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 routing policies in node 1. The objective at this point is to evaluate the queueing behaviour of both buffers under some routing policies and it is not to seek for the optimal packet routing policy in the first node. It is shown how the routing 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 routing 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 routing policies considered. It turns out that a routing policy which best randomizes the packet output process results in the least serious queueing problems in the second node. The latter implies that routing 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 effect of the routing policy considered.

KEY WORDS Queueing systems Buffer behaviour Statistical multiplexing Routing 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 traffic entering the system, and the routing and service policies at the various nodes. For given packet input processes to a large network, different routing 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; \( R_k \) denotes the packet traffic which enters node \( k \) and is to be forwarded to node \( j \), through possibly more than one path, \( R_{kj} \) 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 \( R_{12} \) and a dependent discrete time packet arrival process \( R_{13} \). Notice that no routing is incorporated in node 1 and thus \( \lambda_{12} = \lambda_{12}' \), whereas \( \lambda_{1j} \) and \( \lambda_{2j} \) denote packet rates associated with the packet processes \( R_{1j} \) and \( R_{2j} \), 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 \( R_{12} \) is described. Finally, buffer 2 is analysed by incorporating the (dependent) packet processes \( R_{12} \) and \( R_{13} \).
In the sequel, the behaviours of both buffers of the system in Figure 3 are studied under some simple routing policies applied to node 1, by following the three steps described before. By maintaining a fixed packet output rate, $\lambda^2_1$ (assuming that $\lambda^2_1$ is constant), it is observed that the routing 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 routing policies in node 1 generate statistically different packet output processes $R^2_1$. Although the objective of this paper is not the development of efficient routing strategies, a unified description of the packet output process, which incorporates the routing strategies, is developed and some interesting conclusions on the effect of the routing 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 routing policy, which takes into consideration the queueing problems in the subsequent nodes of the network caused by the packet process generated by the routing policy. This approach to routing is quite different from the traditional ones which are either based only on desired packet rates to be generated by the routing policy, or try to minimize the queueing problems in the node where the routing decision is taken.

2. ANALYSIS OF TWO BUFFERS WITHOUT ROUTING DECISIONS

In this section, the system shown in Figure 4 is analysed. The packet input process to the first node, $R^0_1$, 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, $g(t), 1 \leq j \leq N_b$, 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 $R^0_2$ represents the packet output process from the...
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).

The expected number of packets in the system is given by:

$$Q = \sum_{\gamma \in \mathcal{S}} W_{i}(\gamma)$$

(1)

where $\mathcal{S} = S^{1} \times S^{2} \times \ldots \times S^{N}$ and $W_{i}(\gamma)$, $\gamma \in \mathcal{S}$, 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 $i$th input line, $S^{i}$) given by:

$$W_{i}(\gamma) = \sum_{\bar{\gamma} \in \mathcal{S}} \lambda_{i}(\bar{\gamma})p(\bar{\gamma}, \gamma) + \sum_{\bar{\gamma} \in \mathcal{S}} (\mu_{i} - 1)p(\bar{\gamma})\pi(\bar{\gamma})$$

(2a)

and the linearly independent equation

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

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

The first node. Input $R^{i}_{12}$ is assumed to be described by a Markov modulated generalized Bernoulli process (MMGBP) which is, in general, a dependent process. That is, $R^{i}_{12}$ is described by a discrete time process $\{a_{i}\}_{i=0}^{\infty}$ which depends on an underlying Markov chain $\{z_{i}\}_{i=0}^{\infty}$. The detailed description of $\{a_{i}\}_{i=0}$ is given below.

Let $\{z_{i}\}_{i=0}^{\infty}$ 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 $\{z_{i}\}_{i=0}^{\infty}$. It is assumed that the state of the underlying Markov chain determines (probabilistically) the packet arrival process of the corresponding line. That is, if $a_{i}(z_{i}) : S \rightarrow Z_{0}$, 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 $i$th potential packet arrival instant is given by $\phi(z_{i}, k) = \Pr(a(z_{i}) = k)$. 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. $\phi(x_{i}, 0) = 0$, 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...
\[
\sum_{x \in S} [2(\mu_x - 1)W_1(x) + 2(\mu_x - 1)q_0(x) + (2 + \sigma_x - 3\mu_x)\pi(x)] = 0
\] (2b)

where
\[
\pi(x) = \prod_{i=1}^{N} \pi(x^i), \quad p(\bar{x}, \bar{y}) = \prod_{i=1}^{N} p(x^i, y^i),
\]
\[
q_0(\bar{x}) = (1 - \lambda)p(\bar{x}, \bar{y}),
\]
\[
\mu_x = \sum_{\nu=1}^{R} v g_x(\nu), \quad \sigma_x = \sum_{\nu=1}^{R} \nu^2 g_x(\nu),
\]
\[
g_x(\nu) = \Pr \left( \sum_{i=1}^{N} a(x^i) = \nu \right)
\]
and
\[
\lambda = \sum_{x \in S} \mu_x \pi(x) < 1
\]
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; \( \pi(x) \) and \( p(x^i, y^i) \) are the steady state and the transition probabilities of the Markov chain associated with the \( i \)th input line. The mean packet delay is given by using Little’s formula, i.e.

\[
D = \frac{Q}{\lambda}
\] (3)

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

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

where \( Q_2 \) is the second moment of the buffer occupancy and \( W_2(x) \), \( x \in S \), are the solutions of the following system of linear equations:

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

and the linearly independent equation

\[
\sum_{x \in S} 3(\mu_x - 1)W_2(x) + \sum_{x \in S} [\mu_x^2 \pi(x) + 3\mu_x^2 [W_1(x) + q_0(x)]] + 2\mu_x^2 [W_1(x) + q_0(x)]p(\bar{x}, \bar{y}) = 0
\] (5b)

where

\[
\mu_2^f = \sum_{\nu=0}^{R} (\nu-1)g_2(\nu)
\]

and

\[
\mu_{2}^f = \sum_{\nu=0}^{R} (\nu-1)(\nu-2)g_2(\nu)
\] (5c)

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 MMGBGP 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 MMGBGP consists of a single state, \( x' \). By setting \( \bar{x} = x' \) and since \( \pi(x') = p(x', x') = 1 \) and \( \mu_x = \lambda \), where \( \lambda \) is the packet input rate, the following equations are obtained directly from (1)–(3):

\[
Q_1 = W_1(x'), \quad q_0(x') = 1 - \lambda
\]

\[
2(\lambda - 1)W_1(x') + 2(\lambda - 1)q_0(x') + 2 + \sigma_x^2 - 3\lambda = 0
\]

By manipulating the above equations, the following closed form expression for the mean buffer occupancy, \( Q_1 \), is obtained:

\[
Q_1 = \lambda + \frac{\sigma - \lambda}{2(1 - \lambda)}
\] (6a)

where \( \sigma \) 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,

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

where \( \lambda' \) and \( \sigma' \) are the first and second moments of the packet arrival process associated with the \( i \)th input line. By applying Little’s theorem to (6a), the
following closed form expression for the mean packet delay $D^1$, is obtained:

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

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

$$D^B = 1 + \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \lambda_i \lambda_j$$

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

$$V^1 = Q_1^1 - (Q_1^1)^2 \quad (8a)$$

where $Q_1^1$ is given by (6a), $Q_1^1$ is the second moment of the buffer occupancy given by

$$Q_2^1 = W_2(x') + Q_1^1 \quad (8b)$$

where

$$W_2(x') = \frac{1}{3(1-\lambda)}[\mu_2^2 + 3\mu_2^2(Q_1^1 + 1-\lambda)] \quad (8c)$$

Equations (8) are valid for any number of independent input streams, each of which is modelled as a GBP. In this case, $\lambda$ is given by (6b) and $\mu_2^f$ and $\mu_3^f$ 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 $k$th 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 $k$th 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:

$$Q_1^1 = \sum_{i=0}^{\infty} \pi(i)$$

$$V^1 = Q_2^1 - (Q_1^1)^2, \quad Q_2^1 = \sum_{i=0}^{\infty} \pi(i)^2$$

$$D^1 = Q_1^1$$

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

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

where $(x)^+ = x$ if $x \geq 0$ and is zero otherwise, and $a_j$ denotes the packet arrivals during the $j$th slot, it is clear that $\{d_j\}_{j=0}^{\infty}$ is a Markov chain with transition probabilities given by

$$P(k,j) = \begin{cases} 
1-r, & j=0, \quad 0 \leq k \leq 1 \\
rg(j), & 1 \leq j < K, \quad k=0 \\
rg(j-k+1), & 1 \leq j < K, \quad k \geq 1 \\
0, & \text{otherwise}
\end{cases}$$

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

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

$$\Pi P = \Pi$$

(11b)

where $\Pi$ is the vector of the steady states probabilities and $P$ 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 $\pi(i), 0 \leq i \leq K$, 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 $\pi(i), 0 \leq i \leq K$, 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 $C^\ast$ and $C^L$ which are identical to node 1, except for the capacities of the corresponding buffers. Node $C^\ast$ 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 $A^i$, denote the queueing quantity $A$ associated with node $C^i$, $j=L, \infty$. Then,

\begin{align}
Q^\ast &\leq Q_1 \leq Q^\ast \tag{12a} \\
Q^L &\leq Q_2 \leq Q^L \tag{12b} \\
D^L &\leq D \leq D^\ast \tag{12c} \\
Q^L - (Q^\ast)^2 &\leq V = Q_2 - (Q_1)^2 \\
&\leq Q^L - (Q^\ast)^2 \tag{12d}
\end{align}

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)

\begin{align}
d^L_j &\leq d_j \leq d^\ast_j \tag{13}
\end{align}

where $d^L_j$ denotes the number of packets at the $j$th slot in node $C^j$, $j=L, \infty$. The previous inequality is implied by the fact that, at any time instant, the occupancy of a buffer of capacity $K_j$ cannot be larger than that of a buffer of capacity $K$, 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 $C^\ast$ 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 small so that $\pi(K)$ is extremely small. Under such condition $A = A^\ast, A \in \{Q_1, Q_2, D, V\}$. The smaller the value of $\pi(K)$ the larger the expected accuracy of the approximation of $A$ by $A^\ast$. 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 $A = A^L, A \in \{Q_1, Q_2, D, V\}$. Again, the smaller the value of $\pi(L)$ the larger the expected accuracy of the approximation. If $\pi(L)$ is sufficiently small then $\pi(K) \leq \pi(L)$, is even smaller, and thus $A = A^\ast$, as well. That is, if $A^L$ is a good approximation of $A$ and thus we can write $A = A^L$, then $A^\ast$ is probably a good approximation of $A$ as well, and we can write $A = A^\ast$. 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, $R^*_{12}$

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=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 $j$th 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=1} will serve as the underlying Markov chain of the MMGBP to be used for the description of $R^*_{12}$. If $K$ is very large or infinite, then a new Markov chain, \{d^j\}_{j=0}, with a state space of reduced cardinality $L, L<K$, will be incorporated in the description of $R^*_{12}$, leading to tractable computations of the queueing quantities in node 2. Although the description of $R^*_{12}$ based on \{d^j\}_{j=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 \{d^j\}_{j=0}, \{d_j\} \in \{d_j, d^j\}$, are given by (10) and (11) if $d_{j} = d^L_{j}$ and they are obtained from (11) if $d_{j} = d^\ast_{j}$, where the following probabilities of the reduced state space Markov chain $d^j$ are used:

\begin{align}
\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}
\end{align}

The MMGBP which describes the packet output process $R^*_{12}$ is completely defined by determining the probability distribution $\phi(d^j, k)$. Clearly,

\begin{align}
\phi(0,0) = 1, \phi(d^j, 1) = 1 \text{ for } d_j > 0, \text{ } \\
d_j \in \{d_j, d^j\}
\end{align}

since node 1 outputs one packet when $d_j > 0$ and zero packets otherwise.

Let \{a^j\}_{j=0} and \{a^j\}_{j=0} be the packet output processes determined by the probability distribution given by (14) and the underlying Markov chains \{d^j\}_{j=0} and \{d^j\}_{j=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 \( \{a_j^i\}_{j=0} \) by the process \( \{a_j^i\}_{j=0} \), as long as node 1 operates outside its instability region. Under the latter conditions, states \( i > k_1 \), for some \( k_1 \leq L \), are almost never visited by the true Markov chain \( \{d_j^i\}_{j=0} \). Thus, the Markov chain \( \{d_j^i\}_{j=0} \), for some \( L > k_1 \), is a good approximation of \( \{d_j^i\}_{j=0} \). In addition to that, the number of packets, \( a_j^i \), generated by node 1 in the \( j \)th 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 \( \{d_j^i\}_{j=0} \) by \( \{a_j^i\}_{j=0} \), as seen from the output process \( \{a_j^i\}_{j=0} \), is introduced, as long as \( L > 0 \). Finally, the queuing process in node 2 is a complex one and it may also introduce some smoothing on the differences between the processes \( \{d_j^i\}_{j=0} \) and \( \{a_j^i\}_{j=0} \), and consequently de-emphasize the difference between \( \{d_j^i\}_{j=0} \) and \( \{a_j^i\}_{j=0} \), as inferred from the values of the queuing quantities \( A \) evaluated at node 2. The previous arguments offer some non-rigorous explanations of the expected (and observed) accuracy of the approximation of \( \{d_j^i\}_{j=0} \) by \( \{a_j^i\}_{j=0} \) (and, consequently, of the approximation of \( \{a_j^i\}_{j=0} \) by \( \{a_j^i\}_{j=0} \), as measured by the accuracy of the calculation of the queuing quantities \( A \) at node 2.

2.3. Queueing behaviour of buffer 2

At this point, the queuing behaviour of node 2 is studied under the packet arrival processes \( R_{12} \) and \( R_{23} \), each of which is modelled as an MMGBP; \( R_{12} \) is the packet output process from node 1 and is described in Section 2.2; \( R_{23} \) is an arbitrary MMGBP as defined at the beginning of Section 2. The queuing quantities \( A \) associated with node 2 are calculated when the cardinalities of the Markov chains associated with \( R_{12} \) and \( R_{23} \) 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 queuing 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 \( R_{12} \) 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 queuing quantities of interest can be computed as described in Sections 2.1.2 and 2.1.3, where the Markov chain \( \{d_j^i\}_{j=0} \) is replaced by the two-dimensional Markov chain \( \{x_j^i, y_j^i\}_{j=0} \), \( y_j^i \) denotes the underlying Markov chain of the packet arrival process \( R_{12}^i \), \( x_j^i \) denotes the queuing Markov chain of the packet arrival process \( R_{12}^i \), and \( d_j \) denotes the buffer occupancy of node 2.

3. ANALYSIS OF TWO BUFFERS UNDER SOME ROUTEING POLICIES

In this section, the queuing 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 routing decisions divert some of the packet input traffic to node 1, \( R_{13} \). 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 route from node 1 to node 3 and an indirect route node 1 to node 3.

The queuing processes at node 1 may possibly 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 routing policies, \( P_1 \rightarrow P_4 \), are considered at node 1. The arrival process \( R_{13} \) is assumed to be the same, in probabilistic structure, under all routing policies. Its intensity is modified accordingly to generate a fixed packet output rate under any of the routing policies considered. Then, comparison of the queuing results at node 2 under the various routing policies at node 1 is meaningful.

The packet input process \( R_{13} \) is determined by the parameters \( r_{13} \), \( g(k), 1 \leq k \leq N_B \). Here, \( r_{13} \) 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 routing 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:

- **P₁.** Buffer 12 stores a new message (i.e. all packets arriving over a single slot) only if it is empty.
- **P₂.** Buffer 12 stores a new message (i.e. all packets arriving over a single slot) only if it has at most one packet stored.
- **P₃.** Buffer 12 stores all packets up to a maximum number \( \Theta < \infty \).
- **P₄.** 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 \( \Theta₄ \); \( \Theta₄ \) can be infinite.

Clearly, the above routing policies prevent all traffic \( R_{13} \) 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 \( R_{13} \) 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 \( P_i \), \( P_i \in \mathcal{P} = \{P_1, P_2, P_3, P_4\} \), may be identified. The queueing quantities \( \mathcal{A} \) as induced by node 2 may serve as a measure of the performance of the routing policies and the optimal such policy may be defined with respect to \( \mathcal{A} \). Clearly, the best policy found, \( P_0 \), is optimal in the class \( \mathcal{P} \) considered and it should not be considered as such over all possible routing 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 \( \mathcal{P} \). Although this paper does not provide for the optimal routing policy at node 1 (and this is not the objective here), it does provide some insight on the effect of various routing policies on the intensity of the queueing problems, caused by the packet processes generated by these routing policies. As a result, some conclusions on the characteristics of the optimal routing policy are drawn and a direction for the search of such a routing 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 \( K = N_B \) under policies \( P_1 \) and \( P_2 \) and \( K = \Theta \) 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 \( \rho_{12} \), 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 \( \rho_{12} \) 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 \( \Theta \) under policy \( P_3 \). The transition probabilities \( p(k,j) \) of the Markov chain \( \{d_j\}_{j=0} \) given by

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

under routing policy \( P_1 \), by

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

3.2. The packet output process from node 1, \( R_{12}^\alpha \)

The packet output process \( R_{12}^\alpha \) is modelled as a MMGBP, as described in Section 2.2. The (exact) underlying Markov chain \( \{d_j\}_{j=0} \) is incorporated in the description of the packet output process \( \{a_j\}_{j=0} \) under policies \( P_1, P_2, P_3 \) and \( P_4 \), (if \( \Theta_4 \) is small). The (approximate) underlying Markov chain \( \{d_j\}_{j=0} \) is incorporated in the description of the (approximate) packet output process \( \{a_j\}_{j=0} \), under policy \( P_4 \), when \( \Theta_4 \) 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: \( N_B = 5, g(1) = 0.2, g(2) = 0.3, g(3) = 0.3, g(4) = 0.1, g(5) = 0.1 \) 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), \( 1 \leq K \leq \infty \), 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 routing policies

<table>
<thead>
<tr>
<th>r_{in}</th>
<th>r_{out}</th>
<th>K</th>
<th>Q_1</th>
<th>Q_2</th>
<th>V</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.13</td>
<td>0.0300</td>
<td>0.050</td>
<td>0.050</td>
<td>0.0075</td>
<td>1.000</td>
</tr>
<tr>
<td>0.05</td>
<td>0.13</td>
<td>0.1258</td>
<td>0.280</td>
<td>0.781</td>
<td>0.703</td>
<td>2.176</td>
</tr>
<tr>
<td>0.05</td>
<td>0.13</td>
<td>0.1297</td>
<td>0.289</td>
<td>0.841</td>
<td>0.757</td>
<td>2.223</td>
</tr>
<tr>
<td>0.05</td>
<td>0.13</td>
<td>0.1300</td>
<td>0.291</td>
<td>0.859</td>
<td>0.774</td>
<td>2.237</td>
</tr>
<tr>
<td>0.05</td>
<td>0.13</td>
<td>0.1300</td>
<td>0.291</td>
<td>0.860</td>
<td>0.775</td>
<td>2.237</td>
</tr>
<tr>
<td>0.10</td>
<td>0.26</td>
<td>0.1000</td>
<td>0.100</td>
<td>0.100</td>
<td>0.090</td>
<td>1.000</td>
</tr>
<tr>
<td>0.10</td>
<td>0.26</td>
<td>0.2535</td>
<td>0.579</td>
<td>1.693</td>
<td>1.358</td>
<td>2.288</td>
</tr>
<tr>
<td>0.10</td>
<td>0.26</td>
<td>0.2587</td>
<td>0.621</td>
<td>1.980</td>
<td>1.594</td>
<td>2.404</td>
</tr>
<tr>
<td>0.10</td>
<td>0.26</td>
<td>0.2660</td>
<td>0.636</td>
<td>2.116</td>
<td>1.710</td>
<td>2.450</td>
</tr>
<tr>
<td>0.10</td>
<td>0.26</td>
<td>0.2660</td>
<td>0.638</td>
<td>2.141</td>
<td>1.733</td>
<td>2.455</td>
</tr>
<tr>
<td>0.15</td>
<td>0.39</td>
<td>0.1500</td>
<td>0.150</td>
<td>0.150</td>
<td>0.127</td>
<td>1.000</td>
</tr>
<tr>
<td>0.15</td>
<td>0.39</td>
<td>0.3725</td>
<td>0.897</td>
<td>2.751</td>
<td>1.946</td>
<td>2.412</td>
</tr>
<tr>
<td>0.15</td>
<td>0.39</td>
<td>0.3857</td>
<td>1.008</td>
<td>3.519</td>
<td>2.504</td>
<td>2.620</td>
</tr>
<tr>
<td>0.15</td>
<td>0.39</td>
<td>0.3891</td>
<td>1.063</td>
<td>4.041</td>
<td>2.910</td>
<td>2.733</td>
</tr>
<tr>
<td>0.15</td>
<td>0.39</td>
<td>0.3900</td>
<td>1.078</td>
<td>4.240</td>
<td>3.076</td>
<td>2.765</td>
</tr>
<tr>
<td>0.20</td>
<td>0.52</td>
<td>0.2000</td>
<td>0.200</td>
<td>0.200</td>
<td>0.160</td>
<td>1.000</td>
</tr>
<tr>
<td>0.20</td>
<td>0.52</td>
<td>0.4845</td>
<td>3.124</td>
<td>9.655</td>
<td>2.440</td>
<td>2.550</td>
</tr>
<tr>
<td>0.20</td>
<td>0.52</td>
<td>0.5067</td>
<td>1.457</td>
<td>5.865</td>
<td>3.444</td>
<td>2.877</td>
</tr>
<tr>
<td>0.20</td>
<td>0.52</td>
<td>0.5170</td>
<td>1.610</td>
<td>7.066</td>
<td>4.473</td>
<td>3.177</td>
</tr>
<tr>
<td>0.20</td>
<td>0.52</td>
<td>0.5201</td>
<td>1.676</td>
<td>7.951</td>
<td>5.443</td>
<td>3.225</td>
</tr>
<tr>
<td>0.20</td>
<td>0.52</td>
<td>0.5200</td>
<td>1.685</td>
<td>8.125</td>
<td>5.286</td>
<td>3.241</td>
</tr>
<tr>
<td>0.20</td>
<td>0.52</td>
<td>0.5200</td>
<td>1.687</td>
<td>8.159</td>
<td>5.314</td>
<td>3.243</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.3000</td>
<td>0.300</td>
<td>0.300</td>
<td>0.210</td>
<td>1.000</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.6805</td>
<td>1.949</td>
<td>6.843</td>
<td>3.044</td>
<td>2.868</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.7520</td>
<td>3.228</td>
<td>18.804</td>
<td>8.381</td>
<td>4.293</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.7770</td>
<td>4.293</td>
<td>39.892</td>
<td>18.462</td>
<td>5.527</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.7800</td>
<td>4.540</td>
<td>43.737</td>
<td>23.118</td>
<td>5.825</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.7800</td>
<td>4.598</td>
<td>46.028</td>
<td>24.885</td>
<td>5.895</td>
</tr>
</tbody>
</table>

of $A_i, A \in \{Q_1, Q_2, V, D\}$, for any buffer capacity which is larger than $k_i$. For the values of $r$ considered in Table I, the corresponding probability $\pi(d_i = k_i)$ 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 infinite 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. $\gamma$ is the clusterness coefficient associated with the packet output process from node 1, $R_{12}$, defined as

$$\gamma = \text{P}(0,0) + \text{P}(0,0)$$

where $0$ is the zero packet generating state of the underlying Markov chain of the MMGMB which models the process $R_{12}$ (as described in Section 2.2) and $0$ is the union of all the packet-generating states of this Markov chain. As is observed below, the parameter $\gamma$ affects the values of the queueing quantities of interest associated with node 2. The (independent from $R_{12}$) packet arrival process $R_{13}$ is assumed to be a two-state MMGMB with clusterness coefficient $\gamma' = 0.2$ and packet arrival rate equal to $0.9 - r_{out}$, where $r_{out}$ is the packet rate of $R_{12}$ and 0.9 is the total packet traffic load offered to node 2. A two-state MMGMB is completely determined by $r$ and $\gamma$. Notice that for $r = 0.20$ the behaviour of buffer 1 is identical to that of a buffer with infinite capacity ($r_{out}^\infty = r_{out}$, where $r_{out}^\infty$ is the packet output rate of a buffer of capacity $K$) and that the clusterness coefficient $\gamma$ 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 routing policies at node 1 and buffer 1 capacity $K = 20$

<table>
<thead>
<tr>
<th>r</th>
<th>r_{out}</th>
<th>Bernoulli</th>
<th>Q_1</th>
<th>Q_2</th>
<th>V</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.130</td>
<td>0.615</td>
<td>2.112</td>
<td>3.390</td>
<td>22.963</td>
<td>13.859</td>
</tr>
<tr>
<td>0.10</td>
<td>0.260</td>
<td>0.615</td>
<td>2.849</td>
<td>5.402</td>
<td>59.330</td>
<td>35.554</td>
</tr>
<tr>
<td>0.15</td>
<td>0.390</td>
<td>0.615</td>
<td>3.210</td>
<td>6.897</td>
<td>99.634</td>
<td>58.960</td>
</tr>
<tr>
<td>0.20</td>
<td>0.520</td>
<td>0.615</td>
<td>3.195</td>
<td>7.796</td>
<td>131.000</td>
<td>78.026</td>
</tr>
<tr>
<td>0.30</td>
<td>0.777</td>
<td>0.613</td>
<td>2.060</td>
<td>6.257</td>
<td>87.895</td>
<td>55.000</td>
</tr>
</tbody>
</table>
packet rejection, which would modify the clusterness of the packet output process, takes place and, thus, \( \gamma \) is completely determined by the message size distribution \( g(j) \), and not by the message input rate to node 1, \( r \). For \( r = 0.3 \), some packet rejection takes place \( r_{\text{out}} = 0.777 < 0.78 = r_{\text{out}} \) which leads to a reduced clusterness coefficient \( \gamma \). Notice that the values of the quantities of interest associated with node 2 change as \( r \) increases despite the fact that \( r_{12} + r_{23} \), \( \gamma \) and \( \gamma' \) remain constant. The queueing results in node 2 depend on (a) the probabilistic structure of the processes \( R_{12} \) and \( R_{23} \), (b) the total offered packet load \( r_{12} + r_{23} \) and (c) the ratio \( r_{12}^2/r_{23} \). Clearly, the process \( R_{23} \) and the total offered load \( r_{12}^2 + r_{23} \) remain unchanged. However, despite the invariance of \( \gamma \), the probabilistic structure of \( R_{12} \) changes, since this process depends on a different underlying Markov chain. The change in the structure of \( R_{12} \) (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 \( r_{12}^2/r_{23} \). When \( r_{12}^2 < r_{23} \), \( R_{12} \) acts as a disturbance on a queueing process determined basically by \( R_{23} \). This disturbance, which causes additional queueing problems since more packets may arrive over a single slot, increases as \( r_{12} \) 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 \( R_{23} \) acts as a disturbance (of decreasing intensity, as \( r_{12}^2 \) increases) to a queueing process now dominated by \( R_{12} \).

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 \( R_{12}^0 \) 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 \( \pi(j), 0 < j < \infty \), 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 \( \pi(j) \), \( 0 < j < K \) 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 \( R_{12}^0 \) at \( L = 20 \) gives results which remain, in essence, unchanged for any \( L > 20 \); the latter is observed at \( L = 50 \) when \( r = 0.3 \). 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 \( R_{12}^0 \), as seen from the induced queueing quantities at node 2.

If processes \( R_{12}^0 \) and \( R_{23}^0 \) 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>
<thead>
<tr>
<th>( r )</th>
<th>( r_{in} )</th>
<th>( r_{out} )</th>
<th>( \gamma )</th>
<th>( L )</th>
<th>( Q_1 )</th>
<th>( Q_2 )</th>
<th>( V )</th>
<th>( D )</th>
<th>Bernoulli</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.13</td>
<td>0.13</td>
<td>0.615</td>
<td>1</td>
<td>3.753</td>
<td>29.378</td>
<td>18.947</td>
<td>4.170</td>
<td>2.112</td>
</tr>
<tr>
<td>0.05</td>
<td>0.13</td>
<td>0.13</td>
<td>0.615</td>
<td>5</td>
<td>3.365</td>
<td>21.343</td>
<td>13.382</td>
<td>3.740</td>
<td>2.112</td>
</tr>
<tr>
<td>0.05</td>
<td>0.13</td>
<td>0.13</td>
<td>0.615</td>
<td>10</td>
<td>3.390</td>
<td>21.963</td>
<td>13.859</td>
<td>3.767</td>
<td>2.112</td>
</tr>
<tr>
<td>0.05</td>
<td>0.13</td>
<td>0.13</td>
<td>0.615</td>
<td>20</td>
<td>3.390</td>
<td>21.977</td>
<td>13.872</td>
<td>3.767</td>
<td>2.112</td>
</tr>
<tr>
<td>0.10</td>
<td>0.26</td>
<td>0.26</td>
<td>0.615</td>
<td>1</td>
<td>5.642</td>
<td>64.928</td>
<td>38.734</td>
<td>6.269</td>
<td>2.849</td>
</tr>
<tr>
<td>0.10</td>
<td>0.26</td>
<td>0.26</td>
<td>0.615</td>
<td>5</td>
<td>5.292</td>
<td>55.643</td>
<td>32.932</td>
<td>5.880</td>
<td>2.849</td>
</tr>
<tr>
<td>0.10</td>
<td>0.26</td>
<td>0.26</td>
<td>0.615</td>
<td>10</td>
<td>5.401</td>
<td>59.283</td>
<td>35.516</td>
<td>6.000</td>
<td>2.849</td>
</tr>
<tr>
<td>0.10</td>
<td>0.26</td>
<td>0.26</td>
<td>0.615</td>
<td>20</td>
<td>5.402</td>
<td>59.344</td>
<td>35.565</td>
<td>6.002</td>
<td>2.849</td>
</tr>
<tr>
<td>0.15</td>
<td>0.39</td>
<td>0.39</td>
<td>0.615</td>
<td>1</td>
<td>6.532</td>
<td>83.163</td>
<td>47.032</td>
<td>7.257</td>
<td>3.210</td>
</tr>
<tr>
<td>0.15</td>
<td>0.39</td>
<td>0.39</td>
<td>0.615</td>
<td>5</td>
<td>7.716</td>
<td>104.446</td>
<td>60.124</td>
<td>7.974</td>
<td>3.210</td>
</tr>
<tr>
<td>0.15</td>
<td>0.39</td>
<td>0.39</td>
<td>0.615</td>
<td>10</td>
<td>7.753</td>
<td>128.475</td>
<td>76.121</td>
<td>8.614</td>
<td>3.210</td>
</tr>
<tr>
<td>0.15</td>
<td>0.39</td>
<td>0.39</td>
<td>0.615</td>
<td>20</td>
<td>7.803</td>
<td>131.403</td>
<td>78.317</td>
<td>8.670</td>
<td>3.210</td>
</tr>
<tr>
<td>0.20</td>
<td>0.52</td>
<td>0.52</td>
<td>0.615</td>
<td>1</td>
<td>6.532</td>
<td>83.163</td>
<td>47.032</td>
<td>7.257</td>
<td>3.195</td>
</tr>
<tr>
<td>0.20</td>
<td>0.52</td>
<td>0.52</td>
<td>0.615</td>
<td>5</td>
<td>7.176</td>
<td>104.446</td>
<td>60.124</td>
<td>7.974</td>
<td>3.195</td>
</tr>
<tr>
<td>0.20</td>
<td>0.52</td>
<td>0.52</td>
<td>0.615</td>
<td>10</td>
<td>7.753</td>
<td>128.475</td>
<td>76.121</td>
<td>8.614</td>
<td>3.195</td>
</tr>
<tr>
<td>0.20</td>
<td>0.52</td>
<td>0.52</td>
<td>0.615</td>
<td>20</td>
<td>7.803</td>
<td>131.403</td>
<td>78.317</td>
<td>8.670</td>
<td>3.195</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.78</td>
<td>0.615</td>
<td>1</td>
<td>3.568</td>
<td>21.707</td>
<td>12.547</td>
<td>3.964</td>
<td>2.040</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.78</td>
<td>0.615</td>
<td>5</td>
<td>4.800</td>
<td>44.093</td>
<td>25.854</td>
<td>5.333</td>
<td>2.040</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.78</td>
<td>0.615</td>
<td>10</td>
<td>6.014</td>
<td>78.058</td>
<td>47.201</td>
<td>6.682</td>
<td>2.040</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.78</td>
<td>0.615</td>
<td>20</td>
<td>6.569</td>
<td>102.015</td>
<td>65.422</td>
<td>7.300</td>
<td>2.040</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.78</td>
<td>0.615</td>
<td>30</td>
<td>6.641</td>
<td>106.611</td>
<td>69.145</td>
<td>7.379</td>
<td>2.040</td>
</tr>
<tr>
<td>0.30</td>
<td>0.78</td>
<td>0.78</td>
<td>0.615</td>
<td>50</td>
<td>6.651</td>
<td>107.516</td>
<td>69.922</td>
<td>7.391</td>
<td>2.040</td>
</tr>
</tbody>
</table>
It is easily established that the Bernoulli approximation leads to significant underestimation of the queueing problems induced by node 2.

Finally, when the packet routing policies $P_1$ to $P_4$ 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 $r_{12}'$ is $0.45$; $r_{23}' = 0.45$ and $\gamma = 0.3$. 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 $R_{12}'$ 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 $\Theta$ 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 $\Theta$ increases, although $r_{12}'$ remains constant, due to the increased clusterness $\gamma$ (affecting the value of $A$ associated with node 2) and the increased buffer size $\Theta$ (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 $\Theta$ (denoted by $\Theta$) are given in the same table. The new probability distribution $g'(j)$, $1 \leq j \leq 3$, is given by $g'(1) = g(1) + g(2) = 0.5$, $g'(2) = g(3) + g(4) = 0.4$ and $g'(3) = g(5) = 0.1$. Notice that as the buffer constraint $\Theta$ 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 $\Theta$, 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 $\Theta$ is considered. Thus, these values of $A$ are considered to be equal to the ones obtained for $\Theta = \infty$.

### Table IV. Queueing behaviour of buffer 1 under some routing policies with $r_{out} = 0.45$

<table>
<thead>
<tr>
<th>Policy</th>
<th>$r$</th>
<th>$r_{in}$</th>
<th>$\gamma$</th>
<th>$\Theta$</th>
<th>$Q_1$</th>
<th>$Q_2$</th>
<th>$V$</th>
<th>$D_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>0.3147</td>
<td>0.8182</td>
<td>0.301</td>
<td>5</td>
<td>0.935</td>
<td>2.492</td>
<td>1.619</td>
<td>2.077</td>
</tr>
<tr>
<td>$P_2$</td>
<td>0.2394</td>
<td>0.6224</td>
<td>0.468</td>
<td>5.083</td>
<td>2.492</td>
<td>1.619</td>
<td>2.077</td>
<td></td>
</tr>
<tr>
<td>$P_3$</td>
<td>0.2794</td>
<td>0.7264</td>
<td>0.379</td>
<td>2</td>
<td>0.687</td>
<td>1.160</td>
<td>0.689</td>
<td>1.526</td>
</tr>
<tr>
<td>$P_4$</td>
<td>0.1845</td>
<td>0.4797</td>
<td>0.590</td>
<td>5</td>
<td>1.128</td>
<td>3.571</td>
<td>2.297</td>
<td>2.506</td>
</tr>
<tr>
<td>$P_5$</td>
<td>0.1738</td>
<td>0.4519</td>
<td>0.614</td>
<td>10</td>
<td>1.306</td>
<td>5.311</td>
<td>3.605</td>
<td>2.901</td>
</tr>
<tr>
<td>$P_6$</td>
<td>0.1731</td>
<td>0.4501</td>
<td>0.615</td>
<td>20</td>
<td>1.331</td>
<td>5.718</td>
<td>3.945</td>
<td>2.958</td>
</tr>
<tr>
<td>$P_7$</td>
<td>0.3210</td>
<td>0.5136</td>
<td>0.288</td>
<td>2</td>
<td>0.642</td>
<td>1.024</td>
<td>0.612</td>
<td>1.424</td>
</tr>
<tr>
<td>$P_8$</td>
<td>0.2822</td>
<td>0.4515</td>
<td>0.373</td>
<td>5</td>
<td>0.796</td>
<td>1.784</td>
<td>1.450</td>
<td>1.769</td>
</tr>
<tr>
<td>$P_9$</td>
<td>0.2813</td>
<td>0.4508</td>
<td>0.375</td>
<td>10</td>
<td>0.808</td>
<td>1.882</td>
<td>1.229</td>
<td>1.795</td>
</tr>
</tbody>
</table>

### Table V. Queueing behaviour of buffer 2 under some routing policies with $r_{out} = 0.45$. The mean delay under the Bernoulli model is 3.25

<table>
<thead>
<tr>
<th>Policy</th>
<th>$\gamma$</th>
<th>$r_{in}$</th>
<th>$\Theta$</th>
<th>$Q_1$</th>
<th>$Q_2$</th>
<th>$V$</th>
<th>$D_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>0.3147</td>
<td>0.8182</td>
<td>0.301</td>
<td>5</td>
<td>4.081</td>
<td>27.751</td>
<td>15.178</td>
</tr>
<tr>
<td>$P_2$</td>
<td>0.2394</td>
<td>0.6224</td>
<td>0.468</td>
<td>5</td>
<td>4.992</td>
<td>45.031</td>
<td>25.102</td>
</tr>
<tr>
<td>$P_3$</td>
<td>0.2794</td>
<td>0.7264</td>
<td>0.379</td>
<td>2</td>
<td>4.247</td>
<td>30.810</td>
<td>17.039</td>
</tr>
<tr>
<td>$P_4$</td>
<td>0.1845</td>
<td>0.4797</td>
<td>0.590</td>
<td>5</td>
<td>6.161</td>
<td>73.569</td>
<td>41.771</td>
</tr>
<tr>
<td>$P_5$</td>
<td>0.1738</td>
<td>0.4519</td>
<td>0.614</td>
<td>10</td>
<td>7.223</td>
<td>108.527</td>
<td>63.981</td>
</tr>
<tr>
<td>$P_6$</td>
<td>0.1731</td>
<td>0.4501</td>
<td>0.615</td>
<td>20</td>
<td>7.395</td>
<td>116.005</td>
<td>68.715</td>
</tr>
<tr>
<td>$P_7$</td>
<td>0.3210</td>
<td>0.5136</td>
<td>0.288</td>
<td>2</td>
<td>4.126</td>
<td>28.924</td>
<td>16.028</td>
</tr>
<tr>
<td>$P_8$</td>
<td>0.2822</td>
<td>0.4515</td>
<td>0.373</td>
<td>5</td>
<td>4.959</td>
<td>45.720</td>
<td>27.081</td>
</tr>
<tr>
<td>$P_9$</td>
<td>0.2813</td>
<td>0.4508</td>
<td>0.375</td>
<td>10</td>
<td>5.041</td>
<td>47.912</td>
<td>27.535</td>
</tr>
</tbody>
</table>

### 5. Conclusions

As was mentioned in the summary, the objective of this paper is not the development of an optimal routing policy from node 1 to node 3 (Figure 3), given, for example, the packet input rates $\lambda_3$ and $\lambda_5$. The latter has been the focal point of significant research, where the optimal routing policies are usually based on the assumption of memoryless...
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 routing 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 routing policy, as an MMGBP, the operation of the routing 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 routing 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 routing policies result in different queueing behaviours of node 2, under fixed packet traffics. The latter is the result implies that the probabilistic behaviour of the packet (output) process generated by a routing policy should also be considered in the quest for the optimal routing policies and not only the intensity of that process. Numerical results show that the best packet routing policy (leading to the minimum intensity of the queueing problems at node 2, as reflected by the queueing quantities $A_2$), 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 $\gamma$ 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 routing policy which would continually 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


Author’s biography:

Ioannis Stavvakakis (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. Stavvakakis is a member of Tau Beta Pi and the Technical Chamber of Greece.