Characterizing the departure process of a single server queue from the embedded Markov renewal process at departures

Ping-Cheng Yeh and Jin-Fu Chang

Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
E-mail: pcyeh@ms1.hinet.net, jfchang@fcom.ee.ntu.edu.tw

Received 21 July 1998; revised 6 October 1999

In the literature, performance analyses of numerous single server queues are done by analyzing the embedded Markov renewal processes at departures. In this paper, we characterize the departure processes for a large class of such queueing systems. Results obtained include the Laplace–Stieltjes transform (LST) of the stationary distribution function of interdeparture times and recursive formula for $\{c_n \equiv$ the covariance between interdeparture times of lag $n\}$. Departure processes of queues are difficult to characterize and for queues other than $M/G/1$ this is the first time that $\{c_n\}$ can be computed through an explicit recursive formula. With this formula, we can calculate $\{c_n\}$ very quickly, which provides deeper insight into the correlation structure of the departure process compared to the previous research. Numerical examples show that increasing server irregularity (i.e., the randomness of the service time distribution) destroys the short-range dependence of interdeparture times, while increasing system load strengthens both the short-range and the long-range dependence of interdeparture times. These findings show that the correlation structure of the departure process is greatly affected by server regularity and system load. Our results can also be applied to the performance analysis of a series of queues. We give an application to the performance analysis of a series of queues, and the results appear to be accurate.

Keywords: departure process, $M/G/1$ type, correlation structure, covariance structure

1. Introduction

As the technology for modern communication develops, different networks are integrated or connected together to improve performance. Hence, a communication connection may consist of many communication links connecting intermediate nodes. Thus, the conventional single stage queueing model is no longer sufficient to analyze the performance of such communication connections. Instead, we need to model the connection as a series of queues. This makes the departure process analysis an important issue, since the departure process of a node is the arrival process of some other node in the next stage. But departure process characterization of a queueing system has been difficult unless the queue has very special structure.

Define t_n as the interdeparture time between the nth and $(n + 1)$st departures, and X_n as the number of customers left behind in the system at the nth departure. The departure process is defined as the set $\{t_n\}$, $n = 0, 1, \ldots,$ [8]. In the literature, performance analyses of numerous single server queueing systems have been done by analyzing their embedded Markov renewal processes at departures. Many of these queues are of the $M/G/1$ type, a large class of queues introduced by Neuts [11], which contains many well-known models such as $M/G/1$, *PH*$/G/1$, $M/$*SM*$/1$, $N/G/1$, and etc. (Notice that $M/G/1$ queues is not equivalent to queues of $M/G/1$ type, the former is only a very small subclass of the latter.) In this paper, we characterize the departure process for this large class of single server queueing models. For most cases, the departure process is not renewal [1]). Thus, the distribution function and the correlation structure of interdeparture times are both needed to characterize the departure processes. The Laplace–Stieltjes transform (LST) of the stationary distribution function of interdeparture times and recursive formula for $\{c_n \equiv \mathrm{Cov}(t_0, t_n)\}$ are both derived in this paper. We wish to emphasize that for queues other than $M/G/1$ queues, this is the first time that $\{c_n\}$ can be computed through an explicit recursive formula. (The expression of $\{c_n\}$ of $M/G/1$ queues has been derived by King [8].)

Over the past years, the departure process of queues receiving renewal arrival process has been extensively analyzed in many papers, such as [2,3,5,7,8,12], etc. Saito [13] analyzed the departure process of an $N/G/1$ queue, which was the pioneer work in understanding the departure process of a queue receiving nonrenewal arrival process. Conventionally, the correlation structure of a departure process is studied by deriving the generating function of the sequence $\{c_n\}$. Saito [13] derived the generating function for an $N/D/1$ queue, and this methodology can be generalized to other queues of $M/G/1$ type. However, the generating function is not in explicit form except for a few cases, which makes it very complicated to calculate $\{c_n\}$. Thus, we can only observe the summing-up behavior of $\{c_n\}$ through the generating function. In our paper, the recursive formula of $\{c_n\}$ is derived, which enables us to calculate $\{c_n\}$ very quickly. Thus, we can have a much clearer understanding of the correlation structure of the departure process.

Through numerical examples we observe that increasing server irregularity (i.e., the randomness of the service time distribution) destroys the short-range dependence of interdeparture times, while increasing system load strengthens both the short-range and the long-range dependence of interdeparture times. These findings conclude that server regularity and system load affect the correlation structure of the departure process to a great extent.

The rest of the paper is organized as follows. In section 2, the mathematical model is described. In section 3, we characterize the departure process by deriving the LST of the stationary distribution function of interdeparture times and recursive formula of $\{c_n\}$. In section 4, using numerical experiments, we examine the correlation structure of the departure processes of various *MMPP*$/G/1$ queues (MMPP, the Markov-modulated Poisson process, see [6]). We also give an application of our work to the performance analysis of a series of queues. Finally, conclusions are given in section 5.

2. Model definition

In this paper, we are concerned with single server queues of $M/G/1$ type. Queues of $M/G/1$ type were first introduced by Neuts [11] as an extension of $M/G/1$ queues to contain many well-known, complicated models such as *MMPP*$/G/1$, *PH*$/G/1$, $M/$*SM*$/1$, $N/G/1$, etc. Thus, we shall start the model description from the $M/G/1$ queue. Consider an $M/G/1$ queue with a service time distribution function $\widetilde{H}(x)$. Define X_n as the number of customers left behind in the system at the nth departure and t_n as the interdeparture time between the nth and $(n+1)$st departures. According to Neuts [11], $\{X_n, t_n\}$ forms a Markov renewal sequence on the 2-dimensional state space $\{i \geqslant 0\} \times \{0, \infty\}$, where $i \in \mathbb{Z}$. The transition probability matrix $\widetilde{Q}(x)$ with elements

$$\widetilde{Q}_{ij}(x) = P\{X_n = j,\ t_n \leqslant x \mid X_{n-1} = i\}, \tag{1}$$

for $i \geqslant 0$, $j \geqslant 0$, $x \geqslant 0$, is given by Neuts [11]

$$\widetilde{Q}(x) = \begin{bmatrix} \tilde{b}_0(x) & \tilde{b}_1(x) & \tilde{b}_2(x) & \tilde{b}_3(x) & \tilde{b}_4(x) & \cdots \\ \tilde{a}_0(x) & \tilde{a}_1(x) & \tilde{a}_2(x) & \tilde{a}_3(x) & \tilde{a}_4(x) & \cdots \\ 0 & \tilde{a}_0(x) & \tilde{a}_1(x) & \tilde{a}_2(x) & \tilde{a}_3(x) & \cdots \\ 0 & 0 & \tilde{a}_0(x) & \tilde{a}_1(x) & \tilde{a}_2(x) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{2}$$

where

$$\tilde{a}_v(x) = \int_0^x \mathrm{e}^{-\lambda u} \frac{(\lambda u)^v}{v!} \, \mathrm{d}\widetilde{H}(u), \tag{3}$$

$$\tilde{b}_v(x) = \int_0^x \lambda \, \mathrm{e}^{-\lambda u} \tilde{a}_v(x-u) \, \mathrm{d}u = \int_0^x \left[1 - \mathrm{e}^{-\lambda(x-u)}\right] \mathrm{d}\tilde{a}_v(u). \tag{4}$$

Now, we consider more complicated models in [11], such as *MMPP*$/G/1$ or $M/$*SM*$/1$. For these models, since $\{X_n, t_n\}$ no longer forms a Markov renewal sequence, a new concept of "phase" is employed in the treatment of these queues. For instance, the phase of an *MMPP*$/G/1$ queue is the state of the arrival process during the time that the phase of an $M/$*SM*$/1$ queue is the state of the server (see Neuts [11]). Let the phase of the queue at the nth departure be denoted by J_n. For these complicated models, $\{X_n, J_n, t_n\}$ forms a Markov renewal sequence in the 3-dimensional state space $\{(i,j),\ i,j \geqslant 0\} \times \{0, \infty\}$, where $i, j \in \mathbb{Z}$, and the transition probability matrix $\widetilde{Q}(x)$ has the following form:

$$\widetilde{Q}(x) = \begin{bmatrix} \widetilde{B}_0(x) & \widetilde{B}_1(x) & \widetilde{B}_2(x) & \widetilde{B}_3(x) & \widetilde{B}_4(x) & \dots \\ \widetilde{C}_0(x) & \widetilde{A}_1(x) & \widetilde{A}_2(x) & \widetilde{A}_3(x) & \widetilde{A}_4(x) & \dots \\ 0 & \widetilde{A}_0(x) & \widetilde{A}_1(x) & \widetilde{A}_2(x) & \widetilde{A}_3(x) & \dots \\ 0 & 0 & \widetilde{A}_0(x) & \widetilde{A}_1(x) & \widetilde{A}_2(x) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad x \geqslant 0. \tag{5}$$

Notice that (5) is very similar to (2), except that $\widetilde{B}_0(x)$, $\widetilde{C}_0(x)$, $\{\widetilde{A}_v(x),\ v \geqslant 0\}$, $\{\widetilde{B}_v(x),\ v \geqslant 1\}$ are matrices while $\tilde{a}_v(x)$ and $\tilde{b}_v(x)$ are scalar functions. This is because the state space of $\{X_n, J_n, t_n\}$ has one more dimension than $\{X_n, t_n\}$. Such queues are said to be of $M/G/1$ type, which is self-explanatory from the similarity between (5) and (2). To analyze these models, Neuts used a technique called "matrix–geometric method", which was developed from the technique of $M/G/1$ queue analysis. We note that queues of $M/G/1$ type include multi-server queues, although only single server queues are considered in this paper.

Consider a single server queue of $M/G/1$ type, for which the state space of (X_n, J_n) consists of the states $(0,1), \dots, (0,k)$ and the lattice points (i,j), $i \geqslant 1$, $1 \leqslant j \leqslant m$, where k and m are fixed positive integers. According to Neut's original definition of $M/G/1$ type, k is permitted to be different from m to allow queueing systems such as vacation models to work differently when they become empty. If the arrivals are generated by one MMPP, then $k = m =$ the number of states of the MMPP. The partition of the submatrices in matrix $\widetilde{Q}(x)$ is done according to the number of customers in the system. (Row i of $\widetilde{Q}(x)$ in (5) corresponds to the transition probabilities of the case given that the previous departure left $(i-1)$ customers behind in the system.) The structure of $\widetilde{Q}(x)$ implies that there is one and only one customer leaving at each departure exactly like the $M/G/1$ queue. The dimensions of $\widetilde{B}_0(x)$, $\widetilde{C}_0(x)$, $\{\widetilde{A}_v(x),\ v \geqslant 0\}$, and $\{\widetilde{B}_v(x),\ v \geqslant 1\}$ are $k \times k$, $m \times k$, $m \times m$, and $k \times m$, respectively. The fact that k and m can be different, explains the existence of $\widetilde{C}_0(x)$. (Note that there is no $\tilde{c}_0(x)$ in (2).) From the similarity between (2) and (5), we obtain a useful interpretation of matrices $\widetilde{B}_v(x)$ as follows:

$$\big[\widetilde{B}_v(x)\big]_{ij} = \Pr\big\{\text{given a departure at time 0 (which left behind an empty system and the queue in phase } i\text{), the next departure occurs no later than time } x \text{ with the queue in phase } j\text{, which leaves } v \text{ customers behind in the system}\big\}. \tag{6}$$

Matrices $\widetilde{A}_v(x)$, $\widetilde{C}_0(x)$ can be interpreted similarly. With these interpretations, we have

$$\begin{aligned} \big[\widetilde{A}_v(x)\big]_{ij} &= \begin{cases} \Pr\{X_{n+1} - X_n = -1,\ J_{n+1} = j,\ t_n \leqslant x \mid X_n \geqslant 2,\ J_n = i\}, & \text{for } v = 0, \\ \Pr\{X_{n+1} - X_n = v - 1,\ J_{n+1} = j,\ t_n \leqslant x \mid X_n \geqslant 1,\ J_n = i\}, & \text{for } v \geqslant 1, \end{cases} \\ \big[\widetilde{B}_v(x)\big]_{ij} &= \Pr\{X_{n+1} = v,\ J_{n+1} = j,\ t_n \leqslant x \mid X_n = 0,\ J_n = i\}, \quad \text{for } v \geqslant 0, \\ \big[\widetilde{C}_0(x)\big]_{ij} &= \Pr\{X_{n+1} = 0,\ J_{n+1} = j,\ t_n \leqslant x \mid X_n = 1,\ J_n = i\}. \end{aligned} \tag{7}$$

We note that these matrices are homogeneous in n. Now define the following transform matrices:

$$
\begin{aligned}
&A_v(s) = \int_0^\infty \mathrm{e}^{-sx}\,\mathrm{d}\widetilde{A}_v(x), \qquad B_v(s) = \int_0^\infty \mathrm{e}^{-sx}\,\mathrm{d}\widetilde{B}_v(x),\\
&C_0(s) = \int_0^\infty \mathrm{e}^{-sx}\,\mathrm{d}\widetilde{C}_0(x),\\
&\widetilde{A}(x) = \sum_{v=0}^\infty \widetilde{A}_v(x), \qquad A(s) = \int_0^\infty \mathrm{e}^{-sx}\,\mathrm{d}\widetilde{A}(x) = \sum_{v=0}^\infty A_v(s),\\
&\widetilde{B}(x) = \sum_{v=1}^\infty \widetilde{B}_v(x), \qquad B(s) = \int_0^\infty \mathrm{e}^{-sx}\,\mathrm{d}\widetilde{B}(x) = \sum_{v=1}^\infty B_v(s).
\end{aligned}
\tag{8}
$$

These matrices appear frequently in the analysis following. We note that when the successive service times are independent, identically distributed nonnegative random variables with the common probability distribution function $\widetilde{H}(x)$ and LST $H(s)$, a very useful result can be obtained,

$$A(s)\boldsymbol{e} = H(s)\boldsymbol{e}, \tag{9}$$

where $\boldsymbol{e}$ denotes the vector $(1, 1, \ldots, 1)^{\mathrm{T}}$.

Finally, we introduce the stationary density of the number of customers in the system immediately after a departure. Write $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots)$ where $\boldsymbol{x}_0 = (x_{01}, x_{02}, \ldots, x_{0k})$, and $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})$, $i \geqslant 1$, with

$$
\begin{aligned}
x_{ij} = \Pr\{&\text{the departure leaving the system behind with } i \text{ customers and the phase}\\
&\text{of the queue} = j\}.
\end{aligned}
\tag{10}
$$

Vector $\boldsymbol{x}$ is also the invariant vector of matrix $\widetilde{Q}(\infty)$. Readers can refer to [9–11] for the computation of $\boldsymbol{x}_i$'s. We shall view them as known results in the later sections.

3. Characterizing the departure process

3.1. Stationary distribution of interdeparture times

Let $\widetilde{D}_n(x)$ denote the stationary distribution function of the sum of n-consecutive interdeparture times, and $D_n(s)$ denote its LST. In this section, we focus on $\widetilde{D}_1(x)$. Starting from this definition and (7), (10), we have

$$
\begin{aligned}
\widetilde{D}_1(x) &= \Pr\{t_0 \leqslant x\}\\
&= \boldsymbol{x}_0\left\{\widetilde{B}_0(x)\boldsymbol{e} + \sum_{v=1}^\infty \widetilde{B}_v(x)\boldsymbol{e}\right\} + \boldsymbol{x}_1\left\{\widetilde{C}_0(x)\boldsymbol{e} + \sum_{v=1}^\infty \widetilde{A}_v(x)\boldsymbol{e}\right\}\\
&\quad + \left\{\sum_{i=2}^\infty \boldsymbol{x}_i\right\}\sum_{v=0}^\infty \widetilde{A}_v(x)\boldsymbol{e}
\end{aligned}
$$

$$= \boldsymbol{x}_0\big\{\widetilde{B}_0(x)\boldsymbol{e} + \widetilde{B}(x)\boldsymbol{e}\big\} + \boldsymbol{x}_1\big\{\widetilde{C}_0(x)\boldsymbol{e} + \big[\widetilde{A}(x) - \widetilde{A}_0(x)\big]\boldsymbol{e}\big\} + \left\{\sum_{i=2}^{\infty}\boldsymbol{x}_i\right\}\widetilde{A}(x)\boldsymbol{e}. \tag{11}$$

The derivation in (11) can be explained as follows. According to the Bayes' theorem (see [1]), $\Pr\{t_0 \leqslant x\}$ equals the sum of $\Pr\{X_0 = i,\ t_0 \leqslant x\}$ over i. In the second line of (11), the first term associated with $\boldsymbol{x}_0$ accounts for $\Pr\{X_0 = 0,\ t_0 \leqslant x\}$, while the term associated with $\boldsymbol{x}_i$ accounts for $\Pr\{X_0 = i,\ t_0 \leqslant x\}$. To see this, we initially focus on the first row of $\widetilde{Q}(x)$, i.e., $\widetilde{B}_v(x),\ v = 0, 1, 2, \ldots$. According to (7), it is obvious that $[\widetilde{B}_v(x)]_{ij}$, the entry of $\widetilde{B}_v(x)$, is in fact the joint distribution function of $(X_1 = v,\ J_1 = j,\ t_0 \leqslant x \mid X_0 = 0,\ J_0 = i)$. Since we are only concerned with the marginal distribution function of $(t_0 \mid X_0 = 0,\ J_0 = i)$, we have to sum up $[\widetilde{B}_v(x)]_{ij}$ over J_1 and X_1. Multiplying $\widetilde{B}_v(x)$ by $\boldsymbol{e}$ (to sum over J_1) and summing over v (namely, X_1), we obtain the distribution function of $(t_0 \mid X_0 = 0,\ J_0 = i)$: $\{\widetilde{B}_0(x)\boldsymbol{e} + \widetilde{B}(x)\boldsymbol{e}\}$. Multiplying by $\boldsymbol{x}_0 = (x_{0\,1}, x_{0\,2}, \ldots, x_{0\,k})$, where $x_{0\,i} = \Pr(X_0 = 0,\ J_0 = i)$ from (10), we see that the first term of the third line in (11) is exactly $\Pr\{X_0 = 0,\ t_0 \leqslant x\}$. The remaining terms can be obtained in a similar manner. The LST of $\widetilde{D}_1(x)$ can be obtained as follows:

$$D_1(s) = \boldsymbol{x}_0\big\{B_0(s)\boldsymbol{e} + B(s)\boldsymbol{e}\big\} + \boldsymbol{x}_1\big\{C_0(s)\boldsymbol{e} + \big[A(s) - A_0(s)\big]\boldsymbol{e}\big\} + \left\{\sum_{i=2}^{\infty}\boldsymbol{x}_i\right\}A(s)\boldsymbol{e}. \tag{12}$$

Note that when the successive service times are independent, identically distributed nonnegative random variables with the common probability distribution function $\widetilde{H}(x)$ and LST $H(s)$, we can substitute (9) into (12) to obtain

$$D_1(s) = \boldsymbol{x}_0\big\{B_0(s)\boldsymbol{e} + B(s)\boldsymbol{e}\big\} + \boldsymbol{x}_1\big\{C_0(s)\boldsymbol{e} - A_0(s)\boldsymbol{e}\big\} + (1 - x_0\boldsymbol{e})H(s). \tag{13}$$

3.2. Recursive formula for the covariances of interdeparture times

In this section, we derive the recursive formula for $\{c_n\}$. Let σ_n^2 denote the variance of the distribution $\widetilde{D}_n(x)$, i.e., $D_n''(0) - [D_n'(0)]^2$. The overall methodology is to derive the recursive formula of $\{\sigma_n^2\}$ first, and then obtain $\{c_n\}$ from $\{\sigma_n^2\}$ next.

We first define the column vector function $\widetilde{U}(q, r, x)$, which is essential for deriving $\{\sigma_n^2\}$. The ith element of $\widetilde{U}(q, r, x)$ is defined as follows:

$$\big[\widetilde{U}(q, r, x)\big]_i = \Pr\{t_0 + t_1 + \cdots + t_{r-1} \leqslant x \mid X_0 = q,\ J_0 = i\}, \tag{14}$$

where $q, r \in \mathbb{Z},\ q \geqslant 0,\ r \geqslant 1$ and $x \in \mathbb{R},\ x \geqslant 0$. We further define the following vectors:

$$U(q,r,s) = \int_0^\infty \mathrm{e}^{-sx}\,\mathrm{d}\widetilde{U}(q,r,x),$$
$$U'_{q,r} = \frac{\partial}{\partial s}U(q,r,s)\bigg|_{s=0}, \qquad U''_{q,r} = \frac{\partial^2}{\partial s^2}U(q,r,s)\bigg|_{s=0}. \tag{15}$$

Dimensions of $\widetilde{U}(q,r,x)$ and $U(q,r,s)$ are $k \times 1$ when $q = 0$, and $m \times 1$ otherwise. Note that for all q, r, $U(q,r,0) = \widetilde{U}(q,r,\infty) = \boldsymbol{e}$.

Now we derive $U(q,r,s)$ for all $q, r \in \mathbb{Z}$, $q \geqslant 0$, $r \geqslant 1$. From (7), (14), and (15), it is easy to obtain

(a) when $q > r$,

$$U(q,r,s) = U(r+1,r,s) = \left\{\sum_{v=0}^{\infty} A_v(s)\right\}^r \boldsymbol{e} = \left\{A(s)\right\}^r \boldsymbol{e}, \tag{16}$$

(b) when $q \leqslant r$,

$$U(1,1,s) = C_0(s)\boldsymbol{e} + \sum_{v=1}^{\infty} A_v(s)\boldsymbol{e} = C_0(s)\boldsymbol{e} + \left[A(s) - A_0(s)\right]\boldsymbol{e}, \tag{17}$$

$$U(0,1,s) = \sum_{v=0}^{\infty} B_v(s)\boldsymbol{e} = B_0(s)\boldsymbol{e} + B(s)\boldsymbol{e}, \tag{18}$$

$$\begin{aligned} U(0,r,s) &= \sum_{v=0}^{\infty} B_v(s)U(v,r-1,s) \\ &= \sum_{v=0}^{r-1} B_v(s)U(v,r-1,s) \\ &\quad + \left[B(s) - \sum_{v=1}^{r-1} B_v(s)\right]\left\{A(s)\right\}^{r-1}\boldsymbol{e}, \quad \text{for } r > 1, \end{aligned} \tag{19}$$

$$\begin{aligned} U(1,r,s) &= C_0(s)U(0,r-1,s) + \sum_{v=1}^{\infty} A_v(s)U(v,r-1,s) \\ &= C_0(s)U(0,r-1,s) + \sum_{v=1}^{r-1} A_v(s)U(v,r-1,s) \\ &\quad + \left[A(s) - \sum_{v=0}^{r-1} A_v(s)\right]\left\{A(s)\right\}^{r-1}\boldsymbol{e}, \quad \text{for } r > 1, \end{aligned} \tag{20}$$

$$\begin{aligned} U(r,r,s) &= A_0(s)U(r-1,r-1,s) + \sum_{v=1}^{\infty} A_v(s)U(r+v-1,r-1,s) \\ &= A_0(s)U(r-1,r-1,s) + \left[A(s) - A_0(s)\right]\left\{A(s)\right\}^{r-1}\boldsymbol{e}, \quad \text{for } r > 1, \end{aligned} \tag{21}$$

$$\begin{aligned} U(q,r,s) &= \sum_{v=0}^{\infty} A_v(s) U(q+v-1, r-1, s) \\ &= \sum_{v=0}^{r-q} A_v(s) U(q+v-1, r-1, s) \\ &\quad + \left[A(s) - \sum_{v=0}^{r-q} A_v(s) \right] \{A(s)\}^{r-1} \boldsymbol{e}, \quad \text{for } q, r > 1. \end{aligned} \tag{22}$$

Let us make a few remarks to help understand (16)–(22). First, notice that there are two types of interdeparture times, one is *a pure service time* while the other is *an idle period plus a service time*. The influence of q (the number of customers left behind in the system by the previous departure) on $\widetilde{U}(q,r,x)$ and $U(q,r,s)$ is the probability of the occurrences of idle periods during the next r-consecutive interdeparture times. When $q > r$, it is impossible for any idle period to occur during the next r-consecutive interdeparture times, thus $U(q,r,s) = U(r+1,r,s)$ in (16). Equations (16)–(22) can be understood in a way similar to the derivation of $\widetilde{D}_1(x)$ in section 3.1. For $r = 1$ and $q = 0$ or 1, we choose row $(q+1)$ of $\widetilde{Q}(x)$ in (5) to work with. Multiplying by $\boldsymbol{e}$ and summing over v, we obtain (17) and (18), the LST of the marginal distribution of $(t_0 \leqslant x \mid X_0 = q,\ J_0 = i\}$. For $r \geqslant 2$, we either obtain $U(q,r,s)$ directly as in (16) or relate it of $U(q,r',s)$ of $r' < r$ as in (19)–(22). Hence, $U(q,r,s)$ can be computed through (16)–(22) recursively. Furthermore, the recursive formulae of $U'_{q,r}$ and $U''_{q,r}$ can be obtained easily from (16)–(22).

By (10) and (14), it is simple to see

$$D_n(s) = \sum_{i=0}^{n} \boldsymbol{x}_i U(i,n,s) + \left\{ \sum_{i=n+1}^{\infty} \boldsymbol{x}_i \right\} U(n+1,n,s). \tag{23}$$

Thus, we have

$$D'_n(0) = \sum_{i=0}^{n} \boldsymbol{x}_i U'_{i,n} + \left\{ \sum_{i=n+1}^{\infty} \boldsymbol{x}_i \right\} U'_{n+1,n}, \tag{24}$$

$$D''_n(0) = \sum_{i=0}^{n} \boldsymbol{x}_i U''_{i,n} + \left\{ \sum_{i=n+1}^{\infty} \boldsymbol{x}_i \right\} U''_{n+1,n}, \tag{25}$$

$$\sigma_n^2 = D''_n(0) - \left[D'_n(0) \right]^2. \tag{26}$$

Also since

$$\sigma_{n+1}^2 = \mathrm{Var}(t_0 + t_1 + \cdots + t_n) = (n+1)\sigma_1^2 + 2\sum_{i=1}^{n} (n+1-i) c_i, \tag{27}$$

we have

$$c_n = \frac{1}{2}\left\{\sigma_{n+1}^2 - (n+1)\sigma_1^2 - 2\sum_{i=1}^{n-1}(n+1-i)c_i\right\}. \tag{28}$$

Hence, the methodology to calculate c_n, $n = 1, 2, \dots, l$, is:

1. Compute the vectors $U'_{q,r}$, $U''_{q,r}$ for $q = 0, 1, \dots, r+1$ and $r = 1, 2, \dots, l+1$.

2. Obtain σ_n^2 through (24)–(26) for $n = 1, 2, \dots, l+1$.

3. Calculate c_n recursively by (28) for $n = 1, 2, \dots, l$.

Proceeding in this manner, we can calculate $\{c_n\}$ very quickly. In figure 1, the computation time (expressed in seconds) of $\{c_n\}$ on an ordinary PC with Pentium II CPU (233 MHz) vs. lag n is plotted in logarithmic scale. We see that $\{c_n\}$ can be obtained in only a few seconds when n is small, and the computation time tends to be O($n^{2.5}$) as n increases. This shows the efficiency of our algorithm. In fact, for Markovian models considered in this paper, c_n tends to approach zero as n increases. This is seen in the numerical examples we provide in section 4.1. Furthermore, the most practical usage of c_n is "moment matching", which is introduced in section 4.2. For moment matching, only a few c_n's are needed. Thus, computation of c_n for large n is rarely needed. Finally, we note that γ_n, the correlation coefficient between interdeparture times of lag n, can be obtained easily by c_n/σ_1^2.

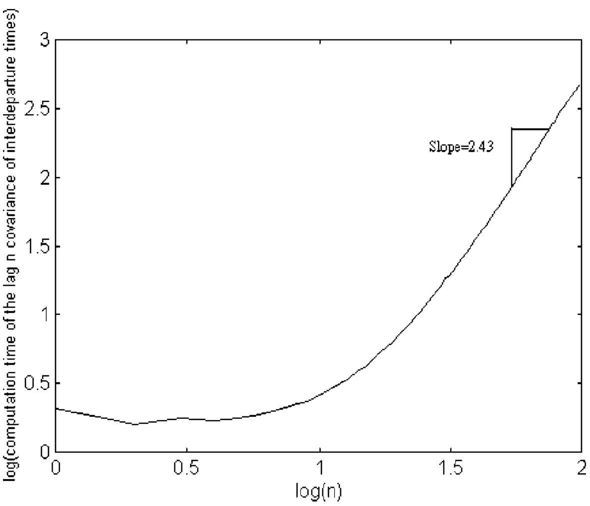

Figure 1. Computation time of c_n vs. lag n of an $MMPP/D/1$ queue in logarithmic scale. The service time = 1, and the MMPP is parameterized by $\lambda_1 = 1.0$, $\lambda_2 = 0.3$, $\sigma_1 = \sigma_2 = 0.1$.

4. Numerical examples and discussions

4.1. Departure processes of various $MMPP/G/1$ queues

In this section, we examine the correlation structure of the departure processes of various $MMPP/G/1$ queues [6] through numerical experiments. In all of these experiments, the arrival process is the same 2-state MMPP parameterized by

$$Q = \begin{bmatrix} -0.1 & 0.1 \\ 0.1 & -0.1 \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} 1.0 & 0 \\ 0 & 0.3 \end{bmatrix}. \tag{29}$$

In the first experiment, we let the service time distribution be the deterministic (D), Erlang type 4 (E4), exponential (Exp), and hyperexponential (H2) distribution with the same service rate 1.0. Note that these are distributions of increasing randomness (here we define randomness as the squared coefficient of variation, i.e., variance/mean^2). In the second experiment, we let the service time distribution be deterministic with service rates 1.25, 1.0, and 0.8, while the corresponding system loads are respectively 0.52, 0.65, and 0.81. In the third experiment, we let the service time distribution be exponential with service rates 1.25, 1.0, and 0.8, while the corresponding system loads are respectively 0.52, 0.65, and 0.81. The correlation coefficient between interdeparture (interarrival) times of lag n vs. lag n of these experiments are plotted in figures 2–4, respectively. For convenience of observation, these point plots are connected by curves.

From figure 2, we observe that as the server irregularity (i.e., the randomness of service time distribution) increases, the dependence of interdeparture times decreases, and the decreasing effect is especially obvious for the short-range dependence. This is reasonable since, for a queue of irregular server, the similarity of the consecutive

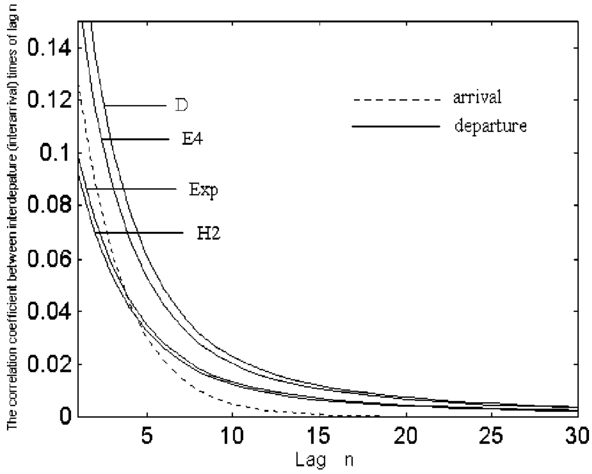

Figure 2. Correlation coefficient between interdeparture (interarrival) times of lag n vs. lag n of $MMPP/G/1$ queues with deterministic, Erlang type 4, exponential, and hyperexponential service time distributions of mean 1.0. The MMPP is parameterized by $\lambda_1 = 1.0$, $\lambda_2 = 0.3$, $\sigma_1 = \sigma_2 = 0.1$.

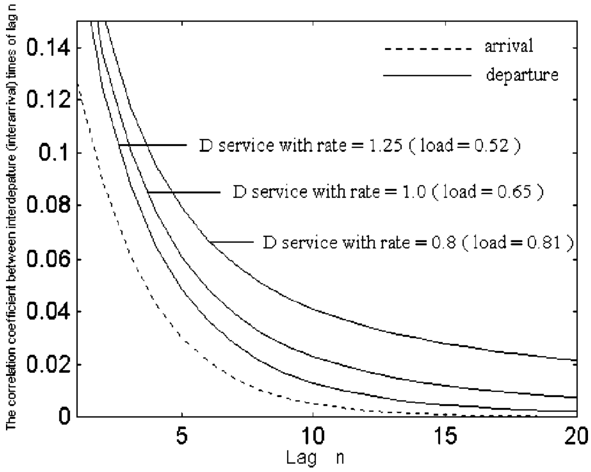

Figure 3. Correlation coefficient between interdeparture (interarrival) times of lag n vs. lag n of *MMPP*$/D/1$ queues with service rates 1.25, 1.0, and 0.8. The MMPP is parameterized by $\lambda_1 = 1.0$, $\lambda_2 = 0.3$, $\sigma_1 = \sigma_2 = 0.1$.

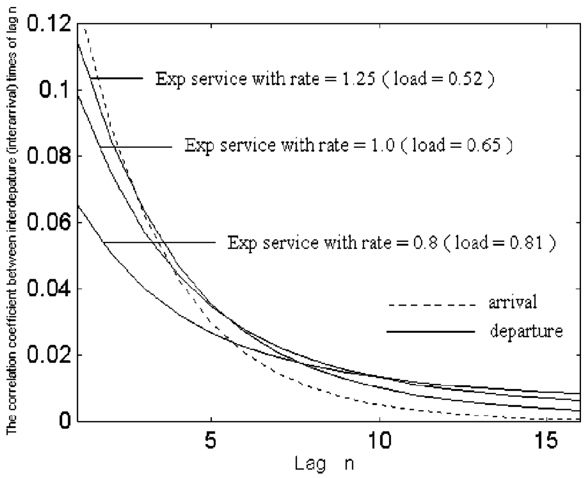

Figure 4. Correlation coefficient between interdeparture (interarrival) times of lag n vs. lag n of *MMPP*$/M/1$ queues with service rates 1.25, 1.0, and 0.8. The MMPP is parameterized by $\lambda_1 = 1.0$, $\lambda_2 = 0.3$, $\sigma_1 = \sigma_2 = 0.1$.

interdeparture times is destroyed by the randomness of the service time distribution, thus the short-range dependence of interdeparture time is much smaller compared to that of regular server. Also since the similarity between interdeparture times of large lag n is small whether the server is regular or not, the decreasing effect of long-range dependence is not as obvious as the short-range dependence.

From figure 3, we observe that for the queue of a regular server (deterministic), the short-range and the long-range dependence of interdeparture times both increase as the system load increases. We note that an interdeparture time of a single server queue is either a pure service time or an idle period plus a service time. As the queue size increases with the system load, there are more interdeparture times of a pure service time type. Thus the similarity of interdeparture times is increased, which also increases the dependence of interdeparture times.

From figure 4, we observe that for the queue of an irregular server (exponential), as the service rate decreases, the short-range dependence of interdeparture times decreases, while the long-range dependence increases. This is, in fact, the compound result of previous experiments. As the service rate decreases, the server irregularity (i.e., the randomness of the service time distribution) increases, thus the short-range dependence decreases. On the other hand, the system load increases as the service rate decreases, thus the long-range dependence increases.

The phenomenon that c_n approaches zero as n grows is widely seen in figures 2–4.

4.2. *Performance analysis of a series of queues*

In this section, we provide an application of our work to the performance analysis of a series of queues.

4.2.1. *A series of queues without cross traffic flows*

We first consider the two-node tandem configuration shown in figure 5. The input of node 1 is a 2-state MMPP, and the servers of both nodes are GI servers. Notice that node 1 is simply an *MMPP*$/G/1$ queue, and the nodal performance can be obtained as in [6]. Whereas in node 2, the input process is not a specified, well-defined process, and thus the nodal performance is difficult to obtain. To overcome the problem, we apply the work developed above in section 3 to obtain the mean, variance, c_1, and c_2 of the departure process from node 1. Then we approximate the departure process by an MMPP, parameterized with its interarrival times statistics matched to the statistics obtained previously. (The parameters of the MMPP can be obtained using ordinary numerical methods for nonlinear equations.) The method is called *moment matching*. Node 2 thus becomes another *MMPP*$/G/1$ queue, and the nodal performance can be obtained easily. Tandem configuration of more than two nodes can be analyzed by iteratively applying this methodology.

We use an example to demonstrate the methodology. First, we parameterize MMPP_1 in figure 5 by

$$Q_1 = \begin{bmatrix} -0.01 & 0.01 \\ 0.02 & -0.02 \end{bmatrix}, \qquad \Lambda_1 = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.1 \end{bmatrix}, \tag{30}$$

Figure 5. Two-node tandem configuration without cross traffic.

and suppose that both servers are deterministic with rates of 1.0 and 0.25, respectively. Examine the delay of MMPP_1 passing through each node. The nodal delay of node 1 is easy to obtain via the $MMPP/D/1$ analysis of Fischer and Meier-Hellstern [6]. For node 2, we apply the moment matching method to obtain MMPP_2, parameterized by

$$Q_2 = \begin{bmatrix} -0.00847 & 0.00847 \\ 0.0175 & -0.0175 \end{bmatrix}, \qquad \Lambda_2 = \begin{bmatrix} 0.0507 & 0 \\ 0 & 0.997 \end{bmatrix}. \tag{31}$$

The covariance between interarrival times of lag n of MMPP_2 and c_n of the departure process from node 1 are both plotted versus n in figure 6, where the two curves nearly overlap. This implies that the correlation structure of MMPP_2 is almost the same as the departure process of node 1, which supports the accuracy of the moment matching method.

Approximating the departure process of node 1 by MMPP_2, we can obtain the nodal delay of node 2 and thus obtain the end-to-end delay. Results are listed in table 1, indicating that the accuracy is good.

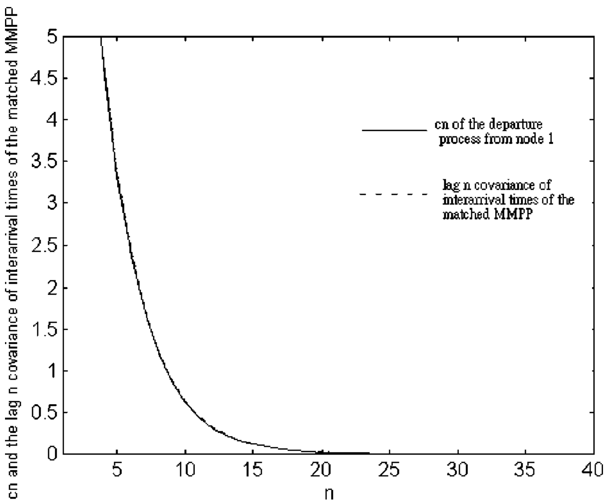

Figure 6. c_n of the departure process from node 1 and the lag n covariance of interarrival times of the matched MMPP vs. lag n.

Table 1
Mean delay analysis of the two-node tandem configuration without cross traffic.

	Node 1	Node 2	Total
Simulation	1.0411 ± 0.0003	4.8533 ± 0.0081	5.8944 ± 0.0082
Analysis	1.0408	4.8847	5.9256

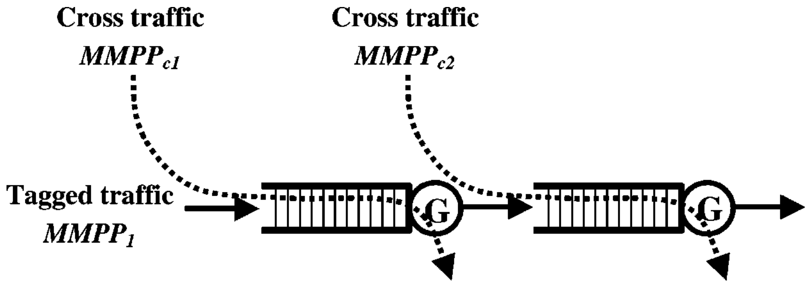

Figure 7. Two-node tandem configuration with cross traffic.

Table 2
Mean delay analysis of the two-node tandem configuration with cross traffic.

	Node 1	Node 2	Total
Simulation	1.1946 ± 0.0012	2.2479 ± 0.1101	3.4425 ± 0.1108
Analysis	1.1952	2.3080	3.5032

4.2.2. A series of queues with cross traffic flows

Now we consider the two-node tandem configuration in figure 7. The tagged traffic and cross traffic are all 2-state MMPP_s. We are here concerned with the performance of the tagged traffic passing through each node. The existence of the cross traffic makes the departure process from node 1 very difficult to characterize, not to mention the difficulty of performance analysis for node 2. Ferng and Chang [4] proposed a scheme, which enables us to eliminate the cross traffic by replacing the server by an equivalent effective server. Thus, the whole configuration can be approximated by another two-node tandem configuration without cross traffic. The performance can then be obtained by the methodology outlined in section 4.2.1. As an example, we let the tagged traffic be the MMPP_1 parameterized by (30). The servers of both nodes remain deterministic with the same service rate 1.0. The cross traffic flows of node 1 and 2 are parameterized by

$$Q = \begin{bmatrix} -0.1 & 0.1 \\ 0.1 & -0.1 \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} 4r/3 & 0 \\ 0 & 2r/3 \end{bmatrix}, \tag{32}$$

where $r = 0.2$ at node 1 and 0.6 at node 2. The nodal and end-to-end delays are listed in table 2, and the accuracy is again good.

5. Conclusions

In this paper, we have characterized the departure process of a single server queue from its embedded Markov renewal process at departures. The LST of the stationary distribution function of interdeparture times and an efficient recursive formula for covariance between interdeparture times of lag n were both obtained. The latter formula enables us to have a much deeper insight to the correlation structure of departure processes compared to previous research. Numerical experiments showed that the server irregularity and the system load both dominate the correlation structure of a departure process. The former destroys the short-range dependence of interdeparture times, while the latter strengthens both the short-range and the long-range dependence of interdeparture times. Finally, we gave an application of our work to the performance analysis of a series of queues and the results appeared to be accurate.

References

[1] E. Cinlar, *Introduction to Stochastic Processes* (Prentice-Hall, Englewood Cliffs, NJ, 1975).

[2] D.J. Daley, The correlation structure of the output process of some single server queueing systems, Ann. Math. Statist. 39 (1968) 1007–1019.

[3] D.J. Daley, Queueing output processes, Adv. in Appl. Probab. 8 (1976) 395–415.

[4] H.W. Ferng and J.F. Chang, Connection-wise end-to-end performance analysis of queueing networks with MMPP inputs, submitted.

[5] P.D. Finch, The output process of the queueing system $M/G/1$, J. Roy. Statist. Soc. Ser. B 21 (1959) 375–380.

[6] W. Fischer and K.S. Meier-Hellstern, The Markov-modulated Poisson process cookbook, Performance Evaluation 18 (1992) 149–171.

[7] J.H. Jenkins, On the correlation structure of the departure process of the $M/E_\lambda/1$ queue, J. Roy. Statist. Soc. Ser. B 28 (1966) 336–344.

[8] R.A. King, The covariance structure of the departure process from $M/G/1$ queues with finite waiting lines, J. Roy. Statist. Soc. Ser. B 33 (1971) 401–406.

[9] D.M. Lucantoni, New results on the single server queue with a batch Markovian arrival process, Comm. Statist. Stochastic Models 7 (1) (1991) 1–46.

[10] M.F. Neuts, A versatile Markovian point process, J. Appl. Probab. 16 (1979) 764–779.

[11] M.F. Neuts, *Structured Stochastic Matrices of $M/G/1$ Type and their Applications* (Marcel Dekker, New York, 1989).

[12] C.D. Pack, The output of an $M/D/1$ queue, Oper. Res. 23 (1975) 750–760.

[13] H. Saito, The departure process of an $N/G/1$ queue, Performance Evaluation 11 (1990) 241–251.