

# An Approximation for Performance Evaluation of Stationary Single Server Queues

Ronald G. Addie

The University of Southern Queensland  
Darling Heights  
Queensland 4350, Australia

Moshe Zukerman

Telecom Research Laboratories  
770 Blackburn Rd., Clayton  
Victoria 3168, Australia

## Abstract

*This paper provides a method for approximating the probability distributions of stationary statistics in FIFO single server queues. The method is based on the Wiener Hopf factorization technique, and is applied to semi-Markov queues where the underlying state space is of unlimited size. A particularly important case is a model for a statistical multiplexer where the net input process forms a stationary ergodic Gaussian discrete-time stochastic process.*

## 1 Introduction

Stationary FIFO single server queues have been used to model a wide range of telecommunications systems and networks serving various bursty traffic streams. Unfortunately, however, the complexity arising from the need to represent real life situations often leads to models with a large (or infinite) underlying state space, which are analytically and computationally intractable by conventional methods. In this paper, we introduce a method for analyzing such queues. The method leads, in some cases, to simple, closed form formulae which are sufficiently accurate for dimensioning purposes.

It has been suggested by Maglaris et al. [15] that an autoregressive process, which may be viewed as a semi-Markov sequence with an underlying process of infinite state space, is a good model for Variable Bit-Rate (VBR) video telephony. However, it is claimed in [10] and [15], that no analytical solution is available for queues involving autoregressive processes. This queueing problem is a special case of the one solved approximately in this paper.

Problems related to bursty traffic characterization and performance evaluation of statistical multiplexers have

been considered amongst the most important teletraffic problems in the last 15 years. See for example [9], [10], [12], [14], [15], [20], [24] and references therein. The traffic models that have been proposed so far have been based on renewal, batch Poisson, and Markov modulated Poisson processes, their discrete counterparts, and on fluid flow approximations. However, none of these models is general enough to capture the full range of second order statistics (i.e. variance and autocovariance), and at the same time lead to explicit performance results.

A general stationary single server queue can be modelled as a semi-Markov queue with an underlying Markov process with infinite state space. Where necessary, such an underlying process can contain information about the entire history of the queue. The analysis of such semi-Markov queues, presented in this paper, can get us closer to solutions for queues fed by a general stationary process which in turn can be applied to the open problems described above.

A general solution for the stationary waiting time distribution for semi-Markov queues based on spectral factorization was originally developed by Miller [16,17,18]. Miller considered cases with finite underlying state space. This work has been extended in [1,5,6,7] to include cases where the state space of the underlying Markov process is infinite. However, no indication of computational methods is given in [5,6,7], while the method in [1] is seldom practical for problems with a large state space.

It has been observed in [18,23] 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. In this paper we extend this result to semi-Markov queues where the underlying state space is infinite. In

[4], an approximate formula for the coefficient  $c$  was derived and checked by means of simulation. In this paper a new method for evaluating  $c$  is introduced. An important application of the method presented here is to a single server queue where the *net input process* to the queue forms a stationary ergodic Gaussian discrete-time process. The net input process is defined, in this paper, as the difference between the arrival process and the process representing the capacity of the server to render service. We derive an exact analytical solution for  $s^*$  and an approximation for  $c$  based on the new method. Numerically, there is little difference between the formula in [4] and the formula presented here, however the new formula is simpler.

## 2 The Model

We consider a single server FIFO queue with an infinite buffer. Let time be divided into consecutive intervals. The endpoints of these intervals are referred to as epochs. We say that the  $n$ th and the  $n+1$ st epochs are the endpoints of the  $n$ th interval. The epochs should correspond to sampling times that see time averages, or to all relevant time instants, e.g. the natural points of arrivals.

A process  $(X_n, Y_n)$  is said to be a *semi-Markov sequence* [19] if

$$\begin{aligned} P\{X_{n+1}, Y_{n+1} | X_n, Y_n, X_{n-1}, Y_{n-1}, \dots\} \\ = P\{X_{n+1}, Y_{n+1} | X_n\}. \end{aligned}$$

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.

Let  $V_n$  represent the unfinished work at the beginning of the  $n$ th sampling interval. Usually, the unfinished work process satisfies the recurrence relation:

$$V_{n+1} = (V_n + Y_n)^+, \quad n \geq 0, \quad (1)$$

where  $V_0 = 0$  and  $Y_n$  is the net input process given by

$$Y_n = A_n - B_n, \quad n \geq 0, \quad (2)$$

in which  $A_n$  denotes the additional new work arriving during the  $n$ th interval (or the  $n$ th customer service time), and  $B_n$  denotes the maximum work which can be processed during the  $n$ th interval (or inter-arrival time between the  $n$ th and the  $n+1$ st customers). A queue in which  $V_n$  satisfies (1) for some semi-Markov sequence  $(X_n, Y_n)_{n \geq 0}$  is referred to as a *semi-Markov queue*. For any stochastic process  $\{R_n\}_{n \geq 0}$ ,  $P\{R_\infty \in A\}$  denotes the stationary probability of the

events  $\{R_n \in A\}$ . Accordingly, the random variable  $V_\infty$  represents the unfinished work in steady state. As usual, define the utilization by  $\rho = E\{A_\infty\}/E\{B_\infty\}$ . We require for stability that  $E\{Y_\infty\} < 0$  or equivalently  $\rho < 1$ .

In this paper we provide a method to obtain the statistics of  $V_\infty$ . Because the recurrence (1) is also applicable to the *queue size* process with  $A_n$  ( $B_n$ ) as the number of customers arriving (can be served) during the  $n$ th interval, an equivalent treatment can be applied to queue size analysis by  $z$ -transform [1].

A general solution for the stationary distribution of  $V_n$  in a semi-Markov queue is given in the next section. This solution is very often extremely difficult to compute, however it leads to the useful approximate solution described in section 4.

## 3 Theoretical Background

In this section we provide background on the theory of Markov and semi-Markov kernels and spectral factorization of Laplace transforms of semi-Markov kernels as applied to the distribution of  $V_\infty$ .

### 3.1 Markov and Semi-Markov Kernels

Denote the state space of  $X_n$  by  $\Xi$ , with associated sigma algebra of measurable sets  $\Sigma$ . Let us denote by  $\mathcal{K}$  the vector space of all bounded kernels, defined on  $\Xi$ . That is to say, any  $K \in \mathcal{K}$  is a function from  $\Xi \times \Sigma$  to  $\mathbb{R}$  (the real line) with the properties that for any  $x \in \Xi$ ,  $K(x, \cdot)$  is a measure on  $\Xi$ , and for any set  $A \in \Sigma$ ,  $K(\cdot, A)$  is a measurable function on  $\Xi$ , and

$$\|K\| = \sup_{x \in \Xi} \|K(x, \cdot)\| < \infty,$$

where the  $\|\cdot\|$  on the Right Hand Side (RHS) denotes total variation (see [11] for a definition of total variation).

If  $K \in \mathcal{K}$  and  $f$  is a bounded, real-valued, measurable function on  $\Xi$ ,  $Kf$  is defined to equal the bounded, real-valued, measurable function

$$x \mapsto \int_{\Xi} K(x, dy) f(y).$$

Here  $x \mapsto g(x)$  is used to denote the function which maps  $x$  to  $g(x)$ , and this integral, and all others in this paper, are Lebesgue integrals [11]. If  $K \in \mathcal{K}$  and  $\mu$  is a finite, signed measure on  $\Xi$ ,  $\mu K$  is defined to equal the signed measure  $A \mapsto \int_{\Xi} K(x, A) \mu(dx)$ .

These definitions allow us to interpret any kernel as a linear operator on the space of bounded measurable functions and on the space of finite signed measures. Notice that we write the kernel on the right of the operand when it is operating on a measure, and on

the left when it is operating on a bounded measurable function.

A bounded measurable function  $f$  induces a linear functional on the space of finite signed measures by the mapping:  $\mu \mapsto \mu f \triangleq \int f d\mu$ . Note that this mapping can be interpreted as defining a product operation between measures and measurable functions.

Finally, we define the product of two kernels,  $K$  and  $L$ , by  $KL(x, A) = \int_{\Xi} K(x, dy)L(y, A)$ .

With this product,  $\mathcal{K}$  forms a Banach algebra [1]. Because all the integrals in question are absolutely convergent, the product operations just defined amongst kernels, measures and functions is associative. We therefore write, for example,  $\mu K f$  instead of either  $(\mu K)f$  or  $\mu(Kf)$ .

The function  $Q(x, A) = P\{X_{n+1} \in A | X_n = x\}$  is a kernel in  $\mathcal{K}$  referred to as the kernel of the Markov process  $\{X_n\}$ . Such kernels are called *Markov kernels*. Using the kernel product just defined, the kernel  $Q^j(x, A)$  has the interpretation

$$P\{X_{n+j} \in A | X_n = x\}.$$

The kernel  $H$  of the semi-Markov process  $(X_n, Y_n)$  is defined by:

$$H(x, A \times B) = P\{X_{n+1} \in A, Y_{n+1} \in B | X_n = x\}.$$

It is a function from  $\Xi \times \Sigma \times \mathcal{R}$  to  $\mathbb{R}$  with the natural measurability properties, in which  $\mathcal{R}$  denotes the sigma algebra of Borel sets on  $\mathbb{R}$ . Kernels of this type are known as *semi-Markov kernels*. The product of two semi-Markov kernels  $H$  and  $J$  is defined by:

$$HJ(x, A \times B) = \int_{\Xi \times \mathbb{R}} H(x, dy \times dz) J(y, A \times (B - z)).$$

The collection of all bounded semi-Markov kernels also forms a Banach algebra [1]. We say that a semi-Markov kernel  $H$  is concentrated on the set  $S \in \mathcal{R}$  if  $H(x, A \times U) = 0$  whenever  $U \cap S = \emptyset$  ( $\emptyset$  denotes the empty set). The Laplace transform of a semi-Markov kernel  $H$  is the kernel-valued function  $\hat{H}(s)$  such that

$$\hat{H}(s)(x, A) = \int_{A \times \mathbb{R}} H(x, dy \times d\xi) e^{-s\xi}. \quad (3)$$

The domain of this function includes the imaginary axis. There exist two real numbers  $\delta_- \leq 0$  and  $\delta_+ \geq 0$ , such that the integral in (3) is absolutely convergent for  $\Re(s)$  (the real part of  $s$ ) in the interval  $[\delta_-, \delta_+]$ , for all  $x \in \Xi$ ,  $A \in \Sigma$ . We shall assume that  $\delta_- \neq 0 \neq \delta_+$ ; this will be the case when the measures  $H(x, A \times \cdot)$  have tails which are uniformly dominated by negative exponential measures with coefficients  $\delta_-$  and  $\delta_+$ . This assumption holds for the example we consider in Section 5. Notice that  $\hat{H}(0) = Q$ , the kernel of  $\{X_n\}_{n \geq 0}$ .

### 3.2 A Formula for the Distribution of $V_\infty$

We say that a kernel valued function  $F(s)$  is analytic for  $s \in D$  if  $\mu F(s)f$  is analytic for  $s \in D$ , for any choice of measure  $\mu$  and bounded measurable function  $f$ . Now suppose  $\Phi(s) = I - \hat{H}(s)$  has the spectral factorization

$$\Phi(s) = \Phi_+(s)\Phi_-(s), \quad \delta_- \leq \Re(s) \leq \delta_+ \quad (4)$$

in which  $\Phi_+(s)$  and its inverse is analytic and bounded as a function of  $s$  in  $\Re(s) \geq \delta_-$  (i.e., the right half of the complex plane plus a strip), and  $\Phi_-(s)$  is analytic and bounded as a function of  $s$  in  $\Re(s) \leq \delta_+$  (i.e., the left half of the complex plane plus a strip). We shall refer to the properties of the function  $\Phi_+(s)$  in the region  $\Re(s) < \delta_-$  also. In this case, we are referring to the analytic continuation of  $\Phi_+(s)$  in this region. As a normalization condition, we further assume that  $\Phi_+(0) = I$ .

Normally a spectral factorization would also have the properties that  $\Phi_-^{-1}(s)$  is analytic and bounded as a function of  $s$  in  $\Re(s) \geq \delta_-$ . In the present case  $\Phi_+^{-1}(s)$  has this sort of property but  $\Phi_-^{-1}(s)$  does not because it has a pole at the origin. This minor transgression does not inhibit the Wiener-Hopf method [22] from applying in the present case.

An alternative characterization of this factorization is that  $\Phi_+(s)$  is the Laplace transform of a semi-Markov kernel which is concentrated on  $[0, \infty)$ , with an inverse which is also concentrated on  $[0, \infty)$ , and  $\Phi_-(s)$  is the Laplace transform of a semi-Markov kernel concentrated on  $(-\infty, 0]$ .

Denote the stationary distribution of the underlying Markov process by  $\pi$ . We shall assume that  $\{X_n\}_{n \geq 0}$  is an ergodic Markov process, so  $\pi$  is the unique solution of the equation  $\pi \hat{H}(0) = \pi$ . Then the solution which we seek [1] (formally equivalent solutions are also presented in [5,6,7] with a different definition of spectral factorization) may be expressed in the form

$$E\{e^{-V_\infty s}; X_\infty \in A\} = (\pi \Phi_+^{-1}(s))(A), \quad \Re(s) \geq 0, \quad A \in \Sigma, \quad (5)$$

where we define  $E\{R; B\} = E\{RX_B\}$ , for any random variable  $R$  and event  $B$ , with  $X_B$  denoting the function (random variable) which takes the value 1 on  $B$  and equals zero elsewhere. Notice that the solution is expressed purely in terms of  $\Phi_+^{-1}(s)$ . Note also that although the domain of (5) is  $\Re(s) \geq 0$ , the singularities of  $\Phi_+^{-1}(s)$  which lies in  $\Re(s) < 0$  will prove to be of critical significance in the sequel.

### 3.3 Elementary Factors

It is shown in [1] that if  $\Phi(s)$  is rational, both spectral factors  $\Phi_+(s)$  and  $\Phi_-(s)$  can be represented as a product of *elementary factors*. We shall refer to this product as the *product representation* of the function  $\Phi_+(s)$  or  $\Phi_-(s)$ . An elementary factor is a kernel valued function  $F(s)$  which has only one pole, at  $\theta$  say, and for which the space  $S = \{f : \text{for any measure } \mu, \mu F(s) \text{ is bounded at } \theta \Rightarrow \mu f = 0\}$  is one-dimensional. Note that the product  $\mu f$  is defined in Subsection 3.1. An elementary factor can have at most one zero, where a zero is defined to be a complex number  $\zeta$  such that  $F(\zeta)$  is singular. Furthermore, in this case, the dimension of the space

$T = \{\mu : \mu F(s) = 0\}$  must also be one. The space  $S$  is referred to as the characteristic space of the pole of  $F$  at  $\theta$ , and the space  $T$  is referred to as the characteristic space of the zero of  $F$  at  $\zeta$ . The definition of the characteristic space of a pole or zero applies to any kernel valued function and not just an elementary factor [1].

The pole or the zero or both can be at  $\infty$  or 0, but in this paper we can restrict consideration to the case where  $0 < |\theta|, |\zeta| < \infty$  and  $\theta \neq \zeta$ . In this case, an elementary factor which takes the value  $I$  at the origin takes the form

$$F(s) = I - \frac{(\zeta - \theta)sf \otimes \mu}{\zeta(s - \theta)\mu f}, \quad s \neq \theta, \quad (6)$$

where  $f \in S$  and  $\mu \in T$ . Since  $S$  and  $T$  are one-dimensional the choice of  $f$  and  $\mu$  does not affect  $F$  because it is normalized by the factor  $\mu f$  (defined in §3.1) in the quotient of its definition. The notation  $f \otimes \mu$  denotes the tensor product of  $f$  and  $\mu$  which is defined in this context to be the kernel

$$f \otimes \mu(x, A) = f(x)\mu(A).$$

An elementary factor is uniquely identified, up to multiplication by a kernel, by the location and characteristic spaces of its zero and pole. The inverse of an elementary factor is also an elementary factor, and it is useful to note that the inverse of a function such as  $F$  in (6) can be obtained simply by interchanging  $\theta$  and  $\zeta$ .

The kernels we consider in this paper are not rational and so we cannot expect  $\Phi_+^{-1}$  to be representable as a finite product of elementary factors. Nevertheless, it will be useful to express  $\Phi_+^{-1}$  as a product of an elementary factor and another *non elementary* factor. If the elementary factor is chosen appropriately it can be regarded as an approximation to  $\Phi_+^{-1}$ .

## 4 The Tail Distribution of $V_\infty$

The exact spectral factorization can seldom be obtained. Nevertheless, the factorization can be used to derive an approximation in several ways. The technique we introduced in [2,3] is suitable for queues with high load. In this section we describe an extension of this approach which is, in the cases we have investigated numerically, sufficiently accurate, for dimensioning purposes, over the full range of server utilization. As discussed in the introduction, it is known that under quite general conditions  $P\{V_\infty \geq t\}e^{-s^*t} \rightarrow c$  as  $t \rightarrow \infty$  for some real number  $c$ . (Note that  $s^*$  is negative.) In this section we outline a proof of this limit, and we show how  $s^*$  can be computed. We shall also describe by a moment matching argument for calculating  $c$ .

In [2,3] we used a single, dominant, elementary factor from a product representation of  $\Phi_+^{-1}(s)$  as an approximation for  $\Phi_+^{-1}(s)$ . In this paper we still use an elementary factor with a pole of similar form (same location and same characteristic space) to the zero of  $\Phi_+(s)$  which lies closest to 0. This zero is referred to as the dominant zero. Note that a zero of  $\Phi_+(s)$  becomes a pole in  $\Phi_+^{-1}(s)$  and conversely. The exact form of the zero of  $\Phi_+(s)$  (its location and its characteristic space) can be determined by examination of  $\Phi(s)$ ; this means that we do not need to compute  $\Phi_+(s)$  in order to completely characterize its zero.

In order to fully define the elementary factor, denoted here by  $E(s)$ , we also need information about its zero. In [2] this information was obtained by matching the zero to a pole of  $\Phi(s)$ . In [4] a truncation of a series expression for  $\Phi_+^{-1}(s)$  was used to choose the parameters of this zero. In this paper a moment matching argument is used to complete the definition of  $E(s)$ .

### 4.1 An Equation for $s^*$

In the finite dimensional case it has been shown [16,17,18] that there is an isolated simple real zero of  $\Phi(s)$  (which has characteristic space of dimension 1) in the region  $\{s : \Re(s) < 0\}$  with the least value of  $|\Re(s)|$  among all other zeros.

The zero occurs when the largest eigenvalue of  $\hat{H}(s)$  (the Perron eigenvalue) takes the value 1, and hence coincides with a solution of

$$\sigma_{\text{rad}}(\hat{H}(s)) = 1, \quad (7)$$

where  $\sigma_{\text{rad}}(A)$  denotes the spectral radius (in the present case, the magnitude of largest eigenvalue) of  $A$ . This zero lies on the real axis. All this also holds

true when the state space is infinite so long as  $\hat{H}(s)$  satisfies appropriate regularity conditions – for example uniform recurrence [1] or recurrence of degree 2 (see [21] for definition) for all  $s \in (-\infty, 0]$ . The location of this zero is denoted by  $s^*$ .

Recalling that  $\Phi_-^{-1}(s)$  is bounded in  $\Re(s) < 0$ , by (4), this zero will also be the zero of  $\Phi_+(s)$  which lies closest to the imaginary axis, and hence the function  $\Phi_+^{-1}(s)$  will have a simple pole at this location with characteristic space of dimension 1 and the function  $s \mapsto (s - s^*)\Phi_+^{-1}(s)$  must be analytic in a region of the form  $\Re(s) > \xi$  for some  $\xi$  such that  $\Re(\xi) < s^*$ . We shall denote by  $\pi^*$  a vector of unit length in the characteristic space of the pole of  $\Phi_+^{-1}(s)$  at  $s^*$ . The existence of this pole of  $\Phi_+^{-1}(s)$  at  $s^*$  explains why this zero of  $\Phi(s)$  has a dominant influence on the tail of the distribution of  $V_\infty$ . Note that  $\pi^*$  is also a vector of unit length in the characteristic space of the zero of  $\Phi(s)$  at  $s^*$ , i.e.  $\pi^*$  satisfies  $\pi^* \Phi(s^*) = 0$ .

## 4.2 The Tail is Exponential

Let  $\mathbf{1}$  denote the function from  $\Xi$  to  $\mathbf{R}$  which is identically 1. The Laplace Transform of the distribution of  $V_\infty$  is  $\pi \Phi_+^{-1}(s) \mathbf{1}$ , so the Laplace transform of  $V(t) = P\{V_\infty \geq t\}$  is  $\hat{V}(s) = (1 - \pi \Phi_+^{-1}(s) \mathbf{1})/s = (1 - \pi \Phi_-^{-1}(s) \Phi_-(s) \mathbf{1})/s$ . Since  $\Phi_-(s)$  is analytic and bounded in  $\Re(s) \leq 0$ , and  $\Phi_+^{-1}(s)$  is analytic in  $\Re(s) > s^*$ , we can conclude that  $\hat{V}(s)$  has an isolated pole at  $s^*$ , it is analytic for  $\Re(s) > s^*$  and, like  $\Phi(s)$ ,  $\inf\{|\Re(s)| : s \neq s^* \text{ is a pole of } \hat{V}(s)\} > |s^*|$ . In other words, the other poles of  $\hat{V}(s)$  are further away from the imaginary axis than  $s^*$ . It follows that

$$\hat{V}(s) = \frac{c}{s - s^*} + \hat{g}(s) \quad (8)$$

for some real number  $c$  where  $\hat{g}(s)$  is the Laplace transform of some bounded function  $g(t)$  which is concentrated on  $[0, \infty)$ . The function  $\hat{g}(s)$  is analytic in  $\Re(s) \geq u$  for some  $u < s^*$ . By (8),  $V(t) = ce^{s^*t} + g(t)$  and from the basic theory of Laplace transforms [8],  $e^{-s^*t}g(t) \rightarrow 0$  as  $t \rightarrow \infty$ , and therefore

$$V(t)e^{-s^*t} \rightarrow c \text{ as } t \rightarrow \infty. \quad (9)$$

## 4.3 An Approximation for $c$

In [3] we simply used (in effect) 1 or  $\rho$  as an estimate for  $c$ . This was shown to be asymptotically accurate for high load because actually  $c \rightarrow 1$  as  $\rho \rightarrow 1$ . Simulation experiments indicate that this overestimates  $c$  for low load.

A method for improving on this estimate by using a truncation of a series for  $\Phi_+^{-1}(s)$  has been described in [4]. In this paper we describe a simpler method (a sort of moment matching technique) which seems to produce a formula of similar accuracy.

In the new method we substitute the approximate solution on the RHS in equation (1) to obtain approximate stationary distributions on the left and the right. By requiring that these two stationary distributions have the *same mean*, we obtain a single equation in a single unknown, which we can then solve to obtain an approximate value for the free parameter.

## 5 A Gaussian Queueing Model

In [4] we introduced the Gaussian traffic model and presented performance results for a single server queue fed by a Gaussian net input process. Here we derive the formula for  $s^*$  given in [4] and use a new approach to derive a formula for  $\bar{c}$ . The resulting formula is simpler but appears to give similar numerical values.

### 5.1 Model Description

Let  $\mathbf{Z}$  denote the set of integers, and assume that the net input process, denoted  $(Y_n)_{n \in \mathbf{Z}}$ , is a stationary, ergodic, Gaussian discrete-time stochastic process. Any such process has a one sided moving average representation [13]:

$$Y_n = m + \sum_{k=0}^{\infty} a_k U_{n-k}, \quad n \in \mathbf{Z}, \quad (10)$$

where the  $U_n$  are mutually independent Gaussian random variables with mean 0 and variance 1, and  $m$  and  $a_k$ ,  $k \geq 0$  are real numbers. It follows that for all  $n$ ,  $E\{Y_n\} = m$  and  $\sigma^2 \triangleq \text{Var}(Y_n) = \sum_{k=0}^{\infty} a_k^2$ . Based on [1], a sufficient and necessary condition for queueing stability is  $m < 0$ . Define the *autocovariance sum* as

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

It can be shown, using (10), that

$$S = \sum_{k=0}^{\infty} a_{k+1} \sum_{j=0}^k a_j. \quad (11)$$

### 5.2 Derivation of a Formula for $s^*$

Let  $X_n$  denote the infinite series  $(U_n, U_{n-1}, \dots)$ ,  $n \in \mathbf{Z}$ . Then  $(X_n, Y_n)_{n \in \mathbf{Z}}$  is a semi-Markov sequence. Let

$(V_n)_{n \in \mathbb{Z}}$  be the stationary process such that

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

By the theory outlined previously we can find an approximation for the tail of the distribution of  $(X_n, V_n)$  if we can solve (7), or, in particular, if we can find a solution of the equation:

$$\nu(s) \hat{H}(s) = \alpha(s) \nu(s) \quad (12)$$

where  $\nu(s)$  is a positive measure on the state space of  $X_n$ ,  $\alpha$  is a real convex function of  $s$  in some neighbourhood of 0, and  $\alpha(s) = 1$  in two places: when  $s = 0$ , and when  $s = s^*$ , for some  $s^* < 0$ .

Let  $\gamma_{b,\sigma}$  denote a Gaussian measure with mean  $b$  and standard deviation  $\sigma$ , let  $\gamma_b$  denote the Gaussian measure with mean  $b$  and variance 1. Let  $\delta_y$  be the measure on  $\mathbb{R}$  with unit mass concentrated at  $y$ . Accordingly,  $\delta_y(U) = 1$  if  $y \in U$ , 0 otherwise. For any measure  $\mu$ , on  $\mathbb{R}$ , define  $\mu^{(u)}$  by

$$\mu^{(u)}(A) = \int_A e^{ux} \mu(dx). \quad (13)$$

In the case where  $\mu = \gamma_{b,\sigma}$  we obtain the more explicit result:

$$\gamma_{b,\sigma}^{(u)} = e^{bu+u^2\sigma^2/2} \gamma_{b+u\sigma^2,\sigma}. \quad (14)$$

For any measure  $\phi$  on  $\mathbb{R}^\omega$ ,  $\phi \hat{H}(s)$  is, by definition, the measure

$$B \mapsto \int_{\mathbb{R}^\omega} E\{e^{-sY_1}; X_1 \in B | X_0 = x\} \phi(dx), \quad B \in \mathcal{R}^\omega. \quad (15)$$

In the case where  $\phi$  is the unit mass concentrated on the point  $(x_0, x_1, \dots) \in \mathbb{R}^\omega$ , we have

$\phi \hat{H}(s) = \psi$  such that  
 $\psi(C_0 \times C_1 \times \dots) = E\{e^{-sY_1}; U_1 \in C_0, U_0 \in C_1, U_{-1} \in C_2, \dots | U_0 = x_0, U_{-1} = x_1, \dots\}$   
 $= E\{\exp\{-s(m + a_0 U_1 + \sum_{k=1}^\infty a_k x_{k-1})\}; U_1 \in C_0, U_0 \in C_1, U_{-1} \in C_2, \dots | U_0 = x_0, U_{-1} = x_1, \dots\}$   
 $= e^{-s(m + \sum_{k=1}^\infty a_k x_{k-1})} E\{e^{-sa_0 U_1}; U_1 \in C_0\}$   
 $\delta_{x_0}(C_1) \delta_{x_1}(C_2) \dots$ . So, using (13),  
 $\psi = e^{-s(m + \sum_{k=1}^\infty a_k x_{k-1})} \gamma_0^{(-a_0 s)} \otimes \delta_{x_0} \otimes \delta_{x_1} \otimes \dots$ .  
 (Note that the use of ‘;’ in expectations was defined in Subsection 3.2.) Hence, using (15),

$$\begin{aligned} & \gamma_{b_0} \otimes \gamma_{b_1} \otimes \dots \hat{H}(s) \\ &= e^{-ms} \gamma_0^{(-a_0 s)} \otimes \gamma_{b_0}^{(-a_1 s)} \otimes \dots \end{aligned} \quad (16)$$

An appropriate form for  $\nu(s)$  is therefore

$$\nu(s) = \gamma_{b_0} \otimes \gamma_{b_1} \otimes \dots \quad (17)$$

The coefficients  $b_0, b_1, \dots$ , can be recursively determined by using (14) and (16):

$$\begin{aligned} b_0 &= -sa_0, \\ b_{k+1} &= b_k - sa_{k+1} = -s \sum_{j=0}^{k+1} a_j. \end{aligned}$$

Also,

$$\begin{aligned} \alpha(s) &= e^{-sm + (sa_0)^2/2} \prod_{k=0}^\infty e^{-sb_k a_{k+1} + (sa_{k+1})^2/2} \\ &= e^{-sm + s^2(S + \frac{1}{2}\sigma^2)}, \end{aligned}$$

where we have used (11). It follows that

$$s^* = 2m/(\sigma^2 + 2S) < 0. \quad (18)$$

As required,  $\alpha(0) = \alpha(s^*) = 1$ , and  $\alpha(s) \neq 0$  for other values of  $s$ .

### 5.3 The Approximation for c

Recalling the definition of  $\pi^*$  at the end of Subsection 4.1, by (12) and (17), we obtain

$$\pi^* = \nu(s^*) = \gamma_{b_0} \otimes \gamma_{b_1} \otimes \dots$$

It follows that  $E(s)$ , our approximation for  $\Phi_+^{-1}$  takes the form

$$E(s) = I - \frac{sf \otimes \pi^*}{(s - s^*)}$$

for some function  $f$  such that  $\pi^* f = \int f(x) \pi^*(dx) \neq 0$ . The precise form of  $f$  is not important in the approximation which results from substituting  $E(s)$  for  $\Phi_+^{-1}$  in (5). All that matters is that the product  $\pi f$ , denoted  $\tilde{c}$ , is correct.

The result which follows from substituting  $E(s)$  for  $\Phi_+^{-1}(s)$  in (5) is the following Laplace-Stieltjes transform of the approximate joint stationary distribution of  $X_n$  and  $V_n$ :

$$\begin{aligned} E\{e^{-sV_n}; X_n \in A\} &\approx \pi(A) - \tilde{c}\pi^*(A) \frac{s}{s - s^*} \\ &= \pi(A) - \tilde{c}\pi^*(A) - \tilde{c}\pi^*(A) \frac{s^*}{s - s^*}, \end{aligned} \quad (19)$$

hence

$$P\{X_n \in A \& V_n = 0\} \approx \pi(A) - \tilde{c}\pi^*(A), \quad (20)$$

and

$$P\{X_n \in A \& V_n > x\} \approx \tilde{c}\pi^*(A) e^{s^* x}, \quad x \geq 0. \quad (21)$$

These formulae appear to assert that (1)  $\tilde{c}$  is an approximation for  $c$ , and (2) the joint probability measure of  $X_n$  and  $V_n$  is a mixture of two separate cases:

## 7c.2.6

one in which  $V_n = 0$  and one in which  $V_n > 0$ . Furthermore, in each separate case  $V_n$  and  $X_n$  are (conditionally) independent. We must recall that (19) is only approximately correct, and the sense in which it is approximately correct is fairly loose. One glaring defect in (20) is the fact that the RHS can take negative values for some choices of the set  $A$ . Equations (20) and (21) cannot be interpreted too literally, however if are treated as exact, it follows that  $V_n$  and  $Y_n$  are conditionally independent given  $V_n > 0$ , and given  $V_n > 0$ ,  $Y_n$  has a Gaussian probability distribution with mean

$$u = m + \sum_{k \geq 0} a_k b_k = \sigma^2 m / (\sigma^2 + 2S), \quad (22)$$

and standard deviation  $\sigma$ . Note that this result has the intuitively reasonable implication that in a system with positive correlation (specifically,  $S > 0$ ), the mean of  $Y_n$ , given  $V_n > 0$ , is larger than the unconditional stationary mean for this same process. Let us now apply the method of subsection 4.3 to derive a formula for  $\tilde{c}$ . It will be useful to introduce the following definition:

$$\begin{aligned} \psi(x) &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty y e^{-\frac{(y+x)^2}{2\sigma^2}} dy \\ &= \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} - \frac{x}{2} \operatorname{erfc}\left(\frac{x}{\sigma\sqrt{2}}\right). \end{aligned}$$

Then

$$\begin{aligned} E\{V_{n+1}\} &= \psi(-m) - \tilde{c}\psi(-u) \\ &\quad + \tilde{c} \int_0^\infty -\frac{xs^*}{2} e^{s^*x} \operatorname{erfc}\left(\frac{-u-x}{\sigma\sqrt{2}}\right) dx \end{aligned}$$

from which it follows that the approximation for the mass of the negative exponential tail is

$$\tilde{c} = -\frac{s^*\psi(-m)}{\operatorname{erf}\left(\frac{-u}{\sigma\sqrt{2}}\right)}. \quad (23)$$

Numerical comparisons show that  $c$  estimates obtained by the latter are close to those obtained by Eq. (14) of [4].

#### 5.4 The Statistics of $V_\infty$

Having formulae for  $s^*$  and  $\tilde{c}$ , we can approximate the distribution of  $V_\infty$  by

$$P\{V_\infty > t\} \approx \tilde{c} e^{s^*t}. \quad (24)$$

Consequently, the mean is approximated by

$$E\{V_\infty\} \approx -\frac{\tilde{c}}{s^*}. \quad (25)$$

Let  $t_p$  be the  $p^{\text{th}}$  percentile of the stationary distribution of  $V_\infty$ . Assuming  $\tilde{c} > 1 - p/100$ , we obtain

$$t_p \approx \frac{1}{s^*} \ln((1 - p/100)/\tilde{c}). \quad (26)$$

If  $\tilde{c} \leq 1 - p/100$  then  $t_p \approx 0$ .

These results have been tested by simulations. Simulation results obtained over a wide range of parameter values agree with their analytical counterparts.

#### 5.5 A Queue with Autoregressive Input

Consider the case where  $A_n$  is a first-order autoregressive process, i.e.,

$$A_n = aA_{n-1} + b\tilde{U}_n, \quad (27)$$

where  $\tilde{U}_n$  is Gaussian with mean  $\eta$  and variance  $\tilde{\sigma}^2$ , and  $|a| < 1$ . This model was proposed in [15] for a VBR traffic stream generated by a single source video telephony. (In [15]  $\tilde{\sigma}^2 = 1$ .) By (27),  $E\{A_\infty\} = \eta b / (1 - a)$ , and  $\operatorname{Var}\{A_\infty\} = b^2 \tilde{\sigma}^2 / (1 - a^2)$ . Assume that the service rate is fixed and let  $\tau = E\{B_n\}$  for all  $n$ . Clearly, in this case,  $\operatorname{Var}\{Y_\infty\} = \operatorname{Var}\{A_\infty\}$ .

As usual,  $\rho = E\{A_\infty\} / E\{B_\infty\} = \eta b / (1 - a) \tau$ . Eq. (27) can be written as  $A_n = b \sum_{k=0}^\infty a^k \tilde{U}_{n-k}$ , where  $\{\tilde{U}_k\}_{k \geq 0}$  is a sequence of i.i.d. Gaussian random variables each with mean  $\eta$  and variance  $\tilde{\sigma}^2$ . Define the net input process as  $Y_n = A_n - \tau$ , and the sequence  $\{U_k\}_{k \geq 0}$ , by  $U_{n-k} = (\tilde{U}_{n-k} - \eta) / \tilde{\sigma}$ . Then  $\{U_k\}_{k \geq 0}$  is a sequence of i.i.d. Gaussian random variables with mean 0 and variance 1. Hence,

$$Y_n = \frac{b\eta}{1-a} - \tau + \sum_{k=0}^\infty b\tilde{\sigma}a^k U_{n-k}, \quad n \in \mathbb{Z}, \quad (28)$$

which is, in fact, a special case of (10) with  $m = b\eta/(1-a) - \tau$  and  $a_k = b\tilde{\sigma}a^k$ . Substituting these  $a_k$  values in (11) we obtain

$$S = \frac{ab^2\tilde{\sigma}^2}{(1-a)^2(1+a)}. \quad (29)$$

Hence, by (18), we obtain

$$s^* = \frac{2(1-a)(\rho-1)\tau}{(1+a)\operatorname{Var}\{A_\infty\}}, \quad (30)$$

which is identical to Eq. (6) in [3] where this case is considered specifically. (Note also that a small error appeared in [2] in this formula.) Applying these formulae for  $m$ ,  $\operatorname{Var}\{Y_\infty\}$ ,  $S$  and  $s^*$ , as well as Eq.

(22) for  $u$ , in (23) through (26), approximations for the statistics of  $V_\infty$  can be obtained.

**Acknowledgement:** The permission of the Director of Research, Telecom Research Laboratories to publish this paper is hereby acknowledged.

## References

- [1] R. G. Addie, *Theory and Applications of Discrete Time Markov Additive Processes*, PhD thesis, Monash University, August 1986.
- [2] R. G. Addie and M. Zukerman, "An approximate solution of a semi-Markov queue with large state space," *Proc., 13th ITC* Copenhagen, June 1991.
- [3] R. G. Addie, and M. Zukerman, "An asymptotically accurate method for performance evaluation of semi-Markov queues with large state space," *Proc., 6th ATRS*, Wollongong, Australia, November 1991.
- [4] R. G. Addie and M. Zukerman, "A Gaussian Traffic Model for a B-ISDN Statistical Multiplexer," *Proc., GLOBECOM '92*, Orlando, Florida, December 1992.
- [5] E. Arjas, "On a fundamental identity in the theory of semi-Markov processes," *Advances in Applied Probability*, vol. 4, pp. 258–270, 1972.
- [6] E. Arjas, "On the use of a fundamental identity in the theory of semi-Markov queues," *Advances in Applied Probability*, vol. 4, pp. 271–284, 1972.
- [7] E. Arjas and T. P. Speed, "Symmetric Wiener-Hopf factorizations in Markov-additive processes," *Z. Wahrscheinlichkeitstheorie und Verwandt Gebiete*, vol. 26, pp. 105–118, 1973.
- [8] G. Doetsch, *Introduction to the Theory and Applications of the Laplace Transforms*, Springer Verlag, 1970.
- [9] K. W. Fendick, V. R. Saksena and W. Whitt, "Dependence in packet queues," *IEEE Trans. Commun.*, vol. 37, no. 11, November 1989.
- [10] O. Gühr and P. Tran-Gia, "A layered description of ATM cell traffic streams and correlation analysis," *ATR*, vol. 24, no. 2, pp. 9–18, 1990.
- [11] P. R. Halmos, *Measure Theory*, Springer-Verlag, New York, 1974.
- [12] H. Heffes and D. Lucantoni, "A Markov modulated characterization of voice and data traffic and related statistical multiplexer performance," *IEEE JSAC*, vol. SAC-4, Sept. 1986.
- [13] J. Lamperti, *Stochastic Processes – a Survey of the Mathematical Theory*, Springer-Verlag, 1977.
- [14] S. Q. Li and C. L. Hwang, "Queue response to input correlation functions: discrete spectral analysis" *Proc., INFOCOM '92*, Florence, May 1992.
- [15] 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, pp. 834–844, July 1988.
- [16] H. D. Miller, "A convexity property in the theory of random variables defined on a finite Markov chain," *Ann. Math. Statist.*, vol. 32, pp. 1260–1270, 1961.
- [17] 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, pp. 268–285, 1962.
- [18] H. D. Miller, "Absorption probabilities for sums of random variables defined on a finite Markov chain," *Proc. Cambridge Philos. Soc. math. phys.*, vol. 58, pp. 286–298, 1962.
- [19] M. F. Neuts, "The single server queue with Poisson input and semi-Markov service times," *J. Appl. Prob.*, vol. 3, pp. 202–230, 1966.
- [20] I. Norros, J. W. Roberts, A. Simonian and J. T. Virtamo, "The superposition of variable bit rate sources in an ATM multiplexer," *IEEE JSAC*, vol. 9, pp. 378–387, April 1991.
- [21] E. Nummelin, *General Irreducible Markov Chains and Non-negative Operators*, Cambridge University Press, 1984.
- [22] R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, American Mathematical Society, 1934.
- [23] 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, pp. 815–852, 1969.
- [24] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *IEEE JSAC*, vol. SAC-4, pp. 833–846, Sept. 1986.
- [25] L. Takacs, "A Banach space of matrix functions and its application in the theory of queues," *Sankhya*, vol. 38, pp 201–211, 1976.