such policy may be defined with respect to A. Clearly, the best policy found, P_o, is optimal in the class P considered and it should not be considered as such over all possible routeing policies.

Modification of the distribution of the number of packets arriving at node 1 over a slot, might also result in a different optimal policy from class P. Although this paper does not provide for the optimal routeing policy at node 1 (and this is not the objective here), it does provide some insight on the effect of various routeing policies on the intensity of the queueing problems, caused by the packet processes generated by these routeing policies. As a result, some conclusions on the characteristics of the optimal routeing policy are drawn and a direction for the search of such a routeing scheme is identified.

The buffer capacity of node 1 is assumed to be such that all packets to be forwarded to node 2 be accommodated. This is guaranteed as long as the capacity of buffer 1, K, is such that $K \geq N_B$ under policies P_1 and P_2 and $K \geq \Theta$ under policy P_3. Although no finite value of K can guarantee the accommodation of all packets destined to node 2 under P_4 (if $\Theta_4 = \infty$), a sufficiently large value of K almost eliminates the probability of buffer overflow. The appropriate value of K is not expected to be large, as long as the traffic load r_{12}^o, offered to the server connected to buffer 12, is not heavy. The latter assumption is realistic if monopolization of the facilities of node 2 by the traffic R_{12}^o is to be avoided.

3.1. *Queueing behaviour of node* 1

The queueing quantities of interest A, $A \in \{Q_1, Q_2, V, D\}$, are computed by applying the analysis presented in Section 2.1.2, where the buffer capacity K is set to be equal to N_B under policies P_1 and P_2 and equal to Θ under policy P_3. The transition probabilities $p(k,j)$ of the Markov chain $\{d_j\}_{j \geq 0}$ are given by

$$P(k,j) = \begin{cases} 1 & , 1 \leq k \leq N_B, j=k-1 \\ rd(j) & , 1 \leq j \leq N_B, k=0 \\ 1-r & , k=j=0 \\ 0 & , \text{otherwise} \end{cases} \tag{15}$$

under routeing policy P_1, by

$$P(k,j) = \begin{cases} 1 & 2 \leq k \leq N_B, j=k-1 \\ rd(j) & 1 \leq j \leq N_B, 0 \leq k \leq 1 \\ 1-r & k=j=0 \text{ or } k=1, j=0 \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

under routeing policy P_2 and by (10) under policy P_3 by setting $K = \Theta$. Finally, the analysis presented in Sections 2.1.2, 2.1.3 or 2.1.1 is applied for the calculation of A, $A \in \{Q_1, Q_2, V, D\}$, under policy P_4, depending on whether the buffer capacity Θ_4 is small, large or infinite, respectively. Notice that the splitting of the packets which arrive over the same slot modifies the message length distribution $g(j)$, as seen from buffer 12, resulting in a better randomized packet output process under this routeing policy.

3.2. *The packet output process from node* 1, R_{12}^o

The packet output process R_{12}^o is modelled as a MMGBP, as described in Section 2.2. The (exact) underlying Markov chain $\{d_j\}_{j \geq 0}$ is incorporated in the description of the packet output process $\{a_j\}_{j \geq 0}$ under policies P_1, P_2, P_3 and P_4, (if Θ_4 is small). The (approximate) underlying Markov chain $\{d_j^L\}_{j \geq 0}$ is incorporated in the description of the (approximate) packet output process $\{a_j^L\}_j$, under policy P_4, when Θ_4 is very large or infinite.

3.3. *Queueing behaviour of node* 2

The queueing quantities of interest A, $A \in \{Q_1, Q_2, V, D\}$ are computed by applying the analysis approach presented in Section 2.3.

4. NUMERICAL RESULTS

The analysis approach developed in the previous two sections is applied to systems such as the ones shown in Figures 2 and 4, operating under specific traffic conditions. Numerical results are obtained and some interesting conclusions are drawn. The input traffic to node 1, which is to be forwarded to node 2, is modelled as a GBP with parameters r and $g(j)$, $1 \leq g(j) \leq N_B$. In this section, the following parameters of this process are considered: $N_B = 5$, $g(1) = 0{\cdot}2$, $g(2) = 0{\cdot}3$, $g(3) = 0{\cdot}3$, $g(4) = 0{\cdot}1$, $g(5) = 0{\cdot}1$ and various values of the message arrival rate r. Let r_{in} and r_{out} be the packet input and the packet output rates associated with node 1.

First, the system shown in Figure 4 is considered. For various values of r and K (the capacity of buffer 1), $1 \leqslant K \leqslant \infty$, the queueing quantities A, $A \in \{Q_1, Q_2, V, D\}$, associated with node 1, are exactly computed. The results are shown in Table I. The queueing quantities are computed from equations (6)–(8) when $K = \infty$ and from equations (9) when $K < \infty$. From these results, the monotonicity of the quantities Q_1, Q_2 and D with respect to the buffer capacity K, as stated in equations (12), is clearly observed. Notice also that the values of A, $A \in \{Q_1, Q_2, V, D\}$, computed for the case of the largest finite value of K, k', shown in Table I, practically coincide with those obtained under infinite buffer capacity. This result illustrates that the closed form expressions (6)–(8) may be used for the computation

Table I. Queueing behaviour of buffer 1 without routeing policies

	r_{in}	r_{out}	K	Q_1	Q_2	V	D
0·05	0·13	0·0500	1	0·050	0·050	0·0475	1·000
0·05	0·13	0·1285	5	0·280	0·781	0·703	2·176
0·05	0·13	0·1297	7	0·289	0·841	0·757	2·223
0·05	0·13	0·1300	10	0·291	0·859	0·774	2·237
0·05	0·13	0·1300	∞	0·291	0·860	0·775	2·237
0·10	0·26	0·100	1	0·100	0·100	0·090	1·000
0·10	0·26	0·253	5	0·579	1·693	1·358	2·288
0·10	0·26	0·258	7	0·621	1·980	1·594	2·404
0·10	0·26	0·260	10	0·636	2·116	1·710	2·450
0·10	0·26	0·260	∞	0·638	2·141	1·733	2·455
0·15	0·39	0·150	1	0·150	0·150	0·127	1·000
0·15	0·39	0·372	5	0·897	2·751	1·946	2·412
0·15	0·39	0·385	7	1·008	3·519	2·504	2·620
0·15	0·39	0·389	10	1·063	4·041	2·910	2·733
0·15	0·39	0·390	∞	1·078	4·240	3·076	2·765
0·20	0·52	0·200	1	0·200	0·200	0·160	1·000
0·20	0·52	0·484	5	1·234	3·963	2·440	2·550
0·20	0·52	0·506	7	1·457	5·565	3·444	2·877
0·20	0·52	0·517	10	1·610	7·066	4·473	3·117
0·20	0·52	0·520	15	1·676	7·951	5·143	3·225
0·20	0·52	0·520	20	1·685	8·125	5·286	3·241
0·20	0·52	0·520	∞	1·687	8·159	5·314	3·243
0·30	0·78	0·300	1	0·300	0·300	0·210	1·000
0·30	0·78	0·680	5	1·949	6·843	3·044	2·868
0·30	0·78	0·752	10	3·228	18·804	8·381	4·293
0·30	0·78	0·777	20	4·293	39·892	18·462	5·527
0·30	0·78	0·780	30	4·540	43·737	23·118	5·825
0·30	0·78	0·780	∞	4·598	46·028	24·885	5·895

of A, $A \in \{Q_1, Q_2, V, D\}$, for any buffer capacity which is larger than k'. For the values of r considered in Table I, the corresponding probability $\pi(d_j = k')$ is less than 10^{-7}.

When the buffer capacity of buffer 1 is finite and equal to 20 and the capacity of buffer 2 is infinitely large, the queueing quantities of interest, A, associated with node 2 are calculated by using equations (1)–(5), and they are shown in Table II; γ is the clusterness coefficient associated with the packet output process from node 1, R^{o}_{12}, defined as

$$\gamma = P(\bar{0},\bar{0}) - P(0,\bar{0})$$

where 0 is the zero packet generating state of the underlying Markov chain of the MMGBP which models the process R^{o}_{12} (as described in Section 2.2) and $\bar{0}$ is the union of all the packet-generating states of this Markov chain. As is observed below, the parameter γ affects the values of the queueing quantities of interest associated with node 2. The (independent from R^{o}_{12}) packet arrival process R^{i}_{23} is assumed to be a two-state MMGBP with clusterness coefficient $\gamma' = 0{\cdot}2$ and packet arrival rate equal to $0{\cdot}9 - r_{out}$, where r_{out} is the packet rate of R^{o}_{12} and 0·9 is the total packet traffic load offered to node 2. A two-state MMGBP is completely determined by r and γ. Notice that for $r \leqslant 0{\cdot}20$ the behaviour of buffer 1 is identical to that of a buffer with infinite capacity ($r^{20}_{out} = r^{\infty}_{out}$, where r^{K}_{out} is the packet output rate of a buffer of capacity K) and that the clusterness coefficient γ is identical to that corresponding to the packet output process of an infinite capacity buffer. The latter is true since no

Table II. Queueing behaviour of buffer 2 without routeing policies at node 1 and buffer 1 capacity $K = 20$

r	r_{out}	γ	Bernoulli	Q_1	Q_2	V	D
0·05	0·130	0·615	2·112	3·390	22·963	13·859	3·767
0·10	0·260	0·615	2·849	5·402	59·330	35·554	6·000
0·15	0·390	0·615	3·210	6·897	99·634	58·960	7·664
0·20	0·520	0·615	3·195	7·796	131·000	78·026	8·662
0·30	0·777	0·613	2·060	6·257	87·895	55·000	6·953

packet rejection, which would modify the clusterness of the packet output process, takes place and, thus, γ is completely determined by the message size distribution $g(j)$, and not by the message input rate to node 1, r. For $r = 0{\cdot}3$, some packet rejection takes place ($r^{20}_{\text{out}} = 0{\cdot}777 < 0{\cdot}78 = r^{\infty}_{\text{out}}$) which leads to a reduced clusterness coefficient γ. Notice that the values of the quantities of interest associated with node 2 change as r increases despite the fact that $r^{o}_{12}+r^{i}_{23}$, γ and γ' remain constant. The queueing results in node 2 depend on (a) the probabilistic structure of the processes R^{o}_{12} and R^{i}_{23}, (b) the total offered packet load $r^{o}_{12}+r^{i}_{23}$ and (c) the ratio r^{o}_{12}/r^{i}_{23}. Clearly, the process R^{i}_{23} and the total offered load $r^{o}_{12}+r^{i}_{23}$ remain unchanged. However, despite the invariance of γ, the probabilistic structure of R^{o}_{12} changes, since this process depends on a different underlying Markov chain. The change in the structure of R^{o}_{12} (which is limited to a small variation in the steady state probabilities of the underlying Markov chain) is one factor affecting the value of A as r increases, probably the least significant. The variations in the values A are believed to be mostly due to the change in the ratio r^{o}_{12}/r^{i}_{23}. When $r^{o}_{12} << r^{i}_{23}$, R^{o}_{12} acts as a disturbance on a queueing process determined basically by R^{i}_{23}. This disturbance, which causes additional queueing problems since more packets may arrive over a single slot, increases as r^{o}_{12} increases up to a point (or a region) where a maximum is achieved. Beyond that point, a decrease in the intensity of the queueing problems is observed as R^{i}_{23} acts as a disturbance (of decreasing intensity, as r^{o}_{12} increases) to a queueing process now dominated by R^{o}_{12}.

When the capacity of the buffer 1 is infinite, then the queueing results in node 2 are obtained from equations (1)–(5), where the MMGBP describing the packet output process R^{o}_{12} is approximately described based on a truncated Markov chain with state space of cardinality L, as described in Section 2.2. The results for the queueing quantities of interest associated with node 2 are shown in Table III, for various message input rates to the first node, r, and different values of L. The steady-state probabilities $\pi(j)$, $0 \leqslant j \leqslant \infty$, are computed by assuming a large but finite value of K instead of an infinite one, and applying equations (11). For the parameters considered in Table III, $\pi(j)$ remains unchanged for any $K > 80$ used in (11). Thus, these values of $\pi(j)$, $0 \leqslant j \leqslant K$ are considered to practically coincide with those under infinite buffer capacity. For message input rates r less than 0·20, a truncation of the true underlying Markov chain associated with R^{o}_{12} at $L = 20$ gives results which remain, in essence, unchanged for any $L > 20$; the latter is observed at $L = 50$ when $r = 0{\cdot}3$. For all values of r, the corresponding value of $\pi(L-1)$ is less than 10^{-5} and, thus, states larger than the $(L-1)$th are visited not often enough to significantly affect the accuracy of the approximation of the process R^{o}_{12}, as seen from the induced queueing quantities at node 2.

If processes R^{o}_{12} and R^{i}_{23} are approximated by a Bernoulli process, then the mean packet delay induced by node 2 is given in Table III, for the various input rates r and total traffic load offered to node 2 equal to 0·9. By comparing the mean delay results in node 2, shown in Table III, under the MMGBP and the Bernoulli process modelling

Table III. Queueing behaviour of buffer 2 without routeing policies at node 1 and infinite capacity of buffer 1

r	r_{in}	r_{out}	γ	L	Q_1	Q_2	V	D	Bernoulli
0·05	0·13	0·13	0·615	1	3·753	29·378	18·947	4·170	2·112
0·05	0·13	0·13	0·615	5	3·365	21·343	13·382	3·740	2·112
0·05	0·13	0·13	0·615	10	3·390	21·963	13·859	3·767	2·112
0·05	0·13	0·13	0·615	20	3·390	21·977	13·872	3·767	2·112
0·10	0·26	0·26	0·615	1	5·642	64·928	38·734	6·269	2·849
0·10	0·26	0·26	0·615	5	5·292	55·643	32·932	5·880	2·849
0·10	0·26	0·26	0·615	10	5·401	59·283	35·516	6·000	2·849
0·10	0·26	0·26	0·615	20	5·402	59·344	35·565	6·002	2·849
0·15	0·39	0·39	0·615	1	6·532	83·163	47·032	7·257	3·210
0·15	0·39	0·39	0·615	5	7·716	104·446	60·124	7·974	3·210
0·15	0·39	0·39	0·615	10	7·753	128·475	76·121	8·614	3·210
0·15	0·39	0·39	0·615	20	7·803	131·403	78·317	8·670	3·210
0·20	0·52	0·52	0·615	1	6·532	83·163	47·032	7·257	3·195
0·20	0·52	0·52	0·615	5	7·176	104·446	60·124	7·974	3·195
0·20	0·52	0·52	0·615	10	7·753	128·475	76·121	8·614	3·195
0·20	0·52	0·52	0·615	20	7·803	131·403	78·317	8·670	3·195
0·30	0·78	0·78	0·615	1	3·568	21·707	12·547	3·964	2·040
0·30	0·78	0·78	0·615	5	4·800	44·093	25·854	5·333	2·040
0·30	0·78	0·78	0·615	10	6·014	78·058	47·201	6·682	2·040
0·30	0·78	0·78	0·615	20	6·569	102·015	65·422	7·300	2·040
0·30	0·78	0·78	0·615	30	6·641	106·611	69·145	7·379	2·040
0·30	0·78	0·78	0·615	50	6·651	107·516	69·922	7·391	2·040

R^o_{12}, it is easily established that the Bernoulli approximation leads to significant underestimation of the queueing problems induced by node 2.

Finally, when the packet routeing policies P_1–P_4 are considered in buffer 1, the queueing quantities of interest associated with nodes 1 and 2 are shown in Tables IV and V, respectively. The message size probability distribution $g(j)$, $1 \leq j \leq 5$, remains the same as before. The message arrival rate is different in each case and is such that the output rate r^o_{12} is 0·45; $r^i_{23} = 0 \cdot 45$ and $\gamma' = 0 \cdot 3$. Notice that the queueing quantities associated with node 1 are the same under policies P_1 and P_2, because of the nature of these policies. On the other hand, policy P_2 results in larger values of A, $A \in \{Q_1, Q_2, V, D\}$ at node 2, as a result of the increased clusterness of R^o_{12} under policy P_2. Note that messages arriving at node 2 under policy P_1 are separated by at least one slot and, as a result, the intensity of the queueing problems in node 2 is reduced. Various values of the buffer limit Θ have been considered under policy P_3. Notice that the values of A, $A \in \{Q_1, Q_2, V, D\}$, increases at both nodes as Θ increases, although r^o_{12} remains constant, due to the increased clusterness γ (affecting the value of A associated with node 2) and the increased buffer size Θ (affecting the value of A associated with node 1), under the same probability distribution $g(j)$, $1 \leq j \leq 5$. Finally, the results under policy P_4 and for various values of the buffer constraint Θ_4 (denoted by Θ) are given in the same table. The new probability distribution $g'(j)$, $1 \leq j \leq 3$, is given by $g'(1) = g(1)+g(2) = 0 \cdot 5$, $g'(2) = g(3)+g(4) = 0 \cdot 4$ and $g'(3) = g(5) = 0 \cdot 1$. Notice that as the buffer constraint Θ increases the values of A, $A \in \{Q_1, Q_2, V, D\}$, also increase for the reasons stated before. Notice that the values of A under policy P_4 are smaller than those under policy P_3 corresponding to the same value of Θ, the reason being that the probability distribution $g'(j)$, $1 \leq j \leq 3$, has been changed under policy P_4 and less clustered packets are generated by $g'(j)$. The values of A obtained for $\Theta = 10$ under policy P_4 remain unchanged if a larger value of Θ is considered. Thus, these values of A are considered to be equal to the ones obtained for $\Theta = \infty$.

5. CONCLUSIONS

As was mentioned in the summary, the objective of this paper is not the development of an optimal routeing policy from node 1 to node 3 (Figure 3), given, for example, the packet input rates λ^i_{13} and λ^i_{23}. The latter has been the focal point of significant research, where the optimal routeing policies are usually based on the assumption of memoryless

Table IV. Queueing behaviour of buffer 1 under some routeing policies with $r_{out} = 0 \cdot 45$

Policy	r	r_{in}	γ	Θ	Q_1	Q_2	V	D
P_1	0·3147	0·8182	0·301	5	0·935	2·492	1·619	2·077
P_2	0·2394	0·6224	0·468	5	0·935	2·492	1·619	2·077
P_3	0·2794	0·7264	0·379	2	0·687	1·160	0·689	1·526
P_3	0·1845	0·4797	0·590	5	1·128	3·571	2·297	2·506
P_3	0·1738	0·4519	0·614	10	1·306	5·311	3·605	2·901
P_3	0·1731	0·4501	0·615	20	1·331	5·718	3·945	2·958
P_4	0·3210	0·5136	0·288	2	0·642	1·024	0·612	1·424
P_4	0·2822	0·4515	0·373	5	0·796	1·784	1·150	1·769
P_4	0·2813	0·4508	0·375	10	0·808	1·882	1·229	1·795

Table V. Queueing behaviour of buffer 2 under some routeing policies with $r_{out} = 0 \cdot 45$. The mean delay under the Bernoulli model is 3·25

Policy	r	r_{in}	γ	Θ	Q_1	Q_2	V	D
P_1	0·3147	0·8182	0·301	5	4·081	27·751	15·178	4·539
P_2	0·2394	0·6224	0·468	5	4·992	45·031	25·102	5·547
P_3	0·2794	0·7264	0·379	2	4·247	30·810	17·039	4·716
P_3	0·1845	0·4797	0·590	5	6·161	73·569	41·771	6·846
P_3	0·1738	0·4519	0·614	10	7·223	108·527	63·581	8·025
P_3	0·1731	0·4501	0·615	20	7·395	116·005	68·715	8·216
P_4	0·3210	0·5136	0·288	2	4·126	28·924	16·028	4·584
P_4	0·2822	0·4515	0·373	5	4·959	45·720	27·081	5·511
P_4	0·2813	0·4508	0·375	10	5·041	47·912	27·535	5·602

packet traffics characterized by their rates. The objectives of this paper have been the following:

1. To develop an exact (or, if not exact, a very accurate) analysis of two interconnected buffers, where the detailed description of the buffer packet output process, rather than simply the buffer packet output rate, is incorporated. Results show that a performance evaluation based on the packet rates and not on the detailed description of the packet process, may result in very inaccurate calculations of the queueing quantities of interest.
2. To investigate the accurate performance of the second node under some routeing decisions at the first node which result in a constant packet output rate from node 1. By describing the packet output process, generated under a certain routeing policy, as an MMGBP, the operation of the routeing policy can be incorporated in the detailed description of the resulting packet output process and, thus, the performance of the system can be accurately evaluated.

Since different routeing policies generate probabilistically different packet output processes, and since the queueing behaviour at node 2 is affected not only by the intensity of the packet arrival process, but also by the probabilistic behaviour of that process, it is concluded (and it is observed in the numerical results) that different routeing policies result in different queueing behaviours of node 2, under fixed packet traffics. The latter important conclusion implies that the probabilistic behaviour of the packet (output) process generated by a routeing policy should also be considered in the quest for the optimal routeing policies and not only the intensity of that pocess. Numerical results show that the best packet routeing policy (resulting in the minimum intensity of the queueing problems at node 2, as reflected by the queueing quantities A), from those which generate a fixed packet output rate from node 1, should be among the policies which best randomize the resulting packet output process. The clusterness coefficient γ has been introduced in this paper, as a measure of the randomization in the packet arrival process. The randomization in the packet arrival process to node 2 reduces the possibility of momentarily overloading of node 2 and, thus, improves the queueing behaviour of this node. The latter observation explains why, for instance, a routeing policy which would continuously forward packets to node 2 over a period T_1 and would divert them over the following T_2 time units, and then repeat the process, would be inefficient with respect to the induced queueing quantities of interest at node 2.

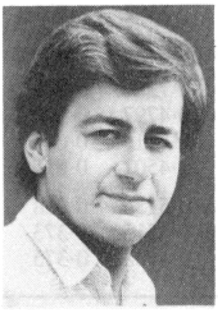
ACKNOWLEDGEMENT

Research partially supported by the U.S. National Science Foundation under Grant NCR-9011962.

REFERENCES

1. I. Stavrakakis, 'An analysis approach to multi-level networking', *International Conference in Communications, ICC90*, Atlanta GA, U.S.A. April 1990.
2. M. Reiser, 'Performance evaluation of data communication networks', *Proceedings of the IEEE*, **70**, (2), 171–196 (1982).
3. H. Kobayashi and A. Konheim, 'Queueing models for computer communications systems analysis', *IEEE Trans. Communications*, **COM-25**, (1), 2–28 (1977).
4. B. Gavish and I. Neumann, 'A system for routing and capacity assignment in computer communication networks', *IEEE Trans. Communications*, **COM-37**, (4), 360–366 (1989).
5. T. G. Yum and M. Schwartz, 'Comparison of routing procedures for circuit-switched traffic in nonhierarchical networks', *IEEE Trans. Communications*, **COM-35**, (5), 535–544 (1987).
6. C. Knessl, B. Matkowsky, Z. Schuss and C. Tier, 'Two parallel queues with dynamic routing', *IEEE Trans. Communications*, **COM-34**, (12), 1170–1175 (1986).
7. C. Knessl, B. Matkowsky, Z. Schuss and C. Tier, 'Two parallel M/G/I queues where arrivals join the system with the smaller buffer contact', *IEEE Trans. Communications*, **COM-35**, (11), 1153–1158 (1987).
8. I. Stavrakakis, 'Analysis of a statistic multiplexer under a general input traffic model', *IEEE INFOCOM'90 Conference*, San Francisco, 5–7 June 1990, pp. 1220–1225.

Author's biography:



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 at present an Assistant Professor. His research interests are in statistical communication theory, multiple-access algorithms, computer communication networks, queueing systems and system modelling and performance evaluation.

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