Journal of the Operations Research Society of Japan
2003, Vol. 46, No. 3, 319-341

STATIONARY QUEUE LENGTH IN A FIFO SINGLE SERVER QUEUE WITH SERVICE INTERRUPTIONS AND MULTIPLE BATCH MARKOVIAN ARRIVAL STREAMS

Hiroyuki Masuyama Tetsuya Takine
Kyoto University

(Received September 2, 2002)

Abstract This paper considers a FIFO single-server queue with service interruptions and multiple batch Markovian arrival streams. The server state (on and off), the type of arriving customers and their batch size are assumed to be governed by a continuous-time Markov chain with finite states. To put it more concretely, the marginal process of the server state is a phase-type alternating Markov renewal process, the marginal arrival process is a batch marked Markovian arrival process, and they may be correlated. Further, service times of arriving customers are allowed to depend on both their arrival stream and the server state on arrival. For such a queue, we derive the vector joint generating function of the numbers of customers from respective arrival streams. Further assuming discrete phase-type batch size distributions, we establish a numerical algorithm to compute the joint queue length distribution at a random point in time. Finally, we show some numerical examples and examine the impact of system parameters on the queue length distribution.

Keywords: Queue, batch marked MAP, FIFO, service interruptions

1. Introduction

This paper considers a FIFO single-server queue with service interruptions. In such a queue, the state of the server changes *on* and *off* alternately. While in on-state, the server is available for service. On the other hand, while in off-state, the server dose not work even if customers are present in the system. Hereafter periods during which the server is in on-state (resp. off-state) are called on-periods (resp. off-periods).

Queues with service interruptions have many applications in the fields of manufacturing, computer and telecommunications systems, and many studies on those queues have been done for a few decades. A detailed survey on queues with service interruptions can be found in the introduction of the paper by Federgruen and Green [1]. They mainly discussed approximation methods for an M/G/1 queue with service interruptions, where on- and off-periods are generally distributed [1]. Further, assuming a phase-type on-period distribution, they established an exact algorithm to compute the steady-state queue length distribution [2].

Recently, more general queues with service interruptions have been studied. Sengupta [9] considered the model where on- and off-period distributions are general, customers arrive according to a Poisson process whose arrival rate depends on the server state, and service times are generally distributed, depending on the server state upon arrival. He showed that the amount of unfinished work in such a queue is closely related to the waiting time in a special GI/G/1 queue. Also, Takine and Sengupta [12] considered a single-server queue with

service interruptions, where both the server state and arrival processes are governed by a finite-state Markov chain. Namely, the marginal processes of the server state and customer arrivals form an alternating phase-type Markov renewal process and a MAP (Markovian arrival process) [5], respectively, and they may be dependent. For this queue, they obtained the steady-state queue length distribution. The crucial assumption posed in [12] was i.i.d. (independent and identically distributed) service times.

This paper considers an extension of the results in [12] to allowing multiple batch Markovian arrival streams. Thus, the marginal arrival process follows a batch marked MAP [3, 4, 6]. Further, service times of customers can depends on both their arrival stream and the server state on arrival. As stated in [12], such a queue cannot be analyzed by the conventional M/G/1 paradigm [8]. To analyze the extended model, therefore, we use a new approach developed in [7, 13–15], which is based on the invariant relationship of the joint queue length distributions at random points in time and at departures [14]. We then derive the vector joint generating function of the numbers of customers from respective arrival streams. Further assuming discrete phase-type batch size distributions as in [7], we provide a computational algorithm for the steady-state joint queue length distribution. We also show some numerical examples and examine the impact of system parameters on the queue length distribution.

The rest of this paper is divided into four sections. In section 2, the mathematical model is described. Section 3 briefly discusses the sojourn time distribution. In section 4, we first derive a general formula for the joint queue length distribution, and assuming discrete phase-type batch size distributions, we show recursive formulas to compute the joint queue length distribution. Finally, in section 5, we show some numerical examples. Throughout the paper, matrices and vectors are denoted by bold capital letters and bold small letters, respectively, and the empty sum is defined as zero.

2. Model

We consider a FIFO single-server queue with service interruptions. The state of the server changes on and off alternately, and while the server is being on, customers are served successively. On the other hand, services of customers stop temporally while the server is being off, and interrupted services are restarted in a preemptive-resume manner when the server becomes on again. In what follows, we call the process of the server state the on-off process.

We assume that both the on-off and arrival processes are governed by an underlying finite-state Markov chain that is assumed to be irreducible. Let $\mathcal{M} = \{1, \ldots, M\}$ denote the state space of the underlying Markov chain, where $M \geq 2$. It stays in state i ($i \in \mathcal{M}$) for an exponential interval of time with mean μ_i^{-1}, and when the sojourn time in state i has elapsed, the underlying Markov chain changes its state to state j with probability $\sigma_{i,j}$ ($j \in \mathcal{M}$), where

$$\sum_{j \in \mathcal{M}} \sigma_{i,j} = 1.$$

The on-off process of the server is defined in the following way. The state space $\mathcal{M}$ is divided into two disjoint sub-spaces, $\mathcal{M}_{\rm on} = \{1, \ldots, M_{\rm on}\}$ and $\mathcal{M}_{\rm off} = \{M_{\rm on}+1, \ldots, M_{\rm on}+M_{\rm off}\}$, where $M_{\rm on} \geq 1$, $M_{\rm off} \geq 1$ and $M_{\rm on} + M_{\rm off} = M$. The server is assumed to be on (resp. off) while the underlying Markov chain is being in state $i \in \mathcal{M}_{\rm on}$ (resp. $i \in \mathcal{M}_{\rm off}$). Thus the on-off process forms a phase-type alternating Markov renewal process.

Next we describe the arrival process of customers. We assume that there are K ($K \geq 1$) arrival streams. Let $\mathcal{K}$ denote a set of class indices, i.e., $\mathcal{K} = \{1, \ldots, K\}$. Customers arriving

from the kth ($k \in \mathcal{K}$) arrival stream are called class k customers. Given a state transition of the underlying Markov chain from state i to state j ($i, j \in \mathcal{M}$), n ($n = 1, 2, \ldots$) customers of class k ($k \in \mathcal{K}$) arrive in batch with probability $\sigma_{k,i,j}(n)/\sigma_{i,j}$, where

$$\sum_{k \in \mathcal{K}} \sum_{n=1}^{\infty} \sigma_{k,i,j}(n) \leq \sigma_{i,j},$$

for all $i, j \in \mathcal{M}$. Note that customers in the same batch belong to the same class. For later use, we define $\sigma_{i,j}(0)$ ($i, j \in \mathcal{M}$) as

$$\sigma_{i,j}(0) = \sigma_{i,j} - \sum_{k \in \mathcal{K}} \sum_{n=1}^{\infty} \sigma_{k,i,j}(n).$$

Note that $\sigma_{i,j}(0)/\sigma_{i,j}$ represents the conditional probability of no arrivals given that a state transition from state i to state j happens. Without loss of generality, we assume $\sigma_{i,i}(0) = 0$ for all i ($i \in \mathcal{M}$).

In terms of service times, class k customers are further classified into two sub-classes based on the server state on arrival. We call class k customers arriving in on-periods (resp. off-periods) class k-on (resp. k-off) customers. Service times of class k-on (resp. k-off) customers are assumed to be i.i.d. according to a distribution function $H_{k,\mathrm{on}}(x)$ (resp. $H_{k,\mathrm{off}}(x)$) with finite mean $h_{k,\mathrm{on}}$ (resp. $h_{k,\mathrm{off}}$).

In the rest of this paper, we impose two assumptions on $\sigma_{k,i,j}(n)$. For each k ($k \in \mathcal{K}$), there exists at least one triad (i, j, n) ($i, j \in \mathcal{M},\ n = 1, 2, \ldots$) such that $\sigma_{k,i,j}(n) > 0$. Thus arrivals of class k customers are certain. Further $\sigma_{k,i,j}(n) = 0$ ($k \in \mathcal{K}$) if $i \in \mathcal{M}_{\mathrm{on}}$ and $j \in \mathcal{M}_{\mathrm{off}}$ or if $i \in \mathcal{M}_{\mathrm{off}}$ and $j \in \mathcal{M}_{\mathrm{on}}$. Thus arrivals of customers and changes of the server state never happen simultaneously.

We now introduce some notations. Let $\boldsymbol{C}$ denote an $M \times M$ matrix whose (i, j)th ($i, j \in \mathcal{M}$) element $C_{i,j}$ is given by

$$C_{i,j} = \begin{cases} -\mu_i, & \text{if } i = j, \\ \sigma_{i,j}(0)\mu_i, & \text{otherwise.} \end{cases}$$

For each $k \in \mathcal{K}$, let $\boldsymbol{D}_k(n)$ ($n = 1, 2, \ldots$) denote an $M \times M$ matrix whose (i, j)th ($i, j \in \mathcal{M}$) element $D_{k,i,j}(n)$ is given by

$$D_{k,i,j}(n) = \begin{cases} \sigma_{k,i,j}(n)\mu_i, & \text{if } i, j \in \mathcal{M}_{\mathrm{on}} \text{ or } i, j \in \mathcal{M}_{\mathrm{off}}, \\ 0, & \text{otherwise.} \end{cases}$$

Then the on-off and arrival processes are characterized by $\boldsymbol{C}$ and $\boldsymbol{D}_k(n)$ ($k \in \mathcal{K}, n = 1, 2, \ldots$). Note here that $\boldsymbol{C}$ and $\boldsymbol{D}_k(n)$ have the following structure:

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_{\mathrm{on}} & \boldsymbol{E}_{\mathrm{on,off}} \\ \boldsymbol{E}_{\mathrm{off,on}} & \boldsymbol{C}_{\mathrm{off}} \end{bmatrix}, \qquad \boldsymbol{D}_k(n) = \begin{bmatrix} \boldsymbol{D}_{k,\mathrm{on}}(n) & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{D}_{k,\mathrm{off}}(n) \end{bmatrix},$$

where $\boldsymbol{C}_{\mathrm{on}}$ and $\boldsymbol{D}_{k,\mathrm{on}}(n)$ are $M_{\mathrm{on}} \times M_{\mathrm{on}}$ matrices, $\boldsymbol{C}_{\mathrm{off}}$ and $\boldsymbol{D}_{k,\mathrm{off}}(n)$ are $M_{\mathrm{off}} \times M_{\mathrm{off}}$ matrices, and $\boldsymbol{E}_{\mathrm{on,off}}$ and $\boldsymbol{E}_{\mathrm{off,on}}$ are $M_{\mathrm{on}} \times M_{\mathrm{off}}$ and $M_{\mathrm{off}} \times M_{\mathrm{on}}$ matrices, respectively.

We define $\boldsymbol{D}_{k,\xi}$ ($k \in \mathcal{K}$, $\xi = \mathrm{on}, \mathrm{off}$) and $\overline{\boldsymbol{D}}_{\xi}$ ($\xi = \mathrm{on}, \mathrm{off}$) as

$$\boldsymbol{D}_{k,\xi} = \sum_{n=1}^{\infty} \boldsymbol{D}_{k,\xi}(n), \qquad \overline{\boldsymbol{D}}_{\xi} = \sum_{k \in \mathcal{K}} \boldsymbol{D}_{k,\xi},$$

respectively. We also define $\boldsymbol{D}$ as

$$\boldsymbol{D} = \begin{bmatrix} \overline{\boldsymbol{D}}_{\rm on} & \boldsymbol{O} \\ \boldsymbol{O} & \overline{\boldsymbol{D}}_{\rm off} \end{bmatrix}.$$

Note that the infinitesimal generator of the underlying Markov chain is given by $\boldsymbol{C}+\boldsymbol{D}$. Note also that $(\boldsymbol{C}+\boldsymbol{D})\boldsymbol{e} = \boldsymbol{0}$, where $\boldsymbol{e}$ denotes a column vector whose elements are all equal to one. We denote the stationary probability vector of the underlying Markov chain by $\boldsymbol{\pi}$. Because of the finite state space $\mathcal{M}$ and the irreducibility of the underlying Markov chain, $\boldsymbol{\pi}$ is uniquely determined so as to satisfy $\boldsymbol{\pi}(\boldsymbol{C}+\boldsymbol{D}) = \boldsymbol{0}$ and $\boldsymbol{\pi}\boldsymbol{e} = 1$. Let $\boldsymbol{\pi}_{\rm on}$ (resp. $\boldsymbol{\pi}_{\rm off}$) denote a $1 \times M_{\rm on}$ (resp. $1 \times M_{\rm off}$) vector representing the conditional stationary probability vector of the underlying Markov chain given that the server is on (resp. off). Note that $\boldsymbol{\pi}_{\rm on}$ and $\boldsymbol{\pi}_{\rm off}$ satisfy

$$\boldsymbol{\pi}_{\rm on}\left[\boldsymbol{C}_{\rm on} + \overline{\boldsymbol{D}}_{\rm on} + \boldsymbol{E}_{\rm on,off}\left[-\boldsymbol{C}_{\rm off} - \overline{\boldsymbol{D}}_{\rm off}\right]^{-1}\boldsymbol{E}_{\rm off,on}\right] = \boldsymbol{0}, \qquad \boldsymbol{\pi}_{\rm on}\boldsymbol{e} = 1,$$
$$\boldsymbol{\pi}_{\rm off}\left[\boldsymbol{C}_{\rm off} + \overline{\boldsymbol{D}}_{\rm off} + \boldsymbol{E}_{\rm off,on}\left[-\boldsymbol{C}_{\rm on} - \overline{\boldsymbol{D}}_{\rm on}\right]^{-1}\boldsymbol{E}_{\rm on,off}\right] = \boldsymbol{0}, \qquad \boldsymbol{\pi}_{\rm off}\boldsymbol{e} = 1,$$

respectively. Let $r_{\rm on}$ and $r_{\rm off}$ denote fractions of time being in on- and off-periods, respectively. We then have

$$r_{\rm on} = \frac{\overline{I}_{\rm on}}{\overline{I}_{\rm on} + \overline{I}_{\rm off}}, \qquad r_{\rm off} = \frac{\overline{I}_{\rm off}}{\overline{I}_{\rm on} + \overline{I}_{\rm off}},$$

where $\overline{I}_{\rm on}$ and $\overline{I}_{\rm off}$ denote the mean lengths of on- and off-periods, respectively, and they are given by

$$\overline{I}_{\rm on} = \frac{\boldsymbol{\pi}_{\rm off}\boldsymbol{E}_{\rm off,on}}{\boldsymbol{\pi}_{\rm off}\boldsymbol{E}_{\rm off,on}\boldsymbol{e}}\left[-\boldsymbol{C}_{\rm on} - \overline{\boldsymbol{D}}_{\rm on}\right]^{-1}\boldsymbol{e}, \qquad \overline{I}_{\rm off} = \frac{\boldsymbol{\pi}_{\rm on}\boldsymbol{E}_{\rm on,off}}{\boldsymbol{\pi}_{\rm on}\boldsymbol{E}_{\rm on,off}\boldsymbol{e}}\left[-\boldsymbol{C}_{\rm off} - \overline{\boldsymbol{D}}_{\rm off}\right]^{-1}\boldsymbol{e}.$$

Note here that $\boldsymbol{\pi}$, $\boldsymbol{\pi}_{\rm on}$ and $\boldsymbol{\pi}_{\rm off}$ are related by

$$\boldsymbol{\pi} = (r_{\rm on}\boldsymbol{\pi}_{\rm on}, r_{\rm off}\boldsymbol{\pi}_{\rm off}).$$

We denote the mean arrival rate of class k ($k \in \mathcal{K}$) customers during on- (resp. off-) periods by $\lambda_{k,\rm on}$ (resp. $\lambda_{k,\rm off}$):

$$\lambda_{k,\xi} = \boldsymbol{\pi}_{\xi}\sum_{n=1}^{\infty} n\boldsymbol{D}_{k,\xi}(n)\boldsymbol{e}, \qquad \xi = \text{on, off.}$$

Let $\lambda_k = r_{\rm on}\lambda_{k,\rm on} + r_{\rm off}\lambda_{k,\rm off}$ ($k \in \mathcal{K}$) denote the mean arrival rate of class k customers. We define ρ as the offered load, i.e.,

$$\rho = r_{\rm on}\sum_{k\in\mathcal{K}}\lambda_{k,\rm on}h_{k,\rm on} + r_{\rm off}\sum_{k\in\mathcal{K}}\lambda_{k,\rm off}h_{k,\rm off}.$$

Further, we define $\rho_{\rm on}$ as the conditional utilization factor given that the server is on, which is given by

$$\rho_{\rm on} = r_{\rm on}^{-1}\rho.$$

In the remainder of this paper, we assume that $\rho_{\rm on} < 1$, and the system is in steady state.

3. Sojourn Time

This section considers sojourn time. Because sojourn time is closely related to the amount of unfinished work in the system, we first discuss the latter.

Let V and S denote generic random variables representing the amount of unfinished work and the state of the underlying Markov chain, respectively, in steady state. With these, we define $\boldsymbol{v}(x)$ as a $1 \times M$ vector whose jth element represents $\Pr[V \le x, S = j]$. Further we define $\boldsymbol{v}_\xi(x)$ ($\xi =$ on, off) as a $1 \times M_\xi$ vector whose jth element represents $\Pr[V \le x, S = j \mid S \in \mathcal{M}_\xi]$. Note here that $\boldsymbol{v}(x)$ is given in terms of $\boldsymbol{v}_{\rm on}(x)$ and $\boldsymbol{v}_{\rm off}(x)$:

$$\boldsymbol{v}(x) = (r_{\rm on}\boldsymbol{v}_{\rm on}(x), r_{\rm off}\boldsymbol{v}_{\rm off}(x)).$$

Thus, we consider $\boldsymbol{v}_{\rm on}(x)$ and $\boldsymbol{v}_{\rm off}(x)$ below.

Let $\boldsymbol{J}_{\rm off,on}(x)$ denote an $M_{\rm off} \times M_{\rm on}$ matrix whose (i,j)th element represents the probability that the amount of work arriving during an off-period is not greater than x and the underlying Markov chain is in state $j \in \mathcal{M}_{\rm on}$ at the beginning of the next on-period, given that the off-period starts in state $i \in \mathcal{M}_{\rm off}$. We define $\overline{\boldsymbol{D}}_\xi(x)$ ($\xi =$ on, off) as

$$\overline{\boldsymbol{D}}_\xi(x) = \sum_{k\in\mathcal{K}} \sum_{n=1}^{\infty} \boldsymbol{D}_{k,\xi}(n) H_{k,\xi}^{(n)}(x),$$

where $H_{k,\xi}^{(1)}(x) = H_{k,\xi}(x)$ and $H_{k,\xi}^{(n)}(x)$ ($n = 2, 3, \ldots$) denotes the n-fold convolution of $H_{k,\xi}(x)$ with itself. According to [10], we define an $M_{\rm on} \times M_{\rm on}$ matrix $\boldsymbol{Q}_{\rm on}$ as an infinitesimal generator of the irreducible Markov chain obtained by observing the underlying Markov chain only when the system is idle in on-periods. Note that $\boldsymbol{Q}_{\rm on}$ satisfies

$$\boldsymbol{Q}_{\rm on} = \boldsymbol{C}_{\rm on} + \int_0^\infty d\overline{\boldsymbol{D}}_{\rm on}(x) \exp(\boldsymbol{Q}_{\rm on}x) + \boldsymbol{E}_{\rm on,off} \int_0^\infty d\boldsymbol{J}_{\rm off,on}(x) \exp(\boldsymbol{Q}_{\rm on}x).$$

Because $\rho_{\rm on} < 1$, $\boldsymbol{Q}_{\rm on}$ is uniquely determined by the above equation [10]. Let $\boldsymbol{\kappa}_{\rm on}$ denote a probability vector satisfying $\boldsymbol{\kappa}_{\rm on}\boldsymbol{Q}_{\rm on} = \boldsymbol{0}$. Then, $\boldsymbol{v}_{\rm on}(0)$ is given by [10]

$$\boldsymbol{v}_{\rm on}(0) = (1 - \rho_{\rm on})\boldsymbol{\kappa}_{\rm on}.$$

Further, the LST $\boldsymbol{v}_{\rm on}^*(s)$ of $\boldsymbol{v}_{\rm on}(x)$ satisfies [10]

$$\boldsymbol{v}_{\rm on}^*(s)\left[s\boldsymbol{I} + \boldsymbol{C}_{\rm on} + \overline{\boldsymbol{D}}_{\rm on}^*(s) + \boldsymbol{E}_{\rm on,off}\left[-\boldsymbol{C}_{\rm off} - \overline{\boldsymbol{D}}_{\rm off}^*(s)\right]^{-1}\boldsymbol{E}_{\rm off,on}\right] = s(1-\rho_{\rm on})\boldsymbol{\kappa}_{\rm on}, \quad (3.1)$$

where $\overline{\boldsymbol{D}}_\xi^*(s)$ ($\xi =$ on, off) denotes the LST of $\overline{\boldsymbol{D}}_\xi(x)$ and $\boldsymbol{I}(m)$ denotes an $m \times m$ identity matrix. We suppress the size m when it is clear from the context. As for the LST $\boldsymbol{v}_{\rm off}^*(s)$ of $\boldsymbol{v}_{\rm off}(x)$, using the same approach as in [12], we readily obtain

$$\boldsymbol{v}_{\rm off}^*(s) = \frac{\boldsymbol{v}_{\rm on}^*(s)\boldsymbol{E}_{\rm on,off}}{\boldsymbol{\pi}_{\rm on}\boldsymbol{E}_{\rm on,off}\boldsymbol{e}} \frac{[-\boldsymbol{C}_{\rm off} - \overline{\boldsymbol{D}}_{\rm off}^*(s)]^{-1}}{\overline{I}_{\rm off}}. \quad (3.2)$$

Next we analyze sojourn time. To do so, we first consider *completion time*, which is a time interval from the beginning of a service to its completion, including service interruptions. Let $T_c(u)$ denote a generic random variable representing the completion time of a service of u units. Note that the completion time $T_c(u)$ depends on the state of the underlying Markov chain at the beginning of the service, as well as the amount of the service. We assume that a service commences at time 0, and let S_t denote the state of the underlying

Markov chain at time t. We denote the number of class k ($k \in \mathcal{K}$) customers arriving in interval $(0,t]$ by $L_k(t)$. We then define $\boldsymbol{P}^{**}(\boldsymbol{z}, s \mid u)$ as an $M_{\rm on} \times M_{\rm on}$ matrix whose (i,j)th element represents

$$\mathrm{E}\left[\prod_{k\in\mathcal{K}} z_k^{L_k(T_c(u))} \exp(-sT_c(u))\, 1\{S_{T_c(u)} = j \in \mathcal{M}_{\rm on}\} \mid S_0 = i \in \mathcal{M}_{\rm on}\right],$$

where $\boldsymbol{z}$ denotes a $1 \times K$ complex vector $(z_1, \ldots, z_K)$ and $1\{\chi\}$ denotes an indicator function of event χ. Further, we define $\boldsymbol{D}^*_{k,\xi}(z_k)$ ($k \in \mathcal{K}$, $\xi = \text{on, off}$) as

$$\boldsymbol{D}^*_{k,\xi}(z_k) = \sum_{n=1}^{\infty} z_k^n \boldsymbol{D}_{k,\xi}(n). \tag{3.3}$$

Following an approach similar to [12], we obtain

$$\boldsymbol{P}^{**}(\boldsymbol{z}, s \mid u) = \exp\left[\boldsymbol{K}(\boldsymbol{z}, s)u\right],$$

where

$$\boldsymbol{K}(\boldsymbol{z}, s) = \boldsymbol{C}_{\rm on} + \sum_{k\in\mathcal{K}} \boldsymbol{D}^*_{k,\rm on}(z_k) - s\boldsymbol{I} + \boldsymbol{E}_{\rm on,off}\left[s\boldsymbol{I} - \boldsymbol{C}_{\rm off} - \sum_{k\in\mathcal{K}} \boldsymbol{D}^*_{k,\rm off}(z_k)\right]^{-1} \boldsymbol{E}_{\rm off,on}.$$

We are now ready to discuss sojourn time.

Let W_k ($k \in \mathcal{K}$) (resp. $W_{k,\xi}$ ($k \in \mathcal{K}$, $\xi = \text{on, off}$)) denote a generic random variable representing the sojourn time of a class k (resp. k-ξ) customer. Also let $W_{k,\xi}(n;m)$ ($k \in \mathcal{K}$, $\xi = \text{on, off}$, $n = 1, 2, \ldots$, $m = 1, \ldots, n$) denote a generic random variable representing the sojourn time of a randomly chosen class k-ξ customer who is a member of a batch of size n and the mth served customer among members in the same batch. For convenience, we assume that if $\lambda_{k,\xi} = 0$, $W_{k,\xi} = 0$, and if $\boldsymbol{D}_{k,\xi}(n) = \boldsymbol{O}$ for some n ($n \geq 1$), $W_{k,\xi}(n;m) = 0$ for all m ($m = 1, \ldots, n$). Further, let $w^*_k(s)$, $w^*_{k,\xi}(s)$ and $w^*_{k,\xi}(s \mid n;m)$ denote the LSTs of the distributions of W_k, $W_{k,\xi}$ and $W_{k,\xi}(n;m)$, respectively. Because a randomly chosen departing customer of class k ($k \in \mathcal{K}$) belongs to class k-ξ ($\xi = \text{on, off}$) with probability $r_\xi \lambda_{k,\xi}/\lambda_k$, we obtain

$$w^*_k(s) = \frac{r_{\rm on}\lambda_{k,\rm on}}{\lambda_k} w^*_{k,\rm on}(s) + \frac{r_{\rm off}\lambda_{k,\rm off}}{\lambda_k} w^*_{k,\rm off}(s).$$

Note here that

$$w^*_{k,\xi}(s) = \sum_{n=1}^{\infty} \frac{n\boldsymbol{\pi}_\xi \boldsymbol{D}_{k,\xi}(n)\boldsymbol{e}}{\lambda_{k,\xi}} \cdot \frac{1}{n} \sum_{m=1}^{n} w^*_{k,\xi}(s \mid n;m), \qquad k \in \mathcal{K},\ \xi = \text{on, off}, \tag{3.4}$$

if $\lambda_{k,\xi} > 0$, and otherwise $w^*_{k,\xi}(s) = 1$. Thus, in what follows, we consider $w^*_{k,\xi}(s \mid n;m)$ ($k \in \mathcal{K}$, $\xi = \text{on, off}$, $n = 1, 2, \ldots$, $m = 1, \ldots, n$).

Let $H_{k,\xi}(n;m)$ ($k \in \mathcal{K}$, $\xi = \text{on, off}$, $n = 1, 2, \ldots$, $m = 1, \ldots, n$) denote a generic random variable representing the service time of a randomly chosen class k-ξ customer who is a member of a batch of size n and the mth served customer among members of the same batch. Because $W_{k,\xi}(n;m) = W_{k,\xi}(n;1) + \sum_{l=2}^{m} T_c(H_{k,\xi}(n;l))$ for $\xi = \text{on, off}$, $n = 1, 2, \ldots$

and $m = 1, \ldots, n$, we obtain for $n = 1, 2, \ldots$ and $m = 1, \ldots, n$,

$$w^*_{k,\mathrm{on}}(s \mid n; m) = \frac{1}{\boldsymbol{\pi}_{\mathrm{on}} \boldsymbol{D}_{k,\mathrm{on}}(n) \boldsymbol{e}} \int_0^\infty d\boldsymbol{v}_{\mathrm{on}}(x) \boldsymbol{D}_{k,\mathrm{on}}(n) \exp\left[\boldsymbol{\Omega}(s)x\right] \cdot \left[\int_0^\infty dH_{k,\mathrm{on}}(y) \exp\left[\boldsymbol{\Omega}(s)y\right]\right]^m \boldsymbol{e}, \quad (3.5)$$

$$w^*_{k,\mathrm{off}}(s \mid n; m) = \frac{1}{\boldsymbol{\pi}_{\mathrm{off}} \boldsymbol{D}_{k,\mathrm{off}}(n) \boldsymbol{e}} \int_0^\infty d\boldsymbol{v}_{\mathrm{off}}(x) \boldsymbol{D}_{k,\mathrm{off}}(n) \left[s\boldsymbol{I} - \boldsymbol{C}_{\mathrm{off}} - \overline{\boldsymbol{D}}_{\mathrm{off}}\right]^{-1} \cdot \boldsymbol{E}_{\mathrm{off,on}} \exp\left[\boldsymbol{\Omega}(s)x\right] \left[\int_0^\infty dH_{k,\mathrm{off}}(y) \exp\left[\boldsymbol{\Omega}(s)y\right]\right]^m \boldsymbol{e}, \quad (3.6)$$

respectively, where $\boldsymbol{\Omega}(s) = \boldsymbol{K}(1, \ldots, 1, s)$. Note here that the (i, j)th $(i, j \in \mathcal{M}_{\mathrm{on}})$ element of $\exp[\boldsymbol{\Omega}(s)x]$ represents $\mathrm{E}[\exp(-sT_c(x))1\{S_{T_c(x)} = j\} \mid$ a service of x units starts at time 0 and $S_0 = i]$. Thus from (3.4) and (3.5), we obtain for $k \in \mathcal{K}$,

$$w^*_{k,\mathrm{on}}(s) = \frac{1}{\lambda_{k,\mathrm{on}}} \sum_{n=1}^\infty \int_0^\infty d\boldsymbol{v}_{\mathrm{on}}(x) \boldsymbol{D}_{k,\mathrm{on}}(n) \exp\left[\boldsymbol{\Omega}(s)x\right] \sum_{m=1}^n \left[\int_0^\infty dH_{k,\mathrm{on}}(y) \exp\left[\boldsymbol{\Omega}(s)y\right]\right]^m \boldsymbol{e},$$

if $\lambda_{k,\mathrm{on}} > 0$, and otherwise $w^*_{k,\mathrm{on}}(s) = 1$. Similarly, from (3.4) and (3.6), we obtain for $k \in \mathcal{K}$,

$$w^*_{k,\mathrm{off}}(s) = \frac{1}{\lambda_{k,\mathrm{off}}} \sum_{n=1}^\infty \int_0^\infty d\boldsymbol{v}_{\mathrm{off}}(x) \boldsymbol{D}_{k,\mathrm{off}}(n) \left[s\boldsymbol{I} - \boldsymbol{C}_{\mathrm{off}} - \overline{\boldsymbol{D}}_{\mathrm{off}}\right]^{-1} \boldsymbol{E}_{\mathrm{off,on}} \exp\left[\boldsymbol{\Omega}(s)x\right] \cdot \sum_{m=1}^n \left[\int_0^\infty dH_{k,\mathrm{off}}(y) \exp\left[\boldsymbol{\Omega}(s)y\right]\right]^m \boldsymbol{e},$$

if $\lambda_{k,\mathrm{off}} > 0$, and otherwise $w^*_{k,\mathrm{off}}(s) = 1$.

4. Joint Queue Length Distribution

This section considers the joint queue length distribution. Let N_k $(k \in \mathcal{K})$ denote a generic random variable representing the number of class k customers in the stationary system. We then define $\boldsymbol{p}(\boldsymbol{n})$ $(\boldsymbol{n} \in \mathcal{Z})$ as a $1 \times M$ vector whose jth element represents $\Pr[N_1 = n_1, \ldots, N_K = n_K, S = j]$, where $\boldsymbol{n}$ denotes a $1 \times K$ nonnegative integer vector $(n_1, \ldots, n_K)$ and $\mathcal{Z} = \{(n_1, \ldots, n_K)\,;\, n_k = 0, 1, \ldots \text{ for all } k \in \mathcal{K}\}$. Further, let $N_\nu^{(\mathrm{D}_k)}$ $(k, \nu \in \mathcal{K})$ and $S^{(\mathrm{D}_k)}$ $(k \in \mathcal{K})$ denote generic random variables that represent the number of class ν customers in the system and the state of the underlying Markov chain, respectively, immediately after departures of class k customers in steady state. We then define $\boldsymbol{q}_k(\boldsymbol{n})$ $(k \in \mathcal{K}, \boldsymbol{n} \in \mathcal{Z})$ as a $1 \times M$ vector whose jth element represents $\Pr[N_1^{(\mathrm{D}_k)} = n_1, \ldots, N_K^{(\mathrm{D}_k)} = n_K, S^{(\mathrm{D}_k)} = j]$. Applying Theorem 1 in [14] to our model, we have the following theorem.

Theorem 4.1 ([14]) *The $\boldsymbol{p}(\boldsymbol{n})$ $(\boldsymbol{n} \in \mathcal{Z})$ is recursively determined by*

$$\boldsymbol{p}(\boldsymbol{0}) = \sum_{k \in \mathcal{K}} \lambda_k \boldsymbol{q}_k(\boldsymbol{0})(-\boldsymbol{C})^{-1},$$

$$\boldsymbol{p}(\boldsymbol{n}) = \sum_{k \in \mathcal{K}} \left[\lambda_k \left(\boldsymbol{q}_k(\boldsymbol{n}) - \boldsymbol{q}_k(\boldsymbol{n} - \boldsymbol{e}_k)\right) + \sum_{m_k=1}^{n_k} \boldsymbol{p}(\boldsymbol{n} - m_k\boldsymbol{e}_k) \boldsymbol{D}_k(m_k)\right] (-\boldsymbol{C})^{-1}, \qquad \boldsymbol{n} \in \mathcal{Z}^+,$$

where $\mathcal{Z}^+ = \mathcal{Z} - \{\boldsymbol{0}\}$, $\boldsymbol{q}_k(\boldsymbol{n}) = \boldsymbol{0}$ for $\boldsymbol{n} \notin \mathcal{Z}$ and $\boldsymbol{e}_k$ $(k \in \mathcal{K})$ denotes the kth unit vector:

$$\boldsymbol{e}_k = (0, \ldots, 0, \underset{k\text{th}}{1}, 0, \ldots, 0).$$

Thus the $\boldsymbol{p}(\boldsymbol{n})$ is given in terms of the $\boldsymbol{q}_k(\boldsymbol{n})$. We then consider the $\boldsymbol{q}_k(\boldsymbol{n})$ in section 4.1. Further, in section 4.2, assuming discrete phase-type batch size distributions, we derive numerically feasible recursions for some quantities required in computing the $\boldsymbol{q}_k(\boldsymbol{n})$.

4.1. Joint queue length distribution immediately after departures

We define $\boldsymbol{q}_k^*(\boldsymbol{z})$ $(k \in \mathcal{K})$ as the vector generating function of the joint queue length distribution immediately after departures of class k customers.

$$\boldsymbol{q}_k^*(\boldsymbol{z}) = \sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{q}_k(\boldsymbol{n}), \qquad |z_k| \le 1 \text{ for all } k \in \mathcal{K}.$$

Let $N_\nu^{(\mathrm{D}_{k,\xi})}$ and $S^{(\mathrm{D}_{k,\xi})}$ $(k, \nu \in \mathcal{K}, \xi = \text{on, off})$ denote generic random variables that represent the number of class ν customers in the system and the state of the underlying Markov chain, respectively, immediately after departures of class k-ξ customers in steady state. We define $\boldsymbol{q}_{k,\xi}(\boldsymbol{n})$ $(k \in \mathcal{K},\ \xi = \text{on, off})$ as a $1 \times M_{\text{on}}$ vector whose jth element represents $\Pr[N_1^{(\mathrm{D}_{k,\xi})} = n_1, \ldots, N_K^{(\mathrm{D}_{k,\xi})} = n_K, S^{(\mathrm{D}_{k,\xi})} = j]$. We also define $\boldsymbol{q}_{k,\xi}^*(\boldsymbol{z})$ $(k \in \mathcal{K},\ \xi = \text{on, off})$ as

$$\boldsymbol{q}_{k,\xi}^*(\boldsymbol{z}) = \sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{q}_{k,\xi}(\boldsymbol{n}).$$

We then have

$$\boldsymbol{q}_k^*(\boldsymbol{z}) = \left(\frac{r_{\text{on}}\lambda_{k,\text{on}}}{\lambda_k} \boldsymbol{q}_{k,\text{on}}^*(\boldsymbol{z}) + \frac{r_{\text{off}}\lambda_{k,\text{off}}}{\lambda_k} \boldsymbol{q}_{k,\text{off}}^*(\boldsymbol{z}), 0, \ldots, 0 \right), \tag{4.1}$$

because all departures always occur in on-periods.

In the rest of this subsection, we derive $\boldsymbol{q}_{k,\text{on}}^*(\boldsymbol{z})$ and $\boldsymbol{q}_{k,\text{off}}^*(\boldsymbol{z})$. We call a randomly chosen class k-ξ $(k \in \mathcal{K},\ \xi = \text{on}, \text{off})$ customer who is a member of a batch of size n and the mth served customer among members in the same batch *the tagged customer*. Further we call the batch to which the tagged customer belongs *the tagged batch*. Let $N_\nu^{(\mathrm{D}_{k,\xi})}(n;m)$ and $S^{(\mathrm{D}_{k,\xi})}(n;m)$ $(k, \nu \in \mathcal{K},\ \xi = \text{on}, \text{off},\ n = 1, 2\ldots,\ m = 1, \ldots, n)$ denote generic random variables that represent the number of class ν customers in the system and the state of the underlying Markov chain, respectively, immediately after the departure of the tagged customer. We then define $\boldsymbol{q}_{k,\xi}^*(\boldsymbol{z} \mid n;m)$ $(k \in \mathcal{K},\ \xi = \text{on}, \text{off},\ n = 1, 2\ldots,\ m = 1, \ldots, n)$ as a $1 \times M_{\text{on}}$ vector whose jth element represents $\mathrm{E}\Big[\prod_{\nu\in\mathcal{K}} z_\nu^{N_\nu^{(\mathrm{D}_{k,\xi})}(n;m)} 1\{S^{(\mathrm{D}_{k,\xi})}(n;m) = j\}\Big]$. It is easy to see that $\boldsymbol{q}_{k,\xi}^*(\boldsymbol{z})$ can be written in terms of $\boldsymbol{q}_{k,\xi}^*(\boldsymbol{z} \mid n;m)$:

$$\boldsymbol{q}_{k,\xi}^*(\boldsymbol{z}) = \sum_{n=1}^{\infty} \frac{n\boldsymbol{\pi}_\xi \boldsymbol{D}_{k,\xi}(n)\boldsymbol{e}}{\lambda_{k,\xi}} \frac{1}{n} \sum_{m=1}^{n} \boldsymbol{q}_{k,\xi}^*(\boldsymbol{z} \mid n;m), \qquad k \in \mathcal{K},\ \xi = \text{on, off}, \tag{4.2}$$

if $\lambda_{k,\xi} > 0$, and otherwise $\boldsymbol{q}_{k,\xi}^*(\boldsymbol{z}) = \boldsymbol{0}$.

Note here that customers who contribute to $\boldsymbol{q}_{k,\text{on}}^*(\boldsymbol{z} \mid n;m)$ can be divided into three types: (i) customers arriving during the completion time of the total unfinished work immediately before the arrival of the tagged batch, (ii) customers arriving during an interval from the beginning of the first service of a member in the tagged batch to the completion of the service of the tagged customer, and (iii) $n - m$ customers who belong to the tagged batch and receive their services after the tagged customer. It then follows that for $k \in \mathcal{K}$, $n = 1, 2, \ldots$ and $m = 1, \ldots, n$,

$$\boldsymbol{q}_{k,\text{on}}^*(\boldsymbol{z} \mid n;m) = z_k^{n-m} \cdot \int_0^\infty \frac{d\boldsymbol{v}_{\text{on}}(x)\boldsymbol{D}_{k,\text{on}}(n)\boldsymbol{N}^*(\boldsymbol{z} \mid x)}{\boldsymbol{\pi}_{\text{on}}\boldsymbol{D}_{k,\text{on}}(n)\boldsymbol{e}} \cdot \left[\int_0^\infty dH_{k,\text{on}}(y)\boldsymbol{N}^*(\boldsymbol{z} \mid y)\right]^m, \tag{4.3}$$

where $\boldsymbol{N}^*(\boldsymbol{z} \mid x) = \boldsymbol{P}^{**}(\boldsymbol{z}, 0 \mid x)$, i.e.,

$$\boldsymbol{N}^*(\boldsymbol{z} \mid x) = \exp\left[\left\{ \boldsymbol{C}_{\text{on}} + \sum_{k\in\mathcal{K}} \boldsymbol{D}_{k,\text{on}}^*(z_k) + \boldsymbol{E}_{\text{on,off}} \left(-\boldsymbol{C}_{\text{off}} - \sum_{k\in\mathcal{K}} \boldsymbol{D}_{k,\text{off}}^*(z_k) \right)^{-1} \boldsymbol{E}_{\text{off,on}} \right\} x \right]. \tag{4.4}$$

Note that $\boldsymbol{N}^*(\boldsymbol{z} \mid x)$ denotes the matrix joint generating function for the numbers of arrivals in respective classes during the completion time of a service of x units. Similarly, we have for $k \in \mathcal{K}$, $n = 1, 2, \ldots$ and $m = 1, \ldots, n$,

$$\boldsymbol{q}^*_{k,\mathrm{off}}(\boldsymbol{z} \mid n; m) = z_k^{n-m} \cdot \int_0^\infty \frac{d\boldsymbol{v}_{\mathrm{off}}(x)\boldsymbol{D}_{k,\mathrm{off}}(n)}{\boldsymbol{\pi}_{\mathrm{off}}\boldsymbol{D}_{k,\mathrm{off}}(n)\boldsymbol{e}} \left[-\boldsymbol{C}_{\mathrm{off}} - \sum_{k\in\mathcal{K}} \boldsymbol{D}^*_{k,\mathrm{off}}(z_k)\right]^{-1} \boldsymbol{E}_{\mathrm{off,on}}\boldsymbol{N}^*(\boldsymbol{z} \mid x) \cdot \left[\int_0^\infty dH_{k,\mathrm{off}}(y)\boldsymbol{N}^*(\boldsymbol{z} \mid y)\right]^m. \quad (4.5)$$

Thus, from (4.2), (4.3) and (4.5), we obtain the following theorem.

Theorem 4.2 *$\boldsymbol{q}^*_{k,\mathrm{on}}(\boldsymbol{z})$ ($k \in \mathcal{K}$) is given by*

$$\boldsymbol{q}^*_{k,\mathrm{on}}(\boldsymbol{z}) = \frac{1}{\lambda_{k,\mathrm{on}}} \sum_{m=1}^\infty \sum_{l=0}^\infty z_k^l \int_0^\infty d\boldsymbol{v}_{\mathrm{on}}(x)\boldsymbol{D}_{k,\mathrm{on}}(m+l)\boldsymbol{N}^*(\boldsymbol{z} \mid x) \left[\int_0^\infty dH_{k,\mathrm{on}}(y)\boldsymbol{N}^*(\boldsymbol{z} \mid y)\right]^m,$$

if $\lambda_{k,\mathrm{on}} > 0$, and otherwise $\boldsymbol{q}^*_{k,\mathrm{on}}(\boldsymbol{z}) = \boldsymbol{0}$. On the other hand, $\boldsymbol{q}^*_{k,\mathrm{off}}(\boldsymbol{z})$ ($k \in \mathcal{K}$) is given by

$$\boldsymbol{q}^*_{k,\mathrm{off}}(\boldsymbol{z}) = \frac{1}{\lambda_{k,\mathrm{off}}} \sum_{m=1}^\infty \sum_{l=0}^\infty z_k^l \int_0^\infty d\boldsymbol{v}_{\mathrm{off}}(x)\boldsymbol{D}_{k,\mathrm{off}}(m+l) \left[-\boldsymbol{C}_{\mathrm{off}} - \sum_{k\in\mathcal{K}} \boldsymbol{D}^*_{k,\mathrm{off}}(z_k)\right]^{-1} \boldsymbol{E}_{\mathrm{off,on}} \cdot \boldsymbol{N}^*(\boldsymbol{z} \mid x) \left[\int_0^\infty dH_{k,\mathrm{off}}(y)\boldsymbol{N}^*(\boldsymbol{z} \mid y)\right]^m,$$

if $\lambda_{k,\mathrm{off}} > 0$, and otherwise $\boldsymbol{q}^*_{k,\mathrm{off}}(\boldsymbol{z}) = \boldsymbol{0}$.

4.2. Recursions for models with discrete phase-type batch sizes

In this subsection, we develop recursive formulas to compute the joint queue length distribution $\boldsymbol{q}_k(\boldsymbol{n})$ immediately after departures of class k customers under the following assumption.

Assumption 4.1 *Batch sizes of class k-ξ ($k \in \mathcal{K}$, ξ = on, off) are independent of the state of the underlying Markov chain and i.i.d. according to a discrete phase-type distribution with representation $(\boldsymbol{\alpha}_{k,\xi}, \boldsymbol{P}_{k,\xi})$, where $\boldsymbol{\alpha}_{k,\xi}$ denotes a $1 \times M_{k,\xi}$ probability vector and $\boldsymbol{P}_{k,\xi}$ denotes an $M_{k,\xi} \times M_{k,\xi}$ substochastic matrix.*

Under Assumption 4.1, $\boldsymbol{D}_{k,\xi}(n)$ ($k \in \mathcal{K}$, ξ = on, off) is given by

$$\boldsymbol{D}_{k,\xi}(n) = g_{k,\xi}(n)\boldsymbol{D}_{k,\xi}, \qquad n = 1, 2, \ldots,$$

where $g_{k,\xi}(n)$ denotes the probability mass function of the batch size of class k-ξ:

$$g_{k,\xi}(n) = \boldsymbol{\alpha}_{k,\xi}\boldsymbol{P}^{n-1}_{k,\xi}(\boldsymbol{I} - \boldsymbol{P}_{k,\xi})\boldsymbol{e}, \qquad n = 1, 2, \ldots.$$

Thus, Theorem 4.2 is reduced to:

Corollary 4.1 *Under Assumption 4.1, $\boldsymbol{q}^*_{k,\mathrm{on}}(\boldsymbol{z})$ ($k \in \mathcal{K}$) is given by*

$$\boldsymbol{q}^*_{k,\mathrm{on}}(\boldsymbol{z}) = \frac{1}{\lambda_{k,\mathrm{on}}} \int_0^\infty d\boldsymbol{v}_{\mathrm{on}}(x)\boldsymbol{D}_{k,\mathrm{on}}\boldsymbol{N}^*(\boldsymbol{z} \mid x) \cdot \left(\boldsymbol{\alpha}_{k,\mathrm{on}} \otimes \int_0^\infty dH_{k,\mathrm{on}}(y)\boldsymbol{N}^*(\boldsymbol{z} \mid y)\right) \left[\boldsymbol{I} - \boldsymbol{P}_{k,\mathrm{on}} \otimes \int_0^\infty dH_{k,\mathrm{on}}(y)\boldsymbol{N}^*(\boldsymbol{z} \mid y)\right]^{-1} \cdot \left[\left\{(\boldsymbol{I} - z_k\boldsymbol{P}_{k,\mathrm{on}})^{-1}(\boldsymbol{I} - \boldsymbol{P}_{k,\mathrm{on}})\boldsymbol{e}\right\} \otimes \boldsymbol{I}(M_{\mathrm{on}})\right], \quad (4.6)$$

if $\lambda_{k,\mathrm{on}} > 0$, *and otherwise* $\boldsymbol{q}^*_{k,\mathrm{on}}(\boldsymbol{z}) = \boldsymbol{0}$. *Similarly,* $\boldsymbol{q}^*_{k,\mathrm{off}}(\boldsymbol{z})$ $(k \in \mathcal{K})$ *is given by*

$$\boldsymbol{q}^*_{k,\mathrm{off}}(\boldsymbol{z}) = \frac{1}{\lambda_{k,\mathrm{off}}} \int_0^\infty d\boldsymbol{v}_{\mathrm{off}}(x) \boldsymbol{D}_{k,\mathrm{off}} \left[-\boldsymbol{C}_{\mathrm{off}} - \sum_{k\in\mathcal{K}} \boldsymbol{D}^*_{k,\mathrm{off}}(z_k) \right]^{-1} \boldsymbol{E}_{\mathrm{off,on}} \boldsymbol{N}^*(\boldsymbol{z} \mid x)$$
$$\cdot \left(\boldsymbol{\alpha}_{k,\mathrm{off}} \otimes \int_0^\infty dH_{k,\mathrm{off}}(y) \boldsymbol{N}^*(\boldsymbol{z} \mid y) \right) \left[\boldsymbol{I} - \boldsymbol{P}_{k,\mathrm{off}} \otimes \int_0^\infty dH_{k,\mathrm{off}}(y) \boldsymbol{N}^*(\boldsymbol{z} \mid y) \right]^{-1}$$
$$\cdot \left[\left\{ (\boldsymbol{I} - z_k \boldsymbol{P}_{k,\mathrm{off}})^{-1} (\boldsymbol{I} - \boldsymbol{P}_{k,\mathrm{off}}) \boldsymbol{e} \right\} \otimes \boldsymbol{I}(M_{\mathrm{on}}) \right], \tag{4.7}$$

if $\lambda_{k,\mathrm{off}} > 0$, *and otherwise* $\boldsymbol{q}^*_{k,\mathrm{off}}(\boldsymbol{z}) = \boldsymbol{0}$.

This corollary can be obtained in the same way as Lemma IV.1 in [7], and therefore we omit the proof.

We define $\boldsymbol{v}_{k,\mathrm{on}}(\boldsymbol{n})$ and $\boldsymbol{v}_{k,\mathrm{off}}(\boldsymbol{n})$ $(k \in \mathcal{K}, \boldsymbol{n} \in \mathcal{Z})$ as $1 \times M_{\mathrm{on}}$ vectors satisfying

$$\sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{v}_{k,\mathrm{on}}(\boldsymbol{n}) = \int_0^\infty d\boldsymbol{v}_{\mathrm{on}}(x) \boldsymbol{D}_{k,\mathrm{on}} \boldsymbol{N}^*(\boldsymbol{z} \mid x), \tag{4.8}$$

$$\sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{v}_{k,\mathrm{off}}(\boldsymbol{n}) = \int_0^\infty d\boldsymbol{v}_{\mathrm{off}}(x) \boldsymbol{D}_{k,\mathrm{off}} \left[-\boldsymbol{C}_{\mathrm{off}} - \sum_{k\in\mathcal{K}} \boldsymbol{D}^*_{k,\mathrm{off}}(z_k) \right]^{-1} \boldsymbol{E}_{\mathrm{off,on}} \boldsymbol{N}^*(\boldsymbol{z} \mid x), \tag{4.9}$$

respectively. We also define $\boldsymbol{A}_{k,\xi}(\boldsymbol{n})$ and $\boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{n})$ $(k \in \mathcal{K}, \xi = \mathrm{on}, \mathrm{off}, \boldsymbol{n} \in \mathcal{Z})$ as $M_{\mathrm{on}} \times M_{\mathrm{on}}$ and $M_{k,\xi} M_{\mathrm{on}} \times M_{k,\xi} M_{\mathrm{on}}$ matrices satisfying

$$\sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{A}_{k,\xi}(\boldsymbol{n}) = \int_0^\infty dH_{k,\xi}(y) \boldsymbol{N}^*(\boldsymbol{z} \mid y), \tag{4.10}$$

$$\sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{n}) = \left[\boldsymbol{I} - \boldsymbol{P}_{k,\xi} \otimes \int_0^\infty dH_{k,\xi}(y) \boldsymbol{N}^*(\boldsymbol{z} \mid y) \right]^{-1},$$

respectively. Then $\boldsymbol{q}^*_{k,\xi}(\boldsymbol{z})$ $(k \in \mathcal{K}, \xi = \mathrm{on}, \mathrm{off})$ in (4.6) and (4.7) are rewritten to be

$$\boldsymbol{q}^*_{k,\xi}(\boldsymbol{z}) = \frac{1}{\lambda_{k,\xi}} \sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \sum_{m=0}^{n_k} \sum_{\substack{\boldsymbol{n}_1,\boldsymbol{n}_2,\boldsymbol{n}_3\in\mathcal{Z} \\ \boldsymbol{n}_1+\boldsymbol{n}_2+\boldsymbol{n}_3 \\ =\boldsymbol{n}-m\boldsymbol{e}_k}} \boldsymbol{v}_{k,\xi}(\boldsymbol{n}_1) [\boldsymbol{\alpha}_{k,\xi} \otimes \boldsymbol{A}_{k,\xi}(\boldsymbol{n}_2)] \boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{n}_3)$$
$$\cdot \left[\left\{ \boldsymbol{P}^m_{k,\xi} (\boldsymbol{I} - \boldsymbol{P}_{k,\xi}) \boldsymbol{e} \right\} \otimes \boldsymbol{I}(M_{\mathrm{on}}) \right], \tag{4.11}$$

if $\lambda_{k,\xi} > 0$, and otherwise $\boldsymbol{q}^*_{k,\xi}(\boldsymbol{z}) = \boldsymbol{0}$. Comparing coefficient vectors of $z_1 \cdots z_K$ on both sides of (4.1) and (4.11), respectively, we obtain the following result.

Theorem 4.3 *Under Assumption 4.1, the* $\boldsymbol{q}_k(\boldsymbol{n})$ *is given by*

$$\boldsymbol{q}_k(\boldsymbol{n}) = \left(\frac{r_{\mathrm{on}} \lambda_{k,\mathrm{on}}}{\lambda_k} \boldsymbol{q}_{k,\mathrm{on}}(\boldsymbol{n}) + \frac{r_{\mathrm{off}} \lambda_{k,\mathrm{off}}}{\lambda_k} \boldsymbol{q}_{k,\mathrm{off}}(\boldsymbol{n}), 0, \ldots, 0 \right), \qquad k \in \mathcal{K},\ \boldsymbol{n} \in \mathcal{Z},$$

where the $\boldsymbol{q}_{k,\xi}(\boldsymbol{n})$ $(k \in \mathcal{K},\ \xi = on,\ off,\ \boldsymbol{n} \in \mathcal{Z})$ *is given by*

$$\boldsymbol{q}_{k,\xi}(\boldsymbol{n}) = \frac{1}{\lambda_{k,\xi}} \sum_{m=0}^{n_k} \sum_{\substack{\boldsymbol{n}_1,\boldsymbol{n}_2,\boldsymbol{n}_3\in\mathcal{Z} \\ \boldsymbol{n}_1+\boldsymbol{n}_2+\boldsymbol{n}_3 \\ =\boldsymbol{n}-m\boldsymbol{e}_k}} \boldsymbol{v}_{k,\xi}(\boldsymbol{n}_1) [\boldsymbol{\alpha}_{k,\xi} \otimes \boldsymbol{A}_{k,\xi}(\boldsymbol{n}_2)] \boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{n}_3)$$
$$\cdot \left[\left\{ \boldsymbol{P}^m_{k,\xi} (\boldsymbol{I} - \boldsymbol{P}_{k,\xi}) \boldsymbol{e} \right\} \otimes \boldsymbol{I}(M_{\mathrm{on}}) \right],$$

if $\lambda_{k,\xi} > 0$, *and otherwise* $\boldsymbol{q}_{k,\xi}(\boldsymbol{n}) = \boldsymbol{0}$ *for all* $\boldsymbol{n} \in \mathcal{Z}$.

Theorem 4.3 implies that the computation of the $\boldsymbol{q}_k(\boldsymbol{n})$ is reduced to those of the $\boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{n})$, $\boldsymbol{A}_{k,\xi}(\boldsymbol{n})$ and $\boldsymbol{v}_{k,\xi}(\boldsymbol{n})$ ($\xi = \text{on, off}$). Note here that the $\boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{n})$ is given in terms of the $\boldsymbol{A}_{k,\xi}(\boldsymbol{n})$.

Lemma 4.1 ([7]) *The $\boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{n})$ ($k \in \mathcal{K}$, $\xi = on, off$, $\boldsymbol{n} \in \mathcal{Z}$) is determined by the following recursion:*

$$\begin{aligned}
\boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{0}) &= [\boldsymbol{I} - \boldsymbol{P}_{k,\xi} \otimes \boldsymbol{A}_{k,\xi}(\boldsymbol{0})]^{-1}, \\
\boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{n}) &= \sum_{\substack{\boldsymbol{0} \le \boldsymbol{l} \le \boldsymbol{n} \\ \boldsymbol{l} \ne \boldsymbol{0}}} \boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{n} - \boldsymbol{l}) \, [\boldsymbol{P}_{k,\xi} \otimes \boldsymbol{A}_{k,\xi}(\boldsymbol{l})] \, \boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{0}), \qquad \boldsymbol{n} \in \mathcal{Z}^+.
\end{aligned}$$

The rest of this subsection therefore discusses the computations of the $\boldsymbol{A}_{k,\xi}(\boldsymbol{n})$ and the $\boldsymbol{v}_{k,\xi}(\boldsymbol{n})$.

We first consider the $\boldsymbol{A}_{k,\xi}(\boldsymbol{n})$. Let $\boldsymbol{F}_m(\boldsymbol{n})$ ($m = 0, 1, \ldots$, $\boldsymbol{n} \in \mathcal{Z}$) denote an $M_{\text{on}} \times M_{\text{on}}$ matrix that satisfies

$$\begin{aligned}
\sum_{\boldsymbol{n} \in \mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{F}_m(\boldsymbol{n}) = \Bigg[\boldsymbol{I} + \theta_{\text{on}}^{-1} \Bigg\{ \boldsymbol{C}_{\text{on}} + \sum_{k \in \mathcal{K}} \boldsymbol{D}^*_{k,\text{on}}(z_k) \\
+ \boldsymbol{E}_{\text{on,off}} \left(-\boldsymbol{C}_{\text{off}} - \sum_{k \in \mathcal{K}} \boldsymbol{D}^*_{k,\text{off}}(z_k) \right)^{-1} \boldsymbol{E}_{\text{off,on}} \Bigg\} \Bigg]^m, \qquad (4.12)
\end{aligned}$$

where $\theta_{\text{on}} = \max_{j \in \mathcal{M}_{\text{on}}} |[\boldsymbol{C}_{\text{on}}]_{j,j}|$.

Lemma 4.2 *The $\boldsymbol{F}_m(\boldsymbol{n})$ is recursively determined by*

$$\boldsymbol{F}_0(\boldsymbol{n}) = \begin{cases} \boldsymbol{I}, & \text{if } \boldsymbol{n} = \boldsymbol{0}, \\ \boldsymbol{O}, & \text{otherwise}, \end{cases} \qquad (4.13)$$

and for $m = 1, 2, \ldots$,

$$\begin{aligned}
\boldsymbol{F}_m(\boldsymbol{n}) = \boldsymbol{F}_{m-1}(\boldsymbol{n})(\boldsymbol{I} + \theta_{\text{on}}^{-1} \boldsymbol{C}_{\text{on}}) + \theta_{\text{on}}^{-1} \sum_{k \in \mathcal{K}} \sum_{l_k = 1}^{n_k} \boldsymbol{F}_{m-1}(\boldsymbol{n} - l_k \boldsymbol{e}_k) \boldsymbol{D}_{k,\text{on}}(l_k) \\
+ \theta_{\text{on}}^{-1} \left[\sum_{\boldsymbol{0} \le \boldsymbol{l} \le \boldsymbol{n}} \boldsymbol{F}_{m-1}(\boldsymbol{n} - \boldsymbol{l}) \boldsymbol{E}_{\text{on,off}} \boldsymbol{N}_{\text{off}}(\boldsymbol{l}) \right] \boldsymbol{E}_{\text{off,on}}, \qquad \boldsymbol{n} \in \mathcal{Z}, \qquad (4.14)
\end{aligned}$$

where $M_{\text{off}} \times M_{\text{off}}$ matrices $\boldsymbol{N}_{\text{off}}(\boldsymbol{n})$'s are determined by the following recursion:

$$\boldsymbol{N}_{\text{off}}(\boldsymbol{0}) = (-\boldsymbol{C}_{\text{off}})^{-1}, \qquad (4.15)$$

$$\boldsymbol{N}_{\text{off}}(\boldsymbol{n}) = \left[\sum_{k \in \mathcal{K}} \sum_{l_k = 1}^{n_k} \boldsymbol{N}_{\text{off}}(\boldsymbol{n} - l_k \boldsymbol{e}_k) \boldsymbol{D}_{k,\text{off}}(l_k) \right] \boldsymbol{N}_{\text{off}}(\boldsymbol{0}), \qquad \boldsymbol{n} \in \mathcal{Z}^+. \qquad (4.16)$$

The proof of Lemma 4.2 is given in Appendix A.

The $\boldsymbol{A}_{k,\xi}(\boldsymbol{n})$ is given in terms of the $\boldsymbol{F}_m(\boldsymbol{n})$ in the following way. It follows from (4.4), (4.10) and (4.12) that

$$\begin{aligned}
\sum_{\boldsymbol{n} \in \mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{A}_{k,\xi}(\boldsymbol{n}) = \sum_{m=0}^{\infty} \int_0^{\infty} dH_{k,\xi}(y) \frac{(\theta_{\text{on}} y)^m}{m!} e^{-\theta_{\text{on}} y} \Bigg[\boldsymbol{I} + \theta_{\text{on}}^{-1} \Bigg\{ \boldsymbol{C}_{\text{on}} + \sum_{k \in \mathcal{K}} \boldsymbol{D}^*_{k,\text{on}}(z_k) \\
+ \boldsymbol{E}_{\text{on,off}} \left(-\boldsymbol{C}_{\text{off}} - \sum_{k \in \mathcal{K}} \boldsymbol{D}^*_{k,\text{off}}(z_k) \right)^{-1} \boldsymbol{E}_{\text{off,on}} \Bigg\} \Bigg]^m \\
= \sum_{\boldsymbol{n} \in \mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \sum_{m=0}^{\infty} \gamma^{(m)}_{k,\xi}(\theta_{\text{on}}) \boldsymbol{F}_m(\boldsymbol{n}), \qquad (4.17)
\end{aligned}$$

where

$$\gamma_{k,\xi}^{(m)}(\theta_{\rm on}) = \int_0^\infty dH_{k,\xi}(y)\frac{(\theta_{\rm on}y)^m}{m!}e^{-\theta_{\rm on}y}, \qquad k \in \mathcal{K},\ \xi = \text{on, off},\ m = 0, 1, \ldots.$$

Comparing coefficient vectors of $z_1^{n_1}\cdots z_K^{n_K}$ on both sides in (4.17), we obtain the following theorem.

Theorem 4.4 *The $\boldsymbol{A}_{k,\xi}(\boldsymbol{n})$ is given by*

$$\boldsymbol{A}_{k,\xi}(\boldsymbol{n}) = \sum_{m=0}^\infty \gamma_{k,\xi}^{(m)}(\theta_{\rm on})\boldsymbol{F}_m(\boldsymbol{n}), \qquad k \in \mathcal{K},\ \xi = \textit{on, off},\ \boldsymbol{n} \in \mathcal{Z},$$

where the $\boldsymbol{F}_m(\boldsymbol{n})$ is given in Lemma 4.2.

Next we consider the $\boldsymbol{v}_{k,\rm on}(\boldsymbol{n})$ in (4.8) and the $\boldsymbol{v}_{k,\rm off}(\boldsymbol{n})$ in (4.9). Expanding $\boldsymbol{N}^*(\boldsymbol{z} \mid x)$ in (4.8) and (4.9), and comparing coefficient vectors of $z_1^{n_1}\cdots z_K^{n_K}$ on both sizes of each equation, we obtain the following theorem.

Theorem 4.5 *The $\boldsymbol{v}_{k,\rm on}(\boldsymbol{n})$ $(k \in \mathcal{K})$ and the $\boldsymbol{v}_{k,\rm off}(\boldsymbol{n})$ $(k \in \mathcal{K})$ are given by*

$$\boldsymbol{v}_{k,\rm on}(\boldsymbol{n}) = \sum_{m=0}^\infty \boldsymbol{v}_{\rm on}^{(m)}(\theta_{\rm on})\boldsymbol{D}_{k,\rm on}\boldsymbol{F}_m(\boldsymbol{n}), \qquad \boldsymbol{n} \in \mathcal{Z},$$

$$\boldsymbol{v}_{k,\rm off}(\boldsymbol{n}) = \sum_{m=0}^\infty \boldsymbol{v}_{\rm off}^{(m)}(\theta_{\rm on})\boldsymbol{D}_{k,\rm off}\sum_{\boldsymbol{0}\le\boldsymbol{l}\le\boldsymbol{n}} \boldsymbol{N}_{\rm off}(\boldsymbol{n}-\boldsymbol{l})\boldsymbol{E}_{\rm off,on}\boldsymbol{F}_m(\boldsymbol{l}), \qquad \boldsymbol{n} \in \mathcal{Z},$$

respectively, where the $\boldsymbol{F}_m(\boldsymbol{n})$ and the $\boldsymbol{N}_{\rm off}(\boldsymbol{n})$ are given in Lemma 4.2, and

$$\boldsymbol{v}_\xi^{(m)}(\theta_{\rm on}) = \int_0^\infty d\boldsymbol{v}_\xi(x)\frac{(\theta_{\rm on}x)^m}{m!}e^{-\theta_{\rm on}x}, \qquad \xi = \text{on}, \text{off},\ m = 0, 1, \ldots.$$

Thus the $\boldsymbol{v}_{k,\xi}(\boldsymbol{n})$ $(\xi = \text{on}, \text{off})$ is given in terms of the $\boldsymbol{v}_\xi^{(m)}(\theta_{\rm on})$ whose computation has already been studied in [15]. In what follows, we summarize the result. Note first that

$$\boldsymbol{v}_\xi^*(\theta_{\rm on} - \theta_{\rm on}z) = \sum_{m=0}^\infty z^m\boldsymbol{v}_\xi^{(m)}(\theta_{\rm on}), \qquad \xi = \text{on}, \text{off}. \tag{4.18}$$

Substituting $\theta_{\rm on} - \theta_{\rm on}z$ for s in (3.1) and using (4.18), we have

$$\sum_{m=0}^\infty z^m\boldsymbol{v}_{\rm on}^{(m)}(\theta_{\rm on})\left[(\theta_{\rm on} - \theta_{\rm on}z)\boldsymbol{I} + \boldsymbol{C}_{\rm on} + \sum_{m=0}^\infty z^m\boldsymbol{D}_{\rm on}^{(m)}(\theta_{\rm on}) + \boldsymbol{E}_{\rm on,off}\sum_{m=0}^\infty z^m\boldsymbol{J}_{\rm off}^{(m)}(\theta_{\rm on})\boldsymbol{E}_{\rm off,on}\right] = (\theta_{\rm on} - \theta_{\rm on}z)(1-\rho_{\rm on})\boldsymbol{\kappa}_{\rm on}, \tag{4.19}$$

where $\boldsymbol{D}_\xi^{(m)}(\theta_{\rm on})$ $(\xi = \text{on}, \text{off})$ and $\boldsymbol{J}_{\rm off}^{(m)}(\theta_{\rm on})$ are matrices satisfying

$$\sum_{m=0}^\infty z^m\boldsymbol{D}_\xi^{(m)}(\theta_{\rm on}) = \overline{\boldsymbol{D}}_\xi^*(\theta_{\rm on} - \theta_{\rm on}z), \tag{4.20}$$

$$\sum_{m=0}^\infty z^m\boldsymbol{J}_{\rm off}^{(m)}(\theta_{\rm on}) = \left[-\boldsymbol{C}_{\rm off} - \overline{\boldsymbol{D}}_{\rm off}^*(\theta_{\rm on} - \theta_{\rm on}z)\right]^{-1},$$

respectively. The computation of the $\boldsymbol{D}_\xi^{(m)}(\theta_{\rm on})$ $(\xi = \text{on}, \text{off})$ has already been studied in [7], while the recursion for the $\boldsymbol{J}_{\rm off}^{(m)}(\theta_{\rm on})$ can be obtained from

$$\sum_{m=0}^\infty z^m\boldsymbol{J}_{\rm off}^{(m)}(\theta_{\rm on})\left[-\boldsymbol{C}_{\rm off} - \sum_{m=0}^\infty z^m\boldsymbol{D}_{\rm off}^{(m)}(\theta_{\rm on})\right] = \boldsymbol{I}.$$

Lemma 4.3 ([7]) *Under Assumption 4.1, the* $\boldsymbol{D}_{\xi}^{(m)}(\theta_{\rm on})$ *(*ξ *= on, off) is given by*

$$\boldsymbol{D}_{\xi}^{(m)}(\theta_{\rm on}) = \sum_{k\in\mathcal{K}} \boldsymbol{d}_{k,\xi}^{(m)}(\theta_{\rm on})\boldsymbol{e}\boldsymbol{D}_{k,\xi}, \qquad m = 0, 1, \ldots,$$

where the $\boldsymbol{d}_{k,\xi}^{(m)}(\theta_{\rm on})$ *(*$k \in \mathcal{K}$*,* ξ *= on, off) is given by the following recursion:*

$$\begin{aligned}
\boldsymbol{d}_{k,\xi}^{(0)}(\theta_{\rm on}) &= \gamma_{k,\xi}^{(0)}(\theta_{\rm on})\boldsymbol{\alpha}_{k,\xi}(\boldsymbol{I} - \boldsymbol{P}_{k,\xi})\left[\boldsymbol{I} - \gamma_{k,\xi}^{(0)}(\theta_{\rm on})\boldsymbol{P}_{k,\xi}\right]^{-1},\\
\boldsymbol{d}_{k,\xi}^{(m)}(\theta_{\rm on}) &= \frac{\gamma_{k,\xi}^{(m)}(\theta_{\rm on})}{\gamma_{k,\xi}^{(0)}(\theta_{\rm on})}\boldsymbol{d}_{k,\xi}^{(0)}(\theta_{\rm on}) + \left[\sum_{l=1}^{m}\gamma_{k,\xi}^{(l)}(\theta_{\rm on})\boldsymbol{d}_{k,\xi}^{(m-l)}(\theta_{\rm on})\right]\\
&\quad\cdot \boldsymbol{P}_{k,\xi}\left[\boldsymbol{I} - \gamma_{k,\xi}^{(0)}(\theta_{\rm on})\boldsymbol{P}_{k,\xi}\right]^{-1}, \qquad m = 1, 2, \ldots.
\end{aligned}$$

Lemma 4.4 *The* $\boldsymbol{J}_{\rm off}^{(m)}(\theta_{\rm on})$ *is recursively determined by the following recursion:*

$$\begin{aligned}
\boldsymbol{J}_{\rm off}^{(0)}(\theta_{\rm on}) &= \left[-\boldsymbol{C}_{\rm off} - \boldsymbol{D}_{\rm off}^{(0)}(\theta_{\rm on})\right]^{-1},\\
\boldsymbol{J}_{\rm off}^{(m)}(\theta_{\rm on}) &= \left[\sum_{l=0}^{m-1}\boldsymbol{J}_{\rm off}^{(l)}(\theta_{\rm on})\boldsymbol{D}_{\rm off}^{(m-l)}(\theta_{\rm on})\right]\boldsymbol{J}_{\rm off}^{(0)}(\theta_{\rm on}), \qquad m = 1, 2, \ldots.
\end{aligned}$$

The $\boldsymbol{v}_{\rm on}^{(m)}(\theta_{\rm on})$ is computed as follows. Comparing the coefficient vectors of z^m ($m = 0, 1, \ldots$) on both sides of (4.19), we can show that the $\boldsymbol{v}_{\rm on}^{(m)}(\theta_{\rm on})$ is identical to the steady-state solution of a Markov chain of M/G/1 type whose transition probability matrix is given by [15]

$$\begin{bmatrix}
\boldsymbol{B}_0 + \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{B}_3 & \boldsymbol{B}_4 & \cdots\\
\boldsymbol{B}_0 & \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{B}_3 & \cdots\\
\boldsymbol{O} & \boldsymbol{B}_0 & \boldsymbol{B}_1 & \boldsymbol{B}_2 & \cdots\\
\boldsymbol{O} & \boldsymbol{O} & \boldsymbol{B}_0 & \boldsymbol{B}_1 & \cdots\\
\boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{B}_0 & \cdots\\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix},$$

where

$$\begin{aligned}
\boldsymbol{B}_0 &= \boldsymbol{I} + \theta_{\rm on}^{-1}\left[\boldsymbol{C}_{\rm on} + \boldsymbol{D}_{\rm on}^{(0)}(\theta_{\rm on}) + \boldsymbol{E}_{\rm on,off}\boldsymbol{J}_{\rm off}^{(0)}(\theta_{\rm on})\boldsymbol{E}_{\rm off,on}\right],\\
\boldsymbol{B}_m &= \theta_{\rm on}^{-1}\left[\boldsymbol{D}_{\rm on}^{(m)}(\theta_{\rm on}) + \boldsymbol{E}_{\rm on,off}\boldsymbol{J}_{\rm off}^{(m)}(\theta_{\rm on})\boldsymbol{E}_{\rm off,on}\right], \quad m = 1, 2, \ldots.
\end{aligned}$$

Thus applying the general theory of Markov chains of M/G/1 type [8], we can compute the $\boldsymbol{v}_{\rm on}^{(m)}(\theta_{\rm on})$.

On the other hand, the $\boldsymbol{v}_{\rm off}^{(m)}(\theta_{\rm on})$ can be computed by the following theorem whose proof is given in Appendix B.

Theorem 4.6 *The* $\boldsymbol{v}_{\rm off}^{(m)}(\theta_{\rm on})$ *is determined by the following recursion:*

$$\boldsymbol{v}_{\rm off}^{(0)}(\theta_{\rm on}) = \frac{\boldsymbol{v}_{\rm on}^{(0)}(\theta_{\rm on})\boldsymbol{E}_{\rm on,off}}{\boldsymbol{\pi}_{\rm on}\boldsymbol{E}_{\rm on,off}\boldsymbol{e}}\frac{\left[-\boldsymbol{C}_{\rm off} - \boldsymbol{D}_{\rm off}^{(0)}(\theta_{\rm on})\right]^{-1}}{\overline{I}_{\rm off}}, \tag{4.21}$$

$$\begin{aligned}
\boldsymbol{v}_{\rm off}^{(m)}(\theta_{\rm on}) &= \left[\frac{\boldsymbol{v}_{\rm on}^{(m)}(\theta_{\rm on})\boldsymbol{E}_{\rm on,off}}{\boldsymbol{\pi}_{\rm on}\boldsymbol{E}_{\rm on,off}\boldsymbol{e}}\frac{1}{\overline{I}_{\rm off}} + \sum_{l=0}^{m-1}\boldsymbol{v}_{\rm off}^{(l)}(\theta_{\rm on})\boldsymbol{D}_{\rm off}^{(m-l)}(\theta_{\rm on})\right]\\
&\quad\cdot\left[-\boldsymbol{C}_{\rm off} - \boldsymbol{D}_{\rm off}^{(0)}(\theta_{\rm on})\right]^{-1}, \qquad m = 1, 2, \ldots,
\end{aligned} \tag{4.22}$$

where the $\boldsymbol{D}_{\rm off}^{(m)}(\theta_{\rm on})$ *is given in Lemma 4.3.*

Among the recursions required in computing the joint queue length distribution, Lemma 4.2 for the $\boldsymbol{F}_m(\boldsymbol{n})$ is the most extensive. In fact, its straightforward implementation will require very huge memory space. Note that an efficient implementation scheme for it is proposed in [7]. All other recursions can be readily implemented as they are. See [7, 11] for details. Note also that from the results in this subsection, we can readily obtain recursions to compute the total queue length distribution, which are much less extensive than those for the joint queue length distribution. The results for the total queue length distribution are summarized in Appendix C.

5. Numerical Examples

In this section, we provide some numerical examples using two-class models. Throughout this section, we assume that the marginal arrival process follows a two-state batch marked MAP $(\widetilde{\boldsymbol{C}}, \widetilde{\boldsymbol{D}}_1(n), \widetilde{\boldsymbol{D}}_2(n))$:

$$\widetilde{\boldsymbol{C}} = \begin{bmatrix} -\lambda g^{-1} - a & a \\ a & -\lambda g^{-1} - a \end{bmatrix}, \tag{5.1}$$

$$\widetilde{\boldsymbol{D}}_1(n) = g(n) \begin{bmatrix} \lambda g^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad \widetilde{\boldsymbol{D}}_2(n) = g(n) \begin{bmatrix} 0 & 0 \\ 0 & \lambda g^{-1} \end{bmatrix}, \qquad n = 1, 2, \ldots, \tag{5.2}$$

where $\lambda, g, a > 0$ and $g(n) = 1$ if $n = g$, and otherwise $g(n) = 0$.

We also assume that the marginal on-off process follows an alternating Markov renewal process whose infinitesimal generator is given by

$$\begin{bmatrix} \boldsymbol{S}_{\rm on} & \boldsymbol{T}_{\rm on,off} \\ \boldsymbol{T}_{\rm off,on} & \boldsymbol{S}_{\rm off} \end{bmatrix},$$

where $\boldsymbol{S}_{\rm on}$ (resp. $\boldsymbol{S}_{\rm off}$) denotes an $m_{\rm on} \times m_{\rm on}$ (resp. $m_{\rm off} \times m_{\rm off}$) matrix representing an infinitesimal generator that governs transitions in on-periods (resp. off-periods), and $\boldsymbol{T}_{\rm on,off}$ (resp. $\boldsymbol{T}_{\rm off,on}$) denotes a transition rate matrix from on-states (resp. off-states) to off-states (resp. on-states). When the on-off and arrival processes are independent of each other, the model is characterized as follows:

$$\begin{aligned} &\boldsymbol{C}_{\rm on} = \boldsymbol{S}_{\rm on} \oplus \widetilde{\boldsymbol{C}}, \quad \boldsymbol{C}_{\rm off} = \boldsymbol{S}_{\rm off} \oplus \widetilde{\boldsymbol{C}}, \quad \boldsymbol{E}_{\rm on,off} = \boldsymbol{T}_{\rm on,off} \otimes \boldsymbol{I}(2), \quad \boldsymbol{E}_{\rm off,on} = \boldsymbol{T}_{\rm off,on} \otimes \boldsymbol{I}(2), \\ &\boldsymbol{D}_{1,\rm on}(n) = \boldsymbol{I}(m_{\rm on}) \otimes \widetilde{\boldsymbol{D}}_1(n), \quad \boldsymbol{D}_{1,\rm off}(n) = \boldsymbol{I}(m_{\rm off}) \otimes \widetilde{\boldsymbol{D}}_1(n), \\ &\boldsymbol{D}_{2,\rm on}(n) = \boldsymbol{I}(m_{\rm on}) \otimes \widetilde{\boldsymbol{D}}_2(n), \quad \boldsymbol{D}_{2,\rm off}(n) = \boldsymbol{I}(m_{\rm off}) \otimes \widetilde{\boldsymbol{D}}_2(n). \end{aligned}$$

5.1. Impact of service time dependency

In this subsection, we discuss the impact of the service time dependency on the queue length. We assume that the on-off and arrival processes are mutually independent. Let $\boldsymbol{S}_{\rm on} = \boldsymbol{S}_{\rm off} = -\alpha$ and $\boldsymbol{T}_{\rm on,off} = \boldsymbol{T}_{\rm off,on} = \alpha$, where $\alpha > 0$. Also let $a = 0.1$ and $\lambda = 0.125$ in (5.1) and (5.2). As for the service time, we consider two cases, Case GD (class-dependent service times) and Case GI (i.i.d. service times):

[**Case GD**] $\quad H_1 = 1$ with probability 1, $\quad H_2 = 5$ with probability 1,

[**Case GI**] $\quad H_k = \begin{cases} 1, & \text{with probability } 1/2, \\ 5, & \text{with probability } 1/2, \end{cases} \qquad k = 1, 2,$

where H_k $(k = 1, 2)$ denotes a generic random variable representing a service time of a class k customer. Note here that the overall service time distributions in both cases are identical and $\rho_{\rm on} = 6\lambda = 0.75$.

Figure 1 plots the expected total queue lengths E[N] in Cases GD and GI as functions of α^{-1}. As α^{-1} goes to 0, the above model gets close to a work-conserving single-server queue (i.e., no service interruptions occur) with the same arrival process and service time distributions, where the processing speed of the server is reduced by half. We observe that the difference in the expected total queue lengths of the two models is kept almost constant regardless of the value of α^{-1} and gets large with constant batch size g.

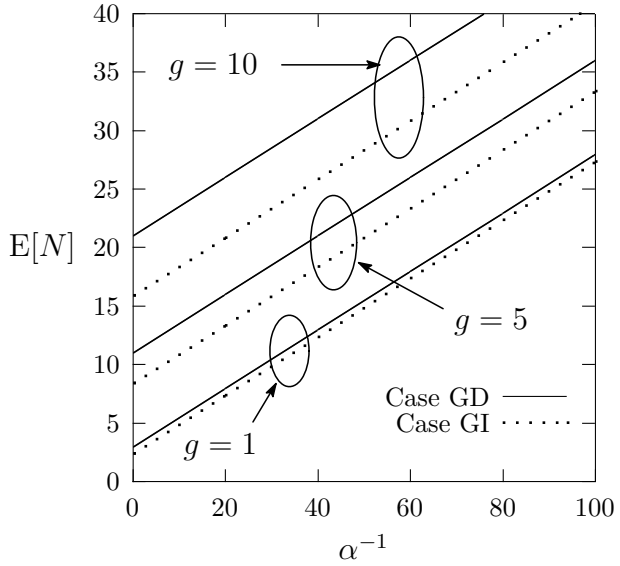

Figure 1: Expected total queue length E[N].

Table 1 shows the joint queue length distribution for $g = 1$. Let $\boldsymbol{p}_{\rm GD}(n_1, n_2)$ (resp. $\boldsymbol{p}_{\rm GI}(n_1, n_2)$) denote $\boldsymbol{p}(n_1, n_2)$ in Case GD (resp. GI). We observe that $\boldsymbol{p}_{\rm GI}(n_1, n_2)\boldsymbol{e} = \boldsymbol{p}_{\rm GI}(n_2, n_1)\boldsymbol{e}$ is due to the symmetry of input parameters of classes 1 and 2 in Case GI. Further, Table 1 shows that $\boldsymbol{p}_{\rm GD}(n_1, n_2)\boldsymbol{e} > \boldsymbol{p}_{\rm GI}(n_1, n_2)\boldsymbol{e}$ for $n_1 < n_2$, and vice versa. Thus for m, n such that $m > n$,

$$\boldsymbol{p}_{\rm GD}(m, n)\boldsymbol{e} < \boldsymbol{p}_{\rm GI}(m, n)\boldsymbol{e} = \boldsymbol{p}_{\rm GI}(n, m)\boldsymbol{e} < \boldsymbol{p}_{\rm GD}(n, m)\boldsymbol{e},$$

in this particular example. We conjecture that this phenomenon is caused by the fact that service times of class 2 customers are larger than those of class 1 customer in Case GD, and while a class 2 customer is being served, succeeding class 2 customers are likely to arrive back to back and stay in the system.

5.2. Impact of variation of on- and off-periods

Next, we discuss the impact of the variation in on- and off-periods on the total queue length N. We assume that the on-off process follows an alternating renewal process, and the on-off and arrival processes are mutually independent. Let $C^2_{v,\rm on}$ and $C^2_{v,\rm off}$ denote the squared coefficients of variation of on-periods and off-periods, respectively. To examine the impact of the variation of on-periods, the off-period distribution is fixed to be exponential with mean 100. For $C^2_{v,\rm on} = k^{-1} \le 1$ ($k = 1, 2, \ldots$), on-periods follow a k-stage Erlang distribution with mean 100, and for $C^2_{v,\rm on} > 1$, they follow a balanced hyper-exponential distribution $\psi(x)$ with mean 100, where

$$\psi(x) = 1 - p\exp(-0.02px) - (1-p)\exp\left[-0.02(1-p)x\right], \qquad 0 < p < 0.5.$$

Note that $C^2_{v,\rm on} = 1/\{2p(1-p)\} - 1$ in this case. On the other hand, in examining the impact of the variation of off-periods on the total queue length, the above on- and off-period distributions are exchanged.

Table 1: Joint queue length distribution $\boldsymbol{p}(n_1, n_2)\boldsymbol{e}$.
(Upper rows for Case GD and lower rows for Case GI)

n_2 \ n_1	0	1	2	5	10	20
0	1.34×10^{-1}	1.74×10^{-2}	5.67×10^{-3}	1.38×10^{-3}	1.79×10^{-4}	3.10×10^{-6}
	1.34×10^{-1}	3.41×10^{-2}	1.41×10^{-2}	2.51×10^{-3}	3.44×10^{-4}	6.80×10^{-6}
1	4.67×10^{-2}	9.37×10^{-3}	5.88×10^{-3}	2.12×10^{-3}	4.02×10^{-4}	1.14×10^{-5}
	3.41×10^{-2}	9.25×10^{-3}	6.25×10^{-3}	2.71×10^{-3}	6.02×10^{-4}	2.07×10^{-5}
2	2.18×10^{-2}	6.87×10^{-3}	5.31×10^{-3}	2.60×10^{-3}	6.48×10^{-4}	2.68×10^{-5}
	1.41×10^{-2}	6.25×10^{-3}	5.27×10^{-3}	3.03×10^{-3}	8.78×10^{-4}	4.38×10^{-5}
5	3.68×10^{-3}	3.26×10^{-3}	3.41×10^{-3}	2.87×10^{-3}	1.30×10^{-3}	1.24×10^{-4}
	2.51×10^{-3}	2.71×10^{-3}	3.03×10^{-3}	2.89×10^{-3}	1.50×10^{-3}	1.70×10^{-4}
10	5.02×10^{-4}	7.92×10^{-4}	1.09×10^{-3}	1.68×10^{-3}	1.54×10^{-3}	4.04×10^{-4}
	3.44×10^{-4}	6.02×10^{-4}	8.78×10^{-4}	1.50×10^{-3}	1.55×10^{-3}	4.73×10^{-4}
20	1.04×10^{-5}	2.97×10^{-5}	6.03×10^{-5}	2.15×10^{-4}	5.42×10^{-4}	6.12×10^{-4}
	6.80×10^{-6}	2.07×10^{-5}	4.38×10^{-5}	1.70×10^{-4}	4.73×10^{-4}	6.09×10^{-4}

As for the arrival process, we set $a = 0.1$, $\lambda = 0.125$ and $g = 1$ in (5.1) and (5.2). Besides, service times of each class are assumed to follow the same service time distribution as in Case GD of the preceding subsection.

Figures 2 and 3 plot the 99.9 percentile (99.9 PT) and expected value E[N] of the total queue length, respectively, as functions of the squared coefficient of variation $C^2_{v,\xi}$ ($\xi = \text{on}, \text{off}$), where the vertical axes are in log-scale. Note that in the case of $C^2_{v,\xi} = 1$ ($\xi = \text{on}, \text{off}$), the two models become identical with exponential on- and off-periods. We observe that both 99.9 PT and E[N] are monotone increasing functions of $C^2_{v,\xi}$ ($\xi = \text{on}, \text{off}$) and $C^2_{v,\text{off}}$ has a more impact on the total queue length N than $C^2_{v,\text{on}}$.

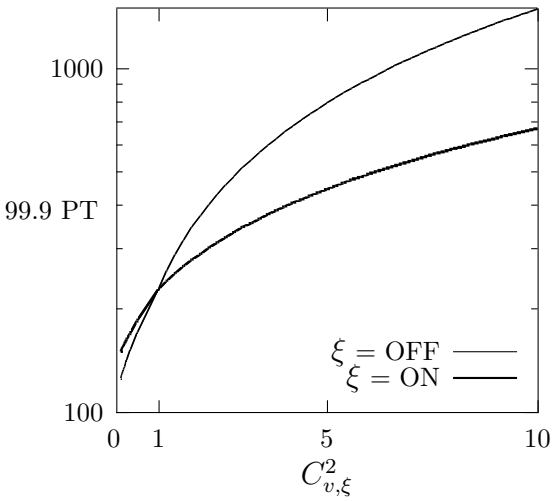

Figure 2: 99.9 percentile (99.9 PT) of the total queue length.

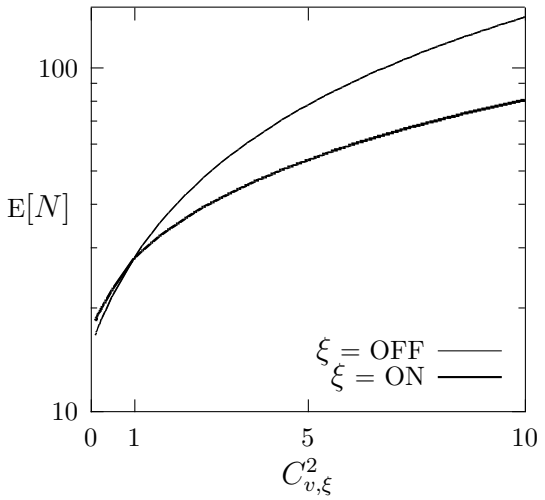

Figure 3: Expected total queue length E[N].

5.3. Impact of correlation in on- and off-periods

In this subsection, we examine the impact of the correlation in on- and off-periods on the total queue length N. For this purpose, we assume that the on-off and arrival processes are

mutually independent and the marginal on-off process is given by

$$\boldsymbol{S}_{\rm on} = \boldsymbol{S}_{\rm off} = \begin{bmatrix} -1/40 & 0 \\ 0 & -1/160 \end{bmatrix}, \qquad \boldsymbol{T}_{\rm on,off} = \boldsymbol{T}_{\rm off,on} = \begin{bmatrix} p/40 & (1-p)/40 \\ (1-p)/160 & p/160 \end{bmatrix},$$

where $0 < p < 1$. Thus the marginal distributions of on- and off-periods follow the same hyper-exponential distribution whose distribution function $\psi(x)$ is given by

$$\psi(x) = 1 - 0.5\exp(-x/40) - 0.5\exp(-x/160).$$

Note that parameter p controls the correlation in consecutive on- and off-periods. Suppose the on-off process starts with an on-period. Let $I_{\rm on}(n)$ and $I_{\rm off}(n)$ $(n = 1, 2, \ldots)$ denote the lengths of the nth on- and off-periods, respectively. Then, both $\mathrm{Cov}[I_{\rm on}(n), I_{\rm off}(n)]$ and $\mathrm{Cov}[I_{\rm off}(n), I_{\rm on}(n+1)]$ are negative for $0 < p < 0.5$, equal to zero for $p = 0.5$, and positive for $0.5 < p < 1$. We set $a = 0.1$, $\lambda = 0.125$ and $g = 1$ in (5.1) and (5.2). Service times of each class follow the same distribution as in Case GD of subsection 5.1.

Figures 4 and 5 plot the 99.9 percentile (99.9 PT) and expected value $\mathrm{E}[N]$ of the total queue length, respectively, as functions of p. From these figures, we observe the followings. As p goes to zero, both 99.9 PT and $\mathrm{E}[N]$ rapidly increase. This phenomenon is due to the fact that once the on-off process is in a long off-period, long off-periods and short on-periods are likely to repeat alternately, and during those intervals, many customers are accumulated in the system. As p becomes large, however, this effect is weakened, and finally, both 99.9 PT and $\mathrm{E}[N]$ take their minimums and turn to increase. This implies that there exists some factor to make the queue length increase with p. In what follows, we examine this phenomenon more closely.

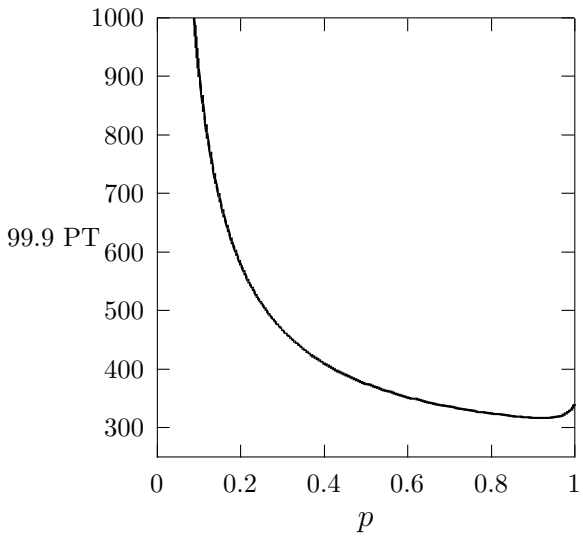

Figure 4: 99.9 percentile (99.9 PT) of the total queue length.

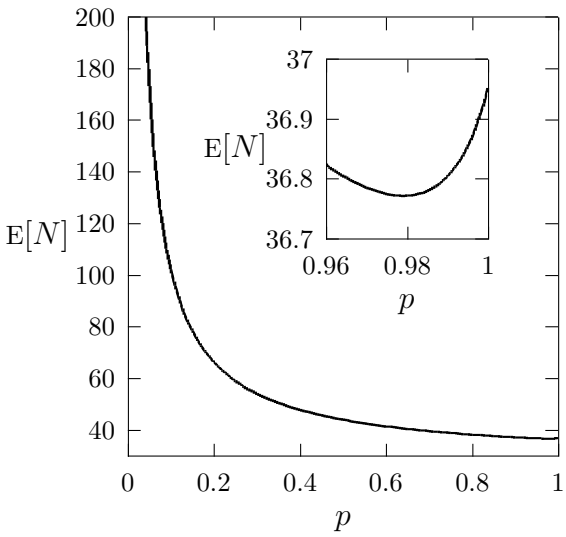

Figure 5: Expected total queue length $\mathrm{E}[N]$.

Let $\Psi_{\rm short-on}$, $\Psi_{\rm long-on}$, $\Psi_{\rm short-off}$ and $\Psi_{\rm long-off}$ denote the events that the on-off process is in a short on-period, long on-period, short off-period and long off-period, respectively. Figures 6 and 7 plot the conditional expected total queue lengths given those events as functions of p. From Figure 6, we observe that as expected, $\mathrm{E}[N \mid \Psi_{\rm long-off}]$ is always larger than $\mathrm{E}[N \mid \Psi_{\rm short-off}]$, so that the total queue length in an on-period following a long off-period is likely to be larger than that in an on-period following a short off-period, regardless

of the value of p. Note here that as p goes to one, the contribution of the total queue length in an on-period following a long off-period to $\mathrm{E}[N \mid \Psi_{\mathrm{long-on}}]$ becomes large, and we conjecture that this factor makes $\mathrm{E}[N \mid \Psi_{\mathrm{long-on}}]$ increase in the region where p is close to one, as shown in Figure 7. Moreover, once $\mathrm{E}[N \mid \Psi_{\mathrm{long-on}}]$ turns to increase, this affects the total queue length in the following off-period, and as p goes to one, off-periods following long on-periods are likely to be long off-periods. Thus $\mathrm{E}[N \mid \Psi_{\mathrm{long-off}}]$ turns to increase after $\mathrm{E}[N \mid \Psi_{\mathrm{long-on}}]$ does, as shown in Figures 6 and 7.

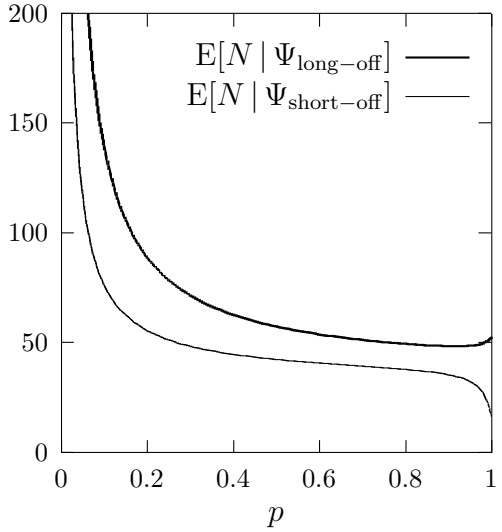

Figure 6: Conditional expected total queue length.

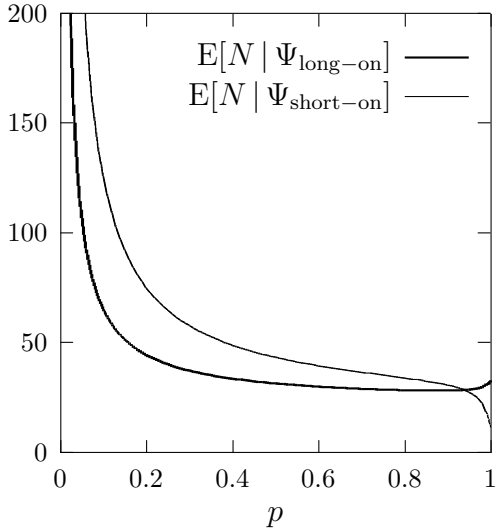

Figure 7: Conditional expected total queue length.

Note here that in this particular example,

$$\Pr(\Psi_{\mathrm{long-on}}) = \Pr(\Psi_{\mathrm{long-off}}) = 0.4, \quad \Pr(\Psi_{\mathrm{short-on}}) = \Pr(\Psi_{\mathrm{short-off}}) = 0.1.$$

Therefore the contributions of $\mathrm{E}[N \mid \Psi_{\mathrm{long-on}}]$ and $\mathrm{E}[N \mid \Psi_{\mathrm{long-off}}]$ to $\mathrm{E}[N]$ are four times as large as those of $\mathrm{E}[N \mid \Psi_{\mathrm{short-on}}]$ and $\mathrm{E}[N \mid \Psi_{\mathrm{short-off}}]$. As a result, $\mathrm{E}[N]$ increases for p near one. A similar observation can be applied to the 99.9 percentile, too.

5.4. Impact of correlation between on-off and arrival processes

Finally, we examine the impact of the correlation between on-off and arrival processes on the total queue length N. We consider the following three models, where service times of each class follow the same distribution as in Case GD of subsection 5.1.

In Model 1, the on-off and arrival processes have a correlation, and they are represented by

$$\boldsymbol{C} = \left[\begin{array}{c|c} -0.125-\alpha & \alpha \\ \hline \alpha & -0.125-\alpha \end{array}\right], \quad \boldsymbol{D}_1(1) = \left[\begin{array}{c|c} 0.125 & 0 \\ \hline 0 & 0 \end{array}\right], \quad \boldsymbol{D}_2(1) = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & 0.125 \end{array}\right],$$

and $\boldsymbol{D}_k(n) = \boldsymbol{O}$ $(k = 1, 2)$ for $n = 2, 3, \ldots$, where $M_{\mathrm{on}} = M_{\mathrm{off}} = 1$. In Model 2, the on-off and arrival processes are mutually independent. As for the arrival process, we set $a = \alpha$, $\lambda = 0.125$ and $g = 1$ in (5.1) and (5.2). The on-off process is the same as that in subsection 5.1. In Model 3, the on-off and arrival processes have a correlation, and they are represented by

$$\boldsymbol{C} = \left[\begin{array}{c|c} -0.125-\alpha & \alpha \\ \hline \alpha & -0.125-\alpha \end{array}\right], \quad \boldsymbol{D}_1(1) = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & 0.125 \end{array}\right], \quad \boldsymbol{D}_2(1) = \left[\begin{array}{c|c} 0.125 & 0 \\ \hline 0 & 0 \end{array}\right],$$

and $\boldsymbol{D}_k(n) = \boldsymbol{O}$ $(k = 1, 2)$ for $n = 2, 3, \ldots$, where $M_{\rm on} = M_{\rm off} = 1$. Note here that the marginal on-off processes in the three models are identical, and so are the marginal arrival processes.

Figures 8 and 9 show the 99.9 percentile (99.9 PT) and expected value E[N] of the total queue length, respectively, as functions of α^{-1}. Note here that as α^{-1} goes to zero, all the three models get close to a work-conserving single-server queue, where the arrival process follows a Poisson process with rate 0.125, and service times are i.i.d. and take 2 or 10 with equal probability. This is a reason why both 99.9 PTs and E[N]'s in all the three models converge the same values, respectively, as $\alpha^{-1} \to 0$. We also observe that 99.9PT and E[N] in Model 1 (resp. Model 3) are always larger (resp. smaller) than those in Model 2. This is due to the fact the amount of work brought into the system during off-periods in Model 1 (resp. Model 3) is likely to be larger (resp. smaller) than that in Model 2.

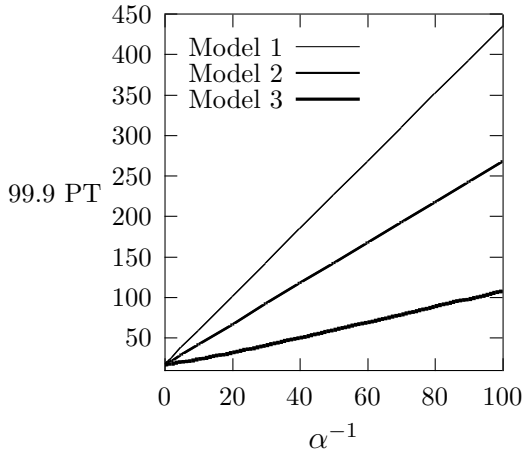

Figure 8: 99.9 percentile (99.9 PT) of the total queue length.

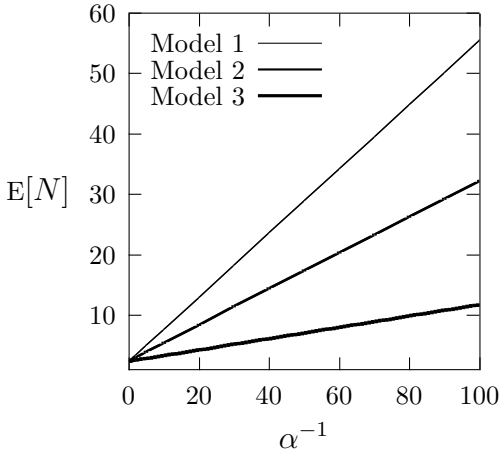

Figure 9: Expected total queue length E[N].

A. Proof of Lemma 4.2

From the definition (4.12) of $\boldsymbol{F}_m(\boldsymbol{n})$, (4.13) is clearly satisfied. (4.14) is proved in the following way. Let $\boldsymbol{N}_{\rm off}(\boldsymbol{n})$ denote an $M_{\rm off} \times M_{\rm off}$ matrix satisfying

$$\sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{N}_{\rm off}(\boldsymbol{n}) = \left(-\boldsymbol{C}_{\rm off} - \sum_{k\in\mathcal{K}} \boldsymbol{D}^*_{k,{\rm off}}(z_k) \right)^{-1}. \tag{A.1}$$

From (3.3), (4.12) and (A.1), we have for $m = 1, 2, \ldots$,

$$\begin{aligned}
&\sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{F}_m(\boldsymbol{n}) \\
&\quad = \sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{F}_{m-1}(\boldsymbol{n}) \\
&\qquad \cdot \left[\boldsymbol{I} + \theta_{\rm on}^{-1} \left\{ \boldsymbol{C}_{\rm on} + \sum_{k\in\mathcal{K}} \sum_{n_k=1}^{\infty} z_k^{n_k} \boldsymbol{D}_{k,{\rm on}}(n_k) + \boldsymbol{E}_{\rm on,off} \sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{N}_{\rm off}(\boldsymbol{n}) \boldsymbol{E}_{\rm off,on} \right\} \right] \\
&\quad = \sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{F}_{m-1}(\boldsymbol{n}) (\boldsymbol{I} + \theta_{\rm on}^{-1} \boldsymbol{C}_{\rm on})
\end{aligned}$$

$$+\sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1}\cdots z_K^{n_K}\theta_{\rm on}^{-1}\sum_{k\in\mathcal{K}}\sum_{l_k=1}^{n_k}\boldsymbol{F}_{m-1}(\boldsymbol{n}-l_k\boldsymbol{e}_k)\boldsymbol{D}_{k,{\rm on}}(l_k)$$
$$+\sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1}\cdots z_K^{n_K}\theta_{\rm on}^{-1}\sum_{\boldsymbol{0}\le\boldsymbol{l}\le\boldsymbol{n}}\boldsymbol{F}_{m-1}(\boldsymbol{n}-\boldsymbol{l})\boldsymbol{E}_{\rm on,off}\boldsymbol{N}_{\rm off}(\boldsymbol{l})\boldsymbol{E}_{\rm off,on}.$$

Comparing coefficient matrices of $z_1^{n_1}\cdots z_K^{n_K}$ on both sides of the above equation, we obtain (4.14). Next, we show (4.15) and (4.16). From (3.3) and (A.1), we have

$$\sum_{\boldsymbol{n}\in\mathcal{Z}} z_1^{n_1}\cdots z_K^{n_K}\boldsymbol{N}_{\rm off}(\boldsymbol{n})\left(-\boldsymbol{C}_{\rm off}-\sum_{k\in\mathcal{K}}\sum_{n_k=1}^{\infty} z_k^{n_k}\boldsymbol{D}_{k,{\rm off}}(n_k)\right)=\boldsymbol{I}.$$

Comparing coefficient matrices of $z_1^{n_1}\cdots z_K^{n_K}$ on both sides of the above equation yields

$$\boldsymbol{N}_{\rm off}(\boldsymbol{0})\,(-\boldsymbol{C}_{\rm off})=\boldsymbol{I},$$
$$\boldsymbol{N}_{\rm off}(\boldsymbol{n})\,(-\boldsymbol{C}_{\rm off})-\sum_{k\in\mathcal{K}}\sum_{l_k=1}^{n_k}\boldsymbol{N}_{\rm off}(\boldsymbol{n}-l_k\boldsymbol{e}_k)\boldsymbol{D}_{k,{\rm off}}(l_k)=\boldsymbol{O},$$

from which (4.15) and (4.16) follow. ∎

B. Proof of Theorem 4.6

Post-multiplying both sides of (3.2) by $-\boldsymbol{C}_{\rm off}-\overline{\boldsymbol{D}}^{*}_{\rm off}(s)$ and substituting $\theta_{\rm on}-\theta_{\rm on}z$ for s, we obtain

$$\sum_{m=0}^{\infty} z^m\boldsymbol{v}_{\rm off}^{(m)}(\theta_{\rm on})\left[-\boldsymbol{C}_{\rm off}-\sum_{m=0}^{\infty} z^m\boldsymbol{D}_{\rm off}^{(m)}(\theta_{\rm on})\right]=\frac{\displaystyle\sum_{m=0}^{\infty} z^m\boldsymbol{v}_{\rm on}^{(m)}(\theta_{\rm on})\boldsymbol{E}_{\rm on,off}}{\boldsymbol{\pi}_{\rm on}\boldsymbol{E}_{\rm on,off}\boldsymbol{e}}\frac{1}{\overline{I}_{\rm off}},\qquad\text{(B.1)}$$

where we use (4.18) and (4.20). Comparing coefficient vectors of z^m $(m=0,1,\ldots)$ on both sides of (B.1) yields

$$\boldsymbol{v}_{\rm off}^{(0)}(\theta_{\rm on})\left[-\boldsymbol{C}_{\rm off}-\boldsymbol{D}_{\rm off}^{(0)}(\theta_{\rm on})\right]=\frac{\boldsymbol{v}_{\rm on}^{(0)}(\theta_{\rm on})\boldsymbol{E}_{\rm on,off}}{\boldsymbol{\pi}_{\rm on}\boldsymbol{E}_{\rm on,off}\boldsymbol{e}}\frac{1}{\overline{I}_{\rm off}},$$

and for $m=1,2,\ldots$,

$$\boldsymbol{v}_{\rm off}^{(m)}(\theta_{\rm on})\left[-\boldsymbol{C}_{\rm off}-\boldsymbol{D}_{\rm off}^{(0)}(\theta_{\rm on})\right]-\sum_{l=0}^{m-1}\boldsymbol{v}_{\rm off}^{(l)}(\theta_{\rm on})\boldsymbol{D}_{\rm off}^{(m-l)}(\theta_{\rm on})=\frac{\boldsymbol{v}_{\rm on}^{(m)}(\theta_{\rm on})\boldsymbol{E}_{\rm on,off}}{\boldsymbol{\pi}_{\rm on}\boldsymbol{E}_{\rm on,off}\boldsymbol{e}}\frac{1}{\overline{I}_{\rm off}}.$$

(4.21) and (4.22) follow the above two equations. ∎

C. Total Queue Length Distribution

This appendix summarizes the recursions for the total queue length distribution. Because they are readily derived from the results in Section 4, we omit the proofs.

We first define $\boldsymbol{p}^{(\rm T)}(n)$ $(n=0,1,\ldots)$ and $\boldsymbol{q}_k^{(\rm T)}(n)$ $(k\in\mathcal{K},n=0,1,\ldots)$ as the stationary total queue length distributions at a random point in time and at immediately after departures of class k, respectively.

$$\boldsymbol{p}^{(\rm T)}(n)=\sum_{\substack{\boldsymbol{n}\in\mathcal{Z}\\|\boldsymbol{n}|=n}}\boldsymbol{p}(\boldsymbol{n}),\qquad \boldsymbol{q}_k^{(\rm T)}(n)=\sum_{\substack{\boldsymbol{n}\in\mathcal{Z}\\|\boldsymbol{n}|=n}}\boldsymbol{q}_k(\boldsymbol{n}),$$

where $|\boldsymbol{n}|=|n_1|+\cdots+|n_K|$. From Theorem 4.1, we obtain the following corollary.

Corollary C.1 *The* $\boldsymbol{p}^{(\mathrm{T})}(n)$ *is recursively determined in the following way:*

$$\boldsymbol{p}^{(\mathrm{T})}(0) = \sum_{k\in\mathcal{K}} \lambda_k \boldsymbol{q}_k^{(\mathrm{T})}(0)(-\boldsymbol{C})^{-1},$$

and for $n = 1, 2, \ldots,$

$$\boldsymbol{p}^{(\mathrm{T})}(n) = \sum_{k\in\mathcal{K}} \left[\lambda_k \left(\boldsymbol{q}_k^{(\mathrm{T})}(n) - \boldsymbol{q}_k^{(\mathrm{T})}(n-1)\right) + \sum_{m=1}^{n} \boldsymbol{p}^{(\mathrm{T})}(n-m)\boldsymbol{D}_k(m)\right](-\boldsymbol{C})^{-1}.$$

In what follows, we show recursions to compute the $\boldsymbol{q}_k^{(\mathrm{T})}(n)$. For this purpose, we introduce the following notations: For $k \in \mathcal{K}$, $\xi = \mathrm{on}, \mathrm{off}$, $n = 0, 1, \ldots$ and $m = 0, 1, \ldots,$

$$\boldsymbol{q}_{k,\xi}^{(\mathrm{T})}(n) = \sum_{\substack{\boldsymbol{n}\in\mathcal{Z}\\ |\boldsymbol{n}|=n}} \boldsymbol{q}_{k,\xi}(\boldsymbol{n}), \quad \boldsymbol{\Gamma}_{k,\xi}^{(\mathrm{T})}(n) = \sum_{\substack{\boldsymbol{n}\in\mathcal{Z}\\ |\boldsymbol{n}|=n}} \boldsymbol{\Gamma}_{k,\xi}(\boldsymbol{n}), \quad \boldsymbol{A}_{k,\xi}^{(\mathrm{T})}(n) = \sum_{\substack{\boldsymbol{n}\in\mathcal{Z}\\ |\boldsymbol{n}|=n}} \boldsymbol{A}_{k,\xi}(\boldsymbol{n}),$$
$$\boldsymbol{v}_{k,\xi}^{(\mathrm{T})}(n) = \sum_{\substack{\boldsymbol{n}\in\mathcal{Z}\\ |\boldsymbol{n}|=n}} \boldsymbol{v}_{k,\xi}(\boldsymbol{n}), \quad \boldsymbol{F}_{m}^{(\mathrm{T})}(n) = \sum_{\substack{\boldsymbol{n}\in\mathcal{Z}\\ |\boldsymbol{n}|=n}} \boldsymbol{F}_{m}(\boldsymbol{n}).$$

From Theorem 4.3, we have the following corollary.

Corollary C.2 *Under Assumption 4.1, the* $\boldsymbol{q}_k^{(\mathrm{T})}(n)$ *(*$k \in \mathcal{K}, n = 0, 1, \ldots$*) is given by*

$$\boldsymbol{q}_k^{(\mathrm{T})}(n) = \left(\frac{r_{\mathrm{on}}\lambda_{k,\mathrm{on}}}{\lambda_k}\boldsymbol{q}_{k,\mathrm{on}}^{(\mathrm{T})}(n) + \frac{r_{\mathrm{off}}\lambda_{k,\mathrm{off}}}{\lambda_k}\boldsymbol{q}_{k,\mathrm{off}}^{(\mathrm{T})}(n), 0, \ldots, 0\right),$$

where the $\boldsymbol{q}_{k,\xi}^{(\mathrm{T})}(n)$ *(*$k \in \mathcal{K}$*,* $\xi =$ *on, off,* $n = 0, 1, \ldots$*) is given by*

$$\boldsymbol{q}_{k,\xi}^{(\mathrm{T})}(n) = \frac{1}{\lambda_{k,\xi}} \sum_{\substack{m_1+m_2+m_3\\ +m_4=n}} \boldsymbol{v}_{k,\xi}^{(\mathrm{T})}(m_1)[\boldsymbol{\alpha}_{k,\xi}\otimes\boldsymbol{A}_{k,\xi}^{(\mathrm{T})}(m_2)]\boldsymbol{\Gamma}_{k,\xi}^{(\mathrm{T})}(m_3)\left[\left\{\boldsymbol{P}_{k,\xi}^{m_4}(\boldsymbol{I}-\boldsymbol{P}_{k,\xi})\boldsymbol{e}\right\}\otimes\boldsymbol{I}(M_{\mathrm{on}})\right],$$

if $\lambda_{k,\xi} > 0$*, and otherwise* $\boldsymbol{q}_{k,\xi}^{(\mathrm{T})}(n) = \boldsymbol{0}$*.*

As for the $\boldsymbol{\Gamma}_{k,\xi}^{(\mathrm{T})}(n)$, the following corollary can be obtained from Lemma 4.1.

Corollary C.3 *The* $\boldsymbol{\Gamma}_{k,\xi}^{(\mathrm{T})}(n)$ *(*$k \in \mathcal{K}$*,* $\xi =$ *on, off,* $n = 0, 1, \ldots$*) is determined by the following recursion:*

$$\boldsymbol{\Gamma}_{k,\xi}^{(\mathrm{T})}(0) = \left[\boldsymbol{I} - \boldsymbol{P}_{k,\xi}\otimes\boldsymbol{A}_{k,\xi}^{(\mathrm{T})}(0)\right]^{-1},$$
$$\boldsymbol{\Gamma}_{k,\xi}^{(\mathrm{T})}(n) = \sum_{l=1}^{n} \boldsymbol{\Gamma}_{k,\xi}^{(\mathrm{T})}(n-l)\left[\boldsymbol{P}_{k,\xi}\otimes\boldsymbol{A}_{k,\xi}^{(\mathrm{T})}(l)\right]\boldsymbol{\Gamma}_{k,\xi}^{(\mathrm{T})}(0), \qquad n = 1, 2, \ldots.$$

As for the $\boldsymbol{A}_{k,\xi}^{(\mathrm{T})}(n)$, the following recursions can be obtained from Lemma 4.2 and Theorem 4.4.

Corollary C.4 *The* $\boldsymbol{A}_{k,\xi}^{(\mathrm{T})}(n)$ *is given by*

$$\boldsymbol{A}_{k,\xi}^{(\mathrm{T})}(n) = \sum_{m=0}^{\infty} \gamma_{k,\xi}^{(m)}(\theta_{\mathrm{on}})\boldsymbol{F}_{m}^{(\mathrm{T})}(n), \qquad k \in \mathcal{K},\ \xi = \mathrm{on}, \mathrm{off},\ n = 0, 1, \ldots,$$

where the $\boldsymbol{F}_m^{(\mathrm{T})}(n)$ *is recursively determined by*

$$\boldsymbol{F}_0^{(\mathrm{T})}(n) = \begin{cases} \boldsymbol{I}, & \text{if } n = 0, \\ \boldsymbol{O}, & \text{otherwise}, \end{cases}$$

and for $m = 1, 2, \ldots,$

$$\begin{aligned} \boldsymbol{F}_m^{(\mathrm{T})}(n) = \boldsymbol{F}_{m-1}^{(\mathrm{T})}(n)(\boldsymbol{I} + \theta_{\mathrm{on}}^{-1}\boldsymbol{C}_{\mathrm{on}}) + \theta_{\mathrm{on}}^{-1} \sum_{l=1}^{n} \boldsymbol{F}_{m-1}^{(\mathrm{T})}(n-l) \sum_{k \in \mathcal{K}} \boldsymbol{D}_{k,\mathrm{on}}(l) \\ + \theta_{\mathrm{on}}^{-1} \left[\sum_{l=0}^{n} \boldsymbol{F}_{m-1}^{(\mathrm{T})}(n-l) \boldsymbol{E}_{\mathrm{on,off}} \boldsymbol{N}_{\mathrm{off}}^{(\mathrm{T})}(l) \right] \boldsymbol{E}_{\mathrm{off,on}}, \qquad n = 0, 1, \ldots, \end{aligned}$$

with the $\boldsymbol{N}_{\mathrm{off}}^{(\mathrm{T})}(n)$*, which is given by the following recursion:*

$$\begin{aligned} \boldsymbol{N}_{\mathrm{off}}^{(\mathrm{T})}(0) &= (-\boldsymbol{C}_{\mathrm{off}})^{-1}, \\ \boldsymbol{N}_{\mathrm{off}}^{(\mathrm{T})}(n) &= \left[\sum_{l=1}^{n} \boldsymbol{N}_{\mathrm{off}}^{(\mathrm{T})}(n-l) \sum_{k \in \mathcal{K}} \boldsymbol{D}_{k,\mathrm{off}}(l) \right] \boldsymbol{N}_{\mathrm{off}}^{(\mathrm{T})}(0), \qquad n = 1, 2, \ldots. \end{aligned}$$

Finally, from Theorem 4.5, we obtain the following result.

Corollary C.5 *The* $\boldsymbol{v}_{k,\mathrm{on}}^{(\mathrm{T})}(n)$ *(*$k \in \mathcal{K}$*) and the* $\boldsymbol{v}_{k,\mathrm{off}}^{(\mathrm{T})}(n)$ *(*$k \in \mathcal{K}$*) are determined by*

$$\begin{aligned} \boldsymbol{v}_{k,\mathrm{on}}^{(\mathrm{T})}(n) &= \sum_{m=0}^{\infty} \boldsymbol{v}_{\mathrm{on}}^{(m)}(\theta_{\mathrm{on}}) \boldsymbol{D}_{k,\mathrm{on}} \boldsymbol{F}_m^{(\mathrm{T})}(n), \qquad n = 0, 1, \ldots, \\ \boldsymbol{v}_{k,\mathrm{off}}^{(\mathrm{T})}(n) &= \sum_{m=0}^{\infty} \boldsymbol{v}_{\mathrm{off}}^{(m)}(\theta_{\mathrm{on}}) \boldsymbol{D}_{k,\mathrm{off}} \sum_{l=0}^{n} \boldsymbol{N}_{\mathrm{off}}^{(\mathrm{T})}(n-l) \boldsymbol{E}_{\mathrm{off,on}} \boldsymbol{F}_m^{(\mathrm{T})}(l), \qquad n = 0, 1, \ldots, \end{aligned}$$

respectively.

Acknowledgement

The research of the second author was supported in part by Grant-in-Aid for Scientific Research (C) of Japan Society for the Promotion of Science under Grant No. 15560328.

References

[1] A. Federgruen and L. Green: Queueing systems with service interruptions. *Operations Research*, **34** (1986) 752–768.

[2] A. Federgruen and L. Green: Queueing systems with service interruptions II. *Naval Research Logistics*, **35** (1988) 345–358.

[3] Q.-M. He: Queues with marked customers. *Advances in Applied Probability*, **28** (1996) 567–587.

[4] Q.-M. He: The versatility of MMAP[K] and the MMAP[K]/G[K]/1 queue. *Queueing Systems*, **38** (2001) 397–418.

[5] D. M. Lucantoni, K. S. Meier-Hellstern and M. F. Neuts: A single-server queue with server vacations and a class of non-renewal arrival processes. *Advances in Applied Probability*, **22** (1990) 676–705.

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

[7] H. Masuyama and T. Takine: Analysis and computation of the joint queue length distribution in a FIFO single-server queue with multiple batch Markovian arrival streams. *Stochastic Models*, **19** (2003) 349–381.

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

[9] B. Sengupta: A queue with service interruptions in an alternating random environment. *Operations Research*, **38** (1990) 308–318.

[10] T. Takine and 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.

[11] T. Takine, Y. Matsumoto, T. Suda and T. Hasegawa: Mean waiting times in nonpreemptive priority queues with Markovian arrival and i.i.d. service processes. *Performance Evaluation*, **20** (1994) 131–149.

[12] T. Takine and B. Sengupta: A single server queue with service interruptions. *Queueing Systems*, **26** (1997) 285-300.

[13] T. Takine: A recent progress in algorithmic analysis of FIFO queues with Markovian arrival streams. *Journal of the Korean Mathematical Society*, **38** (2001) 807–842.

[14] T. Takine: Distributional form of Little's law for FIFO queues with multiple Markovian arrival streams and its application to queues with vacations. *Queueing Systems*, **37** (2001) 31–63.

[15] T. Takine: Queue length distribution in a FIFO single-server queue with multiple arrival streams having different service time distributions. *Queueing Systems*, **39** (2001) 349–375.

Hiroyuki Masuyama
Graduate School of Informatics
Kyoto University
Yoshida-honmachi, Sakyo-ku
Kyoto 606-8501, JAPAN
E-mail: masuyama@amp.i.kyoto-u.ac.jp