ELSEVIER

Performance Evaluation 38 (1999) 85–103

PERFORMANCE EVALUATION
An International Journal

www.elsevier.com/locate/peva

Distributions of the interdeparture times in FCFS and nonpreemptive priority $MMAP[2]/G[2]/1$ queues

Qi-Ming He [a,*], David A. Stanford [b]

[a] *Department of Industrial Engineering, DalTech, Dalhousie University, Halifax, NS, Canada B3J 2X4*

[b] *Department of Statistics and Actuarial Sciences, University of Western Ontario, London, Ont., Canada N6A 5B9*

Received 21 August 1997; received in revised form 21 May 1999

Abstract

This paper deals with single server queueing systems with two classes of customers and a correlated arrival process. The focus is on the interdeparture times of each class of customer. A uniform approach is introduced to find the Laplace–Stieltjes transforms of interdeparture times for queueing systems with equal priority and nonpreemptive priority, respectively. Algorithms can be developed for computing the variances of interdeparture times. The methodology developed in this paper can be used to analyze queueing systems with correlated input processes with special arrival patterns. ©1999 Elsevier Science B.V. All rights reserved.

Keywords: Queueing theory; Matrix analytic methods; Departure process; Interdeparture time

1. Introduction

In open queueing network applications, it is frequently the case that the network must handle several customer classes with possibly differing arrival patterns, service time characteristics and/or routing probabilities among the nodes. In some cases, these differing customer classes will contend with each other on an equal priority basis for the attention of the server. In others, there may be a priority arrangement among them. As the departure process at a given node contributes to the arrival processes elsewhere, it is important to consider how the interactions among the various classes of customers affect the output process of each class. An important element in that study is the analysis of the stationary interdeparture time distribution.

Except for a few very specialized queues such as the $M/M/1$, the output process of most queueing systems is correlated. As a result, it has been practically impossible to analyze even tandem arrangements of non-Markovian queues, because the input to the second stage is already correlated. Until now, no work

* Corresponding author.: Tel.: +1-902-494-6141; fax: +1-902-420-7858.
E-mail address: qi-ming.he@dal.ca (Q.-M. He)

PII: S0166-5316(99)00039-5

we are aware of has addressed the issue of departure processes for queues with correlated arrivals or an arrival process with certain pattern.

This paper deals with single server queueing systems with two classes of customers and a correlated arrival process. The interdeparture time distributions are provided for each class of customers, for both the case where the classes contend on an equal priority basis, and the case where a nonpreemptive priority arrangement exists. Through a corollary, we also find the interdeparture time distribution of a FCFS queue with only a single class of customer. The characterization of these distributions is done in terms of their Laplace–Stieltjes transforms from which we obtain formulas for the variance of the interdeparture time (as the mean is already known).

Interdeparture times of queueing systems and their associated departure processes have been studied extensively. Daley [1], Nain [6], Saito [11], Whitt [17], and many others studied the departure process of queueing systems with a single class of customer. However, the results are not easily extended to queueing systems with several classes of customers. Thus, the interdeparture time, instead of the departure process, became the main topic when several classes of customers are present. Stanford [12], Stanford and Fischer [13,14] studied interdeparture times of queueing systems with several classes of customers. They focused primarily on queueing systems with independent Poisson input processes. (In [14], the input process of one class of customer was allowed to have hyperexponentially distributed interarrival times.) The current paper extends the results of [12–14], by introducing a versatile tractable model of a correlated arrival process with two classes of customers. This generalization makes it possible to analyze a variety of queueing systems whose arrival process may have a special pattern such as cyclic arrivals (see Example 2.2).

The queueing systems we address in this paper have a Markov arrival process with marked transitions (*MMAP*). The advantage of using *MMAP* is that it allows the correlation between input processes of different classes of customers to be captured, while the queueing system is still analytically and numerically tractable (see [3,15,16] and references therein). Compared to Poisson type input processes with multiple classes of customers, the *MMAP* can be used to model bursty stochastic processes with special arrival patterns. Thus, the results obtained in this paper can be useful in analyzing telecommunication systems where the input processes are often bursty. Due to the use of *MMAP* as an input process, an approach which is different from that of [12,13] (and which is more suitable for the matrix environment) is utilized in this paper to find distributions of interdeparture times.

The rest of the paper is organized as follows. Section 2 introduces the queueing systems of interest in detail. In Section 3, the LST of the interdeparture time of each class of customer in a queueing system with equal priority is derived. As a corollary, the LST of the interdeparture time of a $MAP/G/1$ queue is obtained. Sections 4 and 5 give the LSTs of the interdeparture times of the highest priority class and the lowest priority class in a queueing system with nonpreemptive priority respectively. The first and second moments of the interdeparture time of queueing systems with equal priority are given in Section 6. A numerical example is presented to show how the computation of the first two moments can be carried out. Finally, in Section 7, some discussion is given to the results obtained in this paper.

2. The $MMAP[K]/G[K]/1$ queue

The queueing system of interest in this paper is a single server queueing system with a Markov arrival process with marked transitions ($MMAP[K]$). Customers are distinguished into K classes. The service times of each class of customer are independent and identically distributed random variables. The service times of different classes of customers are independent and may have different distribution functions. To define the queueing systems of interest explicitly, the input process $MMAP[K]$ is introduced first and then the service disciplines are specified.

The following definition of the $MMAP[K]$ was given by Marcel Neuts. (See [4] for more details. See [7–10] for the original definition of the Markov arrival process.) First, consider an m-state Markov renewal process with an irreducible embedded Markov chain with transition probability matrix $P = (p_{i,j})$ and exponential distributions for the sojourn time in state i given by $H_i(x) = 1 - \exp\{-\sigma_i x\}$, for $1 \le i, j \le m$. This Markov renewal process is also a Markov process (Markov chain in continuous time). Let D be the infinitesimal generator of this Markov process. The matrix D, matrix P, and parameters σ_i, $1 \le i \le m$, of the Markov process are related to each other as $D_{i,i} = -(1 - p_{i,i})\sigma_i$, for $1 \le I \le m$, and $D_{i,j} = p_{i,j}\sigma_i$, for $1 \le i, j \le m, i \ne j$. Let $J(t)$ denote the state of this Markov process at time t.

An $MMAP[K]$ is obtained by marking the transitions (arrivals) of a Markov arrival process as follows. Define a Markov renewal process $\{(J_n, L_n, \tau_n), n \ge 0\}$ on the state space $\{[\{1, \dots, m\} \times \{1, \dots, K\}] \times [0, \infty)\}$ with the transition probability matrix, for $1 \le i, j \le m$, $1 \le k \le K$, $i \ne j$, $x \ge 0$,

$$\mathbf{P}\{J_n = j, L_n = k, \tau_n \le x | J_{n-1} = i\} = \left[\int_0^x \exp\{D_0 u\} \mathrm{d}u\, D_k\right]_{i,j}, \tag{2.1}$$

where J_n is the state of the Markov chain at the nth arrival epoch, L_n is the marking variable (or the class of the nth arrival), and τ_n is the time between the $(n-1)$st and the nth transitions. The matrices D_k, $1 \le k \le K$, are nonnegative. The matrix D_0 has negative diagonal elements and nonnegative off-diagonal elements. D_0 is assumed to be nonsingular. $\{J(t), t \ge 0\}$ is called the underlying Markov process with an infinitesimal generator D. The relationship between the infinitesimal generator D and matrices $\{D_k, 0 \le k \le K\}$ is

$$D = D_0 + \sum_{k=1}^{K} D_k. \tag{2.2}$$

An arrival is called class k if the arrival is marked by k. The (matrix) marking rate of class k arrivals is D_k. Denote by $\boldsymbol{\theta}$ the stationary probability vector of the matrix D. The stationary arrival rate of class k arrivals is given by

$$\lambda_k = \boldsymbol{\theta} D_k \mathbf{e}, \quad 1 \le k \le K, \tag{2.3}$$

where $\mathbf{e}$ is the column vector with all components one. Let $N_k(t)$ be the total number of class k arrivals in $(0, t)$. The vector $(N_1(t), \dots, N_K(t))$ represents the $MMAP[K]$. Denote by, for $1 \le i, j \le m$,

$$p_{i,j}(n_1, \dots, n_K, t) = \mathbf{P}\{N_1(t) = n_1, \dots, N_K(t) = n_K, J(t) = j | J(0) = i\} \tag{2.4}$$

and $P(n_1, \dots, n_K, t)$ an $m \times m$ matrix with elements $p_{i,j}(n_1, \dots, n_K, t)$. It can be proved that

$$\begin{aligned} P^*(z_1, \dots, z_K, t) &\equiv \mathbf{E} z_1^{N_1(t)} \dots z_K^{N_K(t)} \equiv \sum_{(n_1, \dots, n_K)} P(n_1, \dots, n_K, t) z_1^{n_1} \cdots z_K^{n_K} \\ &= \exp\{(D_0 + z_1 D_1 + \cdots + z_K D_K)t\}. \end{aligned} \tag{2.5}$$

Remark 2.1. Notice that $N(t) = N_1(t) + \cdots + N_K(t)$, which counts all the arrivals in $(0, t)$, is a Markov arrival process (MAP) with a matrix representation $\{D_0, D_1 + \cdots + D_K\}$.

Example 2.1. The superposition process of K independent Poisson processes is the simplest example of an $MMAP[K]$. Suppose that the arrival rates of the K Poisson processes are $\{\lambda_1, \lambda_2, \dots, \lambda_K\}$. Then the matrix representation of their superposition process is $D_0 = -\lambda = -(\lambda_1 + \cdots + \lambda_K)$, $D_1 = \lambda_1, \dots, D_K = \lambda_K$. We will refer to this example frequently to help visualize our results. For instance, for this special case, Eq. (2.5) becomes

$$P^*(z_1, \dots, z_K, t) = \exp\{(-\lambda + z_1\lambda_1 + \cdots + z_K\lambda_K)t\}. \tag{2.6}$$

Example 2.2. Consider an $MMAP[2]$ with $m = 2$ and

$$D_0 = \begin{pmatrix} -10 & 0 \\ 0 & -1 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & 10 \\ 0 & 0 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{2.7}$$

In this example, the two classes of customers arrive cyclically. More complicated $MMAP$s can be constructed to model arrival processes with other arrival patterns. Because of the special arrival pattern, it is difficult to analyze the corresponding queueing systems. In Example 6.1, we shall show how to use the methodology developed in this paper to analyze queueing systems with such arrival processes.

Back to the $MMAP[K]/G[K]/1$ queue. The arrivals of the $MMAP[K]$ correspond to customers in queueing systems of interest. The service times of class k customers have a common distribution function $F_k(x)$ with LST $f_k^*(s)$ and finite mean $1/\mu_k$, $1 \le k \le K$. The traffic intensity of the queueing system is defined as $\rho = \lambda_1/\mu_1 + \cdots + \lambda_K/\mu_K$. Assume that $\rho < 1$ so that the queueing system can reach its steady state. The service times are independent of each other and are independent of the input process. The service disciplines of queueing systems under consideration are specified as follows.

The $MMAP[K]/G[K]/1$ queue with equal priority. In this queueing system, all customers, regardless of their classes, are served on a "first-come-first-served" basis. Such queueing systems were considered in [3,16] for performance measures such as the queue lengths, waiting times, fundamental periods, etc. Stanford and Fischer [13] considered the interdeparture times in such a queueing system with Poisson arrival processes.

The $MMAP[K]/G[K]/1$ queue with nonpreemptive priority. In this queueing system, class 1 customers have the highest priority, then class $2, \dots,$ and finally class K the lowest priority. A higher priority customer enters the server when the server becomes free. For customers of the same class, the "first-come-first-served" service discipline is applied. It is assumed that no service will be interrupted until it is completed. This queueing system was studied in [15]. Stanford [12] considered such a queueing system with Poisson arrival processes.

This paper focuses on the cases where two classes of customers are present, i.e., $K = 2$. The LSTs of the interdeparture times for all classes of customers of interest are obtained. Algorithms are developed for computing the means and variances of interdeparture times.

3. The $MMAP[2]/G[2]/1$ queue without priority

This section considers the case where there are two classes of customers and all customers are being served on a "first-come-first-served" basis. Since there are equal priorities, the two classes of customers are equivalent. Thus, we only need to consider the interdeparture times of class 1 customers. To find the LST of the interdeparture time between two consecutive class 1 customers, the following measures and notations are introduced and discussed first.

Counting process of class 2 *customers*. Using the definition given in Section 2, matrix $P(0, n, t)$ gives the probability that no class 1 customer and *n* class 2 customers arrived in $(0, t)$. Similarly, $P(0, n, t)D_1\mathrm{d}t$ is the (matrix) probability that no class 1 customer and *n* class 2 customers arrived in $(0, t)$ and a class 1 customer arrives in $(t, t + \mathrm{d}t)$. Let

$$A^*(0, n, s) = \int_0^\infty \exp\{-sx\}F_2(\mathrm{d}x)P(0, n, x), \quad n \geq 0. \tag{3.1}$$

The waiting time of an arbitrary class 1 *customer*. Let $W_1(j, x)$ be the distribution of the waiting time of an arbitrary class 1 customer, given that the underlying Markov chain $J(t)$ is in state *j* when the customer arrives, for $1 \leq j \leq m$. Let $\mathbf{W}_1(x) = (W_1(1, x), \ldots , W_1(m, x))$. The LST of $\mathbf{W}_1(x)$ is obtained in [3] as

$$\mathbf{w}_1^*(s) = \frac{s}{\lambda_1}\mathbf{y}_0[s\mathbf{I} + D_0 + D_1 f_1^*(s) + D_2 f_2^*(s)]^{-1}D_1, \tag{3.2}$$

where $\mathbf{y}_0$ is the (vector) probability that the queueing system is empty at an arbitrary time. Matrix $\mathbf{I}$ is the identity matrix. Readers are referred to Section 6 in [3] for more details about Eq. (3.2) and $\mathbf{y}_0$. For the Poisson-arrival case (Example 2.1), Eq. (3.2) reduces to the standard Pollaczek–Khinchine formula for the $M/G/1$ queue for which the service times have an LST $(\lambda_1 f_1^*(s) + \lambda_2 f_2^*(s))/(\lambda_1 + \lambda_2)$.

The busy period involving only class 2 *customers*. Denote by $(G_2^*(s))_{i,j}$ the LST of the length of the busy period in which there are only class 2 customers (no class 1 customers arrive during the busy period) and the underlying Markov chain is in state *j* when the busy period ends, given that the underlying Markov process is in state *i* at the beginning of the busy period. $G_2^*(s)$ is an $m \times m$ matrix which is the minimal nonnegative solution to the matrix equation:

$$G_2^*(s) = \sum_{n=0}^{\infty} A^*(0, n, s)(G_2^*(s))^n = \int_0^\infty \exp\{(-s\mathbf{I} + D_0 + D_2 G_2^*(s))x\}\mathrm{d}F_2(x). \tag{3.3}$$

Eq. (3.3) can be proved by following the approach used in [10] by conditioning that no class 1 customer arrives during such a busy period (abusing notation a little, we say that the input process is a MAP with matrix representation (D_0, D_2)). Eq. (3.3) can also be obtained by setting $z_1 = 0$ and $z_2 = 1$ in Eq. (3.2) in [3]. In the Poisson-arrival case (Example 2.1), $G_2^*(s)$ is identical to the function $\eta_2(s + \lambda_1)$ in [13].

The interdeparture time $V + X_{1,1}$ (see Fig. 1). Consider two consecutive class 1 customers c_0 and c_1. The waiting time and service time of c_0 are $W_{1,0}$ and $X_{1,0}$, respectively. The LST of $W_{1,0}$ is given by Eq. (3.2). Let *V* denote the time between the departure epoch of c_0 and the epoch when c_1 enters service. The service time of c_1 is $X_{1,1}$, which is independent of *V*. The distribution function of $X_{1,0}$ and $X_{1,1}$ is $F_1(x)$ with LST $f_1^*(s)$.

The interdeparture time is $V + X_{1,1}$. To find the LST of $V + X_{1,1}$, there are two cases to be considered (see Fig. 1): (1) c_1 arrives before c_0's departure, and (2) c_1 arrives after c_0's departure. For case (1), we must consider the number of class 2 customers who arrived before c_1, as their service times are components of the interdeparture time. For case (2), we require the number of class 2 customers who arrived during c_0's flow time —$W_{1,0} + X_{1,0}$ —as recursions are developed based on this number that facilitate the analysis. The analysis in this section (and Sections 4 and 5) is based on this decomposition.

Denote by $\Phi_j^*(s)$ the LST of the interdeparture time between c_0 and c_1, given that the underlying Markov process $J(t)$ is in state *j* when customer c_0 arrives. $\mathbf{\Phi}^*(s)$ is a column vector of size *m* with elements $\Phi_j^*(s)$. Conditioning on whether or not c_1 arrives before c_0's departure, the number of class 2

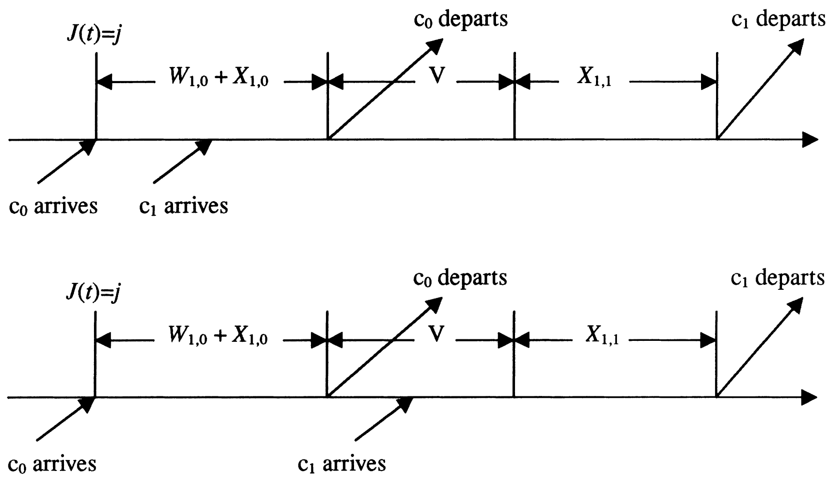


Fig. 1. The departure process.

customers who arrived before c_1, and the number of class 2 customers who arrived before c_0's departure, the following important expression can be derived:

$$\mathbf{\Phi}^*(s) = \int_0^\infty \mathrm{diag}(\mathbf{W}_1 * F_1(\mathrm{d}x)) \left[\int_0^x \sum_{n=0}^\infty P(0,n,u) D_1 \exp\{D(x-u)\}\mathbf{e}\,\mathrm{d}u (f_2^*(s))^n + \sum_{n=0}^\infty P(0,n,x)\mathbf{\Psi}_n^*(s) \right] f_1^*(s), \tag{3.4}$$

where $\mathbf{W}_1 * F_1(x)$ is the convolution of $\mathbf{W}_1(x)$ and $F_1(x)$ and, $\mathbf{\Psi}_n^*(s)$ is the (column vector) LST of V, given the state of the underlying Markov process at the departure epoch of c_0 and n class 2 customers arrived before c_0's departure. The matrix diag ($\mathbf{u}$) has all the elements of the vector $\mathbf{u}$ on its diagonal and other elements zero.

The first part in Eq. (3.4) assumes that c_1 arrives before c_0 departs. Therefore, the interdeparture time is the sum of the service times of class 2 customers who arrived before c_1 plus the service time of c_1 itself. The LST of the conditional interdeparture time in this case is $(f_2^*(s))^n f_1^*(s)$ if n class 2 customers arrived before c_1. The second part in Eq. (3.4) assumes that c_1 arrives after c_0 departs. The period of V starts with n class 2 customers. One needs to find $\{\mathbf{\Psi}_n^*(s), n \geq 0\}$ for the conditional distribution of the interdeparture time.

Similar to Eq. (3.4), the following equations can be established for vectors $\{\mathbf{\Psi}_n^*(s), n \geq 0\}$: (Notice that $\exp\{Dt\}\mathbf{e} = \mathbf{e}$, for $t \geq 0$, since $D\mathbf{e} = 0$.)

$$\mathbf{\Psi}_0^*(s) = (s\mathbf{I} - D_0)^{-1} D_1 \mathbf{e} + (s\mathbf{I} - D_0)^{-1} D_2 \mathbf{\Psi}_1^*(s);$$
$$\mathbf{\Psi}_n^*(s) = \int_0^\infty \exp\{-sx\} F_2(\mathrm{d}x) \left[\int_0^x \sum_{k=0}^\infty P(0,k,u) D_1 \mathbf{e}\,\mathrm{d}u (f_2^*(s))^{n-1+k} + \sum_{k=0}^\infty P(0,k,x)\mathbf{\Psi}_{n-1+k}^*(s) \right], \quad n \geq 1. \tag{3.5}$$

Essentially, Eq. (3.5) is derived by conditioning on what occurred during the service time of the first class 2 customer after c_0's departure (when $n \geq 1$). Now, we are ready to state and prove the main result of this section.

Theorem 3.1. *For an $MMAP[2]/G[2]/1$ queue with equal priority, the LST of the interdeparture time between two consecutive class 1 customers is given by*

$$\begin{aligned}\mathbf{\Phi}^*(s) = &\left[s\int_0^\infty \mathrm{diag}(\mathbf{W}_1 * F_1(\mathrm{d}x))\exp\{(D_0 + D_2G_2^*(s))x\}(s\mathbf{I} - D_0 - D_2G_2^*(s))^{-1} - \mathbf{I}\right] \\ &\times (D_0 + D_2 f_2^*(s))^{-1} D_1 \mathbf{e} f_1^*(s).\end{aligned} \tag{3.6}$$

The LST of the interdeparture time in steady state is obtained by $\boldsymbol{\theta} D_1 \mathbf{\Phi}^*(s)/\lambda_1$, since the stationary distribution of the underlying Markov process $J(t)$ at the arrival epochs of class 1 customers is given by $\boldsymbol{\theta} D_1/\lambda_1$.

Proof. Equalities in Eq. (3.5) are used to find $\{\mathbf{\Psi}_n^*(s), n \geq 0\}$ first. Solutions are then substituted into Eq. (3.4) to obtain Eq. (3.6). Equalities in Eq. (3.5) can be rewritten as follows, for $n \geq 1$,

$$\begin{aligned}\mathbf{\Psi}_n^*(s) &= \int_0^\infty \exp\{-sx\}F_2(\mathrm{d}x)\left[\int_0^x \exp\{(D_0 + D_2 f_2^*(s))u\}D_1\mathbf{e}\mathrm{d}u(f_2^*(s))^{n-1}\right] \\ &\quad + \sum_{k=0}^\infty \int_0^\infty \exp\{-sx\}F_2(\mathrm{d}x)P(0,k,x)\mathbf{\Psi}_{n-1+k}^*(s) \\ &= \int_0^\infty F_2(\mathrm{d}x)[\exp\{(-s\mathbf{I} + D_0 + D_2 f_2^*(s))x\} - \mathrm{e}^{-sx}\mathbf{I}](D_0 + D_2 f_2^*(s))^{-1} \\ &\quad \times D_1\mathbf{e}[f_2^*(s)]^{n-1} + \sum_{k=1}^\infty \int_0^\infty \exp\{-sx\}F_2(\mathrm{d}x)P(0,k,x)\mathbf{\Psi}_{n-1+k}^*(s) \\ &\equiv \mathbf{\Theta}_0(s)[f_2^*(s)]^{n-1} + \sum_{k=0}^\infty A^*(0,k,s)\mathbf{\Psi}_{n-1+k}^*(s).\end{aligned} \tag{3.7}$$

In the derivation of Eq. (3.7), equality $\sum_{k=0}^\infty P(0,k,u)(f_2^*(S))^k = \exp\{(D_0 + D_2 f_2^*(s))u\}$ is used. Function $\mathbf{\Theta}_0(s)$ is a column vector. At this point, we hypothesize a form to the solution of Eq. (3.7). Regardless of later events, it initially entails $(f_2^*(s))^n$, the service times of the n class 2 customers present at c_0's departure epoch. Depending on later events, it may entail the busy period distribution initiated by these n customers. Finally, recalling the form of similar quantities in [13], we propose the solution form:

$$\mathbf{\Psi}_n^*(s) = (f_2^*(s))^n \mathbf{\Lambda}_0(s) + (G_2^*(s))^n \mathbf{\Omega}_0(s), \quad n \geq 0, \tag{3.8}$$

where column vectors $\mathbf{\Lambda}_0(s)$ and $\mathbf{\Omega}_0(s)$ need to be determined. First, substituting Eq. (3.8) into Eq. (3.7), it follows that $\mathbf{\Lambda}_0(s)$ satisfies the following equation:

$$\mathbf{\Lambda}_0(s) f_2^*(s) = \mathbf{\Theta}_0(s) + \sum_{k=0}^\infty A^*(0,k,s)(f_2^*(s))^k \mathbf{\Lambda}_0(s). \tag{3.9}$$

Solving Eq. (3.9) yields (after we have again made use of Eq. (2.4))

$$\mathbf{\Lambda}_0(s) = \left[f_2^*(s)\mathbf{I} - \sum_{k=0}^{\infty} A^*(0,k,s)(f_2^*(s))^k \right]^{-1} \mathbf{\Theta}_0(s) = -[D_0 + D_2 f_2^*(s)]^{-1} D_1\mathbf{e}. \tag{3.10}$$

For vector $\mathbf{\Omega}_0(s)$, the boundary conditions are considered, i.e.,

$$\mathbf{\Psi}_0^*(s) = \mathbf{\Lambda}_0(s) + \mathbf{\Omega}_0(s); \quad \mathbf{\Psi}_1^*(s) = f_2^*(s)\mathbf{\Lambda}_0(s) + G_2^*(s)\mathbf{\Omega}_0(s). \tag{3.11}$$

Using the first equality in Eq. (3.5), (3.10) and (3.11), we obtain

$$\begin{aligned}\mathbf{\Omega}_0(s) &= [s\mathbf{I} - D_0 - D_2 G_2^*(s)]^{-1}[D_1\mathbf{e} + [D_2 f_1^*(s) + D_0 - s\mathbf{I}]\mathbf{\Lambda}_0(s)] \\ &= -s[-s\mathbf{I} + D_0 + D_2 G_2^*(s)]^{-1}(D_0 + D_2 f_2^*(s))^{-1} D_1\mathbf{e}. \end{aligned} \tag{3.12}$$

Substituting Eqs. (3.8), (3.10) and (3.12) into Eq. (3.4), one obtains Eq. (3.6). The last step consists of several substitutions using Eq. (2.4) and the fact that $\exp\{D(x-u)\}\mathbf{e} = \mathbf{e}$. Notice that the invertibility of the matrices involved can be verified routinely.

Finally, we prove that the solution to Eq. (3.5) is unique. In fact, Eq. (3.5) is a linear system which can be rewritten as $\mathbf{\Psi}^*(s) = \mathbf{\Pi}^*(s) + \mathbf{\Xi}^*(s)\mathbf{\Psi}^*(s)$, where $\mathbf{\Psi}^*(s)$ is a column vector obtained by putting vectors $\{\mathbf{\Psi}_n^*(s), n \geq 0\}$ into one column lexicographically, $\mathbf{\Pi}^*(s)$ is a nonnegative vector, and $\mathbf{\Xi}^*(s)$ is a nonnegative matrix:

$$\mathbf{\Xi}^*(s) = \begin{pmatrix} 0 & (s\mathbf{I} - D_0)^{-1} D_2 & 0 & \cdots \\ A^*(0,0,s) & A^*(0,1,s) & A^*(0,2,s) & \cdots \\ & A^*(0,0,s) & A^*(0,1,s) & \cdots \\ & & \ddots & \ddots \end{pmatrix}. \tag{3.13}$$

It is easy to see that $\mathbf{\Xi}^*(s)\mathbf{e} < \mathbf{\Pi}^*(s) + \mathbf{\Xi}^*(s)\mathbf{e} < \mathbf{e}$. Thus, $(\mathbf{\Xi}^*(s))^n \to 0$ as $n \to \infty$. Therefore, $\mathbf{\Psi}^*(s) = \mathbf{\Pi}^*(s)[\mathbf{I} - \mathbf{\Xi}^*(s)]^{-1}$. This implies that the solution to Eq. (3.5) is unique. Hence, Eq. (3.8) gives the unique solution to Eq. (3.5). This completes the proof. □

Remark 3.1. Expression (3.8) can be interpreted as follows. The first part of Eq. (3.8) is the LST of V when c_1 arrives during the busy period with initially n class 2 customers. Further decompose this part into two subparts:
- Function $(f_2^*(s))^n$ is the LST of the sum of the service times of the n class 2 customers initially in the system, and
- Vector $-(D_0 + D_2 f_2^*(s))^{-1} D_1\mathbf{e}$ is the LST of the sum of the service times of the class 2 customers who arrived before c_1.

The second part of Eq. (3.8) gives the LST of V when c_1 does not arrive during the busy period with n class 2 customers initially.
- $(G_2^*(s))^n$ is the LST of the length of the busy period with n class 2 customers initially and no class 1 customer arrives during this period of time.
- $(-s\mathbf{I} + D_0 + D_2 G_2^*(s))^{-1}$ is the LST of the length of the idle periods and busy periods without class 1 arrivals following the initial busy period with n class 2 customers and no class 1 arrival.
- Finally, c_1 arrives during a busy period in which the total service time of all the class 2 customers who arrived before c_1 has LST: $-(D_0 + D_2 f_2^*(s))^{-1} D_1\mathbf{e}$.

Remark 3.2. In the case of Poisson-arrival (Example 2.1), Eq. (3.6) reduces to

$$\Phi^*(s) = \left[1 - \frac{sw_1^*(\lambda - \lambda_2 G_2^*(s)) f_1^*(\lambda - \lambda_2 G_2^*(s))}{s + \lambda - \lambda_2 G_2^*(s)}\right] \frac{\lambda_1 f_1^*(s)}{\lambda - \lambda_2 f_2^*(s)}, \tag{3.14}$$

where $\lambda = \lambda_1 + \lambda_2$. Eq. (3.14) is consistent with Eq. (13) in [13].

Remark 3.3. In this paper, the analysis is based on distinguishing system status at the departure epoch of customer c_0. In [13], the analysis is based on distinguishing system status at the arrival epoch of customer c_1. Both approaches produce the same results. In fact, expression (3.6) can be obtained by using the previous approach, but the proof is much longer. The approach used in this paper is more suitable for the matrix environment.

The LST of the interdeparture time of queueing systems with only one class of customer can be obtained from Theorem 3.1 by setting $D_2 = 0$, i.e., no class 2 arrival at all. Results are given in the following corollary.

Corollary 3.2. *For a $MAP/G/1$ queue with a Markov arrival process (D_0, D_1) and a "first-come-first-served" service discipline, the LST of the interdeparture time between two consecutive customers is given by*

$$\mathbf{\Phi}^*(s) = \left[\mathbf{I} - s\int_0^\infty \mathrm{diag}(\mathbf{W}_1 * F_1(\mathrm{d}x))\exp\{D_0 x\}(s\mathbf{I} - D_0)^{-1}\right]\mathbf{e} f_1^*(s). \tag{3.15}$$

The LST of the interdeparture time in steady state is obtained by $\boldsymbol{\theta} D_1 \mathbf{\Phi}^*(s)/\lambda_1$. For the $M/M/1$ queue, Eq. (3.15) *reduces to the LST of an exponential distribution with parameter λ_1—a well-known result for the interdeparture time of this queueing system.*

4. The $MMAP[2]/G[2]/1$ queue with nonpreemptive priority: the higher priority class

In this case, class 1 customers have higher priorities over class 2 customers. Thus, when the server becomes available, a class 1 customer, if any, will enter the server. If there is no class 1 customer in the system, then the first class 2 customer to arrive enters the server. Since class 1 customers have higher priority, the queueing processes of the two classes of customers are not equivalent. Therefore, their interdeparture times shall be treated separately.

In general, the approach introduced in Section 3 is followed in this section and the next. Also notations used in Section 3 are used in this section and the next, possibly having a different value or expression.

It is important to see that the waiting time distributions of the two classes of customers in the non-preemptive priority queue are different from that of the equal priority queue. Thus, expression (3.2) is no longer useful. The distributions of the waiting times in the nonpreemptive queueing systems can be found in [15]. Readers are referred to Takine [15] for more details about the waiting time process in the queueing system with two classs of customers and nonpreemptive priority. In this section and Section 5, the actual waiting time distributions of the two classes of customers are assumed to be known without being written down explicitly.

This section considers interdeparture times of class 1 customers (the higher priority class). Since the priority class is nonpreemptive, the presence of class 2 customers has influence on the queueing process of class 1 customers. Thus, the interdeparture times of class 1 customers are different from that in Corollary 3.2 for the $MAP/G/1$ queue (see [5] for more details about the $MAP/G/1$ queue).

Consider the interdeparture time between two consecutive class 1 customers c_0 and c_1. Conditioning on whether or not c_1 arrives before c_0 departs and the number of class 2 customers who arrived before c_0's departure, the following equation can be obtained for the LST of the interdeparture time between c_0 and c_1:

$$\mathbf{\Phi}^*(s) = \int_0^\infty \mathrm{diag}(\mathbf{W}_1 * F_1(\mathrm{d}x)) \left[\int_0^x \exp\{(D_0+D_2)u\} D_1 \mathbf{e}\,\mathrm{d}u + \sum_{n=0}^\infty P(0,n,x) \mathbf{\Psi}_n^*(s) \right] f_1^*(s). \tag{4.1}$$

The column vector $\mathbf{\Psi}_n^*(s)$ is, again, the LST of V (the time between the departure epoch of c_0 and the epoch when c_1 enters service), given the state of the underlying Markov process at the departure epoch of c_0 and n class 2 customers arrived before c_0's departure. Expression (4.1) is different from expression (3.4) since, when c_1 arrives before c_0 departs, c_1 enters the server after c_0's departure regardless of the presence of class 2 customers in the priority case.

Similar to Eq. (3.5), the following equations can be established for vectors $\{\mathbf{\Psi}_n^*(s), n \geq 0\}$:

$$\begin{aligned}
\mathbf{\Psi}_0^*(s) &= (s\mathbf{I} - D_0)^{-1} D_1 \mathbf{e} + (s\mathbf{I} - D_0)^{-1} D_2 \mathbf{\Psi}_1^*(s); \\
\mathbf{\Psi}_n^*(s) &= \int_0^\infty \mathrm{e}^{-sx} F_2(\mathrm{d}x) \left[\int_0^x \exp\{(D_0 + D_2)u\} D_1 \mathbf{e}\,\mathrm{d}u + \sum_{k=0}^\infty P(0,k,x) \mathbf{\Psi}_{n-1+k}^*(s) \right] \\
&\equiv \mathbf{\Theta}_0(s) + \sum_{k=0}^\infty A^*(0,k,s) \mathbf{\Psi}_{n-1+k}^*(s), \quad n \geq 1.
\end{aligned} \tag{4.2}$$

Theorem 4.1. *For an $MMAP[2]/G[2]/1$ queue with nonpreemptive priority, the LST of the interdeparture time between two consecutive class 1 customers (the higher priority class) is given by*

$$\begin{aligned}
\mathbf{\Phi}^*(s) = \Bigg\{ & \mathbf{I} + s \int_0^\infty \mathrm{diag}(\mathbf{W}_1 * F_1(\mathrm{d}x)) \exp\{(D_0 + D_2 G_2^*(s))x\} (-s\mathbf{I} + D_0 + D_2 G_2^*(s))^{-1} \\
& - \int_0^\infty \mathrm{diag}(\mathbf{W}_1 * F_1(\mathrm{d}x)) [\exp\{(D_0 + D_2)x\} - \exp\{(D_0 + D_2 G_2^*(s))x\} \\
& \qquad \times (-s\mathbf{I} + D_0 + D_2 G_2^*(s))^{-1} (-s\mathbf{I} + D_0 + D_2)] \\
& \times (1 - f_2^*(s)) \left[\mathbf{I} - \int_0^\infty F_2(\mathrm{d}x) \exp\{(-s\mathbf{I} + D_0 + D_2)x\} \right]^{-1} \Bigg\} \mathbf{e} f_1^*(s).
\end{aligned} \tag{4.3}$$

Similar to Theorem 3.1, the distribution of the interdeparture time in steady state is obtained by $\boldsymbol{\theta} D_1 \mathbf{\Phi}^*(s)/\lambda_1$. Notice that vector $\mathbf{W}_1(x)$ is the distribution function of the waiting time of an arbitrary class 1 customer. Details about $\mathbf{W}_1(x)$ can be found in [15].

Proof. Similar to Theorem 3.1, the key is to solve Eq. (4.2). Suppose that the solution to Eq. (4.2) has the following structure:

$$\mathbf{\Psi}_n^*(s) = \mathbf{\Lambda}_0(s) + (G_2^*(s))^n \mathbf{\Omega}_0(s), \quad n \geq 0. \tag{4.4}$$

First, $\boldsymbol{\Lambda}_0(s)$ satisfies the following equation:

$$\boldsymbol{\Lambda}_0(s) = \boldsymbol{\Theta}_0(s) + \sum_{k=0}^{\infty} A^*(0, k, s)\boldsymbol{\Lambda}_0(s). \tag{4.5}$$

By definition Eq. (3.1), we have $\sum_{k=0}^{\infty} A^*(0, k, s) = \int_0^\infty F_2(\mathrm{d}x)\exp\{(-s\mathbf{I}+D_0+D_2)x\}$. Also by definition (Eq. (4.2)), we have $\boldsymbol{\Theta}_0(s) = [f_2^*(s)\mathbf{I} - \int_0^\infty F_2(\mathrm{d}x)\exp\{-s\mathbf{I} + D_0 + D_2)x\}]\mathbf{e}$. Solving Eq. (4.5) yields,

$$\boldsymbol{\Lambda}_0(s) = \left\{\mathbf{I} - (1 - f_2^*(s))\left[\mathbf{I} - \int_0^\infty F_2(\mathrm{d}x)\exp\{(-s\mathbf{I} + D_0 + D_2)x\}\right]^{-1}\right\}\mathbf{e}. \tag{4.6}$$

For vector $\boldsymbol{\Omega}_0(s)$, the boundary conditions are considered, i.e.,

$$\boldsymbol{\Psi}_0^*(s) = \boldsymbol{\Lambda}_0(s) + \boldsymbol{\Omega}_0(s), \qquad \boldsymbol{\Psi}_1^*(s) = \boldsymbol{\Lambda}_0(s) + G_2^*(s)\boldsymbol{\Omega}_0(s). \tag{4.7}$$

Using Eqs. (4.2), (4.6) and (4.7), yields

$$\begin{aligned}\boldsymbol{\Omega}_0(s) = {} & [-s\mathbf{I} + D_0 + D_2G_2^*(s)]^{-1} \\ & \left\{s\mathbf{I} + (-s\mathbf{I} + D_0 + D_2)(1 - f_2^*(s)) \times \left[\mathbf{I} - \int_0^\infty F_2(\mathrm{d}x)\exp\{(-s\mathbf{I} + D_0 + D_2)x\}\right]^{-1}\right\}\mathbf{e}.\end{aligned} \tag{4.8}$$

Finally, substituting Eqs. (4.2),(4.6) and (4.8) into Eq. (4.1) and after some simplifications, expression (4.3) is obtained. The same argument used in Theorem 3.1 can be applied to show that the solution to Eq. (4.2) is unique. This completes the proof. □

Remark 4.1. Expression (4.4) can be interpreted as follows. The first part of Eq. (4.4) is the LST of the service times of all class 2 customers served before or at the arrival of c_1. The second part of Eq. (4.4) consists of three parts: the busy period (of class 2 customers) with n class 2 customers initially; the idle and busy period before c_1 arrives; plus the total service time of class 2 customers served before and at the arrival epoch of c_1.

Remark 4.2. For the Poisson-arrival case (Example 2.1), Eq. (4.3) reduces to

$$\begin{aligned}\boldsymbol{\Phi}^*(s) = f_1^*(s)\Bigg\{ & 1 - \frac{sw_1^*(\lambda - \lambda_2G_2^*(s))f_1^*(\lambda - \lambda_2G_2^*(s))}{s + \lambda - \lambda_2G_2^*(s)} \\ & - \left[w_1^*(\lambda_1)f_1^*(\lambda_1) - \frac{w_1^*(\lambda - \lambda_2G_2^*(s))f_1^*(\lambda - \lambda_2G_2^*(s))(s + \lambda_1)}{s + \lambda - \lambda_2G_2^*(s)}\right]\frac{(1 - f_2^*(s))}{1 - f_2^*(s + \lambda_1)}\Bigg\}.\end{aligned} \tag{4.9}$$

This expression was obtained in [12].

5. The $MMAP[2]/G[2]/1$ queue with nonpreemptive priority: the lower priority class

Section 4 finds the LST of the interdeparture times of class 1 customers (the higher priority class). This section considers interdeparture times of class 2 customers (the lower priority class). The interdeparture times of class 2 customers are apparently influenced by the presence of class 1 customers. Two consecutive class 2 customers, c_0 and c_1, are considered. The waiting time of c_0 is $W_{2,0}$ and the service time of c_0 is $X_{2,0}$. The service time of c_1 is $X_{2,1}$. The distribution function of $X_{2,0}$ and $X_{2,1}$ is $F_2(x)$ with LST $f_2^*(s)$. The (vector) distribution function of the waiting time of c_0 is denoted by $\mathbf{W}_2(x)$.

The interdeparture time between c_0 and c_1 has to consider all the class 1 customers who arrived before c_0's departure and all class 1 customers who arrived during the subsequent busy period. This makes a difference in the analysis. Another useful fact is that when c_0 enters the server, there is no class 1 customer in the system since class 1 customers have higher priorities. Thus, the waiting time period $W_{2,0}$ and its corresponding service period $X_{2,0}$ are considered separately.

Conditioning on whether or not c_1 arrives during c_0's waiting period, the number of class 1 customers who arrived before c_0's departure and c_1's arrival, and the number of class 1 customers who arrived before c_0's departure and after c_1's arrival, the following equation is obtained:

$$\begin{aligned}\mathbf{\Phi}^*(s) = {} & \int_0^\infty \mathrm{diag}(\mathbf{W}_2(\mathrm{d}x))[\mathrm{e}^{\{Dx\}} - \mathrm{e}^{\{(D_0+D_1)x\}}] \int_0^\infty F_2(\mathrm{d}u) \sum_{n,k\geq 0} P(n,k,u)[G_1^*(s)]^n \mathbf{e} f_2^*(s) \\ & + \int_0^\infty \mathrm{diag}(\mathbf{W}_2(\mathrm{d}x))\mathrm{e}^{\{(D_0+D_1)x\}} \left[\int_0^\infty F_2(\mathrm{d}u) \sum_{n\geq 0, k\geq 1} P(n,k,u)[G_1^*(s)]^n \mathbf{e} \right. \\ & \left. + \int_0^\infty F_2(\mathrm{d}u) \sum_{n=0}^\infty P(n,0,u) \mathbf{\Psi}_n^*(s) \right] f_2^*(s), \end{aligned} \tag{5.1}$$

in which

- the vector $\mathbf{\Psi}_n^*(s)$ is the LST of V, given the state of the underlying Markov process at the departure epoch of c_0, n class 1 customers arrive during c_0's service period, and no class 2 customer arrives during c_0's service period;
- $\exp\{Dx\} - \exp\{(D_0 + D_1)x\}$ is the (matrix) probability that at least one class 2 customer arrives in $(0, x)$;
- $\exp\{(D_0 + D_1)x\}$ is the (matrix) probability that no class 2 customer arrives in $(0, x)$;
- the matrix $G_1^*(s)$ is the LST of the busy period of class 1 customers (the highest priority class) in which class 2 customers may or may not arrive. The matrix $G_1^*(s)$ is the minimal nonnegative solution to the matrix equation

$$\begin{aligned} G_1^*(s) &= \int_0^\infty F_1(\mathrm{d}x)\exp\{(-s\mathbf{I} + D_0 + D_2 + D_1 G_1^*(s))x\} \\ &= \sum_{n=0}^\infty \left[\int_0^\infty \mathrm{e}^{-sx} F_1(\mathrm{d}x) \sum_{k=0}^\infty P(n,k,x) \right] [G_1^*(s)]^n. \end{aligned} \tag{5.2}$$

- Function $\int_0^\infty F_2(\mathrm{d}x)\exp\{(D_0 + D_2 + D_1 G_1^*(s))x\}$ is the LST of V considering class 1 customers who arrived during the service time of c_0 and their corresponding busy period, given that c_1 already arrived during the waiting period of c_0.
- $\mathbf{W}_2(x)$ is the waiting time distribution of an arbitrary class 2 customer, given the state of the underlying Markov chain at the arrival epoch. See [15] for details about $\mathbf{W}_2(x)$.
- The expression of $\mathbf{\Phi}^*(s)$ can be simplified using the following equality:

$$\sum_{k=0}^\infty \sum_{l=0}^\infty P(k,l,x)(G_1^*(s))^k = \exp\{(D_0 + D_2 + D_1 G_1^*(s))x\}. \tag{5.3}$$

The following equations can be established for vectors $\{\mathbf{\Psi}_n^*(s), n \geq 0\}$:

$$\mathbf{\Psi}_0^*(s) = (s\mathbf{I} - D_0)^{-1} D_2 \mathbf{e} + (s\mathbf{I} - D_0)^{-1} D_1 \mathbf{\Psi}_1^*(s);$$

$$\mathbf{\Psi}_n^*(s) = \int_0^\infty \mathrm{e}^{-sx} F_1(\mathrm{d}x) \left[\sum_{k=0}^\infty \sum_{l=1}^\infty P(k,l,x)(G_1^*(s))^{n-1+k}\mathbf{e} + \sum_{k=0}^\infty P(k,0,x)\mathbf{\Psi}_{n-1+k}^*(s) \right]$$
$$\equiv \mathbf{\Theta}_0(s)(G_1^*(s))^{n-1}\mathbf{e} + \sum_{k=0}^\infty A^*(k,0,s)\mathbf{\Psi}_{n-1+k}^*(s), \quad n \geq 0. \tag{5.4}$$

Notice that $A^*(n,0,s)$ is defined similar to $A^*(0,n,s)$ but with function $F_1(x)$ and $P(n,0,x)$.

Theorem 5.1. *For an $MMAP[2]/G[2]/1$ queue with nonpreemptive priority, the LST of the interdeparture time between two consecutive class 2 customers (the lower priority class) is given by*

$$\begin{aligned}\mathbf{\Phi}^*(s) = {} & P_{1,w} \int_0^\infty F_2(\mathrm{d}x) \exp\{(D_0 + D_2 + D_1 G_1^*(s))x\} \mathbf{e} f_2^*(s) \\ & + P_{0,w} \left\{ \int_0^\infty F_2(\mathrm{d}x) \exp\{(D_0 + D_2 + D_1 G_1^*(s))x\} \right. \\ & - \int_0^\infty F_2(\mathrm{d}x) \exp\{(D_0 + D_1 \hat{G}_1^*(s))x\}[-s\mathbf{I} + D_0 + D_1 \hat{G}_1^*(s)]^{-1} \\ & \left. \times [-s\mathbf{I} - D_1 + D_1 G_1^*(s)] \right\} \mathbf{e} f_2^*(s).\end{aligned} \tag{5.5}$$

where $P_{1,w}$ and $P_{0,w}$ are the (matrix) probabilities that c_1 arrives before c_0 enters service or after c_0 enters service, respectively. Matrices $P_{1,w}$ and $P_{0,w}$ are given by

$$P_{1,w} = \int_0^\infty \mathrm{diag}(\mathbf{W}_2(\mathrm{d}x))[\mathrm{e}^{\{Dx\}} - \mathrm{e}^{\{(D_0+D_1)x\}}], \qquad P_{0,w} = \int_0^\infty \mathrm{diag}(\mathbf{W}_2(\mathrm{d}x))\mathrm{e}^{\{(D_0+D_1)x\}}. \tag{5.6}$$

Matrix $\hat{G}_1^*(s)$ is the LST of the busy period of class 1 customers in which no class 2 customer is present. Matrix $\hat{G}_1^*(s)$ is the minimal nonnegative solution to the matrix equation

$$\hat{G}_1^*(s) = \sum_{n=0}^\infty A^*(n,0,s)(\hat{G}_1^*(s))^n = \int_0^\infty F_1(\mathrm{d}x) \exp\{(-s\mathbf{I} + D_0 + D_1 \hat{G}_1^*(s))x\}. \tag{5.7}$$

Similar to Theorem 3.1, the distribution of the interdeparture time in steady state is obtained by $\boldsymbol{\theta} D_2 \mathbf{\Phi}^* s)/\lambda_2$.

Proof. Similar to Theorems 3.1 and 4.1, the key is to solve Eq. (5.4). Suppose that the solution to Eq. (5.4) has the following structure:

$$\mathbf{\Psi}_n^*(s) = \mathbf{\Lambda}_0(s)(G_1^*(s))^n \mathbf{e} + (\hat{G}_1^*(s))^n \mathbf{\Omega}_0(s), \quad n \geq 0. \tag{5.8}$$

Notice that $\mathbf{\Lambda}_0(s)$ is now a matrix. Since $\sum_{k=1}^\infty P(n,k,x) = \sum_{k=0}^\infty P(n,k,x) - P(n,0,x)$, it can be verified that (see Eq. (5.2))

$$\mathbf{\Theta}_0(s) = G_1^*(s) - \sum_{n=0}^\infty A^*(n,0,s)(G_1^*(s))^n. \tag{5.9}$$

Then it is easy to prove that $\mathbf{\Lambda}_0(s) = \mathbf{I}$ and

$$\mathbf{\Omega}_0(s) = -[-s\mathbf{I} + D_0 + D_1\hat{G}_1^*(s)]^{-1}(-s\mathbf{I} - D_1 + D_1G_1^*(s))\mathbf{e}. \tag{5.10}$$

Finally, substituting Eqs. (5.8) and (5.10), and $\mathbf{\Lambda}_0(s) = \mathbf{I}$ into Eq. (5.4), after some simplifications, expression (5.5) is obtained. The same argument used in Theorem 3.1 can be applied to show that the solution to Eq. (5.4) is unique. This completes the proof. □

Remark 5.1. A probabilistic interpretation of expression (5.8) can be given similar to Remark 3.1 for expression (3.8). The first part of expression (5.8) represents the busy period of class 1 customers with initially n class 1 customers. During this busy period, c_1 (a lower priority customer) may have arrived. The second part of expression (5.8) consists of two parts: a busy period without class 2 customers with initially n class 1 customers, plus the time until the end of a busy period of class 1 customers where c_1 has arrived.

Remark 5.2. For the Poisson-arrival case (Example 2.1), Eq. (5.5) reduces to

$$\begin{aligned}\mathbf{\Phi}^*(s) = {} & P_{1,w} f_2^*(\lambda_1 - \lambda_1 G_1^*(s)) f_2^*(s) \\ & + P_{0,w}\left\{ f_2^*(\lambda_1 - \lambda_1 G_1^*(s)) - \frac{f_2^*(\lambda - \lambda_1 \hat{G}_1^*(s))}{s + \lambda - \lambda_1 \hat{G}_1^*(s)}[s + \lambda_1 - \lambda_1 G_1^*(s)] \right\} f_2^*(s).\end{aligned} \tag{5.11}$$

with $P_{1,w} = 1 - w_2^*(\lambda_2)$ and $P_{0,w} = w_2^*(\lambda_2)$. Eq. (5.11) was obtained in [12].

6. Moments of interdeparture times

The LSTs obtained in Sections 3–5 can be used for computing the mean and variance of the interdeparture time. Focusing on the equal priority case (Section 3), this section shows how the computation can be done. First, expressions of the first and second moments of the interdeparture time are derived. Second, a numerical example is presented.

First moment. Differentiating both sides of Eq. (3.6) with respect to s and setting $s = 0$, it yields (notice that $f_2^*(0) = 1$, $G_2 = G_2^*(0)$, and $-(D_0 + D_2)^{-1}D_1\mathbf{e} = \mathbf{e}$)

$$\mathbf{\Phi}^{(1)} = \left.\frac{\mathrm{d}\mathbf{\Phi}^*(s)}{\mathrm{d}s}\right|_{s=0} = \hat{P}_0(D_0 + D_2G_2)^{-1}\mathbf{e} + (D_0 + D_2)^{-1}D_2\mathbf{e}\frac{1}{\mu_2} - \frac{1}{\mu_1}\mathbf{e}, \tag{6.1}$$

where $\hat{P}_0 = \int_0^\infty \mathrm{diag}(\mathbf{W}_1 * F_1(\mathrm{d}x))\exp\{(D_0 + D_2G_2)x\}$. The matrix G_2 is the minimal nonnegative solution to Eq. (3.3) with $s = 0$. An efficient algorithm for computing the matrix G_2 is given in [5]. The matrix $\hat{P}_0$ can be rewritten as

$$\begin{aligned}\hat{P}_0 &= \int_0^\infty \mathrm{diag}(\mathbf{W}_1 * F_1(\mathrm{d}x))\mathrm{e}^{-\gamma x}\exp\left\{\left[\mathbf{I} + (D_0 + D_2G_2)\frac{1}{\gamma}\right]\gamma x\right\}\\ &= \sum_{n=0}^{\infty}\int_0^\infty \mathrm{diag}(\mathbf{W}_1 * F_1(\mathrm{d}x))\frac{\mathrm{e}^{-\gamma x}(\gamma x)^n}{n!}\left[\mathbf{I} + (D_0 + D_2G_2)\frac{1}{\gamma}\right]^n\\ &= \sum_{n=0}^{\infty}\frac{(-\gamma)^n}{n!}\frac{\mathrm{d}^n}{\mathrm{d}s^n}[\mathrm{diag}(\mathbf{w}_1^*(s)f_1^*(s))]\bigg|_{s=\gamma}\left[\mathbf{I} + (D_0 + D_2G_2)\frac{1}{\gamma}\right]^n \end{aligned} \tag{6.2}$$

where γ can be chosen as $\gamma = \max_{\{1\le i\le m\}}\{|(D_0+D_2G_2)_{i,i}|\}+1$. When the derivatives of $f_1^*(s)$ and $\mathbf{w}_1^*(s)$ can be obtained, calculation of $\hat{P}_0$ can be carried out. Then we can calculate vector $\mathbf{\Phi}^{(1)}$ using Eq. (6.1). By Theorem 3.1, the mean interdeparture time of class 1 customer can be obtained as $-\boldsymbol{\theta}D_1\mathbf{\Phi}^{(1)}/\lambda_1$. Since the mean interarrival time must equal the mean interdeparture time in a stable queue, $-\boldsymbol{\theta}D_1\mathbf{\Phi}^{(1)}/\lambda_1 = 1/\lambda_1$.

Second moment. Differentiating both sides of Eq. (3.6) twice with respect to s and setting $s = 0$, an expression of the second moment of the interdeparture time is obtained as

$$\begin{aligned}\mathbf{\Phi}^{(2)} &= 2\sum_{n=1}^{\infty}\frac{\int_0^\infty x^n\mathrm{diag}(\mathbf{w}_1 * F_1)(\mathrm{d}x)}{n!}\left[\sum_{k=0}^{n-1}(D_0 + D_2G_2)^k D_2G_2^{(1)}(D_0 + D_2G_2)^{n-1-k}\right]\\ &\quad\times(D_0 + D_2G_2)^{-1}\mathbf{e} + 2\hat{P}_0(D_0 + D_2G_2)^{-1}(\mathbf{I} - D_2G_2^{(1)})(D_0 + D_2G_2)^{-1}\mathbf{e}\\ &\quad+2\hat{P}_0(D_0 + D_2G_2)^{-1}(D_0 + D_2)^{-1}D_2\mathbf{e}\frac{1}{\mu_2} - 2\hat{P}_0(D_0 + D_2G_2)^{-1}\mathbf{e}\frac{1}{\mu_1}\\ &\quad+2(D_0 + D_2)^{-1}D_2(D_0 + D_2)^{-1}D_2\mathbf{e}\left(\frac{1}{\mu_2}\right)^2 - (D_0 + D_2)^{-1}D_2\mathbf{e}f_2^{(2)}(0)\\ &\quad-2(D_0 + D_2)^{-1}D_2\mathbf{e}\frac{1}{\mu_1\mu_2} + f_1^{(2)}(0)\mathbf{e}, \end{aligned} \tag{6.3}$$

where $f_k^{(2)}(0) = (\partial^2/\partial s^2)f_k^*(s)|_{s=0}$, $k = 1, 2$, and the integration can be computed in the following way:

$$\int_0^\infty x^n\mathrm{diag}(\mathbf{w}_1 * F_1)(\mathrm{d}x) = (-1)^n\,\frac{\mathrm{d}^n}{\mathrm{d}s^n}[\mathrm{diag}(\mathbf{w}_1(s)f_1(s))]\bigg|_{s=0},\quad n \ge 0. \tag{6.4}$$

By definition, the nonnegative matrix $G_2^*(s)$ is decreasing and differentiable with respect to $s \ge 0$ (see Eq. (3.3)). The matrix $G_2^{(1)}$ can be obtained by solving the following equation:

$$G_2^{(1)} = \frac{\mathrm{d}G_2^*(s)}{\mathrm{d}s}\bigg|_{s=0} = \sum_{n=1}^{\infty}\frac{\int_0^\infty x^nF_2(\mathrm{d}x)}{n!}\left[\sum_{k=0}^{n-1}(D_0 + D_2G_2)^k[-\mathbf{I} + D_2G_2^{(1)}](D_0 + D_2G_2)^{n-1-k}\right]. \tag{6.5}$$

Introduce the direct sum of matrix Y as a vector obtained by putting together all the rows of the matrix $\phi : Y \to \phi(Y)$ (Gantmacher [2]). Then it is easy to obtain the following explicit formula for matrix $G_2^{(1)}$:

$$\phi(G_2^{(1)}) = \left\{ \mathbf{I} - \sum_{n=1}^{\infty} \frac{\int_0^\infty x^n F_2(\mathrm{d}x)}{n!} \left[\sum_{k=0}^{n-1} [(D_0 + D_2 G_2)^k]^{\mathrm{T}} \otimes (D_0 + D_2 G_2)^{n-1-k} \right] \right\}^{-1}$$
$$\times \phi\left(-\sum_{n=1}^{\infty} \frac{\int_0^\infty x^n F_2(\mathrm{d}x)}{(n-1)!} (D_0 + D_2 G_2)^{n-1} \right), \tag{6.6}$$

where "$\otimes$" denotes Kronecker product and superscript "T" means matrix transpose [2]. The second moment of the interdeparture time can be calculated by using expression $\boldsymbol{\theta} D_1 \boldsymbol{\Phi}^{(2)}/\lambda_1$. Consequently, the variance of the interdeparture time can be calculated.

As a byproduct, the first and second moments of the interdeparture time of the $MAP/G/1$ queue with one class of customer ($D_2 = 0$) are given as follows:

$$\boldsymbol{\Phi}^{(1)} = \hat{P}_0 D_0^{-1} \mathbf{e} - \frac{1}{\mu_1} \mathbf{e}, \qquad \boldsymbol{\Phi}^{(2)} = 2 \hat{P}_0 D_0^{-2} \mathbf{e} - \frac{2}{\mu_1} \hat{P}_0 D_0^{-1} \mathbf{e} + \int_0^\infty x^2 F_1(\mathrm{d}x) \mathbf{e}. \tag{6.7}$$

The proof of Eq. (6.7) involves only routine calculations.

Example 6.1. We consider a queueing system with an $MMAP[2]$ input process and exponential service times. System parameters are given as

$$D_0 = \begin{pmatrix} -t_1 & 0 \\ 0 & -t_2 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & t_1 \\ 0 & 0 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0 & 0 \\ t_2 & 0 \end{pmatrix}, \qquad t_1 > 0, t_2 > 0, \tag{6.8}$$

$f_1^*(s) = \mu_1/(s + \mu_1)$, and $f_2^*(s) = \mu_2/(s + \mu_2)$. The two classes of customers arrive in the queueing system cyclically. Because of the special arrival pattern, it is difficult to analyze the departure process using existing methods. The computational method developed in this section, however, can be applied to calculate the moments of the interdeparture time. In what follows, we show how the computations can be done.

1. For this special case, $\boldsymbol{\theta} = (t_2/(t_1+t_2), t_1/(t_1+t_2))$, $\lambda_1 = \lambda_2 = t_1 t_2/(t_1+t_2)$, $\boldsymbol{\theta} D_1/\lambda_1 = (0, 1)$, $\int_0^\infty x^n F_1(\mathrm{d}x) = n!/\mu_1^n$, and $\int_0^\infty x^n F_2(\mathrm{d}x) = n!/\mu_2^n$, for $n \geq 0$.
2. Matrix G_2 can be found explicitly as
$$G_2 = \mu_2 \mathbf{I} + D_0 G_2 + D_2 G_2^2 \Rightarrow G_2 = \begin{pmatrix} \dfrac{\mu_2}{(t_1 + \mu_2)} & 0 \\ \dfrac{t_2 \mu_2^2}{(t_2 + \mu_2)(t_1 + \mu_2)^2} & \dfrac{\mu_2}{(t_2 + \mu_2)} \end{pmatrix}. \tag{6.9}$$
3. Eq. (6.5) for matrix $G_2^{(1)}$ can be simplified to
$$G_2^{(1)} = \left(\mathbf{I} - \frac{D_0 + D_2 G_2}{\mu_2} \right)^{-1} \left(\frac{D_2 G_2^{(1)} - \mathbf{I}}{\mu_2} \right) \left(\mathbf{I} - \frac{D_0 + D_2 G_2}{\mu_2} \right)^{-1}. \tag{6.10}$$
4. Vector $\mathbf{y}_0$ satisfies $\mathbf{y}_0 Q = 0$ and $\mathbf{y}_0 \mathbf{e} = 1 - \mu_1/\lambda_1 - \lambda_2/\mu_2$, where matrix Q is an infinitesimal generator satisfying equation:
$$Q = D_0 + D_1 (\mathbf{I} - Q/\mu_1)^{-1} + D_2 (\mathbf{I} - Q/\mu_2)^{-1}. \tag{6.11}$$
Matrix Q can be calculated iteratively using the above equation [16].

Table 1
Mean and variance of interdeparture time ($\mu_1 = 3$, $\mu_2 = 5$, and $\rho = 0.353$)

(t_1, t_2)	(3, 6/7)	(2, 1)	(4/3, 4/3)	(1, 2)	(6/7, 3)
Mean	1.5	1.5	1.5	1.5	1.5
Variance (interdeparture)	1.499	1.29	1.170	1.27	1.47
Variance (interarrival)	1.37	1.25	1.125	1.25	1.37

Table 2
Mean and variance of interdeparture time ($\mu_1 = 2$, $\mu_2 = 4$, and $\rho = 0.5$)

(t_1, t_2)	(3, 6/7)	(2, 1)	(4/3, 4/3)	(1, 2)	(6/7, 3)
Mean	1.5	1.5	1.5	1.5	1.5
Variance (interdeparture)	1.503	1.326	1.215	1.303	1.465
Variance (interarrival)	1.37	1.25	1.125	1.25	1.37

Table 3
Mean and variance of interdeparture time ($\mu_1 = 2$, $\mu_2 = 5$, and $\rho = 0.47$)

(t_1, t_2)	(3, 6/7)	(2, 1)	(4/3, 4/3)	(1, 2)	(6/7, 3)
Mean	1.5	1.5	1.5	1.5	1.5
Variance (interdeparture)	1.531	1.35	1.243	1.337	1.507
Variance (interarrival)	1.37	1.25	1.125	1.25	1.37

5. With routine calculations, an explicit expression of $\mathbf{w}_1^*(s)$ can be obtained and

$$\begin{aligned}\mathbf{w}_1^*(s) f_1^*(s) &= \left(0, \frac{[y_{01}(s+\mu_2)(s-t_2) - y_{02}t_2\mu_2](s+\mu_1)s(t_1+t_2)}{[(s+\mu_1)(s+\mu_2)(s-t_1)(s-t_2) - t_1t_2\mu_1\mu_2]t_2}\right) f_1^*(s) \\ &\equiv (0, u^*(s)).\end{aligned} \tag{6.12}$$

6. Matrix $\hat{P}_0$ can be obtained as

$$\hat{P}_0 = \begin{pmatrix} 0 & 0 \\ \dfrac{t_2\mu_2}{(t_1+\mu_2)} \dfrac{(u^*(t_1) - u^*(t_2))}{(t_2 - t_1)} & u^*(t_2) \end{pmatrix}. \tag{6.13}$$

The above expression can be easily modified to include cases with $t_1 = t_2$.

7. Finally, the following simplification is useful in the computation:

$$\begin{aligned}&\sum_{n=1}^{\infty} \frac{\int_0^\infty x^n \mathrm{diag}(\mathbf{w}_1 * F_1)(\mathrm{d}x)}{n!} \left[\sum_{k=0}^{n-1} (D_0 + D_2G_2)^k D_2 G_2^{(1)} (D_0 + D_2G_2)^{n-1-k}\right] \\ &\quad = \begin{pmatrix} 0 & 0 \\ t_2 (G_2^{(1)})_{1,1} \dfrac{(u^*(t_1) - u^*(t_2))}{(t_2 - t_1)} & 0 \end{pmatrix}\end{aligned} \tag{6.14}$$

Now, we are ready to compute the mean and variance of the interdeparture time of class 1 customers. The results are presented in Tables 1–3. Note that the interarrival times of class 1 customers are the sum of two exponentially distributed random variables with parameter t_1 and t_2, respectively. For all the cases, the mean interarrival times (interdeparture times) are $1/\lambda_1 = 1.5$.

As shown in Table 1, the variances of the interdeparture times of class 1 customers are different for different sets of (t_1, t_2). When the difference between t_1 and t_2 increases, the interdepartures times of class 1 customers become more variable (more unpredictable). Also notice that all the queueing systems have the same traffic intensity $\rho = \lambda_1/\mu_1 + \lambda_2/\mu_2 = 0.353$, which is not large. But the departure process becomes more bursty than the input process. Since class 1 and class 2 customers arrive cyclically, the variance of the interarrival time of class 2 customers is the same as that of class 1 customers. Thus, our discussion applies to class 2 customers.

Table 2 shows again that the interdeparture process is more variable than the input process. Since the system becomes busier, we expected that the variance of interdeparture time would be larger. However, the case (6/7, 3) shows this may not be true. Nonetheless, Tables 1 and 2 show that when the difference between t_1 and t_2 is larger, the variance of the interdeparture time is larger. We further conjecture that the difference between the service rates of the two classes of customers will increase the variance of interdeparture time. This conjecture is supported by Tables 1 and 2, and Table 3. This conjecture is also supported by numerous other examples we have tested.

Notice that the traffic intensity is now $\rho = 0.47$, which is larger than that of Table 1 and smaller than that of Table 2. But the variances of interdeparture times are larger than their counterparts in Tables 1 and 2. This implies that the difference in service times have a stronger impact on the burstiness of the departure process than that of the traffic intensity.

Example 6.1 shows the impact of the difference of the arrival processes on the departure process. It is clear that the influence is strong, especially when the service times of the two classes of customers are dramatically different, and needs to be dealt with in the study of such queueing systems.

7. Discussion

In this paper, the LSTs of the interdeparture times of queueing systems with two classes of customers are obtained, when equal priority and nonpreemptive priority are assumed, respectively. The results are useful in analyzing the departure process of queueing systems of interest, especially when the input process is bursty or nonbursty. Many existing results are special cases of the results obtained in this paper. Numerical examples are presented to show the impact of burstiness and arrival pattern of the input process on the interdeparture times.

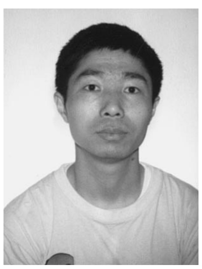
Acknowledgements

Authors would like to thank three anonymous referees for their valuable suggestions and comments that have greatly improved this work. This work has been financially supported through the NSERC operating grants of the two authors.

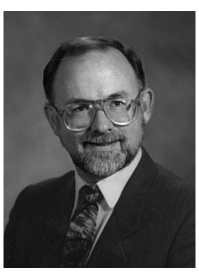
References

[1] D.J. Daley, Queueing output processes, Adv. Appl. Prob. 8 (1976) 395–415.
[2] F.R. Gantmacher, The Theory of Matrices, Chelsea, New York, 1959.
[3] Q-M. HE, Queues with marked customers, Adv. Appl. Prob. 28 (1996) 567–587.
[4] Q-M. HE, M.F. Neuts, Markov arrival processes with marked transitions, Stochastic Processes and their Applications 74 (1998) 37–52.

[5] D.M. Lucantoni, New results on the single server queue with a batch Markovian arrival process, Stochastic Models 7 (1991) 1–46.
[6] P. Nain, Interdeparture times from a queueing system with preemptive resume priority, Performance Evaluation 4 (1984) 93–98.
[7] M.F. Neuts, A versatile Markovian point process, J. Appl. Prob. 16 (1979) 764–779.
[8] M.F. Neuts, Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach, Johns Hopkins University Press, Baltimore, 1981.
[9] M.F. Neuts, Structured Stochastic Matrices of $M/G/1$ Type and Their Applications, Marcel Dekker, New York, 1989.
[10] M.F. Neuts, The fundamental period of the queue with Markov-modulated arrivals, in: Probability, Statistics and Mathematics: Papers in Honor of Samuel Karlin, Academic Press, New York, 1989, pp. 187–200.
[11] H. Saito, The departure process of an N/G/1 queue, Performance Evaluation 11 (1990) 241–251.
[12] D. Stanford, Interdeparture-time distributions for the nonpreemptive priority $\sum M_i/G_i/1$ queue, Performance Evaluation 12 (1991) 43–60.
[13] D. Stanford, W. Fischer, The interdeparture-time distribution for each class in the $\sum M_i/G_i/1$ queue, Queueing Systems 4 (1989) 179–191.
[14] D. Stanford, W. Fischer, Characterising interdeparture times for bursty input streams in the queue with pooled renewal arrivals, Stochastic Models 7 (1991) 311–320.
[15] T. Takine, A nonpreemptive priority $MAP/G/1$ queue with two classes of customers, J. Opns. Res. Soc. Jpn. 39 (1996).
[16] T. Takine, T. Hasegawa, The workload in the $MAP/G/1$ queue with state-dependent services its application to a queue with preemptive resume priority, Stochastic Models 10 (1994) 183–204.
[17] W. Whitt, A light traffic approximation for single-class departure processes from multi-class queues, Manage. Sci. 34 (1988) 1333–1346.



Qi-Ming He is currently an assistant professor in the Industrial Engineering Department of Dalhousie University. He received a Ph.D from the Institute of Applied Mathematics, Chinese Academy of Sciences in 1989 and a Ph.D from the Department of Management Science at the University of Waterloo in 1996. His main research areas are algorithmic methods in applied probability, queueing theory, inventory control, and production management. Recently, he is working on queueing and inventory systems with multiple types of customers (demands).



David Stanford is an associate professor in the Department of Statistical & Actuarial Sciences at the University of Western Ontario. He received his Ph.D. in 1981 from Carleton University, and worked until 1986 at Bell Northern Research in Ottawa. His research interests include queues and risk processes, and applications in call centres and telecommunications. His previous papers include two in Performance Evaluation, while others have appeared in Operations Research, QUESTA, Journal of Applied Probability, INFOR, and other journals. He is a member of the Canadian Operational research Society and INFORMS.