INTERNATIONAL JOURNAL OF DIGITAL AND ANALOG COMMUNICATION SYSTEMS, VOL. 5, 157–165 (1992)

TRAFFIC MODELLING FOR PACKET COMMUNICATION NETWORKS

IOANNIS STAVRAKAKIS

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

SUMMARY

A general model for the description of traffic processes in packet communication networks is presented in this paper. The packet traffic generated by a network component is described in terms of a generalized Bernoulli process whose intensity depends on the state of the component. When the state space of the component is large, an approximate traffic characterization is proposed based on the dominant states. The latter characterization is shown to be well performing for the considered examples of practical network components. Based on the proposed general model, an analysis approach for multi-component communication networks is presented. Numerical results illustrate the inaccuracy of the widely adopted i.i.d. modelling, the effectiveness of the proposed general model and the significant complexity reduction through the adoption of the approximate one.

KEY WORDS Traffic modelling Buffer behaviour

1. INTRODUCTION

Packet communication networks have been widely adopted as an efficient means of transferring information. Such networks extend from small local area networks to large systems of interconnected networks, covering extended geographical areas.[1] The development of local area (or single hop) networks has been the focal point of extended research over the last two decades.[1] The performance evaluation of these networks has been facilitated by the simplicity of the models adopted for the description of the packet generation mechanism. The Bernoulli process has been widely considered as a model for the per user packet generation process over a fixed time interval (slot), in a system with a small (finite) user population. In the case of a large (infinite) user population system, the Poisson model has been adopted for the modelling of the cumulative packet generation process. The memoryless property of these models captures effectively the randomness and unpredictability of the packet generating mechanism of the users.

In a large packet communication network the packet arrival processes to a network component may be not only the direct packet traffic, as generated by the users, but also packet traffic generated by other network components. In this paper, a network component is defined as a network subsystem which generates packets. Such components can be a local area network, a network switch or a repeater, a statistical multiplexer, a link that carries traffic from another large network, or any line carrying packetized information of virtually any kind. We define the first stage component (FSC) to be a network component whose packet input process is determined by simple first-born traffic. The simple first-born traffic is defined as a traffic which is not modulated by any network component. This type of traffic is assumed to be satisfactorily modelled by an independent process.

In a large packet communication network (defined here as a network with more than one component), there is usually a system which serves as the backbone network required for the interconnection of the involved components. Clearly, (some of) the packet processes in this system are the output processes from the supported components. We define the second stage component (SSC) to be a system whose input process is (at least partially) determined by the output processes of other network components. Thus, all network components which are not FSCs are SSCs.

The analysis of FSCs has received significant attention in the past, resulting in the development of many analytical techniques. The simplicity of the simple first-born traffic models has played an important role in the development of these techniques. The analytical tractability provided by Bernoulli/Poisson modelling for the simple first-born traffic has tempted many researchers to adopt these models for the description of the input processes to SSCs as well.[2–6] There are two major problems regarding the adoption of a memoryless model for the input process to an SSC. The first problem arises from the intuitively unpleasing nature of such a model. Although the input process to a FSC may be considered as memoryless, the FSC introduces dependencies in its output process. The second problem is related to the inaccuracy of the performance evaluation of a SSC, under a memoryless model for its input process. This inaccur-

1047–9627/92/030157–09$09.50

Received 20 April 1991
Revised 25 March 1992

acy may lead to erroneous identification of the bottlenecks of the network, to erroneous delay calculations and finally to inaccurate overall network performance evaluation.

The accurate description of a packet process generated by a network component is essential to the performance evaluation of the network. In a large network, the input processes to an SSC may be the output processes of either FSCs or SSCs or a combination of the two. For this reason, the output processes from both types of network components are described in a unified manner. Although it is reasonable to expect that the developed model for the output process of a network component be more tractable in the case of FSCs (due to the simple first-born input traffic), the general model description applies to both cases.

In this paper, an alternative to the i.i.d. (independent and identically distributed) characterization of the packet (output) processes generated by the components of a communication network, is presented. The proposed model is exact and it is presented in the next section. The output process is described as a generalized Bernoulli process whose intensity depends on the state of an underlying Markov chain describing the operation of the component. The cardinality of the state space of this Markov chain may be almost arbitrarily large. To provide for a numerically tractable solution in the later case, the exact model is appropriately modified. For this purpose, a meaningful approximate model based on the output process, based on the dominant states of the underlying Markov chain, is introduced. The approximate model is described by introducing the concept of the $\alpha\%$-exact model.

In section 3, some practical packet communication network components are presented. Their output processes are described by incorporating the proposed models (exact and $\alpha\%$-exact). These examples illustrate the applicability of the general models for the description of the packet processes generated by network components.

A packet communication network which consists of components such as those described in Section 3 is considered in Section 4 and its performance is evaluated. Numerical results are presented in Section 4. The derived numerical results illustrate both the inaccuracy of the i.i.d. models and the effectiveness of the proposed models in generating fairly accurate results with modest complexity. Finally the summary and conclusions of this work are presented in the last section.

2. MODELS FOR THE (DEPENDENT) OUTPUT PROCESSES OF NETWORK COMPONENTS

2.1. *The exact model for the packet generation (output) process*

The discussion in this paper is confined to discrete time network components, as formulated by slotted packet communication systems. The description of the dependent packet process generated by a network component (FSC or SSC) is essential to the analysis of a SSC since the input processes to the latter are determined by the packet processes generated by the feeding network components. The proposed exact model on the output (or packet departure) process of a network component is described through the following definitions. The discussion in this paper is confined to discrete time network components, as formulated by slotted packet communication systems.

Definition 1. The packet generation (output) process of a network component is defined to be the discrete time process of the departing packets, $\{a_j\}_{j\geq 0}$; $a_j = \rho$, $0\leq\rho<\infty$, if ρ packets leave the component at the jth time instant.

Definition 2. Assume that the network component satisfies the following:

(a) There exists an ergodic Markov chain $\{z_j\}_{j\geq 0}$ associated with the description of the state of the component; let $S = \{x_1,x_2,\ldots,x_M\}$, $M<\infty$, be the state space of $\{z_j\}_{j\geq 0}$ and $p(x_k,x_j)$, $\pi(x_k)$, $x_k,x_j\in S$, be the corresponding state transition and steady state probabilities.

(b) There exists a stationary probabilistic mapping $a(z_j)$: $S\rightarrow Z_0$ (where Z_0 is the set of non-negative finite integers), which describes the number of packets departing at the end of the jth time interval (slot). Let $a(z_j) = \rho$, $0\leq\pi\leq\infty$, $z_j\in S$, with probability $\phi_\rho(z_j)$.

Then, the output process of the component is given by

$$\{a_j\}_{j\geq 0} = \{a_j(z_j)\}_{j\geq 0} \tag{1}$$

i.e. it is described as a Markov modulated generalized Bernoulli process. It can also be seen as a random reward process associated with a state transition of a Markov chain.[7] A two-state Markov modulated Poisson process has been adopted in Reference 9 for the approximate characterization of the superimposition of voice and data traffic.

Notice that the process $\{a_j\}_{j\geq 0}$, as given by (1), describes exactly the output process of a network component, provided that the conditions in Definition 2 are satisfied.

2.2. *The approximate model for the packet generation (output) process*

Let $\{z_j\}_{j\geq 0}$ and $\{a_j(z_j)\}_{j\geq 0}$ be two processes, as described in Definition 2, associated with a certain network component. Let $\{z'_j\}_{j\geq 0}$ be a new Markov chain constructed from $\{z_j\}_{j\geq 0}$ in the following way: its state space $S' = \{x'_1, \ldots, x'_{M'}\}$ for some $M'<M$, where $x'_1, \ldots, x'_{M'-1}$ are the $M'-1$ dominant states

and $x'_{M'}$ is the union of the remaining states of the original Markov chain $\{z_j\}_{j\geq 0}$. Dominant states are the states with the largest probability mass. That is, the reduced state space Markov chain $\{z'_j\}_{j\geq 0}$ is formulated by merging all the probabilistically insignificant states of the original Markov chain into a single state $x'_{M'}$. The parameters (probabilities) of the new Markov chain $\{z'_j\}_{j\geq 0}$ are identical to those of the corresponding original processes except from those associated with the new state $x'_{M'}$. In the latter case, appropriate averaging is incorporated (see the Appendix).

Let $\{a'_j\}_{j\geq 0}$ be a process similar to $\{a_j\}_{j\geq 0}$ with a corresponding underlying Markov chain $\{z'_j\}_{j\geq 0}$. That is

$$\{a'_j\}_{j\geq 0} = \{a'_j(z'_j)\}_{j\geq 0}$$

where $a'(z'_j): S' \rightarrow Z_0$ is a probabilistic mapping; $a'(z'_j) = \rho$, $0 \leq \rho < \infty$, $z'_j \in S'$, with probabilities $\phi'_\rho(x'_k)$, $x'_k \in S'$ which are identical to the corresponding ones of the original Markov chain when the dominant states are involved, and they are given by appropriate averaging when the new state $x'_{M'}$ is involved (see the Appendix).

Clearly, the process $\{a'_j\}_{j\geq 0}$ is an approximation on the true packet output process of the network component described by $\{a_j\}_{j\geq 0}$. The approximation is introduced through the reduction of the state space of the original underlying Markov process $\{z_j\}_{j\geq 0}$, which describes completely the operation of the component. In view of the construction of the approximate process $\{a'_j\}$, it is reasonable to expect that as M' increases, $\{a'_j\}_{j\geq 0}$ approaches $\{a_j\}_{j\geq 0}$; that is, the approximate model approaches the exact one. When $M' = M$, the approximate model coincides with the exact one. It is also reasonable to expect that the larger the total probability mass of the (unchanged) $M'-1$ states the better the approximation achieved by the reduced state space process. In view of the previous observations, the following definition-measure of the approximation on the exact process $\{a_j\}_{j\geq 0}$ may be defined.

Definition 3. The process $\{a'_j\}_{j\geq 0}$ defined above is called a%-exact if the cardinality M' of the corresponding approximate underlying Markov $\{z'_j\}_{j\geq 0}$ (defined above) is such that

$$\sum_{1\leq k\leq M'-1} \pi'(x'_k) \geq \alpha \quad \text{and}$$

$$\sum_{1\leq k\leq M'-2} \pi'(x'_k) < \alpha$$

That is, the total probability mass of the unchanged states of the original Markov chain $\{z_j\}_{j\geq 0}$ is at least α and the total probability mass of the $M'-2$ dominant states is less than α.

Definition 3 may be roughly interpreted in the following way, under ergodicity of all processes involved: an α%-exact output process $\{a'_j\}_{j\geq 0}$ is based on the true packet generating mechanism (state of the system) α% of the time; (100-α)% of the time, the output process is based on an average packet generating mechanism. An average packet generating mechanism is the only one assumed present when $\{a_j\}_{j\geq 0}$ is approximated by a (generalized) Bernoulli process. The latter case corresponds to merging all states of the true underlying Markov chain into a single one ($M' = 1$). An average number of outputs is generated under the latter model throughout the time horizon, independently of the true state of the underlying packet generating mechanism. In view of Definition 3, the exact model on the output process $\{a_j\}_{j\geq 0}$ corresponds to $\alpha = 100$. 100-α may also be seen as a measure of the smoothing on the output process introduced by the merging of the states of the true packet generating mechanism.

From the definition of the α%-exact process and the construction of $\{z'_j\}_{j\geq 0}$ turns out that the cardinality of the state space of $\{z'_j\}_{j\geq 0}$ (i.e. M') increases with α. Thus, it is reasonable to expect that the larger the value of α the better the approximation on the packet output process. Assuming that this is generally true, there is a trade-off between the degree of the accuracy of the approximating process and the introduced complexity in its description, as measured by M'. The real concern at this point is not about the complexity in the description of the approximating (or exact) packet output process itself, but it is on the tractability of the analysis of other components of a large network, whose input processes are described by the proposed models.

In many packet communication components the state of the underlying packet generating mechanism (Markov chain) may be defined to be the number of packets currently in the component. Such components may be network nodes, where centralized queues are formulated (the length at the queue may define the state of the component in this case), or multi user (random access) communication networks, where distributed queues are formulated (the number of blocked users may define the state, in this case). In such network components, the transitions of the underlying Markov chains are generally gradual and large jumps have very small probability of occurrence. Also, given an operation point of the network component, there is a state (e.g. number of packets in the component) which dominates probabilistically. Originating from this state, a monotone decrease of the steady-state probabilities is usually observed, by moving toward smaller or larger states. As a result, there are two (asymmetric, in general) tails of the steady state probability distribution. Construction of the process $\{z'_j\}_{j\geq 0}$ as outlined in (2) could lead to the merging of probabilistically insignificant states from both probability tails.

For network components with behaviour as described in the previous paragraph, a slightly different construction procedure for the reduced state space underlying Markov chain $\{z'_j\}_{j\geq 0}$ may be followed. The objective in the new construction approach is to distinguish between the two tails of the steady state probability, whenever states from both tails are included in the last, averaging, state x'_M. By making the reasonable assumptions that (1) neighbouring states determine similar packet generation (output) mechanisms, (2) there are only a few dominant states among those in the probability tail and these states are neighbours, and (3) states in different probability tails determine quite different packet generation (output) mechanisms, it can be seen that excessive smoothing of the output process introduced by the approximating process $\{a'_j\}_{j\geq 0}$ (constructed as in (2)), could be avoided by creating two merging states; each of these merging states contains probabilistically insignificant states of the original Markov chain coming from the same probability tail. The parameters of the new Markov chain and the probability mappings, which are associate with the new merging states, are obtained through appropriate averaging (Appendix).

3. THE OUTPUT PROCESS OF SOME NETWORK COMPONENTS

In this section, the output process of some simple network components is described by incorporating the models developed before. In the next section, the performance of a large packet communication network which consists of such network components is evaluated.

Example 1: *bursty traffic network links*

Consider a link which carries traffic modulated by various other components of a large network and by routeing decisions. The network component in this case is the link and its input and output processes are identical. In Reference 8 it has been found that network packet traffic is bursty. As a result, a first-order Markov model could be adopted in the description of this packet process. If $p(0,1)$ and $p(1,1)$ are the conditional probabilities of a packet arrival (departure) given that 0 or 1 arrivals (departures) occurred in the previous slot, respectively, then the burstiness coefficient is defined[8] as

$$\gamma = p(1,1) - p(0,1)$$

This traffic model can be easily described in terms of the proposed model on the packet process generated by the link (network component). Let z_j be the number of packets in the link at the beginning of the jth slot. Clearly, $\{z_j\}_{j\geq 0}$ is a Markov chain with $S = \{0,1\}$. The transition probabilities are identical to those of the first-order Markov model that describes the bursty traffic. The mapping $a(z_j)$ is, in this case, deterministic and it is given by

$$\phi_1(0) = 0, \phi_1(1) = 1, \phi_0(0) = 1, \phi_0(1) = 0$$

The packet (output) process of the link just described can be the model for an on/off switch with Bernoulli packet arrivals. In this case, when the switch is on the output process is the same Bernoulli process and it is zero otherwise. Markov models can usually be incorporated in the description of the on–off activity of a switch. A switch in the off position could correspond to a failure or to a situation in which it serves other links.

The parameters of the packet output process of a network component which generates bursty traffic are determined from the packet rate $\pi(1)$ and the burstiness coefficient γ. Given these quantities, the rest of the parameters of the Markov model are calculated from the equations

$$\pi(0) = 1 - \pi(1), p(0,0) = 1 - p(0,1),$$
$$p(1,0) = 1 - p(1,1)\, p(0,1) = \pi(1)(1-\gamma),$$
$$p(1,1) = \gamma + p(0,1)$$

Notice that the packet traffic considered in this example describes blocks of packets of geometrically destributed length followed by idle periods of geometrically distributed length.

Example 2: *the single message node*

Consider now a network node which is capable of storing and forwarding a single message at a time. It is assumed that the input process to this component is Bernoulli with intensity μ messages per slot. Each message is assumed to consist of a variable number of packets let $\sigma(i)$ = Pr{message consists of i packets}, $1 \leq i \leq K$. The single message buffering assumption implies that messages which find the component non-empty are either discarded or served by a (buffered) low priority link. In the second scenario, the link served by the node under consideration is reserved for new messages. These messages are given a chance to be transmitted right away (if the line is not busy), before they enter a (probably) first-in–first-out queue formulated in the input of another link. Without loss of generality, it is assumed that a new message is also accepted if there is only one packet (the last of the previous message) in the node. It is assumed that arrivals occur at the beginning of a slot. As a result, a new message may start being served right after the end of the previous message transmission. The packet output process of this component is definitely a non-Bernoulli process. It can be easily described in terms of the processes $\{z_j\}_{j\geq 0}$ and $\{a_j\}_{j\geq 0}$ defined in the previous section. If z_j is the number of packets in the node at the end of the jth slot, then it can be easily shown that $\{z_j\}_{j\geq 0}$ is a Markov chain with state space $S = \{0,1,2, \ldots, K\}$. The transition probabilities are given by

$$p(0,i) = p(1,i) = \mu\sigma(i), 1 \le i \le K$$
$$p(0,0) = p(1,0) = 1 - \mu$$
$$p(k,k-1) = 1, 2 \le k \le K$$
$$p(k,i) = 0, 2 \le k \le K, 1 \le i \le K, i = k-1$$

Given μ and $\sigma(i)$, $1 \le i \le K$, the steady state probabilities $\pi(i)$, $0 \le i \le K$ can be easily computed. The mapping given by (1) is deterministic in this case and has the following parameters:

$$\phi_1(k) = 1, 1 \le k \le K, \phi_1(0) = 0$$
$$\phi_0(k,i) = 1 - \phi_1(k,i), 0 \le k \le K$$

A Bernoulli approximate model on the output process of the node would have intensity

$$\lambda = \sum_{i=1}^{k} \pi(i) = 1 - \pi(0)$$

A better approximate model on the true packet output process could be a first-order Markov model. If 1 and 0 denote one or zero packet outputs, respectively, then the parameters of this Markov model are given by

$$\pi_m(0) = \pi(0), \pi_m(1) = 1 - \pi_m(0)$$
$$p_m(0,0) = 1 - p_m(0,1), p_m(1,0) = 1 - p_m(1,1)$$
$$p_m(0,1) = \mu, p_m(1,1) = 1 - p_m(0,1)\frac{\pi_m(0)}{\pi_m(1)}$$

and the corresponding burstiness coefficient γ is given by

$$\gamma = p_m(1,1) - p_m(0,1)$$

Notice that the packet traffic considered in this example describes blocks of packets of generally distributed length followed by idle periods of geometrically distributed length.

Example 3: *a node with arbitrarily large buffer*

In this case it is assumed that all messages which would fit into the buffer of size $M < \infty$ are received; no message is partially received. Let $g(k)$, $0 \le k \le K$, be the probability that a message with k packets arrives over a slot; $k=0$ corresponds to no message arrival.

Similarly to the previous example, the output process of the finite buffer node can be easily described in terms of the processes $\{z_j\}_{j\ge 0}$ and $\{a_j\}_{j\ge 0}$. If z_j is the number of packets in the node at the end of the jth slot, then $\{z_j\}_{j\ge 0}$ is a Markov chain with state space $S = \{0,1,2, \ldots, M\}$. The transition probabilities are given by (assume $g(k) = 0$ for $k > K$)

$$p(0,j) = g(J), 0 \le j \le M$$
$$p(k,j) = g(j-k+1), 1 \le k \le M, k-1 \le j \le M$$

and the probabilistic mapping is determined by

$$\phi_1(k) = 1, 1 \le k \le M, \phi_1(0) = 0$$
$$\phi_0(k) = 1 - \phi_1(k), 0 \le k \le M$$

The Bernoulli and the Markov approximations on the resulting packet output traffic can be determined as in the previous example.

Notice that the packet traffic considered in this example describes blocks of packets of generally distributed length followed by idle periods of geometrically distributed length.

4. PERFORMANCE ANALYSIS OF A LARGE PACKET COMMUNICATION NETWORK

In this section it is illustrated how the proposed models for the output traffic of a network component can be effectively incorporated in the analysis of a multi-component network. The focus is on the performance evaluation of SSCs, which requires a complete description of their input processes and an appropriate analysis technique. The star topology shown in Figure 1 is considered as an example of a multi-component network, and the mean packet delay in C_0 is calculated. The SSC of this network is a common queueing system which has been analysed under various input processes; the single server is assumed to have a buffer of infinite capacity. The arrival processes $\{a_j^i\}_{j\ge 0}$, $i = 1,2, \ldots, N$, are assumed to be synchronized discrete time processes; at most one arrival can occur in each input line per unit time. The time first-in–first-out (FIFO) policy is adopted and the service time is assumed to be constant and equal to one slot. Arrivals that occur at the same time instant are served in a randomly chosen order.

When each of the output process of the components C_1, C_2, ..., C_N is a first-order Markov process, the mean packet delay in C_0 is given by[8]

$$D_M = 1 + \frac{\sum_{i=1}^{N}\sum_{j>i}^{N} \lambda_i\lambda_j\left(1 + \frac{\gamma_i}{1-\gamma_i} + \frac{\gamma_i}{1-\gamma_j}\right)}{\sum_{i=1}^{N} \lambda_i\left(1 - \sum_{j=1}^{N} \lambda_j\right)} \quad (2)$$

where λ_i, $i = 1,2, \ldots, N$ is the intensity of the output process of component i, and γ_i, $i = 1,2, \ldots, N$ is the corresponding burstiness coefficient. The mean packet delay result under Bernoulli packet arrival processes is also given by (2) by setting $\gamma=0$.

When the output processes of the components $C_1, C_2, \ldots, C_N$ are given by (1) (see Section 2), then the mean packet delay in C_0 is obtained from the solution of $M^1 \times M^2 \times \ldots \times M^N$ (M^i is the cardinality of the underlying Markov chain of the

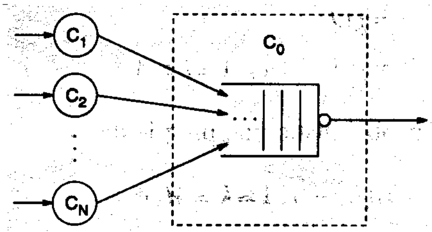


Figure 1. A star topology of a $N + 1$ component network

ith component) linear equations (or by a smaller number of equations, in symmetric cases), provided that the queue stability condition

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

is satisfied; $\bar{x} \in \bar{S} = S^1 \times S^2 \times \ldots \times S^N$ (sup i denotes a quantity associated with the ith component), and $\mu_{\bar{x}}$ is given below. These equations are described by $M^1 \times M^2 \times \ldots \times M^N - 1$ equations (with respect to $W(\bar{x})$, $\bar{x} \in \bar{S}$) from the following set:

$$W(\bar{y}) = \sum_{\bar{x} \in \bar{S}} W(\bar{x}) p(\bar{x},\bar{y}) + \sum_{\bar{x} \in \bar{S}} (\mu_{\bar{x}} - 1) p(\bar{x},\bar{y}) \pi(\bar{x}) + \sum_{\bar{x} \in \bar{S}} p(0;\bar{x}) p(\bar{x},\bar{y}), \bar{y} \in \bar{S} \quad (3a)$$

and the equation

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

(see Reference 10), where

$$\pi(\bar{x}) = \prod_{i=1}^{N} \pi^i(x^i), p(\bar{x},\bar{y}) = \prod_{i=1}^{N} p^i(x^i,y^i)$$

$$\mu_{\bar{x}} = \sum_{v} v g_{\bar{x}}(v), \sigma_{\bar{x}} = \sum_{v} v^2 g_{\bar{x}}(v)$$

$g_{\bar{x}}(v) = \text{pr}\{\text{state } \bar{x} \text{ results in } v \text{ packet arrivals}\}$ and $p(0;\bar{x})$ is the joint probability that the buffer is empty and the underlying Markov chain is in state $\bar{x}$. For the examples considered in this paper, it is only state 0 that results in no packet output and, thus[10] $p(0;\bar{x}) = (1-\lambda)p(\bar{0},\bar{x})$.

The average number of packets in C_0, Q, is given by the sum of the solutions of equations (3). The average time, D, that a packet spends in C_0 can be obtained as the ratio Q/λ, by using Little's formula.

In the sequel, some results are derived for the mean packet delay induced in C_0 for various cases of network components $C_1, \ldots, C_N$ and under exact or approximate descriptions of their packet output processes (Figure 1). These results lead to some conclusions regarding some commonly made over-simplifying assumptions on the packet processes generated by network components. $N=3$ network components (other than C_0) are considered in the following cases.

Case 1

Let C_1 be the network component described in Example 1 with parameters $\pi(1) = 0{\cdot}1$ and $\gamma = 0{\cdot}3$. Let C_2 be the network component described in Example 2 with parameters $K = 5$, $\sigma(0) = 0$, $\sigma(1) = 0{\cdot}1$, $\sigma(2) = 0{\cdot}3$, $\sigma(3) = 0{\cdot}3$, $\sigma(4) = 0{\cdot}2$, $\sigma(5) = 0{\cdot}1$ and $\mu = 0{\cdot}1$, which result in a packet output rate of 0·244 packets per slot. Finally, let C_3 be the network component described in Example 3 with parameters $M = 50$ (buffer size), $r = g(0) = 0{\cdot}8$ (probability of no message arrival in a slot), $g(1) = 0{\cdot}1(1-r)$, $g(2) = 0{\cdot}3(1-r)$, $g(3) = 0{\cdot}3(1-r)$, $g(4) = 0{\cdot}2(1-r)$ and $g(5) = 0{\cdot}1(1-r)$, where $g(k)$, $k=1, \ldots, 5$, is the probability that a message consists of k packets. The output processes of components C_1 and C_2 are exactly described as mentioned in Section 3. For the description of the output process of component C_3, both the exact (which involves an underlying Markov chain with a state space of cardinality 51) and the α%-exact approximate (which involves an underlying Markov chain of cardinality less than 51) models, are adopted (see Section II).

For the network described above, the mean packet delay induced by the SSC C_0 is shown in Table I, for various degrees of approximation α (the range of the corresponding dominant states is also shown) of the output process of C_3. The exact results, obtained by incorporating the exact models on the input processes to C_0, are also shown. It can be easily seen that, e.g., the mean delays induced in C_0 under the 97%-exact model (resulting in a Markov chain of cardinality 10) are very close to the accurate ones, which are obtained by solving five times the number of linear equations. This observation indicates that significant reduction in the numerical complexity of the problem can be achieved by incorporating an approximate model based on the dominant states of a component, at

Table I. Delay results for Case 1

Model	Delay
100% (0–50)	14·655
Bernoulli	4·174
Markov	13·427
85% (0–5)	13·720
90% (0–6)	13·959
95% (0–8)	14·295
97% (0–9)	14·388
99% (0–12)	14·515

the expense of a rather insignificant deviation from the accurate results. Other possible approximations on the true output process, which also simplify the analysis, are the Bernoulli and the (first-order) Markov ones. Results under these approximations are also shown in Table I. As can be easily concluded, the Bernoulli model completely fails to approximate the accurate result, whereas the Markov model is clearly inferior to the α%-exact model for sufficiently large α. Some more conclusions about the performance of the α%-exact model are drawn in the next case.

Case 2

Let C_3 be as in Case 1 and C_1 and C_2 be network components as described in Example 1. Let $\pi(1) = 0{\cdot}17$ and $\gamma = 0{\cdot}3$ for each of the components C_1 and C_2. The delay results in C_0 under various models on the output processes of C_1, C_2 and C_3 are given in Table II, for $\gamma = 0{\cdot}3$, and in Table III for $\gamma = 0{\cdot}0$; the latter value of γ corresponds to Bernoulli traffic.

For a very small value of α only the most significant state will be considered, while the rest of them will be merged into a single state. This situation is reflected for $\alpha = 1$ where the state 0 (the most significant one) is selected. Let 1′ denote the new state generated by merging all other states 1–50. Since, for this particular component, state 0 results in no packet generation and state 1′ (or states 1–50) always results in one packet generation, the 1%-exact model coincides with the first order Markov approximation! Thus, the delay results for both models are the same (10·604). Although one would expect monotone improvement of the accuracy of the obtained delay results as the number of states merged into a single state decreases, this turns out not to be the case. Although the approximation on the output process itself is refined, the effect of this on the improvement of the accuracy of the delay result is not well determined, due to the complexity of the queueing process and the significance of the probability mass which is concentrated in the new compound state. As this probability mass decreases, the delay results improve in a monotone fashion. In the example considered here, the α%-exact model give monotonically improving delay results (which are better than those under the Markov model) when the compound state concentrates less than 20% of the probability mass.

The plot of the induced mean delay D in C_0 for Case 2 (for γ=0·3) is presented in Figure 2 as a function of α, to illustrate better the obtained accuracy as the value of α increases. The previous results are also plotted in Figure 3 as a function of the model complexity C. The model complexity C is defined as the ratio of the cardinality of the reduced state space Markov chain and the original one; that is, $C = M'/M$. From Figure 3 it is clearly concluded that very accurate results can be obtained by adopting a model whose complexity is only a small portion of that of the exact one.

Table II. Delay results for Case 2

Model	Delay
100% (0–50)	12·884
Bernoulli	4·060
Markov	10·604
1% (0–0)	10·604
50% (0–1)	10·171
80% (0–4)	11·595
90% (0–6)	12·387
95% (0–8)	12·705
97% (0–9)	12·793
98% (0–10)	12·850

Table III. Delay results for Case 2

Model	Delay
100% (0–50)	11·404
Bernoulli	4·060
Markov	9·140
1% (0–0)	9·140
50% (0–1)	8·689
80% (0–4)	10·113
90% (0–6)	10·905
95% (0–8)	11·224
97% (0–9)	11·311
98% (0–10)	11·368

5. SUMMARY AND CONCLUSIONS

In this paper, an alternative to the (commonly adopted) independent and identically distributed (i.i.d.) characterization of the packet (output) processes generated by various components of a packet communication network, has been presented. The proposed characterization uses an underlying

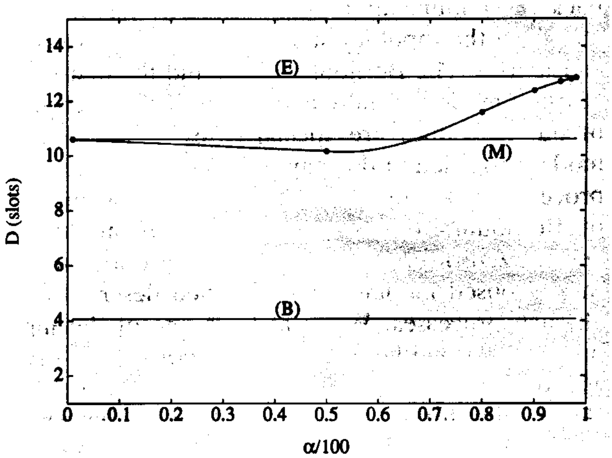


Figure 2. Delay results for case 2 (γ = 0·3) as a function of α under the Bernoulli (B), Markov (M), the exact (E) and the α%-exact models

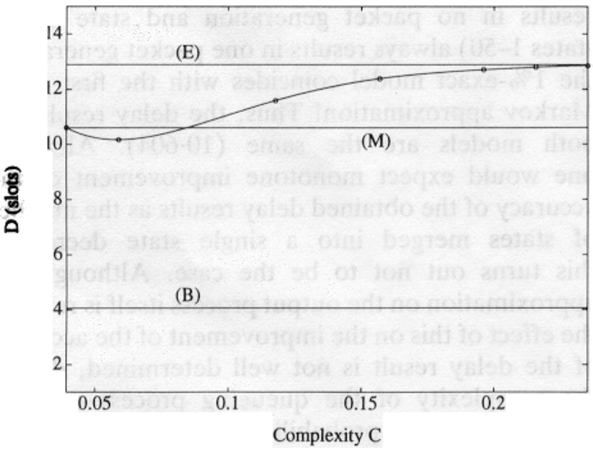


Figure 3. Delay results for case 2 ($\gamma = 0{\cdot}3$) as a function of the model complexity C under the Bernoulli (B), Markov (M), the exact (E) and the α%-exact models

Markov process which describes the behaviour of the component. For many practical components, this Markov process can be identified as the one developed for their study. Based on the state of the component (underlying process), an exact description of its packet output process has been developed based on an appropriate i.i.d. model associated with each state of the component, as opposed to a single one independently of the state (i.i.d. model).

When the underlying Markov chain has a large (or infinite) state space, there are usually only few states which dominate probabilistically, under stability. For such cases, a meaningful aproximate model has been introduced based on the concept of the dominant states of the component, to improve the numerical tractability of the original model. The approximate results obtained under the latter modelling can be almost arbitrarily close to the exact one, for most of the cases, at the expense of increased computational effort.

In the sequel, it has been shown how the developed models for the output process of a network component can be effectively incorporated in the performance evaluation of a multi-component packet communication network. Some cases associated with the topology shown in Figure 1 have been considered. The Bernoulli model and the first-order Markov model (usually adopted for the description of bursty traffic) are special cases of the general model proposed in this paper. Thus, when the input process to a component consists of a mixture of Bernoulli, Markovian and Markov modulated generalized Bernoulli processes (as described here), the proposed models offer a unified description of all such processes. When at least one of the input processes is neither a Bernoulli nor a first-order Markov process, the analysis results based on Bernoulli or first-order Markov modelling have been found to be fairly inaccurate.

Finally it should be noted that the basic ideas behind the proposed analytic models can be adopted for the reduction of the complexity of accurate simulators of large networking structures. By selecting a level of approximation α, significant savings in memory and computation time may be achieved, at the expense of a controlled (through the value of α) deviation from the accurate results.

ACKNOWLEDGEMENT

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

APPENDIX

We give the averaging expressions for the calculation of the parameters of the approximate model associated with the merging state of the reduced space Markov chain:

$$\pi'(x'_{M'}) = \sum_{x_i \in S_{M'}} \pi(x_i) = 1 - \sum_{i=1}^{M'-1} \pi'(x'_i)$$

If (x'_k, x'_j) denotes a transition from x'_k to x'_j and $S_{M'}$ denotes the set of all merged states, then

$$p'(x'_k, x'_{M'}) = \sum_{j \in S_{M'}} p(x'_k, x_j), \; 1 \le k \le M'-1$$

$$p'(x'_{M'}, x'_k) = \frac{p\left(\bigcup_{x_i \in S_{M'}} x_i, x'_k\right)}{p\left(\bigcup_{x_i \in S_{M'}} x_i\right)} = \frac{\sum_{x_i \in S_{M'}} p(x_i, x'_k)\pi(x_i)}{\pi'(x_{M'})}, \; 1 \le k \le M'-$$

$$p'(x'_{M'}, x'_{M'}) = 1 - \sum_{k=1}^{M'-1} p'(x'_{M'}, x'_k)$$

$$\phi'_\rho(x'_{M'}) = \frac{\sum_{x_i \in S_{M'}} \phi_\rho(x_i)\pi(x_i)}{\pi(x'_{M'})}$$

When two merging states, $x'_{M'-1}, x'_{M'}$, are constructed the parameters associated with these new states are given by the following averaging expressions.

Let

$$x'_{M'-1} = \bigcup_{i=1}^{l-1} x_i \quad \text{and} \quad x'_{M'} = \bigcup_{i=u+1}^{M} x_i$$

and

$$a = M'-1, M'a, b = M'-1, M', w_1(M') = u+1,$$
$$w_2(M') = M, w_1(M'-1) = 1, w_2(M'-1) = l-1.$$

Then

$$\pi'(x'_{M'-1}) = \sum_{i=1}^{l-1} \pi(x_i), \; \pi'(x'_{M'}) = \sum_{i=u+1}^{M} \pi(x_i)$$

$$p'(x'_k, x'_j) = p(x_{l+k-1}, x_{l+j-1}),$$
$$1 \leq k \leq M'-2, \; 1 \leq j \leq M'-2$$

$$p'(x'_a, x'_k) = \frac{\sum_{i=w_1(a)}^{w_2(a)} p(x_i, x'_k)\pi(x_i)}{\pi'(x'_a)}, \; 1 \leq k < M'-1$$

$$p'(x'_k, x'_a) = \sum_{i=w_1(a)}^{w_2(a)} p(x'_k, x_i), \; 1 \leq k < M'-1$$

$$p'(x'_a, x'_b) = \frac{\sum_{i=w_1(a)}^{w_2(a)} \sum_{j=w_1(b)}^{w_2(b)} p(x_i, x_j)\pi(x_i)}{\pi'(x'_a)}$$

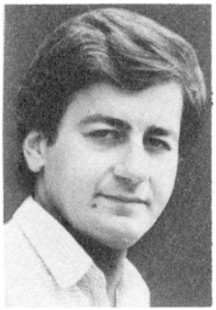
$$\phi'_\rho(x'_{M'-1}) = \frac{\sum_{1 \leq i \leq 1} \phi_\rho(x_i)\pi(x_i)}{\pi'(x'_{M'-1})},$$

$$\phi'_\rho(x'_{M'}) = \frac{\sum_{u \leq i \leq M} \phi_\rho(x_i)\pi(x_i)}{\pi(x'_{M'})}$$

REFERENCES

1. D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, 1987.
2. C. Ko, W. Wong and K. Lye, 'Performance of CSMA/CD networks connected by bridges', *Proceedings of International Conference in Communications ICC88*, Philadelphia, June 1988.
3. H. Takagi and L. Kleinrock, 'Output processes in contention packet broadcasting systems', *IEEE Trans. Communications*, **COM-33**, (11), 1191–1199 (1985).
4. F. A. Tobagi, 'Analysis of a two-hop centralized packet radio network—Part I: slotted ALOHA', *IEEE Trans. Communications*, **COM-28**, (2), 196–207 (1980).
5. F. A. Tobagi, 'Analysis of a two-hop packet radio network-part II: carrier sense multiple access', *IEEE Trans. Communications*, **COM-28**, (2), 208–216 (1980).
6. H. Takagi and L. Kleinrock, 'Throughput-delay characteristics of some slotted ALOHA multihop packet radio network', *IEEE Trans. Communications*, **COM-33**, (11), 1200–1207 (1985).
7. D. Heyman and M. Sobel, *Stochastic Models in Operations Research*, McGraw-Hill, 1982.
8. A. Viterbi, 'Approximate analysis of time-synchronous packet networks', *IEEE Journal on Selected Areas in Communications*, **SAC-4**, (6), 879–890 (1986).
9. H. Heffes and D. Lucantoni, 'A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance', *IEEE Journal on Selected Areas in Communications*, **SAC-4**, (6), 856–868 (1986).
10. I. Stavrakakis, 'Analysis of a statistical multiplexer under a general input traffic model', *INFOCOM'90 Conference*, San Francisco, 5–7 June, 1990.

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.