Performance Evaluation of a Single Server Autoregressive Queue

R.G. ADDIE * and M. ZUKERMAN †

* *University of Southern Queensland*
† *Telecom Australia Research Laboratories*

Abstract: This paper provides an approximation for stationary statistics of the unfinished work in a FIFO single server queue with unlimited buffer fed by a first-order autoregressive input. The method is based on modelling the queue as a semi-Markov queue and using the Wiener-Hopf factorization technique. It has been established that the tail of the unfinished work or waiting time distribution of such queues is negative exponential, and in this paper we provide estimates for the parameters of that exponential term. Simulation results which demonstrate that the formula leads to accurate results are presented.

Keywords: Queueing analysis, Performance evaluation, Traffic modelling, Autoregressive process, Statistical multiplexing, ATM networks

1 INTRODUCTION

The autoregressive process has been proposed in [1] as a model for video telephony traffic. As demonstrated in [1], it is relatively easy to compute the required parameters of an autoregressive traffic model, given stationary statistics of a stationary traffic stream. The reason that it has not been very popular as a traffic model is related to the fact that it has not been considered amenable for queueing analysis [2, 1].

In this paper, we consider a single server First Come First Served (FIFO) queue with an arrival process that follows a 1st order autoregressive process and a constant service rate. In particular, we provide an asymptotically accurate approximation in a closed form for the stationary unfinished work (virtual waiting time) distribution.

Queues where the arrivals are not correlated have been extensively studied. (See for example [3, 4, 5] which involve independent and generally distributed interarrival times.) This paper present results for queues with correlated arrivals.

Many engineers who are involved in design, dimensioning and management of high-speed packet switched networks have realized that, although evaluating the mean values of delay and unfinished work is important, a more important and often harder problem is how to evaluate quantiles of such distributions. Computation of such quantiles is one of the main aims of this paper.

Correspondence to: M. Zukerman, Telecom Australia Research Laboratories, 770 Blackburn Road, Clayton, Victoria 3168, Australia. Email: m.zukerman@trl.oz.au

Paper received 31 August 1993.
Final revision 5 October 1993.

A general solution for the stationary waiting time distribution for semi-Markov queues based on spectral factorization was originally developed by Miller [6, 7, 8]. Miller assumed in his work that the state space of the underlying Markov process is finite. This work has been extended in [9, 10, 11, 12] to include cases where the state space of the underlying Markov process is infinite. However, no indication of computational methods is given in [10, 11, 12], while the method in [9] is seldom practical for problems with a large state space.

It has been observed by Miller [8] and also by Presman [13] that under quite general conditions, the tail of the stationary complementary waiting time distribution (which is the part of most interest in many applications) is of the form ce^{s^*t}, where s^* is obtainable as the negative real root of a certain functional equation which lies closest to the origin. We have demonstrated in [14] [15] that, in the case where the arrival process is Gaussian, which can be modelled as a semi-Markov queue where the state space of the underlying Markov chain is of infinite dimension, an exact analytical solution in closed form of the functional equation for s^* may be obtained. However, calculation of the coefficient c has been a problem, as the only method proposed for semi-Markov queues has been to calculate the complete spectral factorization, which is usually very difficult. Accurate and simple approximations for the coefficient c were proposed in [14] and [15].

The objective of this paper is to apply the method of [15] to the relatively simple case of the 1st order autoregressive queue. While analysis of the Gaussian queue requires solving a semi-Markov queue with an underlying state space of infinite dimension, the case of a queue with 1st order autoregressive input is equivalent to a semi-Markov queue where the underlying state space is one dimensional, i.e. equivalent to R.

In [14, 15, 16] we have shown that performance statistics of a queue fed by traffic which follows a general stationary, ergodic Gaussian process can be accurately evaluated as a function of three parameters: the mean and the variance of the arrival process, and a parameter related to the asymptotic rate of the long term variance. In fact, the formulae for the approximation of c and the computation of s^* can be expressed in terms of these three parameters. For any possible choice of values for these parameters, there exists a first order autoregressive (AR1) process, which has these values for the three parameters. In this sense, to the extent that the asymptotic model of this paper is accurate, the AR1 processes can be regarded as typical of all stationary ergodic Gaussian processes. This makes the AR1 an important process for the purpose of traffic modelling.

The remainder of the paper is organized as follows: first, a short description of the queueing model is presented in section 2. Next, in section 3, we present results for the queueing performance of an autoregressive queue by using results obtained for the more general Gaussian queue. Then, in section 4, we describe a general solution for stationary statistics of semi-Markov queues, based on spectral factorization. In section 5 we use the results for semi-Markov queues and compute the s^* value for a queue with autoregressive input. Finally, in section 6, we present simulation results which demonstrate the accuracy of the solution.

2 MODEL DESCRIPTION

Let time be divided into fixed length intervals. These will henceforth be referred to simply as *frames*. Let A_n be a continuous random variable representing the amount of work entering the system during the n th frame. The variable A_n may represent the number of bits or ATM cells entering the system during the n th frame. Since each cell (or bit) is a fixed quantity of work which is usually small relative to the amount transmitted within a frame, modelling A_n as a continuous random variable will not introduce a significant error.

Assume that A_n is a first-order autoregressive process, i.e.,

$$A_n = aA_{n-1} + b\tilde{U}_n, \tag{1}$$

where $\tilde{U}_n$ is Gaussian with mean η and variance $\tilde{\sigma}^2$ and a and b are real numbers with $|a| < 1$. This model was proposed in [1] for a VBR traffic stream generated by a single source of video telephony. (In [1] $\tilde{\sigma}^2 = 1$.)

The A_n s can be negative with positive probability. This may seem to hinder the application of this model to real traffic processes. However, in modelling traffic we are not necessarily interested in a process which is similar in every detail to the real traffic. What we are interested in is a process which has the property that when it is fed into a queue, the queueing performance is similar to that of the queue fed by the real traffic.

Let τ be a fixed number representing the amount of work which can be processed by the server per frame. We assume here for simplicity that the service takes place at the end of the frame.

Let Z denote the set of integers. Let the sequence of continuous random variables Y_n be the *net input process* defined by

$$Y_n = A_n - \tau, \quad n \in Z. \tag{2}$$

We use notation such as $E\{A_\infty\}$ and $VAR\{A_\infty\}$ to denote the stationary mean and variance of A_n respectively.

Let m be the stationary mean of the net input process, that is

$$m = E\{Y_\infty\}. \tag{3}$$

For stability we require $m = E\{Y_\infty\} < 0$.

Let σ^2 be the stationary variance of the net input process

$$\sigma^2 = VAR\{Y_\infty\}.$$

Using the above notation, the system unfinished work process satisfies Lindley's recurrence relation:

$$V_{n+1} = (V_n + Y_n)^+, \quad n \in Z. \tag{4}$$

where $V_0 = 0$ and where $X^+ = X$ if $X \geq 0$ and $X^+ = 0$ otherwise. Note that although a negative amount of work may join the queue, the queue size always remains nonnegative.

Because the service rate is constant in this case, we have that $VAR\{Y_\infty\} = VAR\{A_\infty\}$. By (1),

$$E\{A_\infty\} = \frac{\eta b}{(1-a)} \tag{5}$$

so

$$m = E\{Y_\infty\} = E\{A_\infty\} - \tau = \frac{\eta b}{(1-a)} - \tau, \tag{6}$$

and

$$\sigma^2 = \text{VAR}\{Y_\infty\} = \text{VAR}\{A_\infty\} = \frac{b^2\tilde{\sigma}^2}{(1-a^2)} \tag{7}$$

As usual, the utilization, denoted ρ, is the ratio between the arrival rate and the service rate, so

$$\rho = \frac{\text{E}\{A_\infty\}}{\tau} = \frac{\eta b}{(1-a)\tau}. \tag{8}$$

It is important to mention that a set of autoregressive processes based on Eq. (1), all with the same value for the first autoregressive coefficient (namely a), is closed under superposition. In particular, if we consider k video sources each generating traffic streams based on the model of Eq. (1) with the same coefficients, their superposition process, denoted by $\{Z_n,\ n \geq 1\}$, will also be an autoregressive process; namely, $Z_n = aZ_{n-1} + b\bar{U}_n$, where $\bar{U}_n$ is Gaussian with mean $k\eta$ and variance $k\tilde{\sigma}^2$. (Note that the variance to mean ratio of such a superposition will be equal to that of each of the traffic streams.) However, if the different streams have different values for the first autoregressive coefficient, then Eq. (1) cannot represent the multiplexed traffic.

In [14] and [15] we have studied a queueing system where the net input process is a stationary, ergodic, Gaussian discrete-time process with mean m and variance σ^2. The autoregressive queue is a special case of this model.

In [14] we introduced the concept of the *autocovariance sum* defined as

$$S = \sum_{k=1}^{\infty} \text{Cov}(Y_n, Y_{n+k}). \tag{9}$$

This is equivalent to the *covariance integral* introduced by Heffes [17]. In [15] we show that, for the autoregressive queue,

$$S = \frac{ab^2\tilde{\sigma}^2}{(1-a)^2(1+a)}. \tag{10}$$

In [16] we introduced the concept of the *asymptotic variance rate* of a stationary and ergodic stochastic process $\{Z_n\}$ denoted by VR $(\{Z_n\})$ for short. It is defined by

$$\text{VR}(\{Z_n\}) \triangleq \lim_{k\to\infty} \frac{\text{VAR}\{\sum_{n=1}^{k} Z_n\}}{k}. \tag{11}$$

In particular, let $v = \text{VR}(\{Y_n\})$ be the asymptotic variance rate of the net input process. Notice that the autocovariance sum is related to the asymptotic variance rate by

$$v = \sigma^2 + 2S. \tag{12}$$

3 SUMMARY OF RESULTS

An important result from [15] is that the tail of the stationary unfinished work distribution for an infinite buffer queue fed by a great variety of stationary and ergodic input processes is exponential. In particular,

$$\text{P}\{V_\infty > t\}e^{-s^* t} \to c \text{ as } t \to \infty \tag{13}$$

for some negative real number s^* and positive (less than 1) real number c. Therefore, if we are able to compute s^* and c, we will have the tail distribution of V_∞.

In [14] and [15] we obtained an exact result for s^* given by

$$s^* = \frac{2m}{\sigma^2 + 2S}. \tag{14}$$

Hence by Eq. (12) we have the fundamental relationship

$$s^* = \frac{2m}{v}. \tag{15}$$

Substituting the results of Eqs. (6), (7), (8) and (10) in (14), we obtain

$$s^* = \frac{2(1-a)(\rho-1)\tau}{(1+a)\sigma^2}. \tag{16}$$

(Note that a small error appeared in [18] in this formula.)

In section 5 we shall derive s^* by a different approach from the one used in [15]. This will confirm the accuracy of Eq. (16) and will provide reassurance to the fundamental result of Eq. (15).

So far we have not been able to obtain an exact result for c as defined by Eq. (13). An approximation for c, denoted by $\tilde{c}$, has been obtained [15] by choosing it so that Lindley's equation (4) is satisfied after taking the expectation on both sides. This results in the following:

$$\tilde{c} = -\frac{s^*\psi(-m)}{\text{erf}\left(\frac{u}{\sigma\sqrt{2}}\right)}, \tag{17}$$

where

$$u = \frac{m\sigma^2}{v} = \frac{1}{2}\sigma^2 s^*$$

and

$$\psi(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_0^\infty y e^{-\frac{(y+x)^2}{2\sigma^2}}\,dy$$

$$= \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} - \frac{x}{2}\,\text{erfc}\left(\frac{x}{\sigma\sqrt{2}}\right). \tag{18}$$

The function $\psi(x)$ represents the mean of the random variable X^+, where X is a normally distributed random variable with mean $-x$ and standard deviation σ.

Having s^* and $\tilde{c}$, we can approximate the distribution of V_∞ by its tail, that is

$$\mathrm{P}\{V_\infty > t\} \approx \tilde{c}e^{s^* t} \quad (19)$$

Consequently, the mean may be approximated by

$$\mathrm{E}\{V_\infty\} \approx -\frac{\tilde{c}}{s^*}. \quad (20)$$

Let t_p be the p^{th} percentile of the stationary distribution of V_∞. Assuming $\tilde{c} > 1 - p/100$, we obtain $p/100 \approx 1 - \tilde{c}e^{s^* t_p}$, or

$$t_p \approx \frac{1}{s^*}\ln((1-p/100)/\tilde{c}). \quad (21)$$

If $\tilde{c} \leq 1 - p/100$ then $t_p \approx 0$. Notice that the AR1 model has four parameters: a, b, η and $\tilde{\sigma}^2$. These parameters are not independent, however. The process is actually determined by the values of $b\eta$, $b^2\tilde{\sigma}^2$, and a. Eqs. (6), (7) and (10) can be used to determine the values of $b\eta$, $b^2\tilde{\sigma}^2$, and a for prescribed values of the mean, variance and asymptotic variance rate of the process. Alternatively, given prescribed values for m, σ^2 and S, any one of the three parameters b, η and $\tilde{\sigma}^2$ can be arbitrarily set to a fixed value (note that the parameter a cannot be freely chosen), and then the other three parameters can be computed by Eqs. (6), (7) and (10).

For example, if we fix $\tilde{\sigma}$, the parameters a, b and η can be calculated using the following equations in the order shown:

$$a = \frac{S}{S+\sigma^2} \quad (22)$$

$$b = \left(\frac{\sigma^2(1-a^2)}{\tilde{\sigma}^2}\right)^{\frac{1}{2}} \quad (23)$$

$$\eta = \frac{(1-a)\lambda}{b} \quad (24)$$

4 SEMI-MARKOV QUEUES

In this section we provide background on the theory of semi-Markov queues. We then show in the next section how this theory is applied to the case of a queue with autoregressive input.

A process (X_n, Y_n) is said to be a *semi-Markov sequence* [15] if

$$\mathrm{P}\{X_{n+1}, Y_{n+1} | X_n, Y_n, X_{n-1}, Y_{n-1}, \ldots\} = \mathrm{P}\{X_{n+1}, Y_{n+1} | X_n\}. \quad (25)$$

The first component of a semi-Markov sequence (X_n, Y_n), namely the (X_n) process, will be called here the *underlying Markov process*. In the present context, the second component, the (Y_n) process, is the abovementioned net input process.

We can in fact take $X_n = Y_n$ because the process $\{Y_n\}$ is actually Markov. Because the results we need come from the theory of semi-Markov queues we shall nevertheless maintain the distinction between X_n and Y_n.

Let $\mathcal{R}$ denote the sigma algebra of Borel sets on $\mathbf{R}$. We denote the state space of $\{X_n\}$ by Ξ (here $\Xi = \mathbf{R}$) and its sigma algebra by $\Sigma = \mathcal{R}$. Let H denote the kernel of the process (X_n, Y_n) [9, 11]. It is a function from $\Xi \times \Sigma \times \mathcal{R}$ to $\mathbf{R}$, defined by:

$$H(x, A \times B) = \mathrm{P}\{X_{n+1} \in A, Y_{n+1} \in B | X_n = x\}, \quad x \in \Xi, \quad A \in \Sigma, \quad B \in \mathcal{R}. \quad (26)$$

This kernel is required to have the following "measurability" properties:

(i) $H(x, C)$ is measurable considered as a function of x to $\mathbf{R}$, for any $C \in \Sigma \times \mathcal{R}$,

(ii) $H(x, C)$ is a positive measure when considered as a function of $C \in \Sigma \times \mathcal{R}$, for any $x \in \mathbf{R}$.

Kernels of this type are known as *semi-Markov* kernels.

The Laplace transform of a semi-Markov kernel H is the Markov kernel-valued function $\widehat{H}(s)$:

$$\widehat{H}(s)(x, A) = \int_{A \times \mathbf{R}} H(x, dy \times d\xi) e^{-s\xi}. \quad (27)$$

The reader is referred to [15] for definitions of products of measures, Markov, semi-Markov kernels, and functions. Now suppose

$$\Phi(s) = I - \widehat{H}(s)$$

has the *improper spectral factorization*

$$\Phi(s) = \Phi_+(s)\Phi_-(s), \quad s \in \mathbf{C}, \quad (28)$$

where $\mathbf{C}$ is the complex plane. By saying that (28) is an *improper spectral factorization*, we mean that $\Phi_+(s)$ and its inverse are analytic and bounded as a function of s in $\Re(s) \geq \delta_-$ for some $\delta_- < 0$, and $\Phi_-(s)$ is analytic and bounded as a function of s in $\Re(s) \leq \delta_+$ for some $\delta_+ > 0$. As a normalization condition, we further assume that $\Phi_+(0) = I$. A *proper* spectral factorization would also have $\Phi_-^{-1}(s)$ analytic and bounded in $\Re(s) \leq \delta_+$. This condition cannot hold in the present case because $\Phi_-(s)$ is singular at the origin.

Denote the stationary distribution of the underlying Markov process by π. We shall assume that $\{X_n\}_{n\geq 0}$ is an ergodic Markov process, so π is the unique solution of the equation $\pi\widehat{H}(0) = \pi$. Then the solution for the stationary distribution which we seek [9] (formally equivalent solutions are also presented in [10, 11, 12] with a different definition of *spectral factorization*) may be expressed in the form

$$\mathrm{E}\{e^{-V_\infty s}; X_\infty \in A\} = \left(\pi\Phi_+^{-1}(s)\right)(A), \quad \Re(s) \geq 0, \ A \in \Sigma, \quad (29)$$

where we define $\mathrm{E}\{R;B\} = \mathrm{E}\{R\,\chi_B\}$, for any random variable R and event B, with χ_B denoting the function (random variable) which takes the value 1 on B and equals zero elsewhere.

It is shown in [15] that s^*, as defined by (13) (which applies to a large class of semi-Markov queues), can be computed as the solution, other than zero, of

$$\sigma_{\mathrm{rad}}(\widehat{H}(s)) = 1, \qquad (30)$$

where $\sigma_{\mathrm{rad}}(A)$ denotes the spectral radius of A. In the present case $\widehat{H}(s)$ is a positive operator and so [19] the spectral radius of $\widehat{H}(s)$ equals the magnitude of its largest eigenvalue. Let us denote the spectral radius of $\widehat{H}(s)$ by $\alpha(s)$. It can be shown to be a convex function of s. From the Perron-Frobenius theory of positive kernels [19], for any s, a value $\alpha(s)$ which satisfies

$$\nu(s)\widehat{H}(s) = \alpha(s)\nu(s) \qquad (31)$$

where $\nu(s)$ is a positive measure on the state space of X_n, must be the maximal eigenvalue of $\widehat{H}(s)$, and therefore its spectral radius. Note that $\alpha(s) = 1$ when $s = 0$, it is convex in its entire domain (this will be shown in the next section), it has slope $E\{Y_\infty\}$ at the origin [20] (the argument in [20] is given for semi-Markov sequences with finite state-space, but it applies without change in the present context), which is positive whenever the queue is stable, and because there is a positive probability that $Y_n > 0$ $\alpha(s) \to \infty$ as $s \to -\infty$. From these properties it follows that equation (30) must have one additional solution for some value of $s < 0$.

5 DETERMINATION OF $\alpha(s)$ AND s^*

The Gaussian measure on $\mathbf{R}$ with mean u and variance w will be denoted by $\gamma_{u,w}$, i.e. for any measurable set $A \subseteq \mathbf{R}$,

$$\gamma_{u,w}(A) = \frac{1}{\sqrt{2\pi w}} \int_A e^{\frac{-(x-u)^2}{2w}}\,dx.$$

The process $(X_n, Y_n)_{n\geq 0}$ is semi-Markov with semi-Markov kernel H of the form:

$$H(x, A\times B) \triangleq P\{X_n \in A, Y_n \in B | X_{n-1} = x\}$$

Now since $X_n = Y_n$ for all n, the probability measure $H(x,\cdot)$ is concentrated on the diagonal (y,y) $y \in \mathbf{R}$. The probability that $(X_n, Y_n) \in A\times B$ is therefore the same as the probability that $X_n \in A \cap B$. Furthermore, by (1) and (2), $X_n - aX_{n-1} = Y_n - aY_{n-1} = b\widetilde{U}_n - (1-a)\tau$ has a Gaussian distribution with mean $\mathrm{E}\{b\widetilde{U}_n\} - (1-a)\tau$ and variance $\mathrm{VAR}\{b\widetilde{U}_n\} = b^2\tilde{\sigma}^2$. Thus

$$H(x, A\times B) = \gamma_{u_1,w_1}((A\cap B) - ax), \quad A \subseteq \mathbf{R},\ B\subseteq \mathbf{R},$$

where $u_1 = b\eta - (1-a)\tau$ and $w_1 = b^2\tilde{\sigma}^2$.

Directly from the definition, the Laplace transform of H takes the form:

$$\begin{aligned}\widehat{H}(s)(x,A) &= \int_A \int_{-\infty}^{\infty} \gamma_{u_1,w_1}((y-ax, y-ax+dy) \\ &\qquad \cap (u-ax, u-ax+du))e^{-su} \\ &= \int_A \gamma_{u_1,w_1}((y-ax, y-ax+dy))e^{-sy}\end{aligned}$$

so

$$\widehat{H}(s)(x,(y,y+dy)) = \gamma_{u_1,w_1}((y-ax, y-ax+dy))e^{-sy}, \quad x \in \mathbf{R}.$$

Proposition 5.1 *The equation:*

$$\nu(s)\widehat{H}(s) = \alpha(s)\nu(s), \qquad s\in(-\infty, 0], \qquad (32)$$

has the solution:

$$\begin{aligned}\alpha(s) &= e^{\frac{s^2b^2\tilde{\sigma}^2}{2(1-a)^2} - s\left(\frac{\eta b}{1-a} - \tau\right)}, \\ \nu(s) &= \gamma_{u_2,w_2},\end{aligned}$$

where $u_2 = \frac{\eta b - \frac{sb^2\tilde{\sigma}^2}{1-a^2}}{1-a} - \tau$ *and* $w_2 = \sigma^2 = \frac{b^2\tilde{\sigma}^2}{1-a^2}$, *for any value of* $s \leq 0$. *It follows that*

$$s^* = \frac{2(1-a)(\rho-1)\tau}{(1+a)\sigma^2}.$$

Notice that $\nu(s)$ is a positive measure and $\alpha(s)$ is a convex function of s.

Proof

To prove (32), it will suffice to substitute for $\alpha(s)$ and $\nu(s)$ and check that equality holds. Now

$$\nu(s)\widehat{H}(s)(A) = \int_{-\infty}^{\infty} \nu(s)(dx)\widehat{H}(s)(x,A)$$

$$= \int_{-\infty}^{\infty} \int_A \nu(s)(dx)\gamma_{u_1,w_1}((y-ax, y-ax+dy))e^{-sy}$$

$$= \int_A \int_{-\infty}^{\infty} \nu(s)(dx)\gamma_{u_1,w_1}((y-ax, y-ax+dy))e^{-sy}.$$

The inner integral here is just the convolution of two Gaussian measures, the first with probability measure $\nu(s)(dx/a)$ (i.e. the probability measure of a random variable Z_0 which is defined to equal aZ_1 where Z_1 is any random variable with probability measure $\nu(s)$) and the second with probability measure γ_{u_1,w_1}, so

Table 1 — Analytic versus simulation results for the mean and 99th percentile of V_∞ [kbyte].

	mean		99th percentile	
utilization	approximation	simulation	approximation	simulation
0.30	0.0	0.0	0.0	0.0
0.40	0.0041	0.0025	0.0	0.0
0.50	0.268	0.163	1.89	2.9
0.60	2.76	2.03	74.6	50.0
0.70	13.0	10.5	180.2	158.4
0.80	44.7	41.0	378.0	360.8
0.90	160.5	157.4	955.0	991.5
0.98	1147.8	1232	5531.2	6160

$$\nu(s)\widehat{H}(s)(A) = \int_A \gamma_{u_3,w_3}(dy)e^{-sy}$$

where

$$\begin{aligned} u_3 &= au_2 + u_1 \\ &= au_2 + b\eta - (1-a)\tau \\ &= a\frac{\eta b - \frac{sb^2\tilde{\sigma}^2}{1-a^2}}{1-a} - a\tau + b\eta - (1-a)\tau \\ &= \frac{\eta b - \frac{asb^2\tilde{\sigma}^2}{1-a^2}}{1-a} - \tau \\ &= u_2 + s\frac{b^2\tilde{\sigma}^2}{1-a^2} \end{aligned}$$

and

$$\begin{aligned} w_3 &= a^2w_2 + w_1 = a^2\frac{b^2\tilde{\sigma}^2}{1-a^2} + b^2\tilde{\sigma}^2 \\ &= \frac{b^2\tilde{\sigma}^2}{1-a^2} = \sigma^2 \end{aligned}$$

(see (7)), so

$$\begin{aligned} \nu(s)\widehat{H}(s) &= e^{-su_3+s^2w_3/2}\gamma_{u_3-sw_3,w_3} \\ &= e^{-s\left(u_2+\frac{sb^2\tilde{\sigma}^2}{1-a^2}\right)+\frac{s^2b^2\tilde{\sigma}^2}{2(1-a^2)}}\nu(s) \\ &= e^{-\frac{s\left(\eta b-\frac{sb^2\tilde{\sigma}^2}{1-a^2}\right)}{1-a}+s\tau-\frac{s^2b^2\tilde{\sigma}^2}{2(1-a^2)}}\nu(s) \\ &= e^{-\frac{s\left(\eta b-\frac{sb^2\tilde{\sigma}^2}{2(1-a)}\right)}{1-a}+s\tau}\nu(s) \\ &= e^{\frac{s^2b^2\tilde{\sigma}^2}{2(1-a)^2}-s\left(\frac{\eta b}{1-a}-\tau\right)}\nu(s) \\ &= \alpha(s)\nu(s) \end{aligned}$$

as required.

This result provides a positive eigenvector of $\widehat{H}(s)$ for every non-positive value of s. Because $\widehat{H}(s)$ is a positive operator for all such values of s, it follows [19] that for any s in $(-\infty, 0]$, $\alpha(s)$ is the spectral radius of $H(s)$. Inspection of the formula for $\alpha(s)$ shows that it is convex, has positive slope at the origin, and $= 1$ for just two values in the region $(-\infty, 0]$, the origin and s^* satisfying

$$\frac{s^*b^2\tilde{\sigma}^2}{2(1-a)^2} = \frac{\eta b}{1-a} - \tau$$

$$\begin{aligned} \Rightarrow \quad s^* &= \frac{2(1-a)(\eta b - \tau(1-a))}{b^2\tilde{\sigma}^2} \\ &= \frac{2(1-a)(\rho-1)\tau}{(1+a)\sigma^2}. \end{aligned}$$

□

6 NUMERICAL RESULTS

To examine the accuracy of our approximation we use the autoregressive video telephony traffic model as proposed in [1]. We also use the parameters computed there based on a video telephony experiment. We compute the mean and the 99th percentile of the unfinished work in an infinite buffer queue fed by this traffic model, using Eqs. (19) and (20), and compare the results with those obtained by simulation.

In particular, each frame in our model will coincide with a video frame of [1], and we set $a = 0.8781$ and $b = 0.1108$. The A_n of Eq. 1 will represent here the amount of work measured in kbytes arriving for service during the n th frame. In [1], A_n is measured in bit/pixel and $\tilde{\sigma}$, the standard deviation of $\tilde{U}$, equals 1. To be consistent with [1] our $\tilde{\sigma}$ should be equal to the conversion factor from bit/pixel to kbyte/frame. Using 250,000 pixels per frame we obtain: $\tilde{\sigma} = 250,000/8000 = 31.25$. Also, considering the conversion factor, we have that the average number of cells arriving within a frame is given by $E\{A_\infty\} = 0.52 \times 31.25 = 16.25$.

As the arrival process is fixed for all simulation runs, we shall use different values for the service rate τ to allow for different utilization values.

A comparison of simulation results and approximate results is given in Table 1. A target of 30% for the range (diameter) of the 95% confidence intervals was used in the simulations. The approximate results are of the same order of magnitude as the simulation results for the entire range of utilization. For utilization of 0.8 and over the analytical results are well within the confidence interval of the simulation.

REFERENCES

[1] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson and J. Robbins, "Performance models of statistical multiplexing in packet video communications", *IEEE Trans. Commun.*, vol. 36, no. 7, July 1988, pp. 834–844.

[2] O. Gihr and P. Tran-Gia, "A layered description of ATM cell traffic streams and correlation analysis", *ATR*, vol. 24, no. 2, 1990, pp. 9–18.

[3] A.A. Fredericks, "A class of approximations for the waiting time distribution in a GI/G/1 queueing system," *Bell Syst. Tech. J.*, Vol. 61, No. 3, pp. 295–325.

[4] J.F.C. Kingman, "The single server queue in heavy traffic," *Proc. Camb. Phil. Soc.*, Vol. 57, 1961, pp. 902–904.

[5] J.F.C. Kingman, "On queues in heavy traffic," *J. Roy. Statist. Soc.*, Vol. B24, 1962, pp. 383–392.

[6] H.D. Miller, "A convexity property in the theory of random variables defined on a finite Markov chain", *Ann. Math. Statist.*, vol. 32, 1961, pp. 1260–1270.

[7] H.D. Miller, "A matrix factorization problem in the theory of random variables defined on a finite Markov chain", *Proc. Cambridge Philos. Soc. math. phys.*, vol. 58, 1962, pp. 268–285.

[8] H.D. Miller, "Absorption probabilities for sums of random variables defined on a finite Markov chain", *Proc. Cambridge Philos. Soc. math. phys.*, vol. 58, 1962, pp. 286–298.

[9] R.G. Addie, *Theory and applications of discrete time Markov additive processes*, PhD thesis, Monash University, August 1986.

[10] E. Arjas, "On a fundamental identity in the theory of semi-Markov processes", *Advances in Applied Probability*, vol. 4, 1972, pp. 258–270.

[11] E. Arjas, "On the use of a fundamental identity in the theory of semi-Markov queues", *Advances in Applied Probability*, vol. 4, 1972, pp. 271–284.

[12] E. Arjas and T.P. Speed, "Symmetric Wiener-Hopf factorizations in Markov-additive processes", *Z. Wahrscheinlichkeitstheorie und Verwandt Gebeite*, vol. 26, 1973, pp. 105–118.

[13] E.L. Presman, "Factorization methods and boundary problems for sums of random variables given on a Markov chain", *Math. USSR Izvestiya*, vol. 3, no. 4, 1969, pp. 815–852.

[14] R. Addie and M. Zukerman, "A Gaussian traffic model for statistical multiplexers", Proceedings, *GLOBECOM '92*, Florida, December 1992.

[15] R. Addie and M. Zukerman, "An approximation for performance evaluation of stationary single server queues", Proceedings, *INFOCOM '93*, San Francisco, March–April 1993.

[16] R.G. Addie and M. Zukerman, "Analysis of connection admission control and multiplexing gain in a B-ISDN based on a Gaussian traffic model", Proceedings, *ABSSS '93*, vol. 1, Wollongong, Australia, July 1993, pp. 1-8.

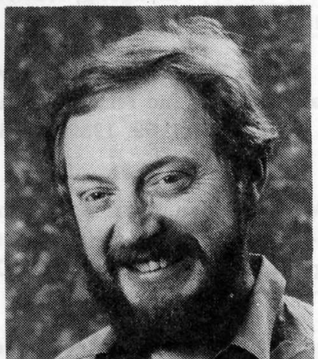
[17] H. Heffes, "A class of data traffic processes – covariance function characterization and related queueing results", *Bell Syst. Tech. J.*, vol. 59, 1980, pp. 897–929.

[18] R.G. Addie and M. Zukerman, "An approximate solution of a semi-Markov queue with large state space", Proceedings, *13th ITC*, Copenhagen, Denmark, June 1991.

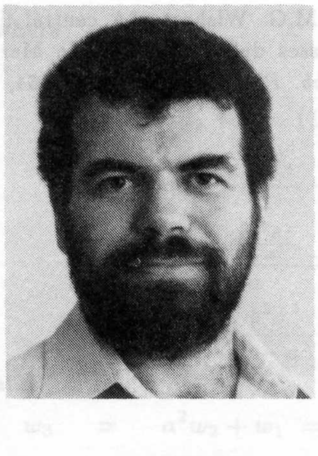
[19] E. Nummelin, *General irreducible Markov chains and non-negative operators*, Cambridge University Press, 1984.

[20] J. Keilson and D.M.G. Wishart, "A central limit theorem for processes defined on a finite Markov chain", *Proc. Camb. Phil. Soc.*, vol. 60, 1964, pp. 547–567.

BIOGRAPHIES



RON ADDIE was born in Melbourne, Australia, in 1950. He received his BSc degree from Monash University in 1972. He completed a PhD at Monash University in the area of semi-Markov queues in 1986. From 1972 to 1992, he worked in the Telecom Australia Research Laboratories where he was involved in research in digital communications, teletraffic, and network analysis and design. He was also a contributor to the concept of virtual path which has since been adopted as part of the B-ISDN standard. In 1992 Ron moved to the University of Southern Queensland, where he is a member of the faculty of science in the department of mathematics and computing. His current research interests include queueing theory, performance evaluation of telecommunications systems and time series modelling.



MOSHE ZUKERMAN received his B.Sc degree in Industrial Engineering and Management and his M.Sc degree in Operations Research from the Technion, Israel. He received his Ph.D degree in Electrical Engineering from the University of California Los Angeles (UCLA) in 1985. During 1985–1986, he held a visiting position in the Electrical Engineering Department at UCLA and he served as a consultant to the IRI corporation. In 1986, he joined the Research Laboratories of Telecom Australia where he is presently a Project Leader in the Network Analysis Section of the Switched Networks Branch. Dr. Zukerman is a member of the Telecommunication Society of Australia, a senior member of the IEEE Communications Society and an editor of the Australian Telecommunication Research Journal. He was awarded the Telecom Research Laboratories Outstanding Achievement award in 1990. His current research interests are: traffic modelling, performance evaluation and enhancements of telecommunications networks and multiple access protocols.