<p style="text-align:center">2</p>

On per-session performance of an ATM multiplexer with heterogeneous speed links

R.I. Balay and A.A. Nilsson
Center for Advanced Computing and Communication
North Carolina State University, Raleigh, NC 27695-7914, USA.
Phone: (919) 515-5130. Fax: (919) 515-2285.
email: {rajesh,nilsson}@eceyv.ncsu.edu

Abstract

Our aim in this paper is to provide insight on a session's performance in the presence of different speed links at an ATM multiplexer. We describe a queueing model for such scenarios considering a finite capacity buffer and bursty traffic with correlated arrivals. We use Markov Modulated Bernoulli Process (MMBP) as the traffic model for the session of interest and model the cross traffic as an MMBP with batch arrivals. We present exact analysis using an evolution method to obtain the steady-state performance measures of the session. To enhance its efficiency, a closed form, partial approximation method is also described. Using numerical results we compare a sessions performance in different scenarios of link speeds, and also study the influence of a few network and traffic characteristics.

Keywords

ATM, quality of service, high-speed networks, performance modeling, per-session performance, discrete-time analysis, Markov Modulated Bernoulli Process, finite capacity queue

1 INTRODUCTION

It is anticipated that Broadband Integrated Services Digital Networks (B-ISDNs) will support data, voice, video and multimedia applications using Asynchronous Transfer Mode (ATM) technology. Most applications will traverse different speed links and will have stringent quality of service (QoS) requirements to be guaranteed by the network. In enabling and efficiently utilizing such networks, per-session performance study becomes a necessity for admission control, network design and for the service provider, who desires to know the behavior of a session given the traffic characteristics and network topology. In this paper we address the above issue by providing discrete-time queueing analysis for per-session performance of a single switching node.

Discrete-time queueing models have received much attention due to increasing focus on ATM networks where packets are transmitted as fixed size units called 'cells'. In literature,
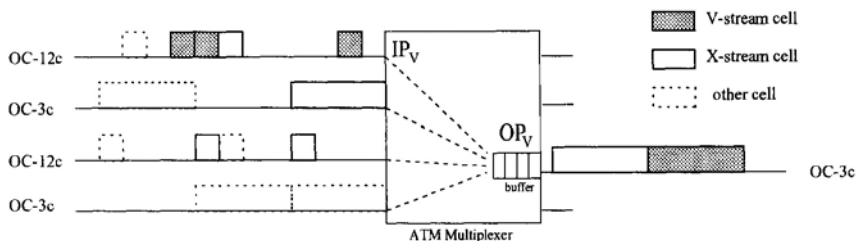
Figure 1 An example ATM switching scenario with SONET links.

most performance studies deal with a multiplexer as a whole and only a few consider per-session performance (Murata, *et al.* (1990), Herrmann (1993)). To model the bursty and correlated nature of ATM traffic many models (e.g. MMBP) have been proposed which are complex and non-renewal. Since closed form analytic solutions are difficult to obtain for systems with such traffic characteristic, algorithmic methods such as matrix-geometric techniques developed by Neuts (1981) are commonly used. We present an alternate algorithmic approach using evolution method which is computationally efficient and has the following advantages: (1) the flexibility to trade computational time for accuracy of results using a closed form, partial approximation and (2) transient behavior of a session can be studied (Balay and Nilsson, 1996).

This paper is organized as follows: In section 2, using a switching scenario we define our analytical model. In section 3, by exploiting the Markov renewal property of an MMBP, we present an exact analysis to derive a session's steady state performance measures such as cell loss probability and delay distribution. In section 4, we discuss the complexity of our method and also compare it with matrix-geometric methods. An approximation technique to enhance the efficiency of our method is described in section 5. We present some numerical results in section 6 and make our concluding remarks in section 7.

2   MODEL DESCRIPTION

The switching node we model is shown in figure 1, where IP_V and OP_V are the arrival and exit ports respectively for the session of interest (V-stream). All other sessions which are routed through OP_V constitute the cross traffic (X-stream). In modeling such scenarios for per-session study, a normal trend is to consider a single server queue of deterministic service time of one slot and assume at most one arrival to occur in a slot; the reasoning being cell transmission time on a link is constant. This model is a gross approximation if the session of interest traverses different speed links (which will be evident when we look at some numerical results). For an accurate model, it is important to capture the nature of cell arrivals with respect to the cell transmission; we do so using two parameters:

N_A: The constant number of slots for a cell arrival from V-stream. It means that two
   consecutive arrivals from V-stream are at least N_A slots apart, and a cell arrival

Input Link (IP_V)	Output Link (OP_V)	Arrival slots (N_A)	Service slots (N_S)
OC-3c	OC-3c	1	1
OC-12c	OC-3c	1	4
OC-12 (4 x OC-3c)	OC-3c	1	1
OC-3c	OC-12c	4	1
OC-48 (4 x OC-12c)	OC-3c	1	4
OC-48 (4 x 4 x OC-3c)		1	1

Table 1 Link speeds and corresponding model parameters

event is considered to occur in the last of the cell arrival slots (i.e. N_Ath slot). Note that this does not apply to cross-traffic cells since it constitutes cells from many sessions which may arrive on different speed links; hence we consider cross-traffic cells to take one slot to arrive.

N_S: The deterministic service time for all cells at the multiplexer.

In our discussion, a 'slot' has a new meaning; it is an interval of time which is the same as the least common multiple of the time to transport a cell on the input and output links traversed by V-stream. To make it more clear, for a few SONET speed links we list the corresponding model parameters in Table 1. Notice that the parameters depend only on the underlying concatenated SONET protocol used, since it determines the rate at which a cell is transported on a link. Some modeling assumptions are: buffer capacity is finite and is of size K cells, a cell cannot receive service in its arrival slot, arrivals occur before the departure event in a slot, an arrival is lost if buffer is full and a cell in service does not occupy buffer space. If a simple cyclic polling scheme is used to transfer the incoming cells to their respective output buffers, the performance of a session depends on the position of the arrival port. To obtain bounds on performance (which is more informative than average) we consider two slot policies: $\mathcal{P}_V$ and $\mathcal{P}_X$, which denote that in a slot the V-stream arrival is considered for buffer allocation before a cross-traffic arrival, and vice-versa, respectively.

2.1   Traffic model

We model V-stream as a Markov Modulated Bernoulli Processes (MMBP), the discrete-time analogue of the Markov Modulated Poisson Process (MMPP) (Fischer and Meier-Hellstern, 1993). These processes capture the notion of burstiness and correlation properties of an arrival stream. An MMBP is a Bernoulli process where the arrival rate is varied by a multiple-state Markov chain. It is characterized by a state transition probability matrix $\mathbf{P}$ and arrival probability descriptor $\mathbf{\Lambda}$, where $[\mathbf{P}]_{i,j}$ denotes the probability of transition from state 'i' to state 'j' and $[\mathbf{\Lambda}]_i$ denotes the probability of arrival in a slot when in state 'i'. We use $\mathbf{\Lambda_V}$ and $\mathbf{P_V}$ to describe the V-stream MMBP. For our analysis we use $a_v(v', j)$, the conditional probability density function (p.d.f.) of inter-arrival time
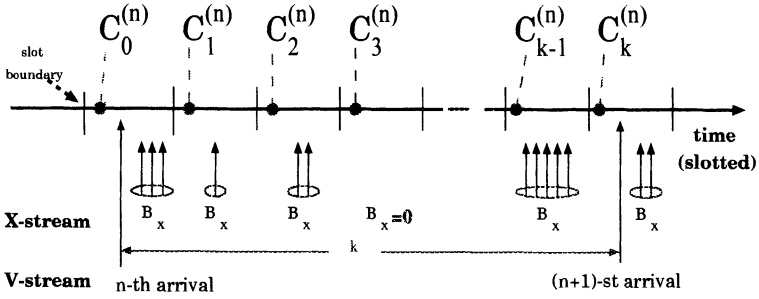
Figure 2 Relating system states observed by two consecutive V-stream arrivals.

of V-stream arrivals, which is defined as

$$a_v(v', j) = P[next\ Vstream\ arrival\ occurs\ in\ state\ v'\ and\ after\ j\ slots$$
$$|\ a\ Vstream\ arrival\ occurred\ in\ state\ v].$$

For the case where an arrival takes more than a slot ($N_A > 1$), a modified MMBP is considered (Balay and Nilsson,1996). The X-stream traffic is modeled as an MMBP with batch arrivals (B-MMBP), described by $\mathbf{P_X}$, $\mathbf{\Lambda_X}$ and a general distribution for the batch size. For our analysis we use:

$$b_x(j) = P[Xstream\ cells\ arriving\ in\ a\ slot = j\ |\ MMBP\ state = x],\ j \geq 0.$$

3   EXACT ANALYSIS

The basis of our queueing analyses is the identification of "embedded points" (observation instants) of the system such that it is "regenerative" (Wolff,1989) at these instants. We follow the evolution path along the embedded points to obtain steady-state information. A high level description of the 'evolution method' is:

1. Identify embedded points and define a regenerative system state at these points.
2. *Evolution step*: obtain the state distribution at the next embedded point using the current state information and the arrival characteristics.
3. Repeat the above evolution step until the state distribution at two consecutive embedded points is the same, i.e., evolve the system until it reaches steady-state.

We consider the V-stream arrival slots as the embedded points. For a regenerative system state, the MMBP state at a V-stream arrival instant should also be a part of the state descriptor, since MMBP is not a *renewal process* but is a *Markov renewal process* (Cinlar,1975); the time interval between consecutive arrivals depends on the MMBP states at these arrival instants. We define $C_k^{(n)}$ as the random variable (r.v.) for the system state at
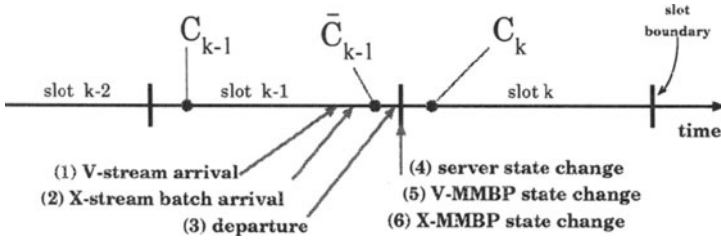
Figure 3 Relative order of events and observations in a slot (policy $\mathcal{P}_V$)

the beginning of the kth slot following the nth V-stream arrival (i.e. in the nth evolution step) given that the $(n+1)^{st}$ V-stream arrival does not occur in the preceding $k-1$ slots. $C_k^{(n)}$ is a 4-tuple, $C_k^{(n)} \equiv (V_k^{(n)}, X_k^{(n)}, S_k^{(n)}, J_k^{(n)})$ (p.d.f. given by $c_k^{(n)}(\cdot)$), $k \geq 0, n \geq 1$, where

$V_k^{(n)}$: Arrival state of V-stream MMBP (V-MMBP) in the nth V-stream arrival slot. Note that $V_k^{(n)}$ does not change with k, i.e. $V_k^{(n)} = V_0^{(n)}$, $\forall k > 0$. It's state space is denoted by $\mathcal{V}$.

$X_k^{(n)}$: State of X-stream MMBP (X-MMBP) in a slot. Its state space is denoted by $\mathcal{X}$.

$S_k^{(n)}$: State of the server. $S_k = i$ implies that the current slot is the ith service slot of cell in service; $S_k = 0$ when server is idle. A departure occurs at the end of slot k if $S_k = N_S$.

$J_k^{(n)}$: Number of cells in system (i.e. queue and server), ($0 \leq J_k \leq K + 1$).

From a known system state at an embedded point, to obtain the system state at the next embedded point we use the statistical distribution of the V-stream arrival process and the system states at each of these possible V-stream arrival slots (see figure 2), for which slot-to-slot system dynamics are required. The events which influence the change in system state between two consecutive slots, say *k-1* and *k*, are shown in figure 3. For ease of derivation we consider them in two phases: **phase(1):** arrivals which occur in slot *k-1*, and **phase(2):** state transitions at the slot boundary and the departure event. $\bar{C}_{k-1}^{(n)} = (\bar{V}_{k-1}^{(n)}, \bar{X}_{k-1}^{(n)}, \bar{S}_{k-1}^{(n)}, \bar{J}_{k-1}^{(n)})$ (p.d.f. represented as $\bar{c}_{k-1}^{(n)}(\cdot)$) is the system state after phase (1), where

$$\bar{V}_{k-1}^{(n)} = V_{k-1}^{(n)}, \quad \bar{X}_{k-1}^{(n)} = X_{k-1}^{(n)}, \quad \bar{S}_{k-1}^{(n)} = S_{k-1}^{(n)}, \text{ and}$$

$$\bar{J}_{k-1}^{(n)} = min(J_{k-1}^{(n)} + B_x + \bar{u}(k-1), K + u(S_{k-1}^{(n)})) \tag{1}$$

for $k \geq 1$, $n > 0$. Here $\bar{u}(n) \stackrel{\text{def}}{=} 1 - u(n)$ and

$$u(n) = \begin{cases} 0 & if \ n = 0 \\ 1 & if \ n \neq 0. \end{cases} \tag{2}$$

Equation (1) accounts for the fact that a V-stream arrival needs to be considered only in slot *0*, and the maximum cells that can be accommodated in the system is K if the server is idle or else it is $K + 1$. The relations to derive $C_k^{(n)}$ from $\bar{C}_{k-1}^{(n)}$ are

$$
\begin{aligned}
V_k^{(n)} &= \bar{V}_{k-1}^{(n)}, \quad J_k^{(n)} = \bar{J}_{k-1}^{(n)} - u(N - S_{k-1}^{(n)}), \\
X_k^{(n)} &= x \text{ with probability } [\mathbf{P_X}]_{i,x} \text{ where } \bar{X}_{k-1}^n = i, \ x \in \mathcal{X}
\end{aligned}
\tag{3}
$$

$$
S_k^{(n)} = \begin{cases}
\bar{S}_{k-1}^n + 1 & if \quad 0 < \bar{S}_{k-1}^n < N \\
1 & if \quad \bar{S}_{k-1}^n = 0, \ \bar{J}_{k-1}^{(n)} > 1 \quad or \quad \bar{S}_{k-1}^n = N, \ \bar{J}_{k-1}^{(n)} > 0 \\
0 & if \quad \bar{S}_{k-1}^n = 0, \ \bar{J}_{k-1}^{(n)} = 1 \quad or \quad \bar{S}_{k-1}^n = N, \ \bar{J}_{k-1}^{(n)} = 0
\end{cases}
\tag{4}
$$

for $k \geq 1$ and $n \geq 1$. Let $\tilde{m}_k \overset{\text{def}}{=} K + u(S_k) - \bar{u}(k)$. In terms of p.d.f's the relations are

$$
\bar{c}_k^{(n)}(v, x, s, j) = \begin{cases}
\displaystyle\sum_{i=0}^{j-\bar{u}(k)} c_k^{(n)}(v, x, s, i) \, b_x(j - i - \bar{u}(k)) & if \quad j < \tilde{m}_k \\
\displaystyle\sum_{l=\tilde{m}_k}^{\infty} \sum_{i=0}^{j-\bar{u}(k)} c_k^{(n)}(v, x, s, i) \, b_x(l - i - \bar{u}(k)) & if \quad j = \tilde{m}_k
\end{cases}
\tag{5}
$$

$$
c_k(v, x, s, j) = \begin{cases}
\displaystyle\sum_{x'}[\mathbf{P_X}]_{x',x}(\bar{c}_{k-1}(v, x', 0, 0) + \bar{c}_{k-1}(v, x', N_S, 1)) \\
\hspace{4cm} if \quad s = 0, \ j = 0 \\
\displaystyle\sum_{x'}[\mathbf{P_X}]_{x',x}(\bar{c}_{k-1}(v, x', 0, j) + \bar{c}_{k-1}(v, x', N_S, j+1)) \\
\hspace{4cm} if \quad s = 1, \ 1 \leq j \leq K \\
\displaystyle\sum_{x'}[\mathbf{P_X}]_{x',x}\bar{c}_{k-1}(v, x', s-1, j+1) \\
\hspace{4cm} if \quad 2 \leq s \leq N_S, \ 1 \leq j \leq K+1 \\
0 \hspace{3cm} otherwise
\end{cases}
\tag{6}
$$

where $k \geq 1$, $n \geq 1$, $x, x' \in \mathcal{X}$, $v \in \mathcal{V}$. The evolution step to obtain the p.d.f. of system state at the next embedded point from the p.d.f. of system state at the nth embedded point is given as

$$
c_0^{(n+1)}(v', x, s, j) = \sum_k \sum_v a_v(v', k) \, c_k^{(n)}(v, x, s, j),
\tag{7}
$$

where $k \geq 1$, $v, v' \in \mathcal{V}$, $x \in \mathcal{X}$, $0 \leq s \leq N_S$ and $0 \leq j \leq K + 1$. If steady state exists then the steady state p.d.f. of the system state at an embedded point is given as

$$
c_0(v, x, s, j) = \lim_{n \to \infty} c_0^{(n)}(v, x, s, j).
\tag{8}
$$

$c_0(\cdot)$ denotes the steady-state p.d.f. of system state at the beginning of a V-stream arrival slot. We use it obtain the V-stream performance measure in the next section; X-stream performance measures can also be obtained (Balay and Nilsson,1994).

3.1   Per-session performance (V-stream)

Let $C^v = (V^v, X^v, S^v, J^v)$ be the r.v. for the system state seen by a V-stream arrival at steady state and let $c^v(\cdot)$ be its p.d.f.

Case 1: $\mathcal{P}_V$ arrival policy. V-stream arrival in a slot is considered first for buffer allocation. Hence, we have $C^v \equiv C_0$.

Case 2: $\mathcal{P}_X$ arrival policy. The system state seen by a V-stream arrival is after the X-stream cells of the same slot. We obtain its p.d.f. as follows:

$$V^v = V_0, \quad X^v = X_0, \quad S^v = S_0, \quad J^v = min(J_0 + B_x, K + u(S_0)). \tag{9}$$

$$c^v(v,x,s,j) = \begin{cases} \displaystyle\sum_{i=0}^{j} c(v,x,s,i)\, b_x(j-i) & if \quad j \ < \ K + u(s) \\[4mm] \displaystyle\sum_{l=K+u(s)}^{\infty} \sum_{i=0}^{j} c(v,x,s,i)\, b_x(j-i) & if \quad j \ = \ K + u(s). \end{cases} \tag{10}$$

Some of the interesting performance measures of V-stream are found as follows:

$$P[buffer\ occupancy = j] = \sum_{v,x} \left[c^v(v,x,0,j) + \sum_{s=1}^{n} c^v(v,x,s,j+1) \right], \quad 0 \le j \le K. \tag{11}$$

$$P[cell\ loss] = \sum_{v,x} \left[c^v(v,x,0,K) + \sum_{s=1}^{N_S} c^v(v,x,s,K+1) \right] \tag{12}$$

$$P[delay = j\ slots] = \begin{cases} \displaystyle\sum_{v,x} c^v(v,x,0,l-1) + \sum_{v,x} c^v(v,x,N_S,l) & if\ j\ mod\ N_S = 0 \\[3mm] \displaystyle\sum_{v,x} c^v(v,x,s,l) & if\ j\ mod\ N_S \ne 0\ and\ N_S \le j < (K+1)N_S \\[3mm] 0 & otherwise \end{cases} \tag{13}$$

where $s = (N_S - j)\ mod\ (N_S)$, $l = \lfloor j/N_S \rfloor$ and delay is the time spent in the system.

4   DISCUSSION ON TIME COMPLEXITY

Let N denote the size of the system state; the main component is the buffer size. Computational time to obtaining slot-to-slot system dynamics is proportional to the system size N and the maximum arrivals which can occur in a slot; we consider the maximum arrivals in a slot to be a constant since it is limited by the number of input links at the multiplexer. Each evolution step requires M_I slot-to-slot operations where M_I denotes the maximum inter-arrival time for the session of interest; it is large for bursty processes. The third factor of computational complexity is the number of evolutions required for convergence which we denoted by N_C. The computational complexity of the arrival-wise method is $O(NM_IN_C)$ to obtain exact steady-state results. Using the approximation technique

described in the next section the latter two factors of complexity are reduced to constants and the complexity of the method is $O(N)$.

For a matrix-geometric solution for the considered system, the transition probability matrix **P** of size N^2 is required where each element (i, j) denotes the transition probability of the system from state i to state j at consecutive embedded points. The computational time to obtain each element is the same as that of an evolution step $(O(NM_I))$. The complexity of generating the transition probability matrix $(O(N^3 M_I))$ prohibits using a matrix-geometric technique for per-session analysis of a multiplexer while an evolution method is much more efficient.

5 APPROXIMATION TECHNIQUE

In this section, we describe an approximate characterization for the change between system states at slots which are N_S slots apart, as a slow changing function. This approximation is applied to partially reduce the evolution step (equation (7)) to a closed-form relation. Similarly, we can approximate the steady-state relation (equation (8)) by considering this approximation for the change between system states at consecutive embedded points.

Let $c_k(q) \ (= P[C_k = q])$, where $q = (q_v, q_x, q_s, q_j)$ be the p.d.f. of system state at the beginning of slot k. Let $\delta_k(q)$ represent the difference in p.d.f. of system states N_S slots apart, and be defined as $\delta_k(q) = c_{k+N_S}(q) - c_k(q)$, $\forall q, k$. For some finite value $T > 0$, the system behavior can be approximately characterized by the following relation

$$\delta_{k+N_S}(q) \ = \ \alpha \, \delta_k(q) \qquad \forall k \geq T \tag{14}$$

where α is a non-negative constant and satisfies $0 < \alpha < 1$. For a valid approximation we choose T such that $0 \leq c_T(q) + \frac{\delta_T(q)}{(1-\alpha)} \leq 1$, $\forall q$. We consider the system states in the tail (i.e. $\geq T$) as N_S geometric streams i.e., $\{c_{T+N_S i}(q)\}, \{c_{T+1+N_S i}(q)\}, \cdots, \{c_{T+N_S-1+N_S i}(q)\}$. Each stream (j) is treated separately with respect to the approximation and hence we have a constant α for each stream which is denoted as α_j where $0 \leq j < N_S$. Using this approximation we can now represent the system state at the beginning of any slot $(\geq T)$ as

$$c_{k+j+N_S i}(q) \ = \ c_{k+j}(q) \ + \ \frac{(1 - \alpha_j^i)}{(1 - \alpha_j)} \, \delta_{k+j}(q), \quad \forall q, \ \forall k \geq T, \ i \geq 0, \ 0 \leq j < N_S \tag{15}$$

For any m-state MMBP with one state as completely bursty (i.e. prob. of arrival in this state is 1), the following is true of its p.d.f.

$$a_v(v', k+1) \ = \ \beta \, a_v(v', k) \qquad \forall k \geq m$$

where β is a constant $(0 < \beta < 1)$. Using this property of an MMBP and the above stated approximation we can reduce and rewrite the evolution step (equation (7)) as

Traffic	V-stream				X-stream					
type	*(offered load* (λN_S)*=0.2)*				*(offered load* $(\lambda N_S \bar{B})$*=0.4)*					
	bursty	CV^2	correlated	ψ_1	bursty	CV^2	correlated	ψ_1	Batch	$\bar{B}$
A	$\checkmark$	16	$\checkmark$	0.1	$\checkmark$ (low)	1	no	0	Geo	3
B	$\checkmark$	16	no	0	no	1	no	0	constant	1
C	no	1	no	0	no	1	no	0	constant	1

Table 2 Traffic parameters for V-stream and X-stream

$$
c_0^{(n+1)}(q') = \sum_{k=1}^{T-1} \sum_{q,v'} a_{q_v}(v',k) \, c_k^{(n)}(q)
$$

$$
+ \sum_{q,v'} \sum_{j=0}^{N_S-1} \frac{a_{q_v}(v',T+j)}{(1-\beta_S^N)} \left(c_{T+j}^{(n)}(q) + \frac{\beta_S^N \, \delta_{T+j}^{(n)}(q)}{(1-\alpha_j^{(n)}\beta_S^N)} \right) \tag{16}
$$

For each evolution step we compute $\delta_{T+j}^{(n)}(q)$ and $\alpha_j^{(n)}$ from $c_{T+j}^{(n)}(q)$, $c_{T+j+N_S}^{(n)}(q)$ and $c_{T+j+2N_S}^{(n)}(q)$ (these are p.d.f.s of system states obtained using exact relations) as $\delta_{T+j}^{(n)}(q) = c_{T+j+N_S}^{(n)}(q) - c_{T+j}^{(n)}(q)$ and $\alpha_j^{(n)}(q) = \frac{c_{T+j+2N_S}^{(n)}(q) - c_{T+j+N_S}^{(n)}(q)}{\delta_{T+j}^{(n)}(q)}$. We consider α_j to be a weighted average which is calculated as $\alpha_j^{(n)} = \sum_q c_{T+j}^{(n)}(q) \, \alpha_j^{(n)}(q)$.

6   NUMERICAL RESULTS

In this section we study the effect of link speeds, traffic characteristics and buffer sizing on the performance of a session. We also validate the accuracy of the approximation technique by comparing results with those obtained from simulation. We consider a scenario where a session's offered load is 20% of an output link capacity while it is 40% for the cross traffic. We compare the session's performance for three cases of link speeds:

1. $N_A = 1, N_S = 1$: both input and output links of V-stream are of same speed.
2. $N_A = 4, N_S = 1$: input link of V-stream is four times slower than the output link.
3. $N_A = 1, N_S = 4$: input link of V-stream is four times faster than the output link.

We use the traffic descriptors $(\lambda N_S, CV^2, \psi_1)$ for V-stream and $(\lambda N_S \bar{B}, CV^2, \psi_1, \bar{B})$ for X-stream where, λ is the mean arrival rate of an MMBP, CV^2 is the burstiness defined as the squared coefficient of variation, ψ_1 is the autocorrelation of lag 1 slot, of the arrival process and $\bar{B}$ is the mean batch size. Given the traffic descriptors we obtain 2-state MMBPs as described by Balay and Nilsson (1994). The traffic descriptors used for numerical results are tabulated in table 2.

6.1   Effect of different speed links

Notice that we use 'offered load' as one of the traffic descriptors; it helps us compare the three systems. In the figures 4(a), 5(a), 5(b), 6(a) and 6(b), we plot some of the

performance measures (buffer occupancy distribution, mean delay and loss probability) of a session. We see that the performance measures are distinctly different for the three systems considered. These results clearly indicate the importance of considering the link speeds for modeling. One way of interpreting these results is as follows: the results for the case of the homogeneous system ($N_A = 1, N_S = 1$) are the same as the results of a heterogeneous system ($N_A \neq N_S$), if modeled ignoring the different speed links. We notice that ignoring the speed of links is a gross approximation and results in very erroneous performance predictions.

6.2 Validating the accuracy of results

To validate the accuracy of the analytical results obtained using approximation, we compare them with simulation results. Figure 4(a) shows the p.d.f. of buffer occupancy seen by V-stream and X-stream, obtained using exact analysis for systems with type A traffic, policy $\mathcal{P}_X$ and buffer size of $K = 64$ cells. For the case of ($N_A = 1, N_S = 4$) system we also plot the results obtained using approximation and simulation. Notice that there is no visible difference between the simulation and analytical results; for a better view of the comparison, we plot the absolute error ($= analytical$ - $simulation$) along with the 95th % confidence regions of simulation results in figure 4(b). For this example, in obtaining exact results the maximum interarrival time for V-stream cells was 2060 slots with an accuracy of 10^{-6}, i.e., if $a(j)$ is the interarrival time p.d.f., $\sum_{j=1}^{2060} a(j) > 1 - 10^{-6}$. We see that, approximation with $T > 150$ slots yields near exact results; a reduction of computational time by a factor of 13. For the case where $N_S = 1$ we obtain results within confidence intervals with $T = 25$ and a gain by a factor of 23. These results shows that by using the approximation technique we can obtain results efficiently while maintaining accuracy. Some observations are that it is computationally much better than simulation, and is more efficient when $N_S = 1$ or when the system size is large; further work is required to engineer the minimum value of T for a given system.

6.3 Effect of burstiness and correlation

First, we illustrate an important phenomenon which occurs in systems with a slower output link and bursty traffic. Notice in figure 4(a), the curve for ($N_A = 1, N_S = 4$) system exhibit an oscillatory behavior for smaller buffer occupancy values; the probability values for occupancies which are multiples of ($N_S - 1$) are higher than its neighboring values. It can be explained using an example scenario: Consider a busy period of the system where the first n ($n > N_S$) arrivals are from V-stream and all of which occur in consecutive slots. The first cell arrives in slot 1, begins service in slot 2, and departs at the end of slot 5 (since $N_S = 4$). The arrivals which occur in slot 5 and 6 respectively see the buffer occupancy to be 4. Also, for each departure i which occurs at the end of slot $iN_S + 1$ in this interval of n arrivals, the occupancy seen by an arrival in slots $iN_S + 1$ and $iN_S + 2$ is $i(N_S - 1)$ cells; hence a higher probability value for $i(N_S - 1)$. Such scenarios are very likely to occur when both streams are bursty. This behavior will also be reflected in the delay distribution and departure process; it introduces periodic (or negative) correlation between departures which makes the characterization of the departure process difficult. In figures 5(a) and 5(b) we plot the loss probability of V-stream for change in V-stream's burstiness and correlation values respectively. We observe that a source which is initially
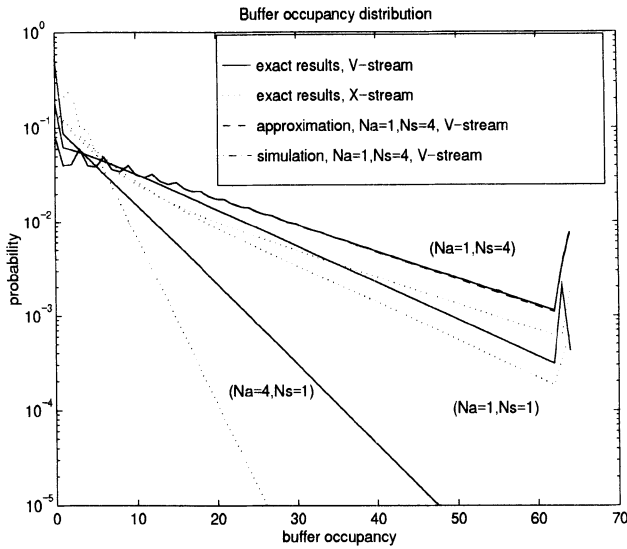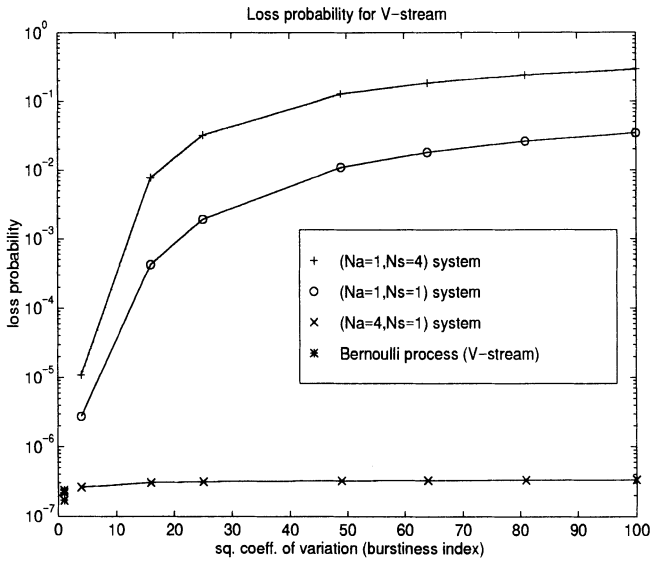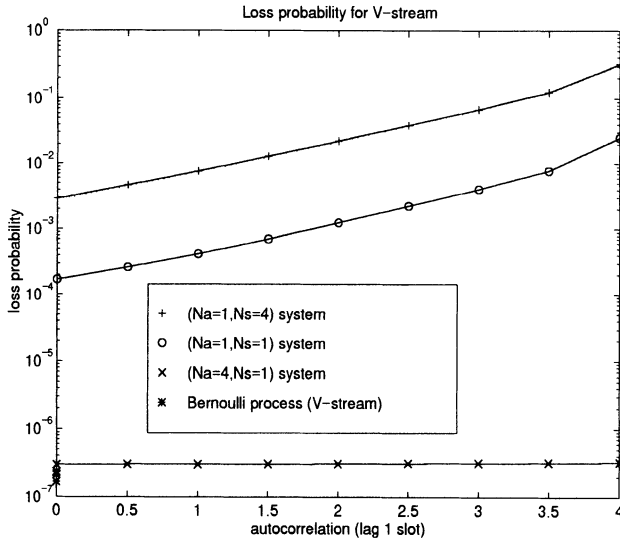
(a) Buffer occupancy distribution
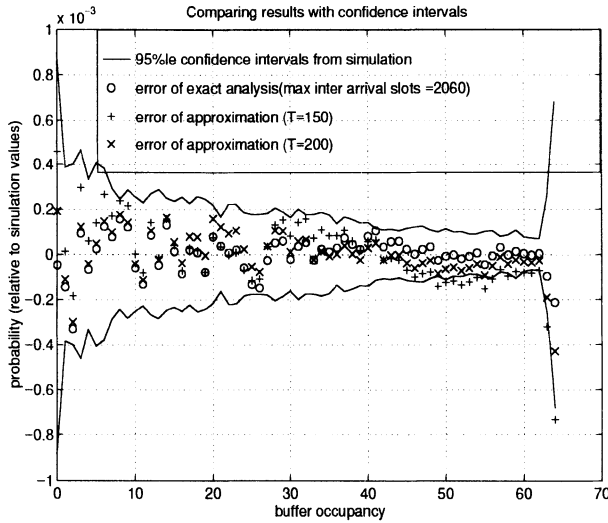
(b) Comparing results with 95 % confidence intervals of simulation: ($N_A = 1, N_S = 4$) system

Figure 4 Buffer occupancy seen by V-stream and X-stream in systems with K=64, Type A traffic and policy $\mathcal{P}_X$.
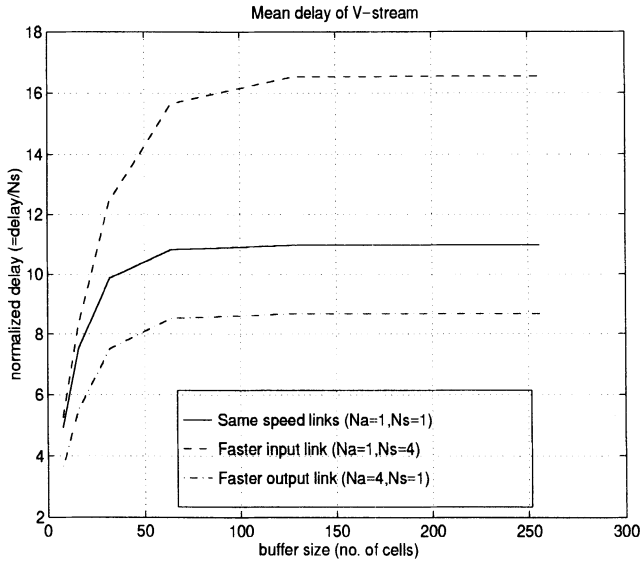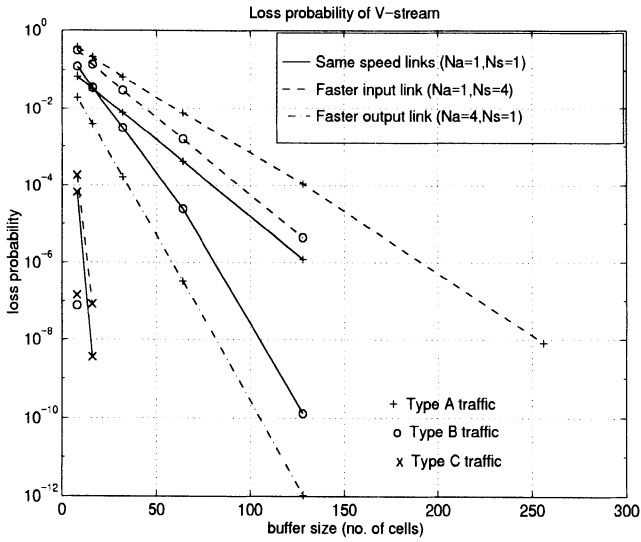
(a) Effect of change in burstiness

(b) Effect of change in autocorrelation

Figure 5 Effect of change in burstiness and autocorrelation of V-stream on its loss probability: systems with K=64, policy $\mathcal{P}_X$ and type Λ traffic.

(a) Effect of change in buffer size on mean delay of V-stream (type A traffic)

(b) Effect of change in buffer size on loss probability of V-stream

Figure 6 Influence of change in buffer size on mean delay and loss probability of V-stream: systems with policy $\mathcal{P}_X$ and all types of traffic.

less bursty is more sensitive to change in burstiness. The converse seems true for the effect of autocorrelation, i.e., correlated streams are more sensitive than non-correlated streams, for change in autocorrelation. We observe that the performance variation due to change in traffic characteristics is less in a $(N_A = 4, N_A = 1)$ system compared to the other systems.

6.4   Influence of slot policy ($\mathcal{P}$) on performance

Consider the curves of figures 5(a) and 5(b) for systems where $N_S = 1$. The loss probabilities plotted are for systems with policy $\mathcal{P}_X$. V-stream cells are not lost when the policy is $\mathcal{P}_V$, since in each slot there is always space for one cell in the buffer which is made available at the beginning of a slot when a cell starts its service. This means the sessions loss probability lies between '0' and the values plotted in the figures, depending on the location of the arrival port. The port of arrival at a switching element is an important factor which influences the performance of a session if $N_S = 1$ when a simple round-robin policy is used for switching the incoming cells. Such variations in performance can be reduced by randomizing the polling policy or randomly distributing the incoming cells to the input ports for switching.

6.5   Influence of buffer size

An important aspect of designing a switch is buffer dimensioning since loss guarantees may be of the order $< 10^{-10}$. Obtaining accurate results of such small magnitude through simulation techniques is clearly a difficult (time consuming) task; one reason why discrete-time analysis is important in studying ATM systems. To illustrate the buffer requirements to provide delay and loss guarantees for systems with different speed links and different traffic characteristic, we plot the mean delay and loss probability for the three systems for change in buffer size in figures 6(a) and 6(b) The values plotted in figure 6(b) are for the three traffic types, buffer size of 2^i, $3 \leq i \leq 8$, and when the loss probability is more than 10^{-12}. We observe that the loss probabilities change almost linearly (on a logarithmic plot) with buffer size while the buffer requirements are very different for the considered system.

7   CONCLUSIONS

In this paper we have shown the importance of considering the speed of links traversed by a session, for studying its performance at an ATM multiplexer. The numerical results presented indicate that a sessions performance is distinctly different when it traverses different speed links; they also show that ignoring the link speeds is a gross approximation and will result in incorrect results.

For studying a session's performance in the presence of different speed links, we have described a queueing model of an ATM multiplexer. For our analysis we considered finite capacity buffer and modeled the session of interest and the cross-traffic as Markov renewal traffic streams. We used an evolution method to obtain the steady-state performance measures, and presented an approximation technique; together they make our analysis much more efficient than a matrix-geometric method. Using numerical results we have also illustrated the effects of traffic characteristics, port of arrival of a session and buffer

size on the steady-state performance measures of a session. The presented analysis also contributes in studying the transient behavior of a session (Balay and Nilsson, 1996). Future work includes characterization of per-session departure process and end-to-end analysis.

8   REFERENCES

Balay, R.I. and Nilsson, A.A. (1994) Analyses of heterogeneous tandem links, part 1: Per-session performance of an ATM multiplexer. Technical Report TR-94/20, Center for Advanced Computing and Communication, NCSU Raleigh, NC.

Balay, R.I. and Nilsson, A.A. (1996) Transient performance of a session at an ATM multiplexer with heterogeneous speed links. *INFOCOM'96*, Submitted for publication.

Blondia, C. and Casals, O. (1992) Performance analysis of statistical multiplexing of VBR sources. *Proceedings, INFOCOM'92*, 828–838.

Cinlar, E. (1975) *Introduction to stochastic processes.* Prentice Hall Inc., NJ.

Fischer, W. and Meier-Hellstern, K. (1993) The mmpp cookbook. *Performance Evaluation*, 18(2):149–171.

Herrmann, C. (1993) Correlation effect on per-stream QoS parameters of ATM traffic superpositions relevant to connection admission control. *Proceedings, ICC*, 1027–1031.

Murata, M. and Oie, Y. and Suda, T. and Miyahara, H. (1990) Analysis of discrete-time single server queue with bursty inputs for traffic control in ATM networks. *IEEE Journal on Selected Areas in Communications*, 8(3):447–458.

Neuts, M. (1981) *Matrix-geometric solution in stochastic models: An algorithmic approach.* John Hopkins University Press.

Wolff, R.W. (1989) *Stochastic Modeling and the Theory of Queues.* Prentice Hall Inc., NJ.

9   BIOGRAPHY

Rajesh I. Balay received an M.Sc.(Tech) degree in Computer Science from the Birla Institute of Technology and Science (BITS Pilani), India, in 1989, and an M.S. degree from the University of Alabama at Birmingham, USA, in 1991. He is currently a doctoral candidate in Computer Science at North Carolina State University, Raleigh, USA. Since 1991 he has been with the Center for Advanced Computing and Communication, at the university, engaged in research on performance evaluation of high speed networks. His research interests include design of efficient network protocols, performance modeling of networks and communication protocols, and computer games of strategy.

Arne A. Nilsson received an M.S. degree in Electrical Engineering in 1968 and a Ph.D degree in Telecommunication Systems in 1976, from Lund University of Technology, Sweden. He is currently a Professor of Electrical and Computer Engineering, and the Technical Director of the Center for Advanced Computing and Communication, at the North Carolina State University, Raleigh, USA. His areas of research include performance modeling of local area networks and computer systems, routing and flow control in computer networks, computer communication synthesis and analysis, medical image networking and packet radio networks.