# QUEUES WITH MARKED CUSTOMERS

**Qi-Ming HE**

Department of Management Sciences
University of Waterloo
Waterloo, Ontario, N2L 3G1, Canada
email: qmhe@mansci.uwaterloo.ca

**Abstract**

Queueing systems with distinguished arrivals are described on the basis of Markov arrival processes with marked transitions. Customers are distinguished by their types of arrivals. Usually, the queues observed by customers of different types are different, especially for queueing systems with bursty arrival processes. We study queueing systems from the point of views of customers of different types. A detailed analysis of the fundamental period, queue lengths and waiting times at the epochs of arrivals are given. The results obtained are the generalizations of the results of the *MAP/G/1* queue.

**Keywords** Queueing theory, Markov arrival process, *MAP/G/1* queue, matrix-analytic method.

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 60K25

The research was supported in part by the K.C.Wang Education Foundation and the National Science Foundation through Grant Nr. DDM-8915235.

## 1. Introduction

To study a queueing system, we usually begin with descriptors such as the busy period, queue lengths and waiting times at the epoch of an arbitrary arrival, or at an arbitrary time $t$, or at the epoch of an arbitrary departure. These descriptors reflect some important characteristics of the queue. However, in some cases, they are inadequate. Let us examine two examples first.

**Example 1.** We consider a single server queue with $K$ types of arrivals, $\sum_k M_k/G_K/1$. The input process consists of several independent *Poisson* processes. Customers of different types have different service times but there is no priority with services. It is clear that the queues observed by arrivals of different types are different. So, customers of different types should be studied separately.

**Example 2.** We consider a *MAP/G/1* queue where the arrival process has coefficient matrices

$$D_0 = \begin{pmatrix} -a_1 & 0 \\ 0 & -a_2 \end{pmatrix} \quad \text{and} \quad \tilde{D}_1 = \begin{pmatrix} 0.9a_1 & 0.1a_1 \\ 0.9a_2 & 0.1a_2 \end{pmatrix},$$

where $a_1$ and $a_2$ are positive and $a_1 >> a_2$.

According to the coefficient matrices, we divide arrivals into four types: type 1 are arrivals at phase 1 without a change of phase, type 2 arrivals occur when the phase changes from 1 to 2, type 3 arrivals occur when the phase changes from 2 to 1, and type 4 arrivals occur in phase 2 without a change of phase. Since the arrival rate $a_2$ is small, there are few arrivals in phase 2. In phase 1, there might be a large number of arrivals in a short time, since $a_1$ is large. So, an arrival of type 3 is the *leading arrival* of a burst of the arrival process and an arrival of type 2 is the *ending arrival* of a burst. The queue observed by an arbitrary leading arrival would be much different from that of an arbitrary ending arrival. Also, the queue observed by an arbitrary arrival in phase 1 would be different from the queue observed by an arbitrary arrival in phase 2. An example of such input process can be found in Chandramouli, Neuts and Ramaswami [2].

Generally, a queueing system may have a bursty input process or an input process with several types of arrivals. To study such queueing systems, we suggest marking all the arrivals of interest and studying the queues observed by marked arrivals. To model the marked point processes, we use the Markov arrival process with marked transitions (*MMAP*.) An *MMAP* is a point process with arrivals of different types. It is a generalization of the *MAP*.

The basic model in this paper is the *MMAP/G/1* queue. We also study the more general models such as $MMAP(K)/G_K/1$, where the service times of customers of different types are different, and $BMMAP(K)/G_K/1$, where batch arrivals are allowed. There is no priority on customers. Therefore, when we do not distinguish the arrivals (and the service times are i.i.d.r.v.s), we obtain a *MAP/G/1* queue, which has been studied extensively. Many important results about the *MAP/G/1* queue will be reviewed in this paper. They provide the basis for the present study.

There is some literature on queues with many types of arrivals. Some of them give priorities to customers (see Cohen [4] and Stanford [17]), while others do not (see Stanford and Fischer [18] and [19].) The models in Stanford and Fischer [18] and [19] are special cases of our model. They studied the interdeparture times of customers of different types for queues of the $\sum_k M_k/G_K/1$ type.

This paper consists of seven parts. Following this introduction, the *MMAP/G/1* queue is introduced and some notation used in this paper is set up in Section 2. In Section 3, the fundamental period of the *MMAP/G/1* queue is discussed and a new formula is given for the expected number of customers served in a busy period of the *MAP/G/1* queue. We study queue lengths at arrival epochs of customers of different types in Section 4 and waiting times in Section 5 for the *MMAP/G/1* queue. In Section 6, we generalize the results in Sections 3 and 5 to the $MMAP(K)/G_K/1$ queue. A short discussion is given to the $MMAP(K)/G_K/1$ queue with batch arrivals. In Section 7, we use time reversal to explain the most important factors which influence the queues observed by customers of different types.

Now, we give the definition of the *MMAP*. The following definition is adopted from He and Neuts [5] with some simplifications.

*The Markov Arrival Process with Marked Transitions:* We consider an $m$-state Markov renewal process with an irreducible embedded Markov chain with transition probability matrix $P$ and exponential sojourn time distributions given by

$$F_{ij}(x) = 1 - \exp(-\sigma_i x),$$

for $1 \le i, j \le m$. This is also a Markov chain in continuous time. Let $D$ be the infinitesimal generator of this Markov chain. The matrix $D$ and the parameters $P$ and $\sigma_i$, $1 \le i \le m$, of the Markov renewal process are related by:

$$D_{ii} = -(1 - P_{ii})\sigma_i, \quad \text{for } 1 \le i \le m,$$

and

$$D_{ij} = P_{ij}\sigma_i, \quad \text{for } 1 \le i,\ j \le m, \quad i \ne j.$$

In what follows we denote the stationary probability vector of $D$ by $\theta$.

We consider a finite set $\{1, \cdots, K\}$. We define a Markov renewal process (*MRP*, see Cinlar [3]) $\{(J_n,\ L_n,\ \tau_n),\ n \geq 0\}$ on the state space $\{[\{1, \cdots, m\} \times \{1, \cdots, K\}] \times [0, \infty)\}$ with the transition probability matrix

$$\mathbf{P}\{J_n = j, L_n = k, \tau_n \leq x | J_{n-1} = i\} = [\int_0^x \exp(D_0 u) du D_k]_{ij},$$

with $1 \leq i, j \leq m$, $1 \leq k \leq K$, $x \geq 0$ and $L_n$ is the marking variable. For convenience, we set $\tau_0 = 0$. The matrices $D_k$, $1 \leq k \leq K$, are nonnegative. The matrix $D_0$ has negative diagonal elements and is assumed to be nonsingular. The latter requirement serves to ensure that, with probability one, infinitely many transitions epochs are marked. In addition, we have that

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

We call an arrival of type $k$ if the arrival is marked by $k$. The stationary arrival rate of arrivals of type $k$ is given by

$$\lambda_k^* = \theta D_k \mathbf{e}, \quad 1 \leq k \leq K, \tag{2}$$

where $\mathbf{e}$ is the column vector with all elements 1.

When we do not distinguish the arrivals of the *MMAP*, we obtain a *MAP* (see Asmussen and Ramaswami [1], Lucantoni [7], Lucantoni, et al. [8] and Neuts [9]) with coefficient matrices $\{D_0, \tilde{D}_1\}$, where

$$\tilde{D}_1 = \sum_{k=1}^{K} D_k. \tag{3}$$

Let $N(t)$ be the number of arrivals in $(0, t]$ and $J(t)$ the phase of the Markov chain with generator $D$ at time $t$. Let

$$p_{jj'}(n, t) = \mathbf{P}\{N(t) = n, J(t) = j' \mid N(0) = 0, J(0) = j\}, \quad 1 \leq j, j' \leq m.$$

$P(n, t)$, $n \geq 0$, are $m \times m$ matrices with $(j, j')$th element $p_{jj'}(n, t)$ and satisfy the differential equations

$$\begin{aligned} P'(0, t) &= P(0, t) D_0; \\ P'(n, t) &= P(n, t) D_0 + P(n-1, t) \tilde{D}_1, \quad n \geq 1. \end{aligned} \tag{4}$$

The generating function of $\{P(n,t), n \geq 0\}$ is denoted by $P^*(z,t)$ and it has been shown that

$$P^*(z,t) = \exp\{(D_0 + z\tilde{D}_1)t\} \tag{5}$$

The total arrival rate $\lambda^*$ (the arrival rate of the *MAP*) and $\{\lambda_k^*, 1 \leq k \leq K\}$ have the following relation,

$$\lambda^* = \sum_{k=1}^{K} \lambda_k^*. \tag{6}$$

**Note:** To show the modelling of point processes whose arrivals are of different types with an *MMAP*, we construct an *MMAP* for the input process in Example 2. We distinguish arrivals into the four types described in Example 2. The coefficient matrices for arrivals of the four types are

$$D_1 = \begin{pmatrix} 0.9a_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0 & 0.1a_1 \\ 0 & 0 \end{pmatrix},$$

$$D_3 = \begin{pmatrix} 0 & 0 \\ 0.9a_2 & 0 \end{pmatrix}, \quad D_4 = \begin{pmatrix} 0 & 0 \\ 0 & 0.1a_2 \end{pmatrix}$$

respectively. $\{D_0, D_1, D_2, D_3, D_4\}$ represents an *MMAP*.

## 2. The Embedded MRP of the *MMAP/G/1* Queue

We consider a single server queue with an *MMAP* as its input process and i.i.d. service times which have distribution function $F(t)$ with Laplace Stieltjes (L.S.) transform $f^*(s)$ and finite mean $\mu^*$. An arrival of type $k$ brings a customer of type $k$ to the system, $1 \leq k \leq K$.

The *MMAP/G/1* queue is simplified to a *MAP/G/1* queue by ignoring the types of arrivals. The *MAP/G/1* queue has been studied in a number of papers such as Lucantoni [7], Neuts [12] and Ramaswami [14], [15] and [16]. Some of the results obtained will be used here.

The study on the *MAP/G/1* queue usually starts with the embedded Markov renewal process at departure epochs of customers. Let $\tau_n$ be the epoch of the $n$th departure from the queue, with $\tau_0 = 0$. $(q_n, J_n)$ is defined as the number of customers in the system and the phase of the arrival process at $\tau_n+$. Then $\{q_n, J_n, \tau_n, \ n \geq 0\}$ is a semi-Markov process on the state space $\{(i,j), i \geq 0, 1 \leq j \leq m\}$. The transition kernel of the embedded $MRP$ is given by

$$Q(s) = \begin{pmatrix} B_0(s) & B_1(s) & B_2(s) & \cdots \\ A_0(s) & A_1(s) & A_2(s) & \cdots \\ & A_0(s) & A_1(s) & \cdots \\ & & A_0(s) & \cdots \\ & & & \ddots \end{pmatrix},$$

where

$$A_n(s) = \int_0^\infty e^{-st} P(n,t) dF(t), \quad Re(s) \geq 0, \quad n \geq 0, \tag{7}$$

and

$$B_n(s) = (sI - D_0)^{-1} \tilde{D}_1 A_n(s), \ n \geq 0.$$

We define

$$A^*(z,s) = \sum_{n=0}^{\infty} z^n A_n(s), \quad |z| \leq 1, \quad Re(s) \geq 0.$$

It is easy to see that

$$A^*(z,s) = \int_0^\infty exp\{(-sI + D^*(z))t\} dF(t), \tag{8}$$

where $D^*(z) = D_0 + z\tilde{D}_1$. Let

$$A^*(z) = A^*(z,0) \ \text{ and } \ A = A^*(1,0).$$

Since $A = \int_0^\infty exp\{Dt\}dF(t)$, we know that $\theta$ is also the left invariant vector of the stochastic matrix $A$. Let

$$\beta = \sum_{n=0}^{\infty} n A_n(0) \mathbf{e},$$

The traffic intensity $\rho$ of this queueing system is defined as

$$\rho = \theta\beta = \lambda^* \mu^*. \tag{9}$$

We assume that $\rho < 1$ so that the embedded $MRP$ is positive recurrent.

## 3. The Fundamental Period

We define the *level* $\mathbf{i}$ to be the set of states $\{(i,1), \cdots, (i,m)\}$, $i \geq 0$. Generally, the fundamental period is the first passage time from level $\mathbf{i+1}$ to level $\mathbf{i}$. It is a generalization of the busy period of the queue. For the embedded $MRP$, let $G^{[r]}_{jj'}(n,x)$,

$n \geq 1$, $x \geq 0$, be the probability that the first passage from the state $(i+r, j)$ to the state $(i, j')$, $i \geq 0$, $1 \leq j, j' \leq m$, $r \geq 1$, occurs at the $n$th transition before time $x$, and that $(i, j')$ is the first state visited in the level $\mathbf{i}$. $G^{[r]}(n, x)$ is the matrix with elements $G^{[r]}_{jj'}(n, x)$. We define the joint transform

$$G^{*[r]}(z, s) = \sum_{n=1}^{\infty} z^n \int_0^{\infty} e^{-sx} G^{[r]}(n, dx), \quad |z| \leq 1, \quad Re(s) \geq 0, \quad r \geq 1. \tag{10}$$

Let $G^*(z, s) = G^{*[1]}(z, s)$. It was proved (see Neuts [10]) that $G^{*[r]}(z, s) = [G^*(z, s)]^r$ and that $G^*(z, s)$ is the unique solution to the nonlinear equation

$$G^*(z, s) = z \sum_{n=0}^{\infty} A_n(s)[G^*(z, s)]^n. \tag{11}$$

It was also known that $G = G^*(1, 0)$ is a stochastic matrix if and only if $\rho \leq 1$. From (11), the mean number of customers served in a fundamental period and the mean length of a fundamental period were derived.

Beginning with Neuts [12] for the *MMPP/G/1* queue, then Ramaswami [16] and Lucantoni [7] for the *BMAP/G/1* queue, it was proved that $G^*(z, s)$ has a matrix exponential form

$$G^*(z, s) = z \int_0^{\infty} exp\{(-sI + D_0 + \tilde{D}_1 G^*(z, s))t\} dF(t). \tag{12}$$

The appearance of the above equation led to a number of interesting consequences, such as the commutativity of the matrix $G$ and $D_0 + \tilde{D}_1 G$ and a substantial simplification in the computation of the matrix $G$. Here, we use (12) to derive a new formula for the mean number of customers served in a fundamental period and, correspondingly, a new formula for the mean length of a fundamental period. Let

$$\tilde{\mu}_1 = (\frac{d}{dz} G^*(z, s)|_{z=1, s=0})\mathbf{e} \text{ and } \hat{\mu}_1 = -(\frac{d}{ds} G^*(z, s)|_{z=1, s=0})\mathbf{e}.$$

$\tilde{\mu}_1$ is a vector for the conditional mean numbers of customers served during a fundamental period and $\hat{\mu}_1$ is a vector for the conditional mean lengths of a fundamental period.

Let $\mathbf{g}$ be the stationary probability vector of the matrix $D_0 + \tilde{D}_1 G$. By (12), we know that $\mathbf{g}$ is also the left invariant vector of the matrix $G$, i.e., $\mathbf{g}G = \mathbf{g}$ and $\mathbf{ge} = 1$. Differentiating both sides of (12) with respect to $z$ and postmultiplying by $\mathbf{e}$, we have, with the fact $(D_0 + \tilde{D}_1 G)\mathbf{e} = 0$,

$$\tilde{\mu}_1 = \mathbf{e} + \int_0^{\infty} \sum_{n=1}^{\infty} \frac{x^n}{n!} (D_0 + \tilde{D}_1 G)^{n-1} \tilde{D}_1 \tilde{\mu}_1 dF(x).$$

For the sum in the integral, we have

$$\sum_{n=1}^{\infty} \frac{x^n}{n!}(D_0 + \tilde{D}_1 G)^{n-1}(\mathbf{eg} - D_0 - \tilde{D}_1 G) = x\mathbf{eg} + I - exp\{(D_0 + \tilde{D}_1 G)x\}.$$

Since the matrix $\mathbf{eg} - D_0 - \tilde{D}_1 G$ is invertible and $\mathbf{eg}(\mathbf{eg} - D_0 - \tilde{D}_1 G) = \mathbf{eg}$, we obtain

$$\sum_{n=1}^{\infty} \frac{x^n}{n!}(D_0 + \tilde{D}_1 G)^{n-1} = (I - exp\{(D_0 + \tilde{D}_1 G)x\})(\mathbf{eg} - D_0 - \tilde{D}_1 G)^{-1} + x\mathbf{eg}.$$

With (12), we have

**Theorem 3.1.** For the *MAP/G/1* queue,

$$\tilde{\mu}_1 = [\mathbf{eg}(I - \mu^* \tilde{D}_1) - D - \tilde{D}_1 G + G\tilde{D}_1]^{-1}\mathbf{e}. \qquad (13)$$

$$\hat{\mu}_1 = \mu^*[\mathbf{eg}(I - \mu^* \tilde{D}_1) - D - \tilde{D}_1 G + G\tilde{D}_1]^{-1}\mathbf{e}. \qquad (14)$$

**Proof.** The only thing left is to prove the invertibility of matrix $\mathbf{eg}(I - \mu^* \tilde{D}_1) - D - \tilde{D}_1 G + G\tilde{D}_1$, which is given in Appendix A. The proof of (14) is similar to that of (13). □

**Note 2.1:** Compared to the classical formula

$$\tilde{\mu}_1 = (I - G + \mathbf{eg})[I - A + (\mathbf{e} - \mathbf{g})\beta]^{-1}\mathbf{e},$$

(13) does not have $A$ and $\beta$ but more elementary parameters $D_0$, $\tilde{D}_1$ and $\mu^*$. This fact is useful in the computations of $\tilde{\mu}_1$ and $\hat{\mu}_1$.

**Note 2.2:** The results in Theorem 3.1 can be generalized to the *BMAP/G/1* queue (see Lucantoni [7]) where batch arrivals are allowed. Let $\{D_0, \tilde{D}_n, n \geq 1\}$ be the coefficient matrices for the *BMAP*, where $\tilde{D}_n$ is for group arrivals of size $n$. (13) becomes, for example,

$$\tilde{\mu}_1 = [\mathbf{eg}(I - \mu^* B(G)) - D - B(G)G + GB(G)]^{-1}\mathbf{e}, \qquad (15)$$

where

$$B(G) = (D - \sum_{n=0}^{\infty} \tilde{D}_n G^n + (\sum_{n=1}^{\infty} n\tilde{D}_n)\mathbf{eg})(I - G + \mathbf{eg})^{-1}.$$

For the *MMAP/G/1* queue, we are interested in the numbers of customers of different types served during a fundamental period. Clearly, the number of customers of each type served in a fundamental period changes with the type of the first customer and the type of the last customer served in the fundamental period. So, we give the following generalization of (12) with the same method used in Neuts [12].

Let $\mathbf{n} = (n_1, \cdots, n_K)$ and $\mathbf{z} = (z_1, \cdots, z_K)$. Similar to the matrix $G^{[1]}(n, x)$, we define the matrix $\tilde{G}_{kk'}(\mathbf{n}, x)$ with elements the conditional probability that the first passage from level **i+1** to level **i** occurs before time $x$, there are total $n_1 + \cdots + n_K$ customers with $n_1$ customers of type 1, $\cdots$, $n_K$ customers of type $K$ arrived during that time and the last arrival before level **i** is reached is of type $k'$, given that the last arrival before time 0 is of type $k$. We define the following joint transforms,

$$\tilde{G}^*_{kk'}(\mathbf{z}, s) = \sum_{\mathbf{n} \geq 0} \mathbf{z}^{\mathbf{n}} \int_0^\infty e^{-sx} \tilde{G}^*_{kk'}(\mathbf{n}, x) dF(x), \quad ||\mathbf{z}|| \leq 1, \quad Re(s) \geq 0, \tag{16}$$

for $1 \leq k, k' \leq K$, where $\mathbf{z}^{\mathbf{n}} = z_1^{n_1} \cdots z_K^{n_K}$. Clearly, $\tilde{G}^*_{kk'}(\mathbf{z}, s)$ is also the joint transform of the length of a busy period and the numbers of customers served during the busy period, where the last arrival in the busy period is of type $k'$, given that the busy period starts with a customer of type $k$.

We define $\Phi_{kk'}(x, \mathbf{n}, y)$ to be a matrix with elements the probability that, given that the last arrival (before 0) is of type $k$ and the residual service time is $x$, the current fundamental period ends before time $y$ with $\mathbf{n}$ new arrivals occurring and the last arrival being of type $k'$. The joint transform of $\Phi_{kk'}(x, \mathbf{n}, y)$, with respect to $\mathbf{n}$ and $y$, is denoted by $\Phi^*_{kk'}(x, \mathbf{z}, s)$, $1 \leq k, k' \leq K$. We denote by $\tilde{G}^*(\mathbf{z}, s)$ and $\Phi^*(x, \mathbf{z}, s)$ the matrices with $(k, k')$th block $\tilde{G}^*_{kk'}(\mathbf{z}, s)$ and $\Phi^*_{kk'}(x, \mathbf{z}, s)$ respectively.

**Theorem 3.2.** For the *MMAP/G/1* queue, the matrix $\Phi^*(x, \mathbf{z}, s)$ is given by

$$\Phi^*(x, \mathbf{z}, s) = e^{-sx} exp\{(T_0 + T_1 \tilde{G}^*(\mathbf{z}, s))x\}, \tag{17}$$

where $T_0$ and $T_1$ are $mK \times mK$ matrices,

$$T_0 = \begin{pmatrix} D_0 & & \\ & \ddots & \\ & & D_0 \end{pmatrix} \quad \text{and} \quad T_1 = \begin{pmatrix} D_1 & \cdots & D_K \\ \vdots & \vdots & \vdots \\ D_1 & \cdots & D_K \end{pmatrix}.$$

The matrix $\tilde{G}^*(\mathbf{z}, s)$ satisfies the equation

$$\tilde{G}^*(\mathbf{z}, s) = \begin{pmatrix} z_1 I & & \\ & \ddots & \\ & & z_K I \end{pmatrix} \int_0^\infty \Phi^*(x, \mathbf{z}, s) dF(x). \tag{18}$$

**Proof.** From the definitions, it is easy to see the following facts.

$$\tilde{G}^*_{kk'}(\mathbf{z}, s) = z_k \int_0^\infty \Phi^*_{kk'}(x, \mathbf{z}, s) dF(x); \tag{19}$$

$$\Phi^*_{kk'}(x_1 + x_2, \mathbf{z}, s) = \sum_{l=1}^{K} \Phi^*_{kl}(x_1, \mathbf{z}, s) \Phi^*_{lk'}(x_2, \mathbf{z}, s), \quad x_1, x_2 \geq 0, \tag{20}$$

for $1 \leq k, k' \leq K$. The second equality follows in a manner similar to (22) in Neuts [12]. We notice that

$$\Phi^*_{kk'}(x, \mathbf{0}, y) = \begin{cases} 0, & x \geq y; \\ \delta(k - k') exp\{D_0 x\}, & x < y, \end{cases}$$

where $\delta(i)$ is an indicator of 0, and for $\mathbf{n} \neq \mathbf{0}$ and $y \geq x$,

$$\Phi^*_{kk'}(x, \mathbf{n}, y)$$

$$= \sum_{l=1}^{K} \int_0^x z_l exp\{D_0 u\} D_l \int_0^{y-x} \Phi_{lk'}(x - u + v, \mathbf{n} - \mathbf{n}_l, y - u)(1 - \delta(n_l)) dF(v) du,$$

where $\mathbf{n}_l$ is a vector of dimension $K$ with the $l$th element 1 and the rest 0. Taking the joint transform of $\Phi_{kk'}(x, \mathbf{n}, y)$ with respect to $\mathbf{n}$ and $y$, we have, after some routine calculations,

$$\Phi^*_{kk'}(x, \mathbf{z}, s)$$

$$= exp\{-(sI - D_0)x\}[I + \sum_{l=1}^{K} \sum_{l'=1}^{K} \int_0^x exp\{(sI - D_0)u\} D_l \tilde{G}^*_{ll'}(\mathbf{z}, s) \Phi^*_{l'k'}(u, \mathbf{z}, s) du],$$

for $1 \leq k, k' \leq K$. Differentiating both sides of the above equation with respect to $x$ we have

$$\frac{\partial \Phi^*_{kk'}(x, \mathbf{z}, s)}{\partial x} = -(sI - D_0)\Phi^*_{kk'}(x, \mathbf{z}, s) + \sum_{l=1}^{K} \sum_{l'=1}^{K} D_l \tilde{G}^*_{ll'}(\mathbf{z}, s) \Phi^*_{l'k'}(x, \mathbf{z}, s), \tag{21}$$

for $1 \leq k, k' \leq K$. From these differential equations, with initial conditions, $\Phi^*_{kk'}(0, \mathbf{z}, s) = \delta(k - k')I$, $1 \leq k, k' \leq K$, the theorem is obtained. $\square$

Theorem 3.2 gives the analogue of the exponential form of the fundamental period. The equation (18) can also be obtained when we use $(T_0, T_1)$ as the coefficient matrices of the input $MAP$, where the types of arrivals are indicated by the phases.

By definition, an *MMAP* is a *MAP* when arrivals are not distinguished. Then the length of a fundamental period of an *MMAP/G/1* queue is independent of the types of the first and the last customers served during the fundamental period. Hence, the length of a fundamental period of an *MMAP/G/1* queue is the same as that of its corresponding *MAP/G/1* queue. Also, we know that the type of the first arrival of a fundamental period affects the number of arrivals of each type in a fundamental period but not the total number of arrivals in the fundamental period. Therefore, we have

$$G^*(z,s) = \sum_{k'=1}^{K} \tilde{G}^*_{kk'}(\mathbf{z},s)|_{z_1=\cdots=z_K=z}, \tag{22}$$

for any $1 \le k \le K$. (22) can be verified from (21). Let $\tilde{G} = \tilde{G}^*(\mathbf{z},s)|_{\{s=0,z_1=\cdots=z_K=1\}}$. Then we have

$$\tilde{G} = \begin{pmatrix} \tilde{G}_1 & \cdots & \tilde{G}_K \\ \vdots & \vdots & \vdots \\ \tilde{G}_1 & \cdots & \tilde{G}_K \end{pmatrix},$$

where $\tilde{G}_{k'} = \tilde{G}^*_{kk'}(\mathbf{z},s)|_{\{s=0,z_1=\cdots=z_K=1\}}$, $1 \le k' \le K$. Let $\tilde{\mathbf{g}}$ be the left invariant vector of $\tilde{G}$. Then $\tilde{\mathbf{g}} = (\tilde{\mathbf{g}}_1, \cdots, \tilde{\mathbf{g}}_K)$, where $\tilde{\mathbf{g}}_k = \mathbf{g}\tilde{G}_k$, $1 \le k \le K$. Let $\tilde{\mu}_1(k)$ be the vector for the expected number of arrivals of type $k$ occurring during a fundamental period (including the one just before the fundamental period). Similar to (13), we have

$$\begin{aligned} \tilde{\mu}_1(k) &= [\mathbf{e}\tilde{\mathbf{g}}(I-\mu^* T_1) - T_0 - T_1 - T_1\tilde{G} + \tilde{G}T_1]^{-1} \\ &\quad \cdot(\mathbf{e}\tilde{\mathbf{g}} - T_0 - T_1\tilde{G})\mathbf{e}(k), \quad 1 \le k \le K, \end{aligned} \tag{23}$$

where $\mathbf{e}(k)$ is the column vector of dimension $mK$ with elements $(k-1)m+1$ to $km$ equal to 1 and all others 0. For a busy period, (23) gives the expected number of customers of type $k$ served in that busy period.

## 4. Queue Lengths

In Neuts [13], Locantoni [7] and Ramaswami [14] and [16], the queue lengths immediately after an arbitrary departure or at an arbitrary time are discussed in detail. We will not concern ourselves with those results except those on the distribution of the queue length at an arbitrary departure. We shall use those results to derive the distributions of the queue length at an arbitrary arrival and the queue lengths at an arbitrary arrival of type $k$, $1 \le k \le K$.

The stationary probability vector of $(q_n, J_n)$ has the form $\mathbf{X} = (\mathbf{x}_0, \mathbf{x}_1, \cdots)$, where $\mathbf{x}_n$, $n \geq 0$, are $m$-vectors. Let $\mathbf{X}^*(z) = \sum_{n=0}^{\infty} z^n \mathbf{x}_n$. Then we have

$$\mathbf{X}^*(z)[zI - A(z)] = -\mathbf{x}_0 D_0^{-1} D(z) A(z). \tag{24}$$

The vector $\mathbf{x}_0$ is given explicitly by

$$\mathbf{x}_0 = \lambda^*(1-\rho)\mathbf{g}(-D_0). \tag{25}$$

First, we discuss the queue length at an arbitrary arrival. Let $x_{a,nj}$ be the stationary probability that at an arbitrary arrival there are $n$ customers in the system and the phase of the arrival process is $j$ right after the arrival, $n \geq 0$, $1 \leq j \leq m$. We define $\mathbf{x}_{a,n}$ by $(x_{a,n1}, \cdots, x_{a,nm})$, $n \geq 0$.

By conditioning on the number of customers in the system and the phase of the arrival process right after the last departure before an arbitrary arrival, we have (ignoring the types of arrivals)

$$\mathbf{x}_{a,0} = \mathbf{x}_0(-D_0^{-1}\tilde{D}_1); \tag{26}$$

$$\begin{aligned} \mathbf{x}_{a,n} &= \mathbf{x}_0(-D_0^{-1}\tilde{D}_1)\int_0^{\infty} P(n-1,t)(1-F(t))dt\tilde{D}_1 \qquad (27) \\ &\quad + \sum_{k=1}^{\infty} \mathbf{x}_k \int_0^{\infty} P(n-k,t)(1-F(t))dt\tilde{D}_1, \quad n \geq 1. \end{aligned}$$

Let $\mathbf{X}_a^*(z) = \sum_{n=0}^{\infty} z^n \mathbf{x}_{a,n}$. Then we have

$$\mathbf{X}_a^*(z) = \mathbf{x}_0(-D_0^{-1}\tilde{D}_1) + [\mathbf{X}^*(z) - \mathbf{x}_0(-D_0^{-1}\tilde{D}_1)z](D_0 + z\tilde{D}_1)^{-1}A(z)\tilde{D}_1.$$

With (24), we have proved

**Theorem 4.1.** For the *MMAP/G/1* queue,

$$\mathbf{X}_a^*(z) = (1-z)\mathbf{x}_0 D_0^{-1} A(z)(zI - A(z))^{-1}\tilde{D}_1, \quad |z| \leq 1. \tag{28}$$

□

We define $\mathbf{x}_{a,n}(k)$ similarly to $\mathbf{x}_{a,n}$ by incorporating the fact that the arrival is of type $k$, $n \geq 0$, $1 \leq k \leq K$.

**Theorem 4.2.** For the *MMAP/G/1* queue, let $\mathbf{X}_a^*(k,z) = \sum_{n=0}^{\infty} z^n \mathbf{x}_{a,n}(k)$, $1 \le k \le K$. Then we have

$$\mathbf{X}_a^*(k,z) = \frac{\lambda^*}{\lambda_k^*}(1-z)\mathbf{x}_0 D_0^{-1} A(z)(zI - A(z))^{-1} D_k, \quad 1 \le k \le K, \quad |z| \le 1. \tag{29}$$

**Proof.** Again, we condition on the the number of customers in the system and the phase of the arrival process after the last departure before an arbitrary arrival of type $k$. Since the probability that an arbitrary arrival is of type $k$ is $\lambda_k^*/\lambda^*$, we obtain equations for $\{\mathbf{x}_{a,n}(k), n \ge 0\}$ by making small changes in (26) and (27), changing $\tilde{D}_1$ to $D_k$ and multiplying $\lambda^*/\lambda_k^*$ on the right hand sides of (26) and (27). The result comes with simplifications. □

**Note:** To derive the results in Theorems 4.1 and 4.2, the following approach was suggested by Professor Marcel Neuts.

Let $\mathbf{y}_n$ be the probability vector that there are $n$ customers in the system at an arbitrary time, $n \ge 0$. The probability that there is an arrival in $(t, t+\delta t)$ is given by $\lambda^* \delta t$. The probability vector that there are $n$ customers in the system and there is an arrival in $(t, t+\delta t)$ is given by $\mathbf{y}_n \tilde{D}_1 \delta t$. So, the probability vector that there are $n$ customers in the system at the epoch of an arbitrary arrival is given by

$$\frac{\mathbf{y}_n \tilde{D}_1 \delta t}{\lambda^* \delta t}. \tag{30}$$

With the relation between $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_n\}$ (see Lucantoni [7]), (30) leads to (28). Similarly, (29) can be derived with this method.

This method is useful not only for queue lengths but also for waiting times. In fact, the results proved in the next section can be obtained by this method. The proofs in this section and the next section provide support for this simple but very intuitive method. See Chapter 5 in Neuts [13] for more examples.

## 5. Waiting Times

The virtual waiting time of the *MAP/G/1* queue can be obtained from a generalization of the classical Pollaczek-Khinchin formula which is for the virtual waiting time in the *M/G/1* queue (see Neuts [11] and Ramaswami [14]). We assume that the service order is FCFS. Let $W_j(x)$ be the joint probability that at an arbitrary time the arrival process is in phase $j$ and the customer who arrives at that time waits at most a time $x$ before entering service, $1 \le j \le m$. Let $\mathbf{W}(x) = (W_1(x), \cdots, W_m(x))$.

The Laplace-Stieltjes transform is defined as

$$\mathbf{W}^*(s) = \int_0^\infty e^{-sx} d\mathbf{W}(x).$$

In Ramaswami [14], it has been shown that

$$\mathbf{W}^*(s) = s\mathbf{y}_0[sI + D_0 + \tilde{D}_1 f^*(s)]^{-1}, \quad Re(s) \geq 0, \tag{31}$$

where $\mathbf{y}_0$ is the probability vector that the queue length is zero at an arbitrary time. It has been shown that

$$\mathbf{y}_0 = -\lambda^* \mathbf{x}_0 D_0^{-1} = (1-\rho)\mathbf{g}. \tag{32}$$

We now derive the L.S. transform of the waiting time of an arbitrary arrival. Then we derive the L.S. transform of the waiting time of an arbitrary arrival of type $k$, $1 \leq k \leq K$. The method used here is from Lindley [6] for the *GI/G/1* queue.

Let $W_{a,j}(x)$ be the joint probability that the waiting time of an arbitrary arrival is at most $x$ before entering service and the phase of the arrival process right after this arrival is $j$, $1 \leq j \leq m$. Let $\mathbf{W}_a(x) = (W_{a,1}(x), \cdots, W_{a,m}(x))$. The Laplace-Stieltjes transform of $\mathbf{W}_a(x)$ is denoted by $\mathbf{W}_a^*(s)$.

**Theorem 5.1.** For the *MMAP/G/1* queue,

$$\mathbf{W}_a^*(s) = \frac{s}{\lambda^*}\mathbf{y}_0(sI + D_0 + f^*(s)\tilde{D}_1)^{-1}\tilde{D}_1. \tag{33}$$

**Proof.** Let $w_n$ be the actual waiting time of the $n$th customer. Then we have

$$w_{n+1} = max\{0,\ w_n - v_n + U_n\}$$

where $v_n$ is the service time of the $n$th customer and $U_n$ is the interarrival time between the $n$th and the $(n+1)$st customer. In steady state, $w_n$ and $w_{n+1}$ have the same distribution vector $\mathbf{W}_a(x)$. Conditioning on the lengths of $v_n$ and $U_n$, we have, for $x \geq 0$,

$$\mathbf{W}_{a,j}(x) = \int_0^\infty \int_0^{x+t} \sum_{i=1}^m \mathbf{W}_{a,i}(x-u+t)(exp\{D_0 t\}D_1)_{ij} dt dF(u).$$

In matrix form, we obtain, for $x \geq 0$,

$$\mathbf{W}_a(x) = \int_0^\infty \int_0^{x+t} \mathbf{W}_a(x-u+t) exp\{D_0 t\} dt D_1 dF(u),$$

we have the integral equation

$$\mathbf{W}_a(x) = \int_{-\infty}^{x} \mathbf{W}_a(x-u)dH(u), \tag{34}$$

where

$$H(u) = \int_0^\infty F(t+u)exp\{D_0 t\}dt\tilde{D}_1, \quad -\infty < u < \infty.$$

We extend the definition of $\mathbf{W}_a(x)$ from $x \geq 0$ to $-\infty \leq x \leq \infty$ to obtain a new function $\mathbf{V}(x)$. For $x < 0$, we have

$$\begin{aligned}
\mathbf{V}(x) &= \int_{-\infty}^{x} \mathbf{W}_a(x-u) \int_0^\infty F(t+du)exp\{D_0 t\}dt\tilde{D}_1 \\
&= \int_0^\infty \int_{x-y}^\infty \mathbf{W}_a(y)dF(u)exp\{D_0(u+y-x)\}dy\tilde{D}_1 \\
&= \int_0^\infty d\mathbf{W}_a(y)exp\{D_0 y\}(-D_0^{-1}) \int_0^\infty dF(u)exp\{D_0 u\}dyexp\{-D_0 x\}\tilde{D}_1 \\
&= \mathbf{c}(-D_0^{-1})exp\{-D_0 x\}\tilde{D}_1,
\end{aligned}$$

where

$$\mathbf{c} = \int_0^\infty d\mathbf{W}_a(y)exp\{D_0 y\} \int_0^\infty exp\{D_0 u\}dF(u).$$

Taking the L.S. transform of $\mathbf{V}(x)$, denoted by $\mathbf{V}^*(s)$, we have, for $Re(s) \geq 0$ and $-Re(s) > sp(D_0)$ (the eigenvalue with the largest real part of $D_0$),

$$\mathbf{V}^*(s) = \mathbf{W}_a^*(s)H^*(s),$$

where $H^*(s)$ is the L.S. transform of $H(u)$ and

$$\begin{aligned}
H^*(s) &= \int_{-\infty}^\infty e^{-st} \int_0^\infty F(du+t)exp\{D_0 t\}dt\tilde{D}_1 \\
&= -f^*(s)(sI+D_0)^{-1}\tilde{D}_1.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\mathbf{V}^*(s) &= [\int_{-\infty}^{0-} + \int_{0-}^{0} + \int_0^\infty]e^{-sx}d\mathbf{V}(x) \\
&= -\mathbf{c}(-D_0^{-1})(sI+D_0)^{-1}(-D_0)\tilde{D}_1 - \mathbf{c}(-D_0^{-1})\tilde{D}_1 + \mathbf{W}_a^*(s).
\end{aligned}$$

By combining the above results, we have

$$\mathbf{W}_a^*(s) = s\mathbf{c}(-D_0^{-1})(sI+D_0+f^*(s)\tilde{D}_1)^{-1}\tilde{D}_1.$$

The constant vector $\mathbf{c}$ is the probability vector that there is no arrival during the waiting period of a customer and there is no arrival when this customer is being served, which is the probability vector that a departure leaves an empty queue, i.e., $\mathbf{x}_0$.

(33) comes from (32). $\square$

We define the vector $\mathbf{W}_a(k,x)$ to be the conditional distributions of the waiting time of an arbitrary arrival of type $k$. $\mathbf{W}_a^*(k,s)$ is the L.S. transform of $\mathbf{W}_a(k,x)$.

**Theorem 5.2.** For the *MMAP/G/1* queue, we have, for $1 \le k \le K$,

$$\mathbf{W}_a^*(k,s) = \frac{s}{\lambda_k^*}\mathbf{y}_0(sI + D_0 + f^*(s)\tilde{D}_1)^{-1}D_k, \quad Re(s) \ge 0. \tag{35}$$

**Proof.** Without loss of generality, we assume that there are only two different types of arrivals, and we discuss the waiting times of arrivals of type 1. We define

$$H_i(u) = \int_0^\infty F(u+t)exp\{D_0t\}dtD_i, \quad -\infty < u < \infty, \quad i = 1,2. \tag{36}$$

Let $H_i^*(s)$ be the L.S.transform of $H_i(u)$, $i = 1,2$. It is easy to derive

$$H_i^*(s) = -f^*(s)(sI + D_0)^{-1}D_i, \quad Re(s) \ge 0, \quad i = 1,2. \tag{37}$$

Since we have two types of arrivals, there may be some arrivals of type 2 between two successive arrivals of type 1. The pattern of arrivals between two successive arrivals of type 1 is of the form $1,2,\cdots,2,1$. We denote by "*" the convolution of two functions. Let $\mathbf{L}_0(x) = \mathbf{W}_a(1,x)$,

$$\mathbf{L}_n(x) = \mathbf{L}_{n-1} * H_2(x) \quad n \ge 1, \quad x \ge 0.$$

Conditioning on the phase of the arrival process after an arrival of type 1 and the number of arrivals of type 2 until the next arrival of type 1, we have, for $x \ge 0$,

$$\mathbf{W}_a(1,x) = \sum_{n=0}^{\infty} \mathbf{L}_n * H_1(x). \tag{38}$$

For $n \ge 1$, let $\tilde{\mathbf{L}}(x)$ be the extension of $\mathbf{L}_n(x)$ from $(0,\infty)$ to $(-\infty,\infty)$. For $x < 0$, we have

$$\tilde{\mathbf{L}}_n(x) = \mathbf{c}_{n-1}(-D_0^{-1})exp\{-D_0x\}D_2,$$

where

$$\mathbf{c}_{n-1} = \int_0^\infty d\mathbf{L}_{n-1}(y)exp\{D_0y\}\int_0^\infty exp\{D_0u\}dF(u), \quad n \ge 1.$$

We denote by $\tilde{\mathbf{L}}_n^*(s)$ and $\mathbf{L}_n^*(s)$ the L.S.transform of $\tilde{\mathbf{L}}_n(x)$ and $\mathbf{L}_n(x)$, respectively, $n \geq 0$. Then we have

$$\tilde{\mathbf{L}}_n^*(s) = \mathbf{L}_{n-1}^*(s)H_2^*(s), \quad n \geq 1.$$

We also have

$$\begin{aligned}
\mathbf{L}_n^*(s) &= \tilde{\mathbf{L}}_n^*(s) - [\int_{-\infty}^{0-} + \int_{0-}^{0}]e^{-sx}d\tilde{\mathbf{L}}_n(x) \\
&= \mathbf{L}_{n-1}^*(s)H_2^*(s) + \mathbf{c}_{n-1}(-D_0^{-1})D_2 + \mathbf{c}_{n-1}(-D_0^{-1})(sI + D_0)^{-1}(-D_0)D_2 \\
&= \mathbf{L}_{n-1}^*(s)H_2^*(s) + s\mathbf{c}_{n-1}(-D_0^{-1})(sI + D_0)^{-1}D_2 \\
&\vdots \\
&= \mathbf{W}_a^*(1,s)[H_2^*(s)]^n + s\sum_{i=0}^{n-1}\mathbf{c}_i(-D_0^{-1})(sI + D_0)^{-1}D_2[H_2^*(s)]^{n-1-i}.
\end{aligned}$$

We extend the function $\mathbf{W}_a(1,x)$ from $[0,\infty)$ to $(-\infty,\infty)$, denoting it by $\mathbf{V}(1,x)$, or equivalently, we extend $\mathbf{L}_n * H_1(x)$ from $[0,\infty)$ to $(-\infty,\infty)$. We have

$$\int_{-\infty}^{\infty} e^{-sx}d\mathbf{V}(1,x)$$

$$= \sum_{n=0}^{\infty}\mathbf{W}_a(1,s)[H_2^*(s)]^n H_1^*(s) + s\sum_{n=1}^{\infty}\sum_{i=0}^{n-1}\mathbf{c}_i(sI + D_0)^{-1}D_2[H_2^*(s)]^{n-1-i}H_1^*(s)$$

$$= [\mathbf{W}_a(1,s) + s\mathbf{c}(-D_0^{-1})(sI + D_0)^{-1}D_2](I - H_2^*(s))^{-1}H_1^*(s)$$

(on the other hand)

$$= [\int_{-\infty}^{0-} + \int_{0-}^{0} + \int_{0}^{\infty}]e^{-sx}d\mathbf{V}(1,x)$$

$$= \mathbf{W}_a(1,s) + \sum_{n=1}^{\infty}[\mathbf{c}_{n-1}(-D_0^{-1})(sI + D_0)^{-1}D_0D_2 - \mathbf{c}_{n-1}(-D_0^{-1})D_2]$$

$$= \mathbf{W}_a(1,s) - s\mathbf{c}(-D_0^{-1})(sI + D_0)^{-1}D_2,$$

where $\mathbf{c} = \sum_{n=0}^{\infty}\mathbf{c}_n$. Then we have

$$\mathbf{W}_a^*(1,s)(I + (sI + D_0 + f^*(s)D_2)^{-1}f^*(s)D_1)$$

$$= s\mathbf{c}(-D_0^{-1})(sI + D_0 + f^*(s)D_2)^{-1}D_1.$$

Therefore,

$$\mathbf{W}_a^*(1,s) = s\mathbf{c}(-D_0^{-1})(sI + D_0 + f^*(s)(D_1 + D_2))^{-1}D_1.$$

Now, we only need to prove that

$$\mathbf{c} = \frac{\lambda^*}{\lambda_1^*}\mathbf{x}_0. \tag{39}$$

To this end, we suppose that the queue is in steady state and starts at 0 with an arrival of type 1. Let $\mathcal{F}_{n,j}$ be the event that the $n$th arrival leaves the system with an empty queue and the phase of the arrival process is $j$. $I_{\mathcal{A}}$ is the indicator of the event $\mathcal{A}$. Let $\eta_0 = 0$ and let

$$\eta_n = \min\{k : \ k > \eta_{n-1} \text{ and the } k\text{th arrival is of type 1}\}, \quad n \geq 1.$$

Let

$$\omega(n) = \max\{k : \ \eta_k \leq n\}, \quad n \geq 0.$$

Then we have

$$\begin{aligned}
\mathbf{x}_0 &= \lim_{n\to\infty} \frac{\sum_{i=0}^{n}(I_{\mathcal{F}_{i,1}}, \cdots, I_{\mathcal{F}_{i,m}})}{n} \\
&= \lim_{n\to\infty} \frac{\omega(n)}{n}\Big[\frac{\sum_{i=0}^{\omega(n)-1}\sum_{l=\eta(i)}^{\eta(i+1)-1}(I_{\mathcal{F}_{l,1}}, \cdots, I_{\mathcal{F}_{l,m}})}{\omega(n)}\Big] \\
&= \frac{\lambda_1^*}{\lambda^*}\,\mathbf{E}\Big[\sum_{n=0}^{\eta(1)-1}(I_{\mathcal{F}_{n,1}}, \cdots, I_{\mathcal{F}_{n,m}})\Big] \\
&= \frac{\lambda_1^*}{\lambda^*}\mathbf{E}\Big[\sum_{n=0}^{\infty}(I_{\mathcal{F}_{n,1}\cap\{\eta(1)>n\}}, \cdots, (I_{\mathcal{F}_{n,m}\cap\{\eta(1)>n\}})\Big] \\
&= \frac{\lambda_1^*}{\lambda^*}\sum_{n=0}^{\infty}\mathbf{c}_n.
\end{aligned}$$

The last equality is true because $\mathbf{c}_n$ is the conditional probability that an arbitrary departure leaves an empty queue and this departure is the $n$th arrival of type 2 after the last arrival of type 1 (if $n = 0$, it is an arrival of type 1.) $\square$

## 6. $MMAP(K)/G_K/1$ Queues

The *MMAP(K)/*$G_K$*/1* queue is defined the same way as an *MMAP(K)/G/1* queue except that the service times of customers of different types are different. This means that an *MMAP(K)/*$G_K$*/1* queue has an *MMAP* input process with $K$ types of arrivals and customers of type $k$ have service time distribution $F_k(x)$ (L.S. transform $f_k^*(s)$), $1 \leq k \leq K$. We assume that the service order is FCFS. We use the same notation as we used in Sections 3 and 5. Here, the traffic intensity is defined as

$$\rho = \sum_{k=1}^{K} \lambda_k^* \mu^*(k), \tag{40}$$

where $\mu^*(k) = -(f_k^*(s))'|_{s=0}$, the mean service time of arrivals of type $k$, $1 \leq k \leq K$. We assume that $\rho < 1$ so that the queue is stable.

Since the input *MMAP(K)* consists of $K$ dependent *MAP*s, the $MMAP(K)/G_K/1$ queue is rather a general model. Many familiar queueing systems, such as the single sever queue with a superposition input process and the *MAP/G/1* queue, can be obtained as special cases of the $MMAP(K)/G_K/1$ queue. Results in Sections 3 (fundamental period) and 5 (waiting times) for the *MMAP(K)/G/1* queue can be generalized. Unfortunately, the extension of the standard analysis of *MAP/G/1* based on the queue length process (Section 4) is impossible, since one needs to keep track of the number of customers in the system, the phase of service, and the type of next service.

### *6.1. The Fundamental Period*

The study of the fundamental period of the $MMAP(K)/G_K/1$ queue is entirely similar to that of the *MMAP/G/1* queue. However, we assume that the customers served during a fundamental period are those which arrived during the fundamental period, except the first one. This additional assumption affects only the fundamental period, but does not make any difference to the busy period with only one customer at the beginning.

We define $\tilde{G}_{kk'}(\mathbf{n}, x)$ to be the matrix with elements the conditional probability that the first passage time from level **i+1** to level **i** occurs before time $x$, there are a total of $n_1 + \cdots + n_K$ customers with $n_1$ customers of type 1, $\cdots$, $n_K$ customers of type $K$ arriving and being served during that time, and the last arrival before level **i** is reached is of type $k'$, given that the last arrival before time 0+ is of type $k$.

**Theorem 6.1.** For the $MMAP(K)/G_K/1$ queue, the matrix $\Phi^*(x, \mathbf{z}, s)$ is given by

$$\Phi^*(x, \mathbf{z}, s) = e^{-sx} exp\{(T_0 + T_1 \tilde{G}^*(\mathbf{z}, s))x\}. \tag{41}$$

The matrix $\tilde{G}^*(\mathbf{z}, s)$ satisfies the equation

$$\tilde{G}^*(\mathbf{z}, s) = \int_0^\infty \begin{pmatrix} z_1 F_1(dx) I & & \\ & \ddots & \\ & & z_K F_K(dx) I \end{pmatrix} \Phi^*(x, \mathbf{z}, s). \tag{42}$$

**Proof.** The proof is the same as in the proof of Theorem 3.2, except that $F(x)$ in Theorem 3.2 should be replaced by $F_k(x)$ accordingly, $1 \le k \le K$. $\square$

Since the commutativity of the matrices $G$ and $T_0 + T_1 G$ is unknown, we can not use (13), (14) and (23) to compute $\tilde{\mu}_1$, $\tilde{\mu}_1(k)$ and $\hat{\mu}_1$. Fortunately, we do have the following simple formulas.

**Corollary 6.2.** For the $MMAP(K)/G_K/1$ queue, let $\bar{\mathbf{g}}$ be the stationary probability vector of $T_0 + T_1 G$. We have

$$\tilde{\mu}_1(k) = [I - (I - G)(\mathbf{e}\bar{\mathbf{g}} - T_0 - T_1 G)^{-1} T_1 - \begin{pmatrix} \mu^*(1)\mathbf{e} \\ \vdots \\ \mu^*(K)\mathbf{e} \end{pmatrix} \bar{\mathbf{g}} T_1]^{-1} \mathbf{e}(k), \tag{43}$$

for $1 \le k \le K$, and

$$\hat{\mu}_1 = [I - (I - G)(\mathbf{e}\bar{\mathbf{g}} - T_0 - T_1 G)^{-1} T_1 - \begin{pmatrix} \mu^*(1)\mathbf{e} \\ \vdots \\ \mu^*(K)\mathbf{e} \end{pmatrix} \bar{\mathbf{g}} T_1]^{-1} \begin{pmatrix} \mu^*(1)\mathbf{e} \\ \vdots \\ \mu^*(K)\mathbf{e} \end{pmatrix}, \tag{44}$$

provided that the inverse in (43) and (44) exists.

**Proof.** Similar to the proof of Theorem 3.1. $\square$

*6.2. Waiting Times*

We study the waiting times for an arbitrary arrival and an arbitrary arrival of type $k$, $1 \le k \le K$.

**Theorem 6.3.** For the $MMAP(K)/G_K/1$ queue, we have, for $Re(s) \ge 0$ and $-Re(s) > sp(D_0)$ (the eigenvalue with the largest real part of $D_0$),

$$\mathbf{W}_a^*(s) = -s\mathbf{x}_0 D_0^{-1}(sI + D_0 + \sum_{i=1}^{K} f_i^*(s) D_i)^{-1} \tilde{D}_1; \tag{45}$$

$$\mathbf{W}_a^*(k, s) = -s\frac{\lambda^*}{\lambda_k^*}\mathbf{x}_0 D_0^{-1}(sI + D_0 + \sum_{i=1}^{K} f_i^*(s) D_i)^{-1} D_k, \quad 1 \le k \le K. \tag{46}$$

**Proof.** The proof is similar to the proofs of Theorems 5.1 and 5.2. However, there are some differences due to the difference of service times of arrivals of different types. Hence, we give the following outline.

First, we prove (45). We define $\bar{\mathbf{W}}_k(x)$ to be the distribution of the waiting times of an arbitrary arrival of type $k$, conditioned on the phase of the arrival process right after the arrival epoch. $\bar{\mathbf{W}}_k^*(s)$ is the L.S.transform of $\bar{\mathbf{W}}_k(x)$. Then we have

$$\mathbf{W}_a^*(s) = \sum_{k=1}^{K} \bar{\mathbf{W}}_k^*(s).$$

Let

$$H_{kk'}(u) = \int_0^\infty F_k(t+u)exp\{D_0 t\}dt D_{k'}, \quad -\infty < u < \infty, \quad 1 \le k, k' \le K.$$

Similarly to (34), we have, for $1 \le k' \le K$,

$$\bar{\mathbf{W}}_{k'}(x) = \sum_{k=1}^{K} \bar{\mathbf{W}}_k * H_{kk'}(x),$$

which yields

$$\bar{\mathbf{W}}_{k'}^*(s) - s\sum_{k=1}^{K} \mathbf{c}(k)(-D_0^{-1})(sI+D_0)^{-1}D_{k'} = -[\sum_{k=1}^{K} \bar{\mathbf{W}}_k^*(s) f_k^*(s)](sI+D_0)^{-1}D_{k'}, \quad (47)$$

where

$$\mathbf{c}(k) = \int_0^\infty d\bar{\mathbf{W}}_k(y)exp\{D_0 y\}\int_0^\infty exp\{D_0 u\}dF_k(u), \quad 1 \le k \le K.$$

Let

$$\mathbf{U}^*(s) = \sum_{k=1}^{K} \bar{\mathbf{W}}_k^*(s) f_k^*(s).$$

With (46), we have

$$\mathbf{U}^*(s) = s(\sum_{k=1}^{K} \mathbf{c}(k))(-D_0^{-1})(sI + D_0 + \sum_{k=1}^{K} f_k^*(s)D_k)^{-1}(\sum_{k=1}^{K} f_k^*(s)D_k).$$

Summing on both sides of (47) with respect to $k$, we obtain (45) except that we need to prove $\sum_{k=1}^{K} \mathbf{c}(k) = \mathbf{x}_0$. But the same reasoning in the proof of Theorem 3.4 to the vector $\mathbf{c}$ is still true in this case.

For (46), the case when $K = 2$ is exactly the same as the case in Theorem 5.2. For $K \geq 3$, we need to consider more complicated patterns of arrivals between two successive arrivals of type 1, since we have multi-type of arrivals with different service times. For example, arrivals between two successive arrivals of type 1 can occur only in the pattern $(1, 2, \cdots, 2, 1)$ when $K = 2$, while we have $(1, k_1, \cdots, k_n, 1)$, $2 \leq k_1, \cdots, k_n \leq K$, for $K \geq 3$. So, instead of $\mathbf{L}_n(x)$, we have to define functions $\mathbf{L}_{1,k_1,\cdots,k_n}(x)$ for all possible patterns of $(k_1, \cdots, k_n)$. The derivation is similar to Theorem 5.2 and the generalization is straightforward. We omit the details. $\square$

For the vector $\mathbf{x}_0$, we have the following theorem.

**Theorem 6.4.** For the $MMAP(K)/G_K/1$ queue,

$$\lambda^* \mathbf{x}_0 D_0^{-1} \mathbf{e} = 1 - \rho. \tag{48}$$

**Proof.** We denote by $\eta(s)$ the eigenvalue with the largest real part of $D_0 + \sum_{k=1}^{K} f_k(s) D_k$. $\mathbf{u}(s)$ and $\mathbf{v}(s)$ are the corresponding left and right eigenvectors of $\eta(s)$ respectively. $\mathbf{u}(s)$ and $\mathbf{v}(s)$ can be chosen to be differentiable, $\mathbf{u}(s)\mathbf{v}(s) = 1$ and $\mathbf{u}(s)\mathbf{e} = 1$. It is obvious that $\lim_{s\to 0} \eta(s) = 0$, $\lim_{s\to 0} \mathbf{u}(s) = \theta$ and $\lim_{s\to 0} \mathbf{v}(s) = \mathbf{e}$. By a routine method (see Lemma 1.3.3 in Neuts [10]), we have

$$\begin{aligned}
\eta'(0) &= \lim_{s\to 0} \mathbf{u}(s)[D_0 + \sum_{k=1}^{K} f_k^*(s) D_k]' \mathbf{v}(s) \\
&= -\sum_{k=1}^{K} \mu^*(k) \lambda_k^* \\
&= -\rho.
\end{aligned}$$

Since $D$ is irreducible, $D_0 + \sum_{k=1}^{K} f_k^*(s) D_k$ is irreducible. Therefore, the largest eigenvalue of $D_0 + \sum_{k=1}^{K} f_k^*(s) D_k$ has algebraic multiplicity 1 and geometric multiplicity 1. By using the Jordan canonical form of $D_0 + \sum_{k=1}^{K} f_k^*(s) D_k$, we have

$$\lim_{s\to 0} (I + \frac{D_0 + \sum_{k=1}^{K} f_k^*(s) D_k}{s})^{-1} = \mathbf{e}\theta \frac{1}{(1 + \eta'(0))}.$$

Therefore, we have

$$\begin{aligned}
\mathbf{W}_a^*(0)\mathbf{e} &= -\lim_{s\to 0} \mathbf{x}_0 D_0^{-1} (I + \frac{D_0 + \sum_{k=1}^{K} f_k^*(s) D_k}{s})^{-1} \tilde{D}_1 \mathbf{e} \\
&= -\mathbf{x}_0 D_0^{-1} \mathbf{e} \lambda^* / (1 + \eta'(0)) \\
&= 1,
\end{aligned}$$

which yields (48). $\square$

Let $p_{kj,k'j'}$ be the probability that the embedded *MRP* at departures reaches level $\mathbf{0}$ for the first time, does so in state $(0, j')$ and the last arrival before the level $\mathbf{0}$ is reached is of type $k'$, given that the process started in phase $(0, j)$ and the last arrival before starting is of type $k$. $P$ is the matrix with elements $p_{kj,k'j'}$. Similar to Ramaswami [14], we have

$$P = -T_0^{-1} T_1 G. \tag{49}$$

Let $\mathbf{x}_0(k)$ be the probability that the last arrival before an arbitrary departure is of type $k$ and the queue length after the departure is zero, $1 \leq k \leq K$. Clearly, we have $\mathbf{x}_0 = \sum_{k=1}^{K} \mathbf{x}_0(k)$. Let $\tilde{\mathbf{x}}_0 = (\mathbf{x}_0(1), \cdots, \mathbf{x}_0(K))$. By Theorem 6.4, the vector $\tilde{\mathbf{x}}_0$ satisfies

$$\tilde{\mathbf{x}}_0 P = \tilde{\mathbf{x}}_0 \quad \text{and} \quad -\lambda^* \tilde{\mathbf{x}}_0 T_0^{-1} \mathbf{e} = 1 - \rho. \tag{50}$$

This gives a method for computing the vector $\mathbf{x}_0$. If we still denote by $\mathbf{y}_0$ the probability vector of seeing an empty queue at an arbitrary time, (32) is still true. Therefore, the method for the computation of $\mathbf{y}_0$ is given.

### *6.3. The $BMMAP(K)/G_K/1$ Queue*

We give a short discussion on the $BMMAP(K)/G_K/1$ queue where batch arrivals are allowed for the input process. The *BMMAP(K)* with batch arrivals can be defined with coefficient matrices $\{D_0, D_{\mathbf{n}}, \mathbf{n} \in \Omega\}$, where $\Omega$ is a finite set of nonnegative vectors of dimension $K$. $D_{\mathbf{n}}$ is the coefficient matrix for arrivals where each arrival corresponds to a group of customers, including $n_1$ customers of type 1, ..., $n_K$ customers of type $K$.

We now modify the $BMMAP(K)/G_K/1$ queue with batch arrivals to an ordinary $MMAP(K)/G_K/1$ queue by remarking the group arrivals with the transition matrix $D_{\mathbf{n}}$ as type $\mathbf{n}$. The service times for arrivals of type $\mathbf{n}$ in the modified queueing system are i.i.d.r.v.s with distribution function $F_1^{(n_1)} * \cdots * F_K^{(n_K)}(x)$, $\mathbf{n} \in \Omega$. The busy periods of the two queueing systems are the same. The least waiting time (FCFS) for customers in a group for the $BMMAP(K)/G_k/1$ queue is the same as the waiting time of an arbitrary arrival in the modified system. We omit the details.

## 7. Probabilistic Interpretations with the Time Reversal

So far, we have discussed the distributions of queue lengths and waiting times of customers of different types and at different times. As we mentioned in Section 1, those distributions are different for customers of different types. It is natural to inquire about the source of those differences and what factors influence those distributions. In this section, we use time reversal method to analyze the problem. What we found is that the phase of the arrival process plays an important role. Customers of different types have different views of the queueing system because the probabilities that customers of a certain type come from a certain phase may be different from others.

We define a new $MMAP^*$, associated with the $MMAP$, by coefficient matrices $\{\Lambda^{-1}D_0^T\Lambda,\ \ \Lambda^{-1}D_k^T\Lambda,\ 1 \le k \le K\}$, where $\Lambda = diag(\theta)$ and the superscript $T$ is transpose. The $MMAP$ and the $MMAP^*$ have the same arrival rate and the same stationary distribution vector. Let

$$\mathbf{d}(k) = \frac{\theta\Lambda^{-1}D_k^T\Lambda}{\lambda_k^*}, \quad 1 \le k \le K. \tag{51}$$

$d_i(k)$ is the probability that the $MMAP^*$ is in phase $i$ after an arrival of type $k$, so $d_i(k)$ is the probability that an arrival of type $k$ of the $MMAP$ is from phase $i$.

We define $\hat{W}_i(x) = W_i(x)/W_i(\infty)$, $1 \le i \le m$, and

$$\hat{\mathbf{W}}(x) = (\hat{W}_1(x), \cdots, \hat{W}_m(x)).$$

With (35), it can be proved that

$$\mathbf{W}_a(k,x)\mathbf{e} = \sum_{i=1}^{m} \hat{W}_i(x) d_i(k). \tag{52}$$

Then, for a *MAP/G/1* queue, the waiting times of arrivals of type $k$ are determined by the vector $\mathbf{d}(k)$ and conditional waiting times at an arbitrary time. When the conditional distributions of the waiting time at an arbitrary time are greatly different from one phase to another, waiting times of customers of different types may be significantly different if the weights $\{d_i(k)\}$ are different for different $k$.

From (28) and (29), we know that the same reasoning goes through for queue lengths.

For the relations between the queue expected to be seen by an arbitrary customer and the queue expected to be seen by an arbitrary customer of a special type, we have the following equations from (28), (29), (33) and (35), for the *MMAP/G/1* queue.

$$\mathbf{X}_a^*(z) \;=\; \sum_{k=1}^{K} \frac{\lambda_k^*}{\lambda^*}\mathbf{X}_a^*(k,z); \tag{53}$$

$$\mathbf{W}_a^*(s) \;=\; \sum_{k=1}^{K} \frac{\lambda_k^*}{\lambda^*}\mathbf{W}_a^*(k,s). \tag{54}$$

From the above discussions and equations, we know that the queue length (or the waiting time) at an arbitrary arrival is the average of the queue lengths (or the waiting times) at the arrivals of different types. This shows why in some cases the classical descriptors are not enough to characterize the queueing system.

**Acknowledgements**

The author wishes to thank Professor Marcel Neuts for his encouragement in this research and Dr. David Rauschenberg for proofreading of the paper. Also, the author thanks the referee for his valuable comments on this paper.

## REFERENCES

[1] Asmussen, S. and Ramaswami, V., Probabilistic interpretation of some duality results of the matrix paradigms in queueing theory, *Stochastic Models*, **6**, 715-734, 1990.

[2] Chandramouli, Y., Neuts, M.F. and Ramaswami, V., A queueing model with service interruptions with applications to meteor burst packet communication systems, *IEEE Transaction on Communication*, **Vol 37**, 1024-1030, 1989.

[3] Cinlar, E., Markov Renewal Theory, *Adv. Appl. Prob.*, **1**, 123-187, 1969.

[4] Cohen, J. W., *The Single Server Queue*, North-Holland Publishing Company, 1982.

[5] He, Qi-Ming and Neuts, M. F., Markov Arrival Processes with Marked Transitions, (submitted for publication.)

[6] Lindley, D. V., The theory of queues with a single server, *Proc. Cambridge Phil. Soc.*, **48**, 227-289, 1952.

[7] Lucantoni, D. M., New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, **7**, 1-46, 1991.

[8] Lucantoni, D. M., Meier-Hellstern, K. S., and Neuts, M. F., A single server queue with server vacations and a class of non-renewal arrival processes, *Adv. Appl. Prob.*, **22**, 676-705, 1990.

[9] Neuts, M. F., A versatile Markovian point process. *J.Appl. Prob.*, **16**, 764-79, 1979.

[10] Neuts, M. F., *Matrix-Geometric Solutions in Stochastic Models: An algorithmic Approach.* Baltimore: The Johns Hopkins University Press, 1981.

[11] Neuts, M. F., Generalizations of the Pollaczek-Khinchin integral equation in the theory of queues. *Adv. Appl. Prob.*, **18**, 952-90, 1986.

[12] Neuts, M. F., The fundamental period of the queue with Markov-modulated arrivals, *Probability, statistics and Mathematics: Papers in Honor of Samuel Karlin*, Academic Press, 187-200, 1989.

[13] Neuts, M. F., *Structured Stochastic Matrices of $M/G/1$ type and Their Applications.* New York: Marcel Dekker Inc., 1989.

[14] Ramaswami, V., The *N/G/1* queue and its detailed analysis, *Adv. Appl. Prob.*, **12**, 222-61, 1980.

[15] Ramaswami, V., A duality theorem for the matrix paradigms in queueing theory, *Stochastic Models*, **6**, 151-163, 1990.

[16] Ramaswami, V., From the matrix-geometric to the matrix-exponential, *Queueing Systems,*, **6**, 229-260, 1990.

[17] Stanford, D. A., Interdeparture-time distributions in the non-preemptive priority $\sum_i M_i/G_i/1$ queue, *Performance Evaluation*, **12**, 43-60, 1991.

[18] Stanford, D. A. and Fischer, W., The interdeparture time distribution for each class in the $\sum_i M_i/G_i/1$ queue, *Queueing Systems*, **4**, 179-191, 1989.

[19] Stanford, D. A. and Fischer, W., Characterising interdeparture times for bursty input streams in the queue with pooled renewal arrivals, *Stochastic Models*, **7**, 311-320, 1991.

## APPENDIX A

**The invertibility of the matrix $M = \mathbf{eg}(I - \mu^* \tilde{D}_1) - D - \tilde{D}_1 G + G\tilde{D}_1$**

**Proof.** Suppose that $M$ is singular. Then there exists a nonzero vector $\mathbf{u}$ satisfying $\mathbf{u}M = 0$.

If $\mathbf{ue} = 0$, we have

$$\mathbf{u}(D_0 + \tilde{D}_1 G + (I - G)\tilde{D}_1) = 0. \tag{55}$$

Multiplying $G$ on both sides of the above equation, we obtain

$$\mathbf{u}(D_0 + \tilde{D}_1 G)G + \mathbf{u}(I - G)(\tilde{D}_1 G + D_0) - \mathbf{u}(I - G)D_0 = 0. \tag{56}$$

With (55) and the commutativity of $G$ and $D_0 + \tilde{D}_1 G$, the above equation is simplified as

$$\mathbf{u}(I - G)D = 0. \tag{57}$$

If $\mathbf{u}(I - G) = 0$, then $\mathbf{u} = c\mathbf{g}$, where $c$ is a nonzero constant, which contradicts to $\mathbf{ue} = 0$.

If $\mathbf{u}(I - G) \neq 0$, then $\mathbf{u}(I - G) = c\theta$, where $c$ is a nonzero constant. Then $c\theta\mathbf{e} = 0$, which contradicts to $\theta\mathbf{e} = 1$.

If $\mathbf{ue} \neq 0$, we assume that $\mathbf{ue} = 1$. Then we have

$$\mathbf{g}(I - \mu^* \tilde{D}_1) = \mathbf{u}(D_0 + \tilde{D}_1 G + (I - G)\tilde{D}_1). \tag{58}$$

Multiplying $G$ on both sides of the above equation, we obtain

$$\mathbf{g}(I - \mu^* \tilde{D}_1 G) = \mathbf{g}(I - \mu^* \tilde{D}_1) - \mathbf{u}(I - G)D. \tag{59}$$

Since $\mathbf{g}(D_0 + \tilde{D}_1 G) = 0$, we have $\mathbf{g}\tilde{D}_1 G = -\mathbf{g}D_0$. (59) is simplified as

$$[\mathbf{g}\mu^* - \mathbf{u}(I - G)]D = 0. \tag{60}$$

If $\mathbf{g}\mu^* - \mathbf{u}(I - G) = 0$, then $\mu^*\mathbf{ge} = 0$, which contradicts to $\mathbf{ge} = 1$.

If $\mathbf{g}\mu^* - \mathbf{u}(I - G) \neq 0$, then $\mathbf{g}\mu^* - \mathbf{u}(I - G) = c\theta$. We have $c = \mu^*$ and

$$\mu^*(\mathbf{g} - \theta) + \mathbf{u}(I - G) = 0. \tag{61}$$

With (58) and (61), we have

$$\begin{aligned} 1 - \mu^* \mathbf{g}\tilde{D}_1\mathbf{e} &= \mathbf{u}(I - G)\tilde{D}_1\mathbf{e} \\ &= \mu^*(-\mathbf{g}\tilde{D}_1\mathbf{e} + \theta\tilde{D}_1\mathbf{e}). \end{aligned}$$

The last equation implies $1 = \mu^* \lambda^*$, which contradicts to $\rho < 1$.

Therefore, $M$ is invertible.

$\square$