BLOCK RENEWAL PROCESSES AND
THEIR APPLICATION TO MODELING

Hiroshi Toyoizumi   J. George Shanthikumar   Ronald W. Wolff
NTT Laboratories   The University of Califoria at Berkeley

Abstract   A new stochastic process called block renewal process (or $G(m)$ process) is proposed for modeling auto-correlated arrival processes of queueing systems. This process can be modeled easily by the observable marginal distribution and global auto-correlation index of auto-correlated arrival process. This new process is very simple and tractable, so it is good for both simulation and numerical computation. We also discuss how to obtain analytically the steady-state distribution of queue length of a $G(2)/M/1$ queue.

1.   Introduction

Many traffic engineers and researchers are starting to pay much more attention to the importance of the auto-correlation of arrival processes in queueing systems (see [1], [2] and [3] for example). It is known that a stronger positive correlation results in a worse waiting time. Moreover, in telecommunication traffic engineering, one encounters various kinds of auto-correlated arrival processes. In particular, in a multimedia network, such as an ATM network, one must deal with many different types of arrival processes like data, voice, still images and moving images. All of them have different auto-correlation structures. Thus, one must consider good methods of modeling auto-correlated processes.

Several methods have already been proposed. The best-known auto-correlated process in queueing problems is probably the Markov-Modulated Poisson Process (MMPP) [3]. This process is a special case of Markovian Arrival Process (MAP) [4] or more general semi-Markov processes. Using phase space, we can impose a correlation structure on these semi-Markov processes. By tracking transitions in the phase space, we can treat these processes as a Markov process, which is tractable through the Matrix-Geometric method [5]. However, we need some skill to make a good MMPP model given empirical data of arrival process [6]. One of the biggest problems in building a good MMPP type model is that we need to deal with many parameters that usually have no physical meaning and are not observable. In old telephone traffic theory, the only parameter we have to observe is the mean traffic volume (or erlang), so modeling traffic processes and designing the network are relatively easy.

Other methods for modeling auto-correlated processes have recently appeared in the literatures. Jagerman and Melamed [7] proposed the Transform Expanded Sample (TES) method. In this method, we first construct a discrete-time stochastic process, called a background process, which has uniformly-distributed marginals and tunable correlated structure. The background process is then transformed to have the exact target marginal distribution and approximate target autocorrelation function. TES queueing process are difficult to solve analytically, but recently Melamed et al. reported a numerical solution for quantized TES/PH/1 [8].

Both MMPP modeling and TES modeling require a complicated procedure for obtaining a model from observed data of the arrival process. Here, we propose a relatively simple auto-correlated stochastic process. We can construct an auto-correlated arrival process that matches observed parameters with important meaning in the context of queueing systems. This process has a renewal structure, so it is relatively easy to get analytical results by using the Matrix-Geometric method. Indeed, for the simplest case, we do not have to use the Matrix-Geometric method at all to get numerical results (see Section 4).

This paper is organized as follows. In section 2, we define the so-called $G(m)$ and $G(m, \alpha)$ classes of processes, which are designed to approximate auto-correlated processes. In section 3 we propose a procedure for building a $G(m, \alpha)$ process from empirical data. We then solve the queueing problem of a single-server queue in section 4. In section 5, we examine some examples of the $G(2)$ process. In section 6 we treat the global auto-correlation behavior of MAP, which will be used in section 7. In section 7, we discuss the accuracy of our method through some numerical examples.

2.   The $G(m)$ and $G(m, \alpha)$ Processes

Let $X_1, X_2, \cdots$ be a sequence of random variables (possibly auto-correlated). We say that the process $X_1, X_2, \cdots$ is a block renewal process with a renewal length of m, or that $X_1, X_2, \cdots$ is a $G(m)$ process for short, when the sequence of m-dimensional random vectors $(X_{mn+1}, X_{mn+2}, \cdots, X_{m(n+1)})_{n=0,1,2,\cdots}$ for $m = 1, 2, \cdots$ is i.i.d. Note that in this notation, the term GI (commonly used for a renewal process in queueing theory) is replaced by $G(1)$. The $G(m)$ process is a natural extension of ordinary renewal processes and we can impose any type of correlation within renewal cycles.

In the sequel, we define a process $G(m, \alpha)$ by restricting the correlation structure of the $G(m)$ process. First however, we prove the next lemma concerning the probability mixtures of auto-correlated random vectors.

Lemma 1. *Let X_1, Y and Z be correlated random variables that have the same marginal mean $E[X]$ and variance $Var[X]$, Let*

$$\rho(X_1, Y) = \rho_y \quad and \quad \rho(X_1, Z) = \rho_z, \tag{2-1}$$

where $\rho(A, B) = \{E[AB] - E[A]E[B]\}/\{Var[A]Var[B]\}^{1/2}$ is the correlation coefficient of random variable A and B.

   Define a random variable X_2 by

$$X_2 = \begin{cases} Y & \text{with probability} \quad p \\ Z & \text{with probability} \quad (1-p) \end{cases} \tag{2-2}$$

then we have

$$\rho(X_1, X_2) = p\rho_y + (1-p)\rho_z. \tag{2-3}$$

Proof.   By the definition of X_2, we have

$$
\begin{aligned}
E[X_1 X_2] &= pE[X_1 Y] + (1-p)E[X_1 Z] \\
&= p\{\rho_y Var[X] + E^2[X]\} + (1-p)\{\rho_z Var[X] + E^2[X]\} \\
&= \{p\rho_y + (1-p)\rho_z\}Var[X] + E^2[X].
\end{aligned}
$$

Theorem 1. *Let X_1 and X_2 be random variables with identical distributions. Suppose that $X_1 = X_2$ with probability ρ, and X_1 and X_2 are independent and identically distributed with probability $1 - \rho$. Then*

$$\rho(X_1, X_2) = \rho. \tag{2-4}$$

Proof. Substitute Y and Z in Lemma 1 as follows: set $Y = X_1$ and Z and X_1 be i.i.d. Thus,

$$\rho_z = \rho(X_1, Z) = 0 \quad \text{and} \quad \rho_y = \rho(X_1, Y) = 1.$$

By (2-3) we have

$$\rho(X_1, X_2) = p \cdot 1 + (1 - p) \cdot 0 = p = \rho.$$

Remark. This theorem shows that we can construct a pair of random variables having both some prescribed marginal distribution and some prescribed positive correlation coefficient ρ. Furthermore, we can construct a pair of random variables that has a negative correlation coefficient by using the following random variables.

$$P(X \le x, Y \le y) = \text{Max}\{0, \{P(X \le x) + P(Y \le y) - 1\}\}. \tag{2-5}$$

(See Tchen [12] for a detailed description of this type of random variables.)

Let $G(x)$ be any given distribution. Let the m-dimensional random vector $X = (X_1, \cdots, X_m)$ have the following correlation structure: (A) with probability α, all the m random variables in X are identical and distributed as $G(x)$, and (B) with probability $1 - \alpha$, all the m random variables in X are i.i.d. distributed as $G(x)$.

The m-dimensional random vector has the following properties;

(1) The $X_1, \cdots, X_m$ have an identical marginal distribution,

$$G(x) = P\,(X_i \le x).$$

(2) The correlation coefficients of $(X_1, \cdots, X_m)$ are

$$\rho(X_i, X_j) = \alpha \quad \text{for} \quad 1 \le i \ne j \le m.$$

Let $\{X^{(k)}\}_{k=1}^{\infty}$ be a sequence of i.i.d. random vectors, where each $X^{(k)} = (X_1^{(k)}, \cdots, X_m^{(k)})$ is constructed as X above. Construct discrete-time stochastic process $X = \{X_i\}_{i=1}^{\infty}$ such that $X_i = X_j^{(k)}$, provided $i = (k - 1)m + j$. In other words, we construct X out of the vector "blocks", $X^{(k)}$. Clearly, this constitutes a subclass of $G(m)$ processes, which we call a $G(m, \alpha)$ process. The merit of $G(m, \alpha)$ processes is the ability to endow it with any prescribed marginal distribution. Moreover, by varying the pair of values (m, α), we can tune its auto-correlation structure.

Figure 1 shows an example of a sample path of a $G(m, \alpha)$ process with uniform marginal distribution on the interval $[0,10]$, $m = 5$, and $\alpha = 0.5$.

Next, we define the index of global auto-correlation, β, as

$$\beta = \sum_{i=1}^{\infty} \rho(X_1, X_i), \tag{2-6}$$

provided the sum exists. The index of global auto-correlation is a global measure of the "amount" of the auto-correlation of a stochastic process; it plays an important role in the subsequent queueing processes.
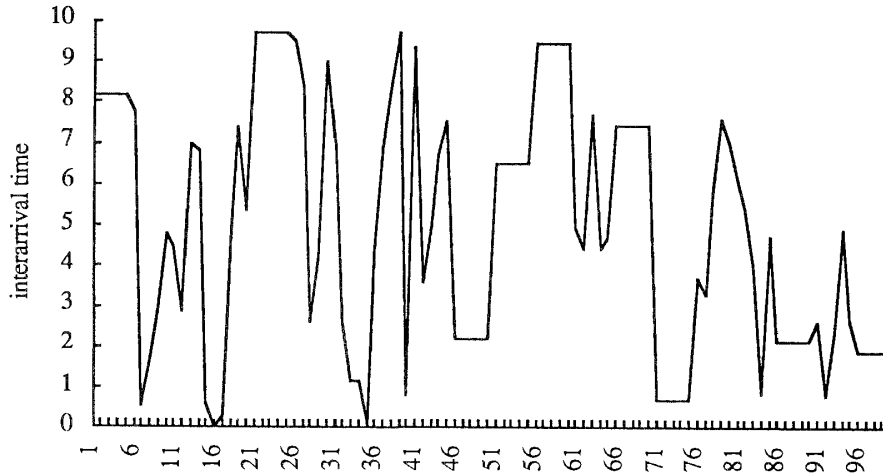
Figure 1. A sample path of a $G(m, \alpha)$ process

By extending the ordinary heavy traffic argument to the $G(m)$ process, we can easily find that if two processes have identical marginal distributions and the same index of global auto-correlation, they are asymptotically equivalent in single-server queues. Also, the index of global auto-correlation plays an important role for calculating the tail probability of buffer content of Gaussian-type input processes [14].

The following theorem about the index of global auto-correlation of the $G(m, \alpha)$ process is useful.

Theorem 2. *The index of global auto-correlation β of an equilibrium $G(m, \alpha)$ process is*

$$\beta = 1 + \frac{\alpha(m-1)}{2}. \tag{2-7}$$

Proof. Since we assume an equilibrium $G(m, \alpha)$ process, the selection of any of the m interarrival times is equiprobable. Thus,

$$\beta = \sum_{j=1}^{m} \rho[X_1, X_j] = 1 + \frac{1}{m}\{\alpha + 2\alpha + \cdots + (m-1)\alpha + 0\} = 1 + \frac{\alpha(m-1)}{2},$$

since correlations across blocks vanish by construction.

3. Procedure for Modeling the Auto-correlated Process

In this section we summarize the procedure for constructing a $G(m, \alpha)$ model from empirical data (in our case, interarrival times). Assume we wish to construct a stochastic sequence $X_1, X_2, \cdots$ that models a target stochastic sequence $Y_1, Y_2, \cdots$ with its empirical data $(y_1, \cdots, y_N)$. Note that the marginal distribution and the parameters (m, α) completely determine a $G(m, \alpha)$ process $X_1, X_2, \cdots$.

Model Construction Procedure:

(1) Compute an empirical marginal distribution $G(x)$ from the path $(y_1, \cdots, y_N)$ (typically, a histogram).

(2) Compute an empirical index of global auto-correlation β from the path $(y_1, \cdots, y_N)$.

(3) Calculate the model's renewal length, m, as the smallest integer greater than $2\beta - 1$, i.e.

$$m - 1 < 2\beta - 1 \le m. \tag{3-1}$$

(4) Compute the corresponding α by

$$\alpha = \frac{2(\beta - 1)}{m - 1}.\tag{3-2}$$

(5) The requisite model of the process $Y_1, Y_2, \cdots$ is given by the $G(m, \alpha)$ process with marginal distribution $G(x)$.

From (3-2) and Theorem 2, we can easily find that this $G(m, \alpha)$ process has the same index of global auto-correlation β as its empirical counterpart.

Remark. We do not have to select m as the smallest integer greater than $2\beta - 1$. We can make the renewal length longer by using smaller α, but for computational and simulation efficiency, it is preferable to use the smallest possible m.

Compared with the moment-matching method of MMPP, our procedure emphasizes matching the marginal distribution and the global behavior of the auto-correlation structure, which can be captured by the index of global auto-correlation. We believe that this results in a better approximation. The test of accuracy using real-life data is under further study. Also, compared with TES, the above procedure has the advantage of simplicity. This makes it easy to perform both simulation and analytical computation.

Taking m identical copies of interarrival time with probability α often results in a $G(m, \alpha)$ process whose simulated sample paths look different from the empirical sample path. However, there exists real life processes for which $G(m, \alpha)$ is a good model, notably the output cell streams from leaky bucket, which is commonly used in ATM for traffic shaping. Leaky bucket has constant interarrival time and their multiplexed streams often resemble the sample path of a $G(m, \alpha)$ process.

4. The $G(2)/M/1$ queue

Since $G(m)$ processes (including the $G(m, \alpha)$ process) have a special renewal structure, we can solve analytically the simplest $G(2)/M/1$ queue by using an embedded Markov process.

Consider a $G(2)/M/1$ queue with a $G(2)$ arrival process $X_1, X_2, \cdots$ and iid exponential service times with mean μ^{-1}. Let $G(x, y) = P\{X_1 \le x, X_2 \le y\}$ be the joint distribution of (X_1, X_2), which completely determines the $G(2)$ process. We assume the 0-th customer arrived at time 0. Let $T_i = \sum_{n=1}^{i} X_n$ be the arrival epoch of the i-th customer and L_i be the number of customers in the system at T_i^- (the number of customers seen by arrivals). The process L_{2n} forms a Markov process with transition matrix

$$P = \begin{bmatrix} C_2 & B_1 & A_0 & & & 0 \\ C_3 & B_2 & A_1 & A_0 & & \\ C_4 & B_3 & A_2 & A_1 & A_0 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.\tag{4-1}$$

Here, A_k is the probability that k customers departed during $[T_{2m}, T_{2(m+1)})$ and more than two customers remained in the system. B_k is the probability that k customers departed during $[T_{2m}, T_{2(m+1)})$ and one customer remained in the system. C_k is the probability that k customers departed during $[T_{2m}, T_{2(m+1)}]$ and no customers remained in the system. The explicit form of A_k is

$$A_k = \int_0^\infty \frac{\{\mu(x + y)\}^k}{k!} e^{-\mu(x+y)} G(dx, dy).\tag{4-2}$$

Letting $P_n = \lim\limits_{m \to \infty} P\{L_{2m} = n\}$, we get the steady-state equations

$$P_n = \sum_{k=0}^{\infty} P_{n-2+k} A_k \quad \text{for} \quad n \geq 2, \tag{4-3}$$

$$P_1 = \sum_{k=0}^{\infty} P_k B_{k+1},$$

$$P_0 = \sum_{k=0}^{\infty} P_k C_{k+2}. \tag{4-4}$$

Define the generating function of P_n by

$$P^*(z) = \sum_{k=0}^{\infty} P_k z^k. \tag{4-5}$$

Then, from (4-3), we have

$$\sum_{k=2}^{\infty} P_k z^k = z^2 G^*(\mu(1 - z^{-1}), \mu(1 - z^{-1})) P^*(z) - \sum_{k=0}^{\infty} \sum_{n=0}^{k-1} P_n A_k z^{2-k+n}, \tag{4-6}$$

where $G^*(s,t)$ is the LST of the joint distribution of $G(x,y)$. A usual manipulation yields,

$$P^*(z) = \frac{-\sum_{k=0}^{\infty} \sum_{n=0}^{k-1} P_n A_k z^{2-k+n} + z P_1 + P_0}{1 - z^2 G^*(\mu(1 - z^{-1}), \mu(1 - z^{-1}))}. \tag{4-7}$$

We now proceed to use an argument similar to the one for the bulk arrival queue [9, 10].

Lemma 2. *If $\mu\{E[X_1] + E[X_2]\} > 2$, then inside the unit disc there are two roots, σ_1 and σ_2, of the equation*

$$\sigma^2 = G^*(\mu(1 - \sigma), \mu(1 - \sigma)). \tag{4-8}$$

Moerover, one root, σ_1, is positive real and the other, σ_2, is negative real.

Remark. The condition

$$\mu\{E[X_1] + E[X_2]\} > 2 \tag{4-9}$$

is equivalent to the traffic intensity less than 1.

Proof. Let $f(z) = -z^2$ and $g(z) = G^*(\mu(1 - z), \mu(1 - z))$. From the definition of $G(\cdot)$, we have

$$g'(1) = \frac{d}{d\sigma} G^*(\mu(1 - \sigma), \mu(1 - \sigma)) \mid_{\sigma=1} = \mu\{E[X_1] + E[X_2]\}.$$

On the circle $\{|z| = 1 - \varepsilon\}$, we have

$$|f(z)| = (1 - \varepsilon)^2 = 1 - 2\varepsilon + \varepsilon^2. \tag{4-10}$$

Since $\mu\{E[X_1] + E[X_2]\} > 2$,

$$|f'(1)| = 2 < \mu\{E[X_1] + E[X_2]\} = |g'(1)|$$

Thus, we have

$$|g(z)| \leq 1 - \varepsilon g'(1) + o(\varepsilon) < 1 - 2\varepsilon + o(\varepsilon) \tag{4-11}$$

Hence,

$$|g(z)| < |f(z)| \quad \text{on the set} \quad \{|z| - 1 - \varepsilon\} \tag{4-12}$$

for some small positive ε.

Using Rouche's theorem, we know that $f(z) = -z^2$ and $f(z) + g(z) = -z^2 + G^*(\mu(1 - z), \mu(1 - z))$ have the same number of zeros inside $\{|z| = 1 - \varepsilon\}$. Thus, equation (4-8) has two roots.

On the other hand, since $g(0) = G^*(\mu, \mu) > 0$ and $g(z)$ is a continuous function, there should be at least one intersection of $y = z^2$ and $y = G^*(\mu(1 - z), \mu(1 - z))$ in the region $\{(z, y) \in R^2 : z > 0, y > 0\}$. Hence, one root, σ_1, of (4-8) must be positive real. Similarly, $y = z^2$ and $y = G^*(\mu(1 - z), \mu(1 - z))$ intersects in the region $\{(z, y) \in R^2 : z < 0, y > 0\}$, whence σ_2 is negative real.

Theorem 3. *Let P_n be the steady-state probability in a $G(2)/M/1$ queue that the number of customers seen by an even-numbered customer is n, i.e.,*

$$P_n = \lim_{m \to \infty} P\{L_{2m} = n\},$$

where L_{2m} is the number of customers seen by the $2m$-th customer.

If $\mu\{E[X_1] + E[X_2]\} > 2$, then the generating function of $\{P_n\}$, $P^*(z) = \sum_{k=0}^{\infty} P_k z^k$, can be explicitly written as

$$P^*(z) = \frac{C_1 z + C_2}{(1 - \sigma_1 z)(1 - \sigma_2 z)}, \tag{4-13}$$

where σ_1 and σ_2 are the solutions of the equation (4-8), and

$$
\begin{aligned}
C_1 &= (1 - \sigma_1)(1 - \sigma_2) - C_2 \\
&= \frac{-g_1 \sigma_1 \sigma_2 (1 - \sigma_2) + g_2 \sigma_1 \sigma_2 (1 - \sigma_1) + \sigma_1 \sigma_2 g_0 (\sigma_1 - \sigma_2)}{g_1 \sigma_2 - g_2 \sigma_1}, \tag{4-14} \\
C_2 &= \frac{g_1 \sigma_2 (1 - \sigma_2) - g_2 \sigma_1 (1 - \sigma_1) - \sigma_1 \sigma_2 g_0 (\sigma_1 - \sigma_2)}{g_2 \sigma_1 - g_1 \sigma_2}, \tag{4-15} \\
g_0 &= G^*(0, \mu), \\
g_1 &= G^*(\mu(1 - \sigma_1), \mu), \tag{4-16} \\
g_2 &= G^*(\mu(1 - \sigma_2), \mu).
\end{aligned}
$$

Proof. Since the denominator of (4-7) has only two zeros, $1/\sigma_1$ and $1/\sigma_2$, outside the unit disc,

$$A(z) = \frac{C(z)(1 - \sigma_1 z)(1 - \sigma_2 z)}{1 - z^2 G^*(\mu(1 - z^{-1}), \mu(1 - z^{-1}))}$$

is regular for all z, where

$$C(z) = -\sum_{k=0}^{\infty} \sum_{n=0}^{k-1} P_n A_k z^{2-k+n} + z P_1 + P_0.$$

It is easy to see that

$$\lim_{|z| \to \infty} \frac{A(z)}{z^2} = 0.$$

Hence we can write $A(z) = C_1 z + C_2$ for some constants C_1 and C_2. Consequently, we have

$$P^*(z) = \frac{C_1 z + C_2}{(1 - \sigma_1 z)(1 - \sigma_2 z)}. \tag{4-17}$$

Since $\sum_{n=0}^{\infty} P_n = 1$,

$$C_1 + C_2 = (1 - \sigma_1)(1 - \sigma_2). \tag{4-18}$$

From the asympktotic behavior of $A(z) = C_1 z + C_2$, we have

$$
\begin{aligned}
C_1 &= \lim_{|z| \to \infty} \frac{A(z)}{z} = \lim_{|z| \to \infty} \frac{(1 - \sigma_1 z)(1 - \sigma_2 z)C(z)/z}{1 - z^2 G^*(\mu(1 - z^{-1}), \mu(1 - z^{-1}))} \\
&= -\frac{\sigma_1 \sigma_2}{G^*(\mu, \mu)} \left\{ \sum_{k=0}^{\infty} (B_{k+1} - A_{k+1}) P_k \right\}.
\end{aligned} \tag{4-19}
$$

Since $B_{k+1} = \mathrm{P}\{(k+1)$ Poisson events occur$\} + \mathrm{P}\{$at least $(k+2)$ Poisson events occur but only $(k+1)$ events are effective departures$\}$, we have

$$B_{k+1} - A_{k+1} = \sum_{l=k+2}^{\infty} \int \left(\frac{x}{x+y}\right)^l \left(\frac{y}{x+y}\right)^0 \frac{\{\mu(x+y)\}^l}{l!} e^{-\mu(x+y)} G(dx, dy).$$

Hence,

$$\sum_{k=0}^{\infty} (B_{k+1} - A_{k+1}) P_k = \int \sum_{l=2}^{\infty} \sum_{k=0}^{l-2} \frac{(\mu x)^l}{l!} P_k e^{-\mu(x+y)} G(dx, dy). \tag{4-20}$$

We can calculate $\sum_{l=2}^{\infty} \sum_{k=0}^{l-2} \frac{(\mu x)^l}{l!} P_k$ by using the following Lemma concerning the product of generating functions.

Lemma 3. (See Vich [11] for proof.)   *Let* $F(z) = \sum_{l=0}^{\infty} a_l z^{-l}$ *and* $G(z) = \sum_{l=0}^{\infty} b_l z^{-l}$. *Then the generating function of the product is*

$$H(z) = \sum_{l=0}^{\infty} a_l b_l z^{-l} = \frac{1}{2\pi i} \oint_c \frac{F(\xi)G(z/\xi)}{\xi} d\xi, \tag{4-21}$$

where the integral path c *can be chosen to have all poles of* $F(z)$ *inside* c *and all poles of* $G(z/\xi)$ *outside* c.

Setting $a_l = \sum_{k=0}^{l-2} P_k$ and $b_l = 1/l!$ in Lemma 3, we have

$$F(z^{-1}) = \frac{z^2 P^*(z)}{1 - z} = \frac{z^2(C_1 z + C_2)}{(1 - z)(1 - \sigma_1 z)(1 - \sigma_2 z)}$$

and

$$G(z^{-1}) = e^z.$$

Hence,

$$H(z) = \sum_{l=2}^{\infty} \sum_{k=0}^{l-2} \frac{z^{-l}}{l!} P_k = \frac{1}{2\pi i} \oint_{|z|=1+\varepsilon} \frac{(C_1 + C_2 \xi)e^{\xi/z}}{\xi(\xi - 1)(\xi - \sigma_1)(\xi - \sigma_2)} d\xi. \tag{4-22}$$

By the residue theorem and (4-17), we can deduce

$$H(z) = -\frac{C_1}{\sigma_1\sigma_2} + \frac{(C_1 + C_2)e^{1/z}}{(1-\sigma_1)(1-\sigma_2)} + \frac{(C_1 + C_2\sigma_1)e^{\sigma_1/z}}{\sigma_1(\sigma_1 - 1)(\sigma_1 - \sigma_2)} + \frac{(C_1 + C_2\sigma_2)e^{\sigma_2/z}}{\sigma_2(\sigma_2 - 1)(\sigma_2 - \sigma_1)}.$$

Since $\sum_{l=2}^{\infty}\sum_{k=0}^{l-2}\frac{(\mu x)^l}{l!}P_k = H((\mu x)^{-1})$, we have

$$\sum_{k=0}^{\infty}(B_{k+1} - A_{k+1})P_k = -\frac{C_1}{\sigma_1\sigma_2}G^*(\mu,\mu) + G^*(0,\mu) + \frac{(C_1 + C_2\sigma_1)}{\sigma_1(\sigma_1 - 1)(\sigma_1 - \sigma_2)}G^*(\mu(1-\sigma_1),\mu)$$

$$+\frac{(C_1 + C_2\sigma_2)}{\sigma_2(\sigma_2 - 1)(\sigma_2 - \sigma_1)}G^*(\mu(1-\sigma_2),\mu).$$

Hence the constants C_1 and C_2 are determined by the following two linear equations:

$$G^*(0,\mu) + \frac{(C_1 + C_2\sigma_1)}{\sigma_1(\sigma_1 - 1)(\sigma_1 - \sigma_2)}G^*(\mu(1-\sigma_1),\mu)$$

$$+\frac{(C_1 + C_2\sigma_2)}{\sigma_2(\sigma_2 - 1)(\sigma_2 - \sigma_1)}G^*(\mu(1-\sigma_2),\mu) = 0$$

$$C_1 + C_2 = (1-\sigma_1)(1-\sigma_2), \tag{4-23}$$

which can be explicitly solved, yielding (4-14) and (4-15).

Theorem 4. *The mean queue length seen by an odd-numbered customer in the $G(2)/M/1$ queue is given by*

$$\begin{aligned}E[L_{2n+1}] = \ & E[L_{2n}] + 1 - \frac{1}{1-\sigma_1} - \frac{1}{1-\sigma_2} + \\ & +\frac{C_2}{(1-\sigma_1)(1-\sigma_2)} + \frac{C_1 + C_2\sigma_1}{(1-\sigma_1)^2(\sigma_1 - \sigma_2)}G^*(\mu(1-\sigma_1),0) \\ & +\frac{C_1 + C_2\sigma_2}{(1-\sigma_2)^2(\sigma_2 - \sigma_1)}G^*(\mu(1-\sigma_2),0).\end{aligned} \tag{4-24}$$

Proof. Let $p_m = P\{L_{2n} = m\}$ and $q_m = P\{L_{2n+1} = m\}$. Conditioning on the interarrival times, for $n \geq 1$ we have

$$P\{L_{2k+1} = n \mid (X_1, X_2) = (x,y)\} = \sum_{m=n-1}^{\infty}\frac{(\mu x)^{m+1-n}}{(m+1-n)!}e^{-\mu x}p_m = \sum_{m=0}^{\infty}\frac{(\mu x)^m}{m!}e^{-\mu x}P_{m+n-1}.$$

Hence, q_n can be obtained as

$$q_n = \int_0^{\infty}\sum_{m=0}^{\infty}\frac{(\mu x)^m}{m!}e^{-\mu x}p_{m+n-1}G(dx,dy) \quad \text{for} \quad n \geq 1,$$

$$q_0 = \int_0^{\infty}\sum_{n=0}^{\infty}\sum_{k=n+1}^{\infty}\frac{(\mu x)^k}{k!}e^{-\mu x}p_n G(dx,dy).$$

Let $Q^*(z)$ be the generating function of q_m. The mean queue length seen by odd-numbered customers can be calculated by

$$\begin{aligned}E[L_{2k+1}] = \ & Q^{*\prime}(1) = \int_0^{\infty}\sum_{n=1}^{\infty}\sum_{m=0}^{\infty}n\frac{(\mu x)^m}{m!}e^{-\mu x}P_{m+n-1}G(dx,dy) \\ = \ & E[L_{2k}] + 1 - \mu E[X_1] - \int_0^{\infty}\sum_{m=0}^{\infty}\frac{(\mu x)^m}{m!}e^{-\mu x}\left(\sum_{n=0}^{m-1}(n+1-m)p_n\right)G(dx,dy).\end{aligned}$$

$$\tag{4-25}$$

We will use complex analysis to get an explicit form of the third term. Define the sequences $\{a_m\}$ and $\{b_m\}$ by

$$a_m = \sum_{n=0}^{m-1} (n+1-m)p_n, \quad \text{and} \quad b_m = 1/m!. \tag{4-26}$$

Using ordinary generating function arguments, we have

$$\sum_{m=0}^{\infty} \sum_{n=0}^{m-1} np_n z^m = \frac{z^2 P^{*\prime}(z)}{1-z}$$

and

$$\sum_{m=0}^{\infty} (1-m) \sum_{n=0}^{m-1} p_n z^m = -\frac{z^2 P^*(z)}{(1-z)^2} - \frac{z^2 P^{*\prime}(z)}{1-z}.$$

Letting $A(z)$ be the generating function of a_m, and using (4-13), we have

$$A(z) \equiv \sum_{m=0}^{\infty} a_m z^m = \frac{-z^2 P^*(z)}{(1-z^2)} = \frac{-z^2(C_1 z + C_2)}{(1-z)^2(1-\sigma_1 z)(1-\sigma_2 z)} \tag{4-27}$$

Now letting

$$F(z) \equiv A(z^{-1}) = \frac{-z(C_1 + C_2 z)}{(z-1)^2(z-\sigma_1)(z-\sigma_2)}$$

and

$$G(z) \equiv B(z^{-1}) \equiv \sum_{m=0}^{\infty} b_m z^{-m} = e^{1/z}$$

in Lemma 3, the generating function of the product of a_m and b_m becomes

$$\begin{aligned} H(z) &\equiv \sum_{l=0}^{\infty} a_l b_l z^{-l} = -\frac{1}{2\pi i} \oint_c \frac{F(\xi)G(z/\xi)}{\xi} d\xi \\ &= -\sum_i \operatorname*{Re}_{\xi=\xi_i} s \frac{\xi(C_1 + C_2 \xi)}{(\xi-1)^2(\xi-\sigma_1)(\xi-\sigma_2)} e^{\xi/z}, \end{aligned} \tag{4-28}$$

where $\xi_1 = 1, \xi_2 = \sigma_1$ and $\xi_3 = \sigma_2$. Calculating the residue value at each pole of the denominator, we obtain

$$\begin{aligned} H(z) = {} & \frac{e^{1/z}}{1-\sigma_1} + \frac{e^{1/z}}{1-\sigma_2} - \frac{C_2 e^{1/z}}{(1-\sigma_1)(1-\sigma_2)} - \frac{e^{1/z}}{z} \\ & - \frac{(C_1 + C_2\sigma_1)e^{\sigma_1/z}}{(1-\sigma_1)^2(\sigma_1-\sigma_2)} - \frac{(C_1 + C_2\sigma_2)e^{\sigma_2/z}}{(1-\sigma_2)^2(\sigma_2-\sigma_1)}. \end{aligned}$$

Integrating both sides with respect to $G(dx, dy)$ yields

$$\begin{aligned} \text{the forth term of (4-25)} = {} & \frac{1}{1-\sigma_1} + \frac{1}{1-\sigma_2} - \frac{C_2}{(1-\sigma_1)(1-\sigma_2)} - \mu E[X_1] \\ & - \frac{(C_1 + C_2\sigma_1)}{(1-\sigma_1)^2(\sigma_1-\sigma_2)} G^*(\mu(1-\sigma_1), 0) - \frac{(C_1 + C_2\sigma_2)}{(1-\sigma_2)^2(\sigma_2-\sigma_1)} G^*(\mu(1-\sigma_2), 0). \end{aligned}$$

We next propose a numerical procedure for finding roots σ_1 and σ_2. We consider the following iteration:

$$\begin{aligned} \sigma_1^{(n)} &= \{G^*(\mu(1-\sigma_1^{(n-1)}), \mu(1-\sigma_1^{(n-1)}))\}^{1/2}, \\ \sigma_2^{(n)} &= -\{G^*(\mu(1-\sigma_2^{(n-1)}), \mu(1-\sigma_2^{(n-1)}))\}^{1/2}, \end{aligned} \tag{4-29}$$

with $\sigma_1^{(0)}, \sigma_2^{(0)} \in [-1, 1)$. Then we can expect

$$\sigma_1 = \lim_{n \to \infty} \sigma_1^{(n)} \quad \text{and} \quad \sigma_2 = \lim_{n \to \infty} \sigma_2^{(n)}. \tag{4-30}$$

Remark. Alternatively, the $G(m)/M/1$ queue can be solved by using the Matrix-Geometric method analogously to the bulk arrival queue (see [5] section 1.5). We believe that the above procedure is efficient for $m = 2$, but for higher m, we recommend using Matrix-Geometric methods for simplicity.

5. Examples of $G(2)/M/1$ queue

In this section we exhibit some examples and relate them to existing models.

Example 1. (GI/M/1 queue)   In this case, the LST of the joint distribution of interarrival time is just the product, i.e.,

$$G^*(s, t) = G^*(s)G^*(t), \tag{5-1}$$

where $G^*(s)$ is the LST of the marginal interarrival time distribution. The roots σ_1 and σ_2 are determined by

$$\sigma_1 = G^*(\mu(1 - \sigma_1)) \quad \text{and} \quad \sigma_2 = -G^*(\mu(1 - \sigma_2)), \tag{5-2}$$

and, solving (4-23), yields

$$C_1 = -(1 - \sigma_1)\sigma_2 \quad \text{and} \quad C_2 = 1 - \sigma_1. \tag{5-3}$$

Substituting (5-3) into (4-13), we have

$$P^*(z) = \frac{1 - \sigma_1}{(1 - \sigma_1 z)}. \tag{5-4}$$

This coincides with the solution of ordinary GI/M/1 queues.

Example 2. (Bulk arrival $G^{[2]}/M/1$ queue)   In this case, the LST of the joint distribution of interarrival times is independent of the first argument, i.e.

$$G^*(s, t) = G^*(t), \tag{5-5}$$

and

$$C_1 = 0 \quad \text{and} \quad C_2 = (1 - \sigma_1)(1 - \sigma_2). \tag{5-6}$$

Substituting (5-6) into (4-13), we have

$$P^*(z) = \frac{(1 - \sigma_1)(1 - \sigma_2)}{(1 - \sigma_1 z)(1 - \sigma_2 z)}, \tag{5-7}$$

which coincides with the solution of ordinary bulk-arrival $G^{[2]}/M/1$ queues.

6. The Marginal Distribution and Index of Global Auto-correlation of MAP

In this section we discuss the marginal distribution and the index of global auto-correlation, β, of Markovian Arrival Process (MAP) (see [3] for details). Consider a MAP and let D_0 and D_1 be the transition rate matrix of phases without arrival and with arrival. Then, the LST of marginal distibution of interarrival time has the form

$$G^*(s) = \pi_a(sI - D_0^{-1})D_1 e, \tag{6-1}$$

where π_a is the stationary probability vector of the phase at arrival.

Let $Q(t)$ be the semi-Markov kernel of MAP, that is, the (i, j)-element of matrix $Q(t)$ is the probability that the first transition takes place by time t and jumps to phase j, given phase i. Then

$$Q(t) = \int_0^t e^{D_0 u} du D_1, \tag{6-2}$$

and

$$Q(\infty) = -D_0^{-1} D_1 \tag{6-3}$$

The LST of $Q(t)$ is defined by

$$Q^*(s) = \int_0^\infty e^{-st} Q(dt) = (sI - D_0^{-1}) D_1,$$

and its derivatives are

$$Q^{*'}(s) = -D_0^{-2} D_1, \tag{6-4}$$

and

$$Q^{*''}(s) = -2D_0^{-3} D_1. \tag{6-5}$$

The auto-covariance of interarrival time sequence $(X_i)_{i=1,2,\ldots}$ can be computed for $m \geq 2$ as

$$Cov(X_1, X_m) = \pi Q^{*'}(0) Q(\infty) Q^{*'}(0) e - [\pi Q^{*'}(0)e]^2, \tag{6-6}$$

where π is the stationary distribution of the phase at arrival epochs. For $m \geq 3$, we can rewrite (6-6) as

$$Cov(X_1, X_m) = \pi Q^{*'}(0)\{Q(\infty) - Q^\infty(\infty)\}^{m-1} Q^{*'}(0)e, \tag{6-7}$$

and for $m = 2$,

$$Cov(X_1, X_2) = \pi Q^{*'}(0)\{Q(\infty) - Q^\infty(\infty)\}^0 Q^{*'}(0)e - (\pi Q^{*'}(0)e)^2. \tag{6-8}$$

Moreover, we have

$$Var[X] = \pi Q^{*''}(0)e - (\pi Q^{*'}(0)e)^2. \tag{6-9}$$

Hence, β can be calculated as

$$\beta = 1 + \frac{\pi Q^{*'}(0)\{I - Q(\infty) + Q^\infty(\infty)\}^0 Q^{*'}(0)e - (\pi Q^{*'}(0)e)^2}{\pi Q^{*''}(0)e - (\pi Q^{*'}(0)e)^2}. \tag{6-10}$$

7.   Example of an Approximation of MMPP(2)

Consider a two-state MMPP (MMPP(2)) as an example of an auto-correlated arrival process. The transition rate matrix without arrival, D_0, and with arrival, D_1, can be written respectively as

$$D_0 = \begin{pmatrix} -r_0 - \lambda_0 & r_0 \\ r_1 & -r_1 - \lambda_1 \end{pmatrix} \tag{7-1}$$

and

$$D_1 = \begin{pmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{pmatrix}. \tag{7-2}$$

Recalling results of the MMPP(2)/M/1 queue, the infinitesimal generator of the queue length process is

$$Q = \begin{pmatrix} B_0 & A_0 & & & 0 \\ B_1 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & A_2 & A_1 & \ddots \\ 0 & & & \ddots & \ddots \end{pmatrix}, \tag{7-3}$$

where

$$A_0 = \begin{pmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{pmatrix}, \tag{7-4}$$

$$A_1 = \begin{pmatrix} -(\mu + r_0 + \lambda_0) & r_0 \\ r_1 & -(\mu + r_1 + \lambda_1) \end{pmatrix}, \tag{7-5}$$

and

$$A_2 = \begin{pmatrix} \mu & 0 \\ 0 & \mu \end{pmatrix}. \tag{7-6}$$

Let $\pi = (\pi_1, \pi_2, \cdots) = ((\pi_1^1, \pi_1^2), (\pi_2^1, \pi_2^2), \cdots)$ be the stationary probability vector of Q, i.e.

$$\pi Q = 0. \tag{7-7}$$

Then by the usual Matrix Geometric argument (see Neuts[4]), there exists a rate matrix R such that

$$\pi_n R = \pi_{n-1}, \tag{7-8}$$

which satisfies the equation

$$R^2 A_2 + R A_1 + A_0 = 0. \tag{7-9}$$

The stationary probability vector of phases, ν, is

$$\nu = (\nu_1, \nu_2) = \left(\frac{r_0}{r_0 + r_1}, \frac{r_1}{r_0 + r_1} \right), \tag{7-10}$$

whence the stationary vector of the queue length of MMPP(2)/M/1 can be written as
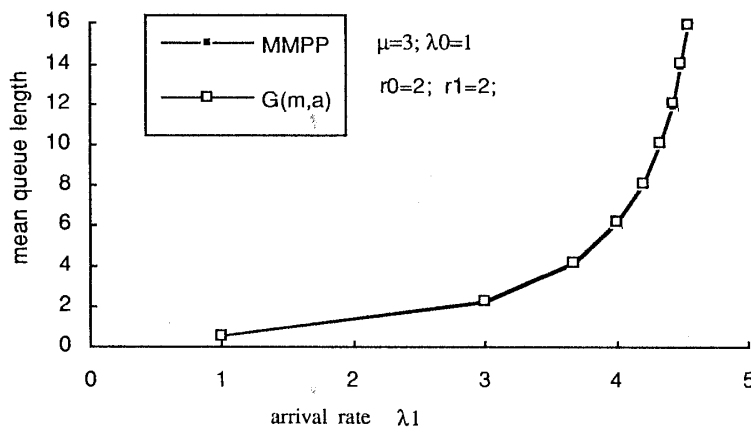
$$\pi_n = \nu(I - R)^{-1} R^n. \tag{7-11}$$

Figure 2. One arrival rate and mean queue length

Moreover, the mean queue length is given by

$$E[L] = \nu R(I - R)^{-1}e,$$ (7-12)

where $e = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Note that R can be obtained by numerical iteration.

Next we examine the efficacy of our approximation method when applied to known target models. Let MMPP(2) be the target process, and follow the procedure described in section 3 to construct an appropriate $G(m, \alpha)$ model. The following figures compare the mean queue lengh of various MMPP(2)/M/1s and the corresponding $G(m, \alpha)$/M/1.
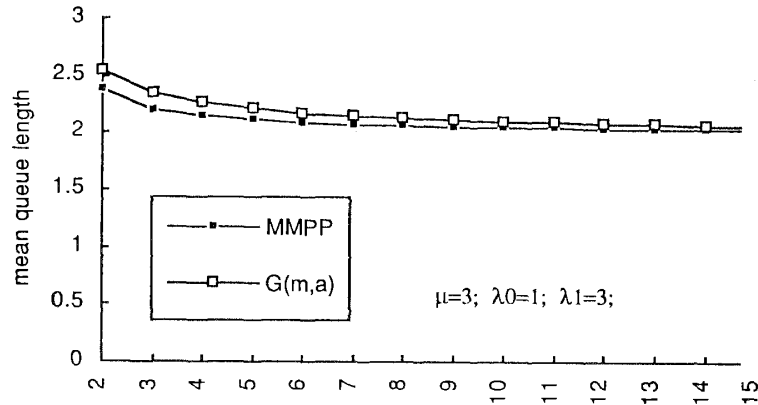
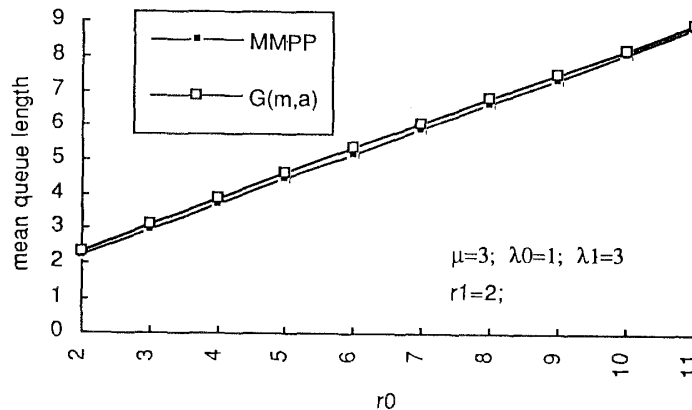Figure 3. Phase transition rates and mean queue length

Figure 4. One of the phase transition rates and mean queue length

Figure 2 depicts the mean queue length of both processes for various arrival rates of phase 1, λ_1. Obviously for larger λ_1, the mean queue length of MMPP(2)/M/1 increases since the overall arrival rate becomes larger. This figure shows that the corresponding $G(m, \alpha)$ approximation tracks this trend perfectly. In Figure 3, we set the phase transition rates r_0 and r_1 to a common value r. As r increases, the correlation between arrivals becomes smaller, so the mean queue length of MMPP(2)/M/1 becomes shorter. We can see that the $G(m, \alpha)$-approximation tracks this trend. In Figure 4, we increase the value of r_0, which affects on the mean queue length in MMPP(2)/M/1 non-intuitively. For larger r_0, the overall arrival rate increases but the correlation between arrivals decreases. In this case, the queue length of MMPP(2)/M/1 becomes larger. Again, we see that the $G(m, \alpha)$/M/1 queue approximates well the MMPP(2)/M/1 queue.

References

[1] S.Q. Li and C.L. Hwang, "Queue response to input correlation function: continuous spectral analysis," IEEE/ACM Trans. on Networking, Vol. 1, No. 6, pp. 678–693, 1993.

[2] R. Grunenfelder and S. Robert, "Which arrival law parameters are decisive for queueing system performance," ITC14, pp. 377–386, 1994.

[3] H. Heffes and D.M. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," IEEE J. on Selected Areas in Communication, SAC-4, 6, pp. 856–868, 1986.

[4] D.M. Lucantoni, "New results on the single server queue with a batch Markovian Arrival Process," Communication Statist. Stochastic Models, 7(1), pp. 1–46, 1991.

[5] M.F. Neuts, "Matrix-Geometric solutions in Stochastic Models: An Algorithmic approach," The Johns Hopkins University Press, Baltimore, 1981.

[6] A. Baiocche, N. Belfari Melazzi, M. Listanti, A. Roveri and R. Winkler, "Loss performance analysis of an ATM Multiplexer loaded with high-speed on-off sources," IEEE JSAC, Vol. 9, No. 3, pp. 388–393, 1991.

[7] Jagerman and B. Melamed, "The transition and autocorrelation structure of TES Processes Part I," Stat. Stoch. Mod, Vol. 8, No. 2, pp. 193–219, 1992.

[8] B. Melamed, Q. Ren, B. Sengupta, "The QTES/PH/1 queue," Performance Evaluation, Vol. 26, pp. 1–20, 1996.

[9] F.G. Foster, "Queues with batch arrival 1," Acta Mathematica Acad. Sci. Hung., Vol. 12, pp. 1–10, 1961.

[10] M.L. Chaundhry and J.G.C. Templeton, "A First Course in Bulk Queues," John Wiley & Sons, New York, 1983.

[11] R. Vich, "Z-transform theory and Applications," D. Reidel Publishing, London, 1987.

[12] A.H. Tchen, "Inequalities for Distributions with Given Marginals," Annal. of Prob., Vol. 8, No. 4, pp. 814–827, 1980.

[13] M. Livny, B. Melamed, A.K. Tsiolis, "The Impact of Autocorrelation on Queueing systems", Management Science, Vol. 39, No. 3, 1993.

[14] K. Kobayashi, Y. Takahashi, "The tail behavior of the stationary distribution of fluid queue with a Gaussian-type input rate process," JORSJ, Vol. 40, No. 1, 1997.

Hiroshi Toyoizumi
NTT Multimedia Networks
Laboratories
Midori-cho 3-9-11, Musashino-shi
Tokyo 180, Japan
toyo@hashi.tnl.ntt.co.jp