Hindawi Publishing Corporation
Journal of Applied Mathematics and Stochastic Analysis
Volume 2008, Article ID 396871, 20 pages
doi:10.1155/2008/396871

Research Article

Analysis of MAP/PH(1), PH(2)/2 Queue with Bernoulli Vacations

B. Krishna Kumar,[1] R. Rukmani,[2] and V. Thangaraj[3]

[1] *Department of Mathematics, College of Engineering, Anna University, Chennai 600 025, India*
[2] *Department of Mathematics, Pachaiyappa's College, Chennai 600 030, India*
[3] *Ramanujan Institute for Advanced Study in Mathematics, University of Madras, Chennai 600 005, India*

Correspondence should be addressed to B. Krishna Kumar, drbkkumar@hotmail.com

Received 13 June 2008; Accepted 5 October 2008

Recommended by Ho Lee

We consider a two-heterogeneous-server queueing system with Bernoulli vacation in which customers arrive according to a Markovian arrival process (MAP). Servers returning from vacation immediately take another vacation if no customer is waiting. Using matrix-geometric method, the steady-state probability of the number of customers in the system is investigated. Some important performance measures are obtained. The waiting time distribution and the mean waiting time are also discussed. Finally, some numerical illustrations are provided.

1. Introduction

Queueing systems that allow servers to take vacation have a wide range of applications in many engineering systems such as flexible manufacturing environments, production, computers, communication networks, and telecommunication systems. Servers' vacations are useful for the systems in which the servers want to utilize their idle times for different purposes. For instance, servers' vacations may be due to lack of work, servers' failure, or some other tasks being assigned to the servers which occur in applications like computer maintenance and testing, preventive maintenance job in a production system, priority queue, and so forth (see [1]).

In general, queueing systems with vacations can be classified as systems with a single server or multiple servers involving single vacation or multiple vacations. The servers may take vacations at a random time, after serving utmost k customers (k-limited) or after all the customers in the queue are served (exhaustive). Also, depending on the applications, when the server finishes a vacation and there is no customer to be served, the server may take another vacation (multiple vacation model) or it may wait ready to serve until a new customer arrives (single vacation model). The queueing systems with single and multiple vacations have been first investigated by Levy and Yechiali [2].

Another interesting and important vacation model is Bernoulli vacation scheduling service. The Bernoulli vacation scheduling service discipline has been proposed by Keilson and Servi [3]. In this service discipline, when the server visits a queue, at least one customer, if any, is served. After the completion of its service, the server switches to the next vacation if there are no customers. If customers remain, however, in the queue, the next customer is served with probability q, and the server repeats this procedure, or the server switches to the next vacation with probability $p(= 1 - q)$. The important merit of the Bernoulli vacation scheduling service discipline is the existence of a control parameter q. By adjusting the value of q, we can control the congestion of the system. We note that when $q = 1$, the Bernoulli vacation scheduling service discipline is equivalent to the exhaustive service discipline, and $q = 0$ corresponds to the 1-limited service discipline. Servi [4] has proposed an approximate procedure to calculate the average waiting time of an asymmetric polling system under the Bernoulli scheduling service discipline. Further, Tedijanto [5] has investigated in detail the stochastic behavior of a polling system under the Bernoulli vacation scheduling service discipline and has obtained the average waiting time of a symmetric system. For comprehensive and excellent surveys/monographs on queueing systems with server vacations, see [1, 6–8], and the references therein.

In the literature, there are only a limited number of studies on multiple server vacation models. The multiserver queue with exponentially distributed vacations is first studied by Levy and Yechiali [9]. Using partial generating function technique, the system size has been obtained. Vinod [10] and Kao and Narayanan [11] have discussed the $M/M/c$ queue with multiple vacation of the servers using matrix geometric approach. Further, Chao and Zhao [12] have investigated the multiserver vacation models of both synchronous (servers taking the same vacation together) and asynchronous types and provided some algorithms for computing the stationary probability distributions and expected performance measures. The $M/M/c$ queue with phase-type synchronous vacations has been analyzed by Tian and Li [13]. An $M/M/c$ queue with multiple vacations and 1-limited service has been discussed by Tyagi et al. [14]. Recently, Krishna Kumar and Pavai Madheswari [15] have analyzed an $M/M/c$ queue with Bernoulli vacation scheduling service.

The study on multiserver queueing system generally assumes the servers to be homogeneous in which the individual service rates are the same for all the servers in the system. This assumption may be valid only when the service process is mechanically or electronically controlled. In a queueing system with human servers, the above assumptions can hardly be realized. It is common to observe servers rendering service to identical jobs at different service rates, that is, the service time distributions may be different for different servers. As noted earlier, the study of vacation queueing system incorporates the secondary jobs in the modelling. The analysis of queueing systems with heterogeneous servers and related vacation models helps to study the impact of secondary jobs on system performance.

Neuts and Takahashi [16] have pointed out that for the queueing systems with more than two heterogeneous servers, analytical results are intractable and one may have to use algorithmic approach to study even the asymptotic behavior of the performance measures like stationary distribution of system size and tail probability of waiting time of a customer in the system.

Based on the above observations, in this paper, we consider a vacation queueing system in which customers arrive according to a Markovian arrival process (MAP) involving only two heterogeneous servers availing Bernoulli vacations. The service times follow phase-type distributions (PH distributions), and vacation times of servers follow exponential distributions with heterogeneous vacation rates.

The organization of the paper is as follows. In Section 2, after recalling the definition of the MAP used as input process, we describe the model under investigation. We also study the steady-state probability of system size and some important and interesting performance measures of the system in this section. Next, the waiting time distribution and its related characterizations are discussed in Section 3. Section 4 presents some illustrative numerical examples of the system performance measures. Conclusions are given in Section 5.

2. Source characterization and model description

In this section, we provide information about the Markovian arrival process (MAP) which has been assumed for the customers arrival process. We also describe the model under study.

2.1. Markovian arrival process

In order to allow bursty type traffic in our system, we have chosen the input process as the MAP. The MAP is particularly a tractable point process which is in general nonrenewal and which lends itself very well to modelling bursty arrival processes commonly arising in communications (see [17]). Further, the MAP has made it much easier to develop numerically tractable queueing models which take correlation into account (see [18, 19]). The MAP is a rich class of processes which includes the phase-type renewal process and the Markov-modulated Poisson process as special cases.

An MAP can be considered as a Markov process $\{N(t), \mathbf{J}(t)\}$ on the state space $\{(i, j);\ i \geq 0,\ 1 \leq j \leq l\}$ with an infinitesimal generator $\mathbf{Q}^*$ having the structure

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{2.1}$$

where $\mathbf{D}_0$ and $\mathbf{D}_1$ are $l \times l$ matrices, $\mathbf{D}_0$ has negative diagonal elements and nonnegative off-diagonal elements, $\mathbf{D}_1$ is nonnegative, and $(\mathbf{D}_0 + \mathbf{D}_1)\mathbf{e}_l = 0$, where $\mathbf{e}_l$ is an l-dimensional column vector of ones. An arrival process can be associated with this Markov process as follows. An arrival occurs whenever there is a state transition into a state corresponding to a $\mathbf{D}_1$ block, and there is no arrival otherwise. Here, $\mathbf{N}(t) = i$ represents the number of arrivals in $(0, t)$, and $\mathbf{J}(t) = j$ represents an axillary state or phase variable. Let $\boldsymbol{\eta}$ be the stationary probability vector of the generator $\mathbf{D} = \mathbf{D}_0 + \mathbf{D}_1$. That is, $\boldsymbol{\eta}$ is the unique vector satisfying $\boldsymbol{\eta}(\mathbf{D}_0 + \mathbf{D}_1) = \mathbf{0}$ and $\boldsymbol{\eta}\mathbf{e}_l = 1$. The fundamental arrival rate of this process is given by $\delta = \boldsymbol{\eta}\mathbf{D}_1\mathbf{e}_l$. The constant δ is the expected number of arrivals per unit time in the stationary version of the MAP.

2.2. Model description and analysis

Consider a two-server vacation queueing system in which the service time distributions of the servers are not identical. Customers arrive singly, according to MAP, to an infinite waiting space and form a single waiting line. Customers are served under the first-come-first-served (FCFS) discipline. There are two heterogeneous servers. The service times of server-1 and server-2 are assumed to be phase-type distributions (PH-distributions) PH(1) and PH(2) with representations $(\boldsymbol{\alpha}, \mathbf{S})$ of order m with $\mathbf{S}\mathbf{e}_m = -\mathbf{S}^0$ and $(\boldsymbol{\beta}, T)$ of order n with $\mathbf{T}\mathbf{e}_n = -\mathbf{T}^0$, respectively. We follow matrix formalism of PH-distributions as represented by Neuts [20]. The servers take Bernoulli vacation scheduling service as described by Keilson and Servi [3], that is, after each service completion, the server-i $(i = 1, 2)$ takes a vacation with probability p_i and with probability $q_i = 1 - p_i$, it starts serving the next customer, if any, in the system. If the system is empty, the servers always take vacations. At the end of a vacation period, service commences if a customer is present in the queue. Otherwise, the server takes another vacation immediately and continues in the same manner until it finds at least one customer waiting upon returning from a vacation (multiple vacations). This process holds good for both servers 1 and 2.

The length of vacations' duration $\{V_{i,r};\ r = 1, 2, 3, \ldots\}$ of the two servers is assumed to be independent and identically distributed exponential random variables with parameters, θ_i for $i = 1, 2$, and is independent of the length of the service times and the arrival process. The Bernoulli vacation scheduling service queueing model with heterogeneous servers under consideration can be formulated as a continuous time Markov chain (CTMC). By appropriately keeping track of the various states, such as the number of customers in the system, the phase of the arrival process, the phases of the service processes of server-1 and server-2, and the status of the server-1 and server-2, the state space of the Markov chain describing the model is given as follows.

The set of states $\{(k, 0, j_0);\ k \geq 0;\ 1 \leq j_0 \leq l\}$ represents that there are k customers in the system, both servers are on vacation and the phase of the arrival process is j_0; the set of states $\{(k, 1, j_0, j_1);\ k \geq 1;\ 1 \leq j_0 \leq l;\ 1 \leq j_1 \leq m\}$ indicates that there are k customers in the system, server-1 is busy in the system serving a customer in phase j_1 while server-2 is on vacation, and j_0 represents the phase of the arrival process; the set of states $\{(k, 2, j_0, j_2);\ k \geq 1;\ 1 \leq j_0 \leq l;\ 1 \leq j_2 \leq n\}$ indicates that there are k customers in the system, server-2 is busy in the system serving a customer in phase j_2 while server-1 is on vacation, and the arrival process is in phase j_0; the set of states $\{(k, 3, j_0, j_1, j_2);\ k \geq 2;\ 1 \leq j_0 \leq l;\ 1 \leq j_1 \leq m;\ 1 \leq j_2 \leq n\}$ represents that there are k customers in the system, both servers 1 and 2 are busy in the system serving the customers in phases j_1 and j_2, respectively, and the arrival process is in phase j_0.

We define levels $\mathbf{0}, \mathbf{1}, \mathbf{2}, \ldots, \mathbf{k}, \ldots$ as the set of states:

$$\begin{aligned}
\mathbf{0} &= \{(0, j_0);\ 1 \leq j_0 \leq l\}, \\
\mathbf{1} &= \{(1, 0, j_0) \cup (1, 1, j_0, j_1) \cup (1, 2, j_0, j_2);\ 1 \leq j_0 \leq l;\ 1 \leq j_1 \leq m;\ 1 \leq j_2 \leq n\}, \\
\mathbf{k} &= \{(k, 0, j_0) \cup (k, 1, j_0, j_1) \cup (k, 2, j_0, j_2) \cup (k, 3, j_0, j_1, j_2); \\
&\qquad k \geq 2;\ 1 \leq j_0 \leq l;\ 1 \leq j_1 \leq m;\ 1 \leq j_2 \leq n\},
\end{aligned} \tag{2.2}$$

where the elements of the sets are arranged in lexicographical order.

In the sequel, we use the following notations.

$\mathbf{e}_n$: a column vector of order $n \times 1$ with all its elements equal to 1.

$\mathbf{e} = \mathbf{e}_{l+lm+ln+lmn}$: a column vector of order $(l + lm + ln + lmn) \times 1$ with all its elements equal to 1.

$\mathbf{e}_{l+lm+ln}(l)$: a column vector of order $(l + lm + ln) \times 1$ with the 1st l elements equal to 1, and other elements are zeros.

$\mathbf{e}_{l+lm+ln}(lm)$: a column vector of order $(l+lm+ln)\times 1$ with the $(l+1)$st to $(l+lm)$th elements equal to 1, and other elements are zeros.

$\mathbf{e}_{l+lm+ln}(ln)$: a column vector of order $(l+lm+ln)\times 1$ with $(l+lm+1)$st to $(l+lm+ln)$th elements equal to 1, and other elements are zeros.

$\mathbf{e}_{l+lm+ln+lmn}(l)$: a column vector of order $(l+lm+ln+lmn)\times 1$ with the 1st to l elements equal to 1, and other elements are zeros.

$\mathbf{e}_{l+lm+ln+lmn}(lm)$: a column vector of order $(l+lm+ln+lmn)\times 1$ with the $(l+1)$st to $(l+lm)$th elements equal to 1, and other elements are zeros.

$\mathbf{e}_{l+lm+ln+lmn}(ln)$: a column vector of order $(l+lm+ln+lmn)\times 1$ with $(l+lm+1)$st to $(l+lm+ln)$th elements equal to 1, and other elements are zeros.

$\mathbf{e}_{l+lm+ln+lmn}(lmn)$: a column vector of order $(l+lm+ln+lmn)\times 1$ with $(l+lm+ln+1)$st to $(l+lm+ln+lmn)$th elements equal to 1, and other elements are zeros.

Using elementary arguments, the infinite-state Markov chain for the model under study has a transition rate matrix $\mathbf{Q}$ which has a block-tridiagonal structure given by

$$\mathbf{Q}=\begin{bmatrix} \mathbf{B}_{00} & \mathbf{B}_{01} & & & & \\ \mathbf{B}_{10} & \mathbf{B}_{11} & \mathbf{B}_{12} & & & \\ & \mathbf{B}_{21} & \mathbf{A}_1 & \mathbf{A}_0 & & \\ & & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \\ & & & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 \\ & & & & \ddots & \ddots & \ddots \end{bmatrix}. \tag{2.3}$$

The entries of $\mathbf{Q}$ are given by the following block matrices. The boundary matrices are defined by

$$\begin{aligned} \mathbf{B}_{00} &= \mathbf{D}_0, \qquad \mathbf{B}_{01} = [\mathbf{D}_1, \mathbf{0}, \mathbf{0}], \qquad \mathbf{B}_{10} = \left[\mathbf{0}, \mathbf{S}^0\otimes \mathbf{I}_l, \mathbf{T}^0\otimes\mathbf{I}_l\right]^T, \\ \mathbf{B}_{11} &= \begin{bmatrix} \mathbf{D}_0-(\theta_1+\theta_2)\mathbf{I}_l & \theta_1\boldsymbol{\alpha}\otimes\mathbf{I}_l & \theta_2\boldsymbol{\beta}\otimes\mathbf{I}_l \\ \mathbf{0} & \mathbf{S}\oplus\mathbf{D}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}\oplus\mathbf{D}_0 \end{bmatrix}, \\ \mathbf{B}_{12} &= \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m\otimes\mathbf{D}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_n\otimes\mathbf{D}_1 & \mathbf{0} \end{bmatrix}, \\ \mathbf{B}_{21} &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ p_1\mathbf{S}^0\otimes\mathbf{I}_l & q_1\mathbf{S}^0\boldsymbol{\alpha}\otimes\mathbf{I}_l & \mathbf{0} \\ p_2\mathbf{T}^0\otimes\mathbf{I}_l & \mathbf{0} & q_2\mathbf{T}^0\boldsymbol{\beta}\otimes\mathbf{I}_l \\ \mathbf{0} & \mathbf{T}_0\otimes(\mathbf{I}_m\otimes\mathbf{I}_l) & \mathbf{I}_n\otimes(\mathbf{S}^0\otimes\mathbf{I}_l) \end{bmatrix}. \end{aligned} \tag{2.4}$$

The square matrices $\mathbf{A}_0$, $\mathbf{A}_1$, and $\mathbf{A}_2$ are of order $l + lm + ln + lmn$ and are given by

$$\mathbf{A}_0 = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \otimes \mathbf{D}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_n \otimes \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_n \otimes \mathbf{I}_m \otimes \mathbf{D}_1 \end{bmatrix},$$

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{D}_0 - (\theta_1 + \theta_2)\mathbf{I}_l & \theta_1 \boldsymbol{\alpha} \otimes \mathbf{I}_l & \theta_2 \boldsymbol{\beta} \otimes \mathbf{I}_l & \mathbf{0} \\ \mathbf{0} & (\mathbf{S} - \theta_2 \mathbf{I}_m) \oplus \mathbf{D}_0 & \mathbf{0} & \theta_2 \boldsymbol{\beta} \otimes (\mathbf{I}_m \otimes \mathbf{I}_l) \\ \mathbf{0} & \mathbf{0} & (\mathbf{T} - \theta_1 \mathbf{I}_n) \oplus \mathbf{D}_0 & \mathbf{I}_n \otimes (\theta_1 \boldsymbol{\alpha} \otimes \mathbf{I}_l) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & (\mathbf{T} \oplus \mathbf{S}) \oplus \mathbf{D}_0 \end{bmatrix}, \tag{2.5}$$

$$\mathbf{A}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ p_1 \mathbf{S}^0 \otimes \mathbf{I}_l & q_1 \mathbf{S}^0 \boldsymbol{\alpha} \otimes \mathbf{I}_l & \mathbf{0} & \mathbf{0} \\ p_2 \mathbf{T}^0 \otimes \mathbf{I}_l & \mathbf{0} & q_2 \mathbf{T}^0 \boldsymbol{\beta} \otimes I_l & \mathbf{0} \\ \mathbf{0} & p_2 \mathbf{T}^0 \otimes (\mathbf{I}_m \otimes \mathbf{I}_l) & \mathbf{I}_n \otimes (p_1 S^0 \otimes I_l) & (q_2 T^0 \boldsymbol{\beta} \oplus q_1 S^0 \boldsymbol{\alpha}) \otimes \mathbf{I}_l \end{bmatrix},$$

where $\otimes$ and $\oplus$ are the Kronecker product and Kronecker sum, respectively (see [21]).

We now derive the condition for the system to reach steady state. To accomplish this, we define $\mathbf{A} = \mathbf{A}_0 + \mathbf{A}_1 + \mathbf{A}_2$. Then, the matrix $\mathbf{A}$ can be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{D} - (\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2) I_l & \boldsymbol{\theta}_1 \boldsymbol{\alpha} \otimes I_l & \boldsymbol{\theta}_2 \boldsymbol{\beta} \otimes I_l & \mathbf{0} \\ p_1 \mathbf{S}^0 \otimes \mathbf{I}_l & (\mathbf{S} + q_1 \mathbf{S}^0 \boldsymbol{\alpha} - \boldsymbol{\theta}_2 I_m) \oplus \mathbf{D} & \mathbf{0} & \boldsymbol{\theta}_2 \boldsymbol{\beta} \otimes (\mathbf{I}_m \otimes I_l) \\ p_2 \mathbf{T}^0 \otimes I_l & \mathbf{0} & (\mathbf{T} + q_2 T^0 \boldsymbol{\beta} - \boldsymbol{\theta}_1 I_n) \oplus \mathbf{D} & \mathbf{I}_n \otimes (\boldsymbol{\theta}_1 \boldsymbol{\alpha} \otimes \mathbf{I}_l) \\ \mathbf{0} & p_2 \mathbf{T}^0 \otimes (\mathbf{I}_m \otimes \mathbf{I}_l) & \mathbf{I}_n \otimes (p_1 \mathbf{S}^0 \otimes \mathbf{I}_l) & (\mathbf{T} + q_2 \mathbf{T}^0 \boldsymbol{\beta}) \oplus (\mathbf{S} + q_1 \mathbf{S}^0 \boldsymbol{\alpha}) \oplus \mathbf{D} \end{bmatrix}, \tag{2.6}$$

where $\mathbf{D} = \mathbf{D}_0 + \mathbf{D}_1$.

It is clear that the order of the square matrix $\mathbf{A}$ is $l+lm+ln+lmn$, and it is an irreducible infinitesimal generator matrix [20, page 82], and so there exists $1 \times (l+lm+ln+lmn)$ stationary probability vector $\boldsymbol{\pi}$ of $\mathbf{A}$ satisfying $\boldsymbol{\pi} \boldsymbol{A} = \mathbf{0}$ and $\boldsymbol{\pi} \boldsymbol{e} = 1$. The vector $\boldsymbol{\pi}$ is denoted by $\boldsymbol{\pi} = (\boldsymbol{\pi}(0), \boldsymbol{\pi}(1), \boldsymbol{\pi}(2), \boldsymbol{\pi}(3))$ whose components are

$$\begin{aligned} \boldsymbol{\pi}(0) &= (\pi(k,0,1), \pi(k,0,2), \ldots, \pi(k,0,j_0), \ldots, \pi(k,0,l)) \quad \text{for } k \geq 0, \\ \boldsymbol{\pi}(1) &= (\pi(k,1,1), \pi(k,2,1), \ldots, \pi(k,j_0,j_1), \ldots, \pi(k,l,m)) \quad \text{for } k \geq 1, \\ \boldsymbol{\pi}(2) &= (\pi(k,1,1), \pi(k,2,1), \ldots, \pi(k,j_0,j_2), \ldots, \pi(k,l,n)) \quad \text{for } k \geq 1, \\ \boldsymbol{\pi}(3) &= \pi(k,1,1,1), \pi(k,2,1,1), \ldots, \pi(k,j_0,j_1,j_2), \ldots, \pi(k,l,m,n) \quad \text{for } k \geq 2. \end{aligned} \tag{2.7}$$

Here, $\boldsymbol{\pi}(0)$ is the stationary probability vector when both servers are on vacation, $\boldsymbol{\pi}(1)$ is the stationary probability vector when server-1 is busy and server-2 is on vacation, $\boldsymbol{\pi}(2)$ is the stationary probability vector when server-2 is busy and server-1 is on vacation and $\boldsymbol{\pi}(3)$ is the stationary probability vector when both servers are busy. Further, $\pi(k,0,j_0)$ is the stationary probability that there are k $(k \geq 0)$ customers in the system, both servers are on vacation, and the underlying arrival process MAP is in phase j_0 $(1 \leq j_0 \leq l)$; $\pi(k, j_0, j_1)$ is the stationary probability that there are k $(k \geq 1)$ customers in the system with the underlying arrival process MAP in phase j_0 $(1 \leq j_0 \leq l)$, the server-1 is busy serving a customer with the underlying service process PH(1) in phase j_1 $(1 \leq j_1 \leq m)$, and server-2 is on vacation; $\pi(k, j_0, j_2)$ is the stationary probability that there are k $(k \geq 1)$ customers in the system with the underlying arrival process MAP in phase j_0 $(1 \leq j_0 \leq l)$, server-2 is busy serving a customer with the underlying service process PH(2) in phase j_2 $(1 \leq j_2 \leq n)$, and server-1 is on vacation. Finally, $\pi(k, j_0, j_1, j_2)$ is the stationary probability that there are k $(k \geq 2)$ customers in the system with the underlying arrival process MAP in phase j_0 $(1 \leq j_0 \leq l)$, both server-1 and server-2 are busy serving customers with the underlying service processes PH(1) in phase j_1 $(1 \leq j_1 \leq m)$ and PH(2) in phase j_2 $(1 \leq j_2 \leq n)$, respectively.

As the Markov process has the QBD structure, it is well known [20, page 83] that the standard drift condition

$$\boldsymbol{\pi}\mathbf{A}_0\mathbf{e} < \boldsymbol{\pi}\mathbf{A}_2\mathbf{e} \tag{2.8}$$

is the necessary and sufficient condition for the stability of a QBD process.

After some algebraic manipulation, the stability condition turns out to be

$$\begin{aligned}
&\boldsymbol{\pi}(0)D_1\mathbf{e}_l + \boldsymbol{\pi}(1)(\mathbf{e}_m \otimes \mathbf{D}_1\mathbf{e}_l) + \boldsymbol{\pi}(2)(\mathbf{e}_n \otimes \mathbf{D}_1\mathbf{e}_l) + \boldsymbol{\pi}(3)(\mathbf{e}_{nm} \otimes \mathbf{D}_1\mathbf{e}_l) \\
&\quad < \boldsymbol{\pi}(1)(\mathbf{S}^0 \otimes \mathbf{e}_l) + \boldsymbol{\pi}(2)(T^0 \otimes \mathbf{e}_l) + \boldsymbol{\pi}(3)\left[(T^0 \otimes \mathbf{e}_m) \oplus (\mathbf{e}_n \otimes \mathbf{S}^0)\right] \otimes \mathbf{e}_l.
\end{aligned} \tag{2.9}$$

Remark 2.1. As the LHS of (2.9) is the rate of flow into the system, and RHS of (2.9) is the maximum rate of flow out of the system, (2.9) should be a necessary and sufficient condition for positive recurrence.

Under stability condition (2.9) of the system, there exists the steady-state probability vector $\mathbf{X}$ satisfying $\mathbf{XQ} = \mathbf{0}$, $\mathbf{X}\mathbf{e}_\infty = 1$. The stationary probability vector $\mathbf{X}$, partitioned as $\mathbf{X} = (\mathbf{X}(0), \mathbf{X}(1), \mathbf{X}(2), \ldots)$, is given by

$$\mathbf{X}(0)\mathbf{B}_{00} + \mathbf{X}(1)B_{10} = \mathbf{0}, \tag{2.10}$$

$$\mathbf{X}(0)\mathbf{B}_{01} + \mathbf{X}(1)B_{11} + \mathbf{X}(2)B_{21} = \mathbf{0}, \tag{2.11}$$

$$\mathbf{X}(1)\mathbf{B}_{12} + \mathbf{X}(2)(A_1 + RA_2) = \mathbf{0}, \tag{2.12}$$

$$\mathbf{X}(k) = \mathbf{X}(2)R^{k-2}, \quad k = 3, 4, 5, \ldots \tag{2.13}$$

and by the normalizing equation

$$\mathbf{X}(0)\mathbf{e}_l + \mathbf{X}(1)\mathbf{e}_{l+lm+ln} + \mathbf{X}(2)(I - R)^{-1}\mathbf{e}_{l+lm+ln+lmn} = 1, \tag{2.14}$$

where

$$\begin{aligned}
\mathbf{X}(0) &= (x(0,0,1), x(0,0,2), \dots, x(0,0,j_0), \dots, x(0,0,l)), \\
\mathbf{X}(1) &= (x(1,0,1), x(1,0,2), \dots, x(1,0,j_0), \dots, x(1,0,l), \\
&\qquad x(1,1,1,1), x(1,1,2,1), \dots, x(1,1,j_0,j_1), \dots, x(1,1,l,m), \\
&\qquad x(1,2,1,1), x(1,2,2,1), \dots, x(1,2,j_0,j_2), \dots, x(1,2,l,n)),
\end{aligned} \tag{2.15}$$

and for $k \geq 2$,

$$\begin{aligned}
\mathbf{X}(k) = (&x(k,0,1), x(k,0,2), \dots, x(k,0,j_0), \dots, x(k,0,l), \\
&x(k,1,1,1), x(k,1,2,1), \dots, x(k,1,j_0,j_1), \dots, x(k,1,l,m), \\
&x(k,2,1,1), x(k,2,2,1), \dots, x(k,2,j_0,j_2), \dots, x(k,2,l,n), \\
&x(k,3,1,1,1), x(k,3,2,1,1), \dots, x(k,3,j_0,j_1,j_2), \dots, x(k,3,l,m,n)).
\end{aligned} \tag{2.16}$$

Here, $x(k,0,j_0)$ refers to the joint probability that there are k $(k \geq 0)$ customers in the system, both servers are on vacation and j_0 $(1 \leq j_0 \leq l)$ corresponds to the phase of the arrival process; $x(k,1,j_0,j_1)$ refers to the joint probability that there are k $(k \geq 1)$ customers in the system, server-1 is busy serving a customer, server-2 is on vacation, j_0 $(1 \leq j_0 \leq l)$ corresponds to the phase of the arrival process, and j_1 $(1 \leq j_1 \leq m)$ corresponds to the phase of the service process PH(1) provided by server-1; $x(k,2,j_0,j_2)$ refers to the joint probability that there are k $(k \geq 1)$ customers in the system, server-2 is busy serving a customer, server-1 is on vacation, j_0 $(1 \leq j_0 \leq l)$ corresponds to the phase of the arrival process, and j_2 $(1 \leq j_2 \leq n)$ corresponds to the phase of the service process PH(2) provided by server-2; $x(k,3,j_0,j_1,j_2)$ refers to the joint probability that there are k $(k \geq 2)$ customers in the system, both servers are busy serving customers, j_0 $(1 \leq j_0 \leq l)$ corresponds to the phase of the arrival process, j_1 $(1 \leq j_1 \leq m)$ and j_2 $(1 \leq j_2 \leq n)$ correspond to phases of the underlying service processes PH(1) and PH(2) provided by server-1 and server-2, respectively. $\mathbf{I}$ is the identity matrix of order $l + lm + ln + lmn$, and the matrix R is the minimal nonnegative solution with spectral radius less than 1, that is, $\mathbf{S}p(R) < 1$ of the matrix quadratic equation [20, pages 82-83]:

$$\mathbf{R}^2\mathbf{A}_2 + \mathbf{R}\mathbf{A}_1 + \mathbf{A}_0 = \mathbf{0}. \tag{2.17}$$

Since the system is stable, and the square matrices $\mathbf{A}_0$, $\mathbf{A}_1$, and $\mathbf{A}_2$ are of order $l + lm + ln + lmn$, $\mathbf{R}$ is also a square matrix of order $l + lm + ln + lmn$ and is obtained from the above matrix quadratic equation and from the following relation:

$$\mathbf{R}\mathbf{A}_2\mathbf{e} = \mathbf{A}_0\mathbf{e}. \tag{2.18}$$

Equation (2.18) implies that the rate of transition from a state with k customers to a state with $k + 1$ matches the transition rate from k to $k - 1$.

The matrix $\mathbf{R}$ is approximated by the following iterations :

$$\begin{aligned} \mathbf{R}(0) &= \mathbf{0}, \\ \mathbf{R}(n+1) &= -\mathbf{A}_0\mathbf{A}_1^{-1} - \mathbf{R}^2(n)\mathbf{A}_2\mathbf{A}_1^{-1}, \quad \text{for } n \geq 0. \end{aligned} \tag{2.19}$$

The values of $\mathbf{R}$ will converge, since $-\mathbf{A}_1^{-1}$ and $(\mathbf{A}_0 + \mathbf{R}^2\mathbf{A}_2)$ are positive. Hence, after each iteration, the elements of $\mathbf{R}$ will increase monotonically. Iteration will be continued until $\max_{i,j}[R_{ij}(n+1) - R_{ij}(n)] < \epsilon$ is satisfied, where $\mathbf{R}(n)$ is the value of $\mathbf{R}$ in the nth iteration, and ϵ is the degree of accuracy required. The accuracy may be checked by (2.18) with $\epsilon = 10^{-12}$ [20].

The boundary probability vectors $\mathbf{X}(0)$, $\mathbf{X}(1)$, $\mathbf{X}(2)$ and the probability vectors $\mathbf{X}(k)$, $k \geq 3$, can be obtained form (2.10) to (2.14). These steady-state joint probability vectors are then used to find the following system performance measures.

2.3. Performance measures

We will list some important performance measures which are used to bring out the qualitative behavior of the queueing model under study.

(1) The mean and second moments of the number of customers in the system can be obtained exactly as

$$\begin{aligned} E(\mathbf{N}) &= \mathbf{X}(1)\mathbf{e}_{l+lm+ln} + \mathbf{X}(2)\left[2(\mathbf{I}-\mathbf{R})^{-1} + \mathbf{R}(\mathbf{I}-\mathbf{R})^{-2}\right]\mathbf{e}, \\ E(\mathbf{N}^2) &= \mathbf{X}(1)\mathbf{e}_{l+lm+ln} + 4\mathbf{X}(2)\mathbf{e} - \mathbf{X}(2)\left[(\mathbf{I}-\mathbf{R})^{-1} + (\mathbf{I}-\mathbf{R})^{-2} + 2(\mathbf{I}-\mathbf{R})^{-3} - 4\mathbf{I}\right]\mathbf{e}, \end{aligned} \tag{2.20}$$

where $\mathbf{N}$ is the system size at an arbitrary time. The variance of the system can also be found.

(2) The probability that no customer in the system and both servers are on vacation is given by

$$P_0 = \mathbf{X}(0)\mathbf{e}_l. \tag{2.21}$$

(3) The probability that both the servers are on vacation is given as

$$V_0 = \mathbf{X}(0)\mathbf{e}_l + \mathbf{X}(1)\mathbf{e}_{l+lm+ln}(l) + \mathbf{X}_2(\mathbf{I}-\mathbf{R})^{-1}\mathbf{e}_{l+lm+ln+lmn}(l). \tag{2.22}$$

(4) The probability that server-1 is busy serving a customer and server-2 on vacation is given by

$$P_1 = \mathbf{X}(1)\mathbf{e}_{l+lm+ln}(lm) + \mathbf{X}(2)(I-R)^{-1}\mathbf{e}_{l+lm+ln+lmn}(lm). \tag{2.23}$$

(5) The probability that server-2 is busy serving a customer and server-1 on vacation is given as

$$P_2 = \mathbf{X}(1)\mathbf{e}_{l+lm+ln}(ln) + \mathbf{X}(2)(I-R)^{-1}\mathbf{e}_{l+lm+ln+lmn}(ln). \tag{2.24}$$

(6) The probability that both servers are busy is obtained as

$$P_{12} = \mathbf{X}(2)(\mathbf{I}-\mathbf{R})^{-1}e_{l+lm+ln+lmn}(lmn). \tag{2.25}$$

(7) The mean number of customers $E(N_V)$ present in the system when both the servers are on vacation is given as

$$E(\mathbf{N}_V) = \mathbf{X}(1)e_{l+lm+ln}(l) + \mathbf{X}(2)\left[2(\mathbf{I}-\mathbf{R})^{-1} + R(I-R)^{-2}\right]\mathbf{e}_{l+lm+ln+lmn}(l). \tag{2.26}$$

(8) The mean number of customers $E(\mathbf{N}_B)$ present in the system when both the servers are busy serving customers is given as

$$E(\mathbf{N}_B) = \mathbf{X}(2)\left[2(\mathbf{I}-\mathbf{R})^{-1} + \mathbf{R}(\mathbf{I}-\mathbf{R})^{-2}\right]\mathbf{e}_{l+lm+ln+lmn}(lmn). \tag{2.27}$$

(9) The mean number of customers $E(\mathbf{N}_1)$ present in the system when server-1 is busy and server-2 on vacation is given as

$$E(\mathbf{N}_1) = \mathbf{X}(1)\mathbf{e}_{l+lm+ln}(lm) + \mathbf{X}(2)\left[2(\mathbf{I}-\mathbf{R})^{-1} + \mathbf{R}(\mathbf{I}-\mathbf{R})^{-2}\right]\mathbf{e}_{l+lm+ln+lmn}(lm). \tag{2.28}$$

(10) The mean number of customers $E(\mathbf{N}_2)$ present in the system when server-1 on vacation and server-2 being busy is obtained as

$$E(\mathbf{N}_2) = \mathbf{X}(1)\mathbf{e}_{l+lm+ln}(ln) + \mathbf{X}(2)\left[(\mathbf{I}-\mathbf{R})^{-1} + 2(\mathbf{I}-\mathbf{R})^{-2}\right]\mathbf{e}_{l+lm+ln+lmn}(ln). \tag{2.29}$$

3. Waiting time distribution

We now analyze the waiting time of an arriving customer in the queue by first-passage time analysis. We then show how the mean waiting time can be determined.

Let $\mathbf{W}(t)$, $t \geq 0$, be the distribution function for the waiting time in the queue of an arriving (tagged) customer, that is, $\mathbf{W}(t)$ denotes the probability vector of order $1 \times mn$ that an arriving tagged customer has to wait utmost t time units until it is served. For the multiserver queue with Bernoulli vacation scheduling service, we have $\mathbf{W}(0+) = \mathbf{0}$, because an arrival must either wait for a service completion or a vacation termination of servers. States corresponding to the number of customers are present in the system $\{0, \mathbf{1}, \mathbf{2}, \ldots\}$, and an absorbing state $\{*\}$forms the state space of the CTMC. Thus, the state space of the CTMC is $\{*\} \cup \{0, \mathbf{1}, \mathbf{2}, \mathbf{3}, \ldots\}$. Note that the customer arrival process is not required, since we are discussing an FCFS queue discipline, and hence, customers that arrive after the tagged customer do not have any impact on the waiting time of the tagged customer. Therefore, the absorbing state $\{*\}$ is a vector of order $1 \times mn$ given by $* = ((*,1,1),(*,2,1),\ldots,(*,j_1,j_2),\ldots(*,m,n))$, level 0 has the single component $(0,0)$, level $\mathbf{1}$ is a vector of order $1 \times (1+m+n)$ given by

$$\begin{aligned}\mathbf{1} = \{&(1,0),(1,1,1),(1,1,2),\ldots,(1,1,j_1),\ldots,(1,1,m),\\ &(1,2,1),(1,2,2),\ldots,(1,2,j_2),\ldots,(1,2,n)\},\end{aligned} \tag{3.1}$$

and levels $\mathbf{i} \geq 2$ are vectors of order $1 \times (1+m+n+mn)$ given by

$$\begin{aligned}\mathbf{i} = \{&(i,0),(i,1,1),(i,1,2),\ldots,(i,1,j_1),\ldots,(i,1,m),\\ &(i,2,1),(i,2,2),\ldots,(i,2,j_2),\ldots,(i,2,n),\\ &(i,3,1,1),(i,3,2,1),\ldots,(i,3,j_1,j_2),\ldots,(i,3,m,n)\},\end{aligned} \tag{3.2}$$

where j_1 and j_2 are the phases of the service time distributions PH(1) and PH(2), respectively.

On entering the absorbing state $*$, a tagged customer starts receiving service. Clearly, this happens at the arrival of a server either from vacation or after the completion of a service when the customer is at the head of the queue. The transition rate matrix Q_1 for this absorbing Markov chain is given by

$$Q_1 = \begin{array}{c} \\ * \\ 0 \\ \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccc} * & 0 & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \cdots \end{array} \\ \begin{bmatrix} 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{C}_0 & \mathbf{E}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{C}_1 & \mathbf{E}_{10} & \mathbf{E}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{C}_2 & \mathbf{0} & \mathbf{E}_{21} & \mathbf{E} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{E}_2 & \mathbf{E} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{E}_2 & \mathbf{E} & \\ \vdots & \vdots & \vdots & \vdots & & \ddots & \ddots \end{bmatrix} \end{array}, \tag{3.3}$$

where

$$\begin{aligned}
\mathbf{C}_0 &= e'_n \otimes \boldsymbol{\theta}_1 \alpha + \boldsymbol{\theta}_2 \beta \otimes e'_m,\\
\mathbf{C}_1 &= \left[\mathbf{0},\, e'_n \otimes q_1 S^0 \alpha + \boldsymbol{\theta}_2 \boldsymbol{\beta} \otimes \mathbf{e}_m \mathbf{e}'_m,\, e_n e'_n \otimes \boldsymbol{\theta}_1 \boldsymbol{\alpha} + q_2 T^0 \boldsymbol{\beta} \otimes e'_m\right]^T,\\
\mathbf{C}_2 &= \left[\mathbf{0},\, \mathbf{0},\, \mathbf{0},\, \mathbf{e}_n e'_n \otimes q_1 S^0 \alpha + q_2 T^0 \boldsymbol{\beta} \otimes \mathbf{e}_m e'_m\right]^T,\\
\mathbf{E}_0 &= -(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2), \qquad \mathbf{E}_{10} = \left[0, p_1 S^0, p_2 T^0\right]^T,
\end{aligned}$$

$$\mathbf{E}_1 = \begin{bmatrix} -(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2) & \boldsymbol{\theta}_1\boldsymbol{\alpha} & \boldsymbol{\theta}_2\boldsymbol{\beta} \\ \mathbf{0} & S - \boldsymbol{\theta}_2\mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & T - \boldsymbol{\theta}_1\mathbf{I}_n \end{bmatrix},$$

$$\mathbf{E}_{21} = \begin{bmatrix} 0 & \mathbf{0} & \mathbf{0} \\ p_1\mathbf{S}^0 & q_1\mathbf{S}^0\boldsymbol{\alpha} & \mathbf{0} \\ p_2\mathbf{T}^0 & \mathbf{0} & q_2\mathbf{T}^0\boldsymbol{\beta} \\ 0 & p_2\mathbf{T}^0 \otimes \mathbf{I}_m & \mathbf{I}_n \otimes p_1\mathbf{S}^0 \end{bmatrix},$$

$$\mathbf{E}_2 = \begin{bmatrix} 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ p_1S^0 & q_1S^0\boldsymbol{\alpha} & \mathbf{0} & \mathbf{0} \\ p_2T^0 & \mathbf{0} & q_2T^0\boldsymbol{\beta} & \mathbf{0} \\ \mathbf{0} & p_2\mathbf{T}^0 \otimes \mathbf{I}_m & \mathbf{I}_n \otimes p_1S^0 & q_2T^0\boldsymbol{\beta} \oplus q_1\mathbf{S}^0\boldsymbol{\alpha} \end{bmatrix},$$

$$\mathbf{E} = \begin{bmatrix} -(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2) & \boldsymbol{\theta}_1\boldsymbol{\alpha} & \boldsymbol{\theta}_2\boldsymbol{\beta} & \mathbf{0} \\ \mathbf{0} & (\mathbf{S} - \boldsymbol{\theta}_2\mathbf{I}_m) & \mathbf{0} & \boldsymbol{\theta}_2\boldsymbol{\beta} \otimes \mathbf{I}_m \\ \mathbf{0} & \mathbf{0} & (\mathbf{T} - \boldsymbol{\theta}_1\mathbf{I}_n) & \mathbf{I}_n \otimes \boldsymbol{\theta}_1\boldsymbol{\alpha} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{T} \oplus \mathbf{S} \end{bmatrix}. \tag{3.4}$$

It is observed that the customer arrival rates have not appeared in matrix Q_1.

To obtain the distribution of waiting time $\mathbf{W}(t)$, $t \geq 0$, of the tagged customer, the first step is to determine the stationary probability distribution of the system's state immediately after its arrival. This is actually the stationary distribution of the number of customers in the system as seen by this tagged customer at its arrival time. Denote by $\mathbf{Y}(0) = (\mathbf{Y}_0(0), \mathbf{Y}_1(0), \mathbf{Y}_2(0), \ldots)$ the probability distribution which can be obtained from $\mathbf{X} = (\mathbf{X}(0), \mathbf{X}(1), \mathbf{X}(2), \ldots)$ by a standard method, that is, $Y(0)$ can be interpreted as a conditional probability distribution of the system's state conditioned on the occurrence of the tagged customer arrival. Due to the Markovian property of the arrival process, it is seen that the arrival-stationary probability distribution of the number of customers in the system is given by

$$\mathbf{Y}_0(0) = \mathbf{X}(0)\frac{\mathbf{D}_1\mathbf{e}_l}{\delta}, \qquad \mathbf{Y}_1(0) = \mathbf{X}(1)\left(\mathbf{I}_{l+lm+ln} \otimes \frac{\mathbf{D}_1\mathbf{e}_l}{\delta}\right),$$
$$\mathbf{Y}_k(0) = \mathbf{X}(k)\left(\mathbf{I}_{l+lm+ln+lmn} \otimes \frac{D_1e_l}{\delta}\right) \quad \text{for } k \geq 2, \tag{3.5}$$

where δ is the fundamental arrival rate of the MAP as given in Section 2.1.

Now we define $\mathbf{Y}(t) = (\mathbf{Y}_*(t), \mathbf{Y}_0(t), \mathbf{Y}_1(t), \ldots)$, where $\mathbf{Y}_i(t)$ is a row vector of order $1 \times (1 + m + n + mn)$, when $i \geq 2$, $Y_0(t)$ is of order 1×1 and $\mathbf{Y}_1(t)$ is a row vector of order $1 \times (1 + m + n)$, and its elements represent the probability that at time t, the CTMC with generator $\mathbf{Q}_1$ is in the respective state of level $\mathbf{i}$. Clearly, $\mathbf{Y}_*(t)$ gives the probability that the tagged customer is in the absorbing state at time t. Hence, the vector waiting time distribution is $\mathbf{W}(t) = \mathbf{Y}_*(t)$ for $t \geq 0$. In finding the waiting time distribution of the tagged customer at arrival time, it is assumed that the process starts with initial probability vector $\mathbf{Y}(0) = (\mathbf{Y}_0(0), \mathbf{Y}_1(0), \mathbf{Y}_2(0), \ldots)$.

The differential equation $\mathbf{Y}'(t) = \mathbf{Y}(t)\mathbf{Q}_1$ for $t \geq 0$ reduces to

$$\begin{aligned}
\mathbf{Y}'_*(t) &= \sum_{k=0}^{2} \mathbf{Y}_k(t)\mathbf{C}_k, \\
\mathbf{Y}'_0(t) &= \mathbf{Y}_0(t)\mathbf{E}_0 + Y_1(t)E_{10}, \\
\mathbf{Y}'_1(t) &= \mathbf{Y}_1(t)\mathbf{E}_1 + Y_2(t)E_{21}, \\
\mathbf{Y}'_k(t) &= \mathbf{Y}_k(t)\mathbf{E} + Y_{k+1}(t)E_2, \quad k \geq 2,
\end{aligned} \tag{3.6}$$

where the prime denotes the derivative of the function concerned with respect to t.

The tagged customer at arrival time finds the system in level $\mathbf{k}$ with probability $\mathbf{Y}_k(0)$, for $k \geq 2$, the Laplace-Stieltjes transform (LST) of the first passage time to level $\mathbf{2}$ is given by the row vector $\chi(s)$.

As in [20], we get

$$\chi(s) = \sum_{i=2}^{\infty} Y_i(0) \left[(s\mathbf{I} - E)^{-1} E_2\right]^{i-2}. \tag{3.7}$$

Let $\Phi(i, s)$ be the LST of the absorbing time to the state $\{*\}$ given that the process starts from level $\mathbf{i} = 0, \mathbf{1}, \mathbf{2}$. On the basis of Q_1, we can write the following relations for the matrices $\Phi(i, s)$:

$$\Phi_{1\times mn}(0, s) = (s\mathbf{I} - \mathbf{E}_0)^{-1}\mathbf{C}_0, \tag{3.8}$$

$$\Phi_{(1+m+n)\times mn}(1, s) = (s\mathbf{I} - \mathbf{E}_1)^{-1}\mathbf{E}_{10}\Phi(0, s) + (s\mathbf{I} - \mathbf{E}_1)^{-1}\mathbf{C}_1, \tag{3.9}$$

$$\Phi_{(1+m+n+mn)\times mn}(2, s) = (s\mathbf{I} - \mathbf{E})^{-1}\mathbf{E}_{21}\Phi(1, s) + (s\mathbf{I} - \mathbf{E})^{-1}\mathbf{C}_2\,. \tag{3.10}$$

Finally, it can be seen that the LST $\widetilde{\mathbf{W}}(s)$ for the waiting time distribution is given by

$$\widetilde{\mathbf{W}}(s) = \mathbf{Y}_0(0)\boldsymbol{\Phi}(0, s) + \mathbf{Y}_1(0)\boldsymbol{\Phi}(1, s) + \chi(s)\boldsymbol{\Phi}(2, s). \tag{3.11}$$

3.1. Mean waiting time

The mean waiting time of an arriving customer is computed from $\widetilde{\mathbf{W}}(s)$

$$\mathbf{E}(\mathbf{W}) = -\mathbf{Y}_0(0)\boldsymbol{\Phi}'(0,0)\mathbf{e}_{mn} - \mathbf{Y}_1(0)\boldsymbol{\Phi}'(1,0)\mathbf{e}_{mn} - \chi'(0)\mathbf{e}_{1+m+n+mn} - \chi(0)\boldsymbol{\Phi}(2,0)\mathbf{e}_{mn}. \tag{3.12}$$

The first two terms give the mean time to reach an absorbing state $\{*\}$ by the tagged customer if the system is in a level ≤ 1. On its arrival, the third and fourth terms represent the time to reach the absorbing state $\{*\}$ if the system is in a level ≥ 2.

To compute the mean waiting time of the tagged customer, we must calculate the value of each term in (3.12). Differentiating (3.8) with respect to s and setting $s = 0$, we get

$$\mathbf{\Phi}(0,0) = (-1)(-\mathbf{E}_0)^{-1}\mathbf{C}_0. \tag{3.13}$$

Similarly, differentiating (3.9) and (3.10) with respect to s and setting $s = 0$, we obtain

$$\mathbf{\Phi}'(1,0) = (-1)(-\mathbf{E}_1)^{-2}\mathbf{E}_{10}\mathbf{\Phi}(0,0) + (-\mathbf{E}_1)^{-1}\mathbf{E}_{10}\mathbf{\Phi}'(0,0) + (-1)(-\mathbf{E}_1)^{-2}\mathbf{C}_1, \tag{3.14}$$

$$\mathbf{\Phi}'(2,0) = (-1)(-\mathbf{E})^{-2}E_{21}\mathbf{\Phi}(1,0) + (-E)^{-1}E_{21}\mathbf{\Phi}'(1,0) + (-1)(-E)^{-2}C_2. \tag{3.15}$$

Thus, formulae (3.13)–(3.15) permit the recursive computation of the matrices $\mathbf{\Phi}'(i,0)$, $0 \le i \le 2$. Using (3.13) to (3.15) and absorbing the initial condition $\mathbf{Y}(0) = (\mathbf{Y}_0(0), \mathbf{Y}_1(0), \mathbf{Y}_2(0), \ldots)$, we can compute the first two terms of (3.12).

The value of $\chi(0) = \sum_{i=2}^{\infty} Y_i(0)\mathbf{U}^{i-2}$, where $\mathbf{U} = (-E)^{-1}E_2$, is obtained by substituting $s = 0$ in (3.7), and it needs to be evaluated numerically. Since $\mathbf{U}\mathbf{e}_{1+m+n+mn} = \mathbf{e}_{1+m+n+mn}$, owing to the relation $E_2\mathbf{e}_{1+m+n+mn} + E\mathbf{e}_{1+m+n+mn} = \mathbf{0}$, we have that $\chi(0)\mathbf{e}_{1+m+n+mn} = 1 - Y_0(0) - Y_1(0)\mathbf{e}_{1+m+n}$. The value of $\chi(0)\mathbf{e}_{1+m+n+mn}$ can also be used, as mentioned in [11], to evaluate an approximate value of $\chi(0)$ by finite summation. Using (3.13) and (3.14), we get the value of $\Phi'(2,0)$, for the fourth term of (3.12).

To compute the third term of (3.12), differentiating (3.7) with respect to s and setting $s = 0$, we obtain

$$\chi'(0) = -\sum_{i=1}^{\infty} Y_{2+i}(0)\sum_{j=0}^{i-1}\mathbf{U}^j(-E)^{-1}\mathbf{U}^{i-j}. \tag{3.16}$$

Since $\mathbf{U}$ is a stochastic matrix, and using the relationship $\mathbf{U}\mathbf{e}_{1+m+n+mn} = \mathbf{e}_{1+m+n+mn}$, it follows that

$$-\chi'(0)\mathbf{e}_{1+m+n+mn} = \sum_{i=1}^{\infty} Y_{2+i}(0)\sum_{j=0}^{i-1}\mathbf{U}^j(-E)^{-1}\mathbf{e}_{1+m+n+mn}. \tag{3.17}$$

To obtain the value of $-\chi'(0)\mathbf{e}_{1+m+n+mn}$ from (3.17), we modify the technique used by Kao and Narayanan [11] and Neuts and Lucantoni [22]. Now, construct a stochastic matrix $\mathbf{U}_2$ such that $\mathbf{I}-\mathbf{U}+\mathbf{U}_2$ is nonsingular and generalized inverse of $(\mathbf{I}-\mathbf{U})$. In cases where $\mathbf{U}$ is irreducible, the matrix $\mathbf{U}_2$ may be chosen as $\mathbf{U}_2 = \mathbf{e}_{1+m+n+mn}\mathbf{u}_0$, where $\mathbf{u}_0$ is the invariant probability vector of $\mathbf{U}$, that is, $\mathbf{u}_0\mathbf{U} = \mathbf{u}_0$ and $\mathbf{u}_0\mathbf{e}_{1+m+n+mn} = 1$. This follows from the classical theorem on finite Markov chains given in the work of Kemeny and Snell [23, page 100].

Further, the following relation is satisfied owing to the property $\mathbf{U}\mathbf{U}_2 = \mathbf{U}_2\mathbf{U} = \mathbf{U}_2$:

$$\sum_{j=0}^{i-1}\mathbf{U}^j(\mathbf{I}-\mathbf{U}+\mathbf{U}_2) = \mathbf{I}-\mathbf{U}^i + i\mathbf{U}_2, \quad \text{for } i \ge 1. \tag{3.18}$$

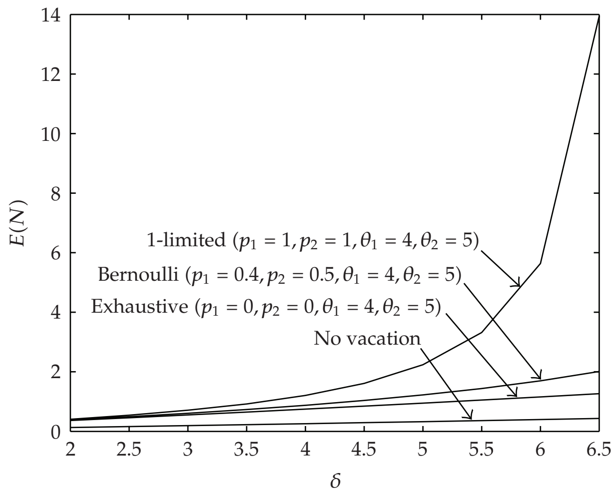

Figure 1: $E(N)$ versus δ for $\mu_1 = 31.94$, $\mu_2 = 10$.

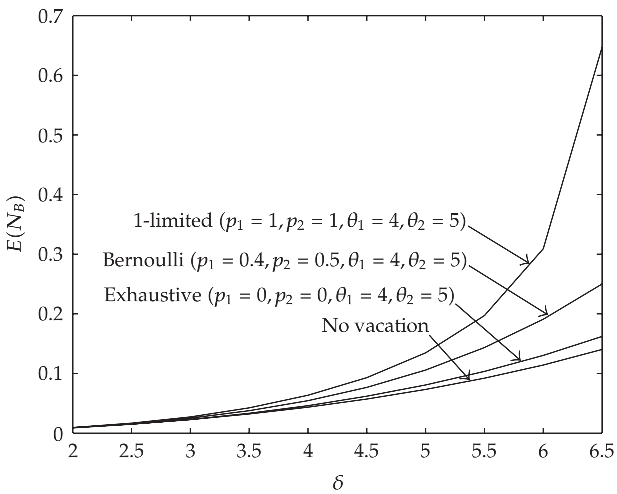

Figure 2: $E(N_B)$ versus δ for $\mu_1 = 31.94$, $\mu_2 = 10$.

Substituting (3.18) in (3.17) and simplifying, we obtain

$$
\begin{aligned}
-\chi'(0)\mathbf{e}_{1+m+n+mn} = {} & \left\{ Y_2(0)(\mathbf{I}-\mathbf{R})^{-1} + Y_2(0)\left[(\mathbf{I}-\mathbf{R})^{-2} - (\mathbf{I}-\mathbf{R})^{-1}\right]\mathbf{U}_2 - \chi(0) \right\} \\
& \times (\mathbf{I}-\mathbf{U}-\mathbf{U}_2)^{-1}(-E)^{-1}\mathbf{e}_{1+m+n+mn}.
\end{aligned}
\tag{3.19}
$$

Thus, all four terms of (3.12) have been computed, and the mean waiting time can be obtained.

Hence, we are able to compute the values of steady-state joint probabilities and mean waiting time using algorithms for this queueing system with Bernoulli vacation scheduling service.

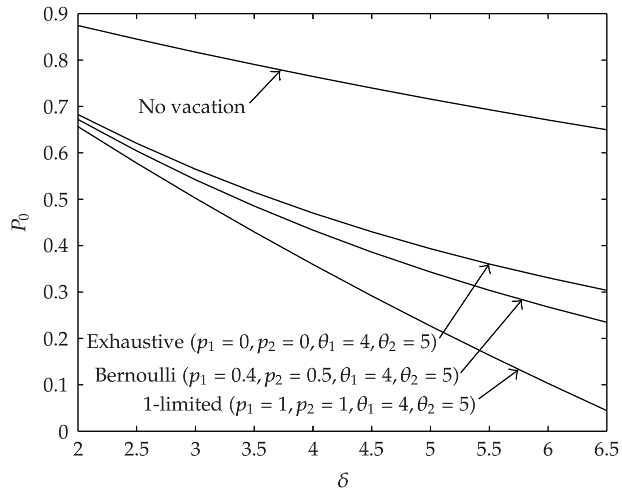

Figure 3: P_0 versus δ for $\mu_1 = 31.94$, $\mu_2 = 10$.

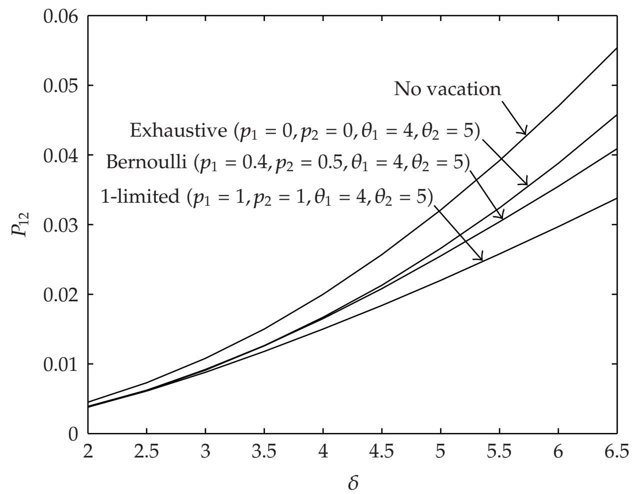

Figure 4: P_{12} versus δ for $\mu_1 = 31.94$, $\mu_2 = 10$.

4. Numerical examples

In this section, we discuss some interesting numerical examples that qualitatively describe the performance of the queueing system under investigation. To gain an understanding of the performance measures of the Bernoulli vacation scheduling service queueing model, we study the effect of the system parameters on the following items:

(i) the expected number $E(N)$ of customers in the system,

(ii) the expected number $E(N_B)$ of customers in the system when both servers are busy,

(iii) the probability P_0 of no customers in the system,

(iv) the probability P_{12} that both servers are busy, and

(v) the mean waiting time $E(\mathbf{W})$ of an arriving customer.

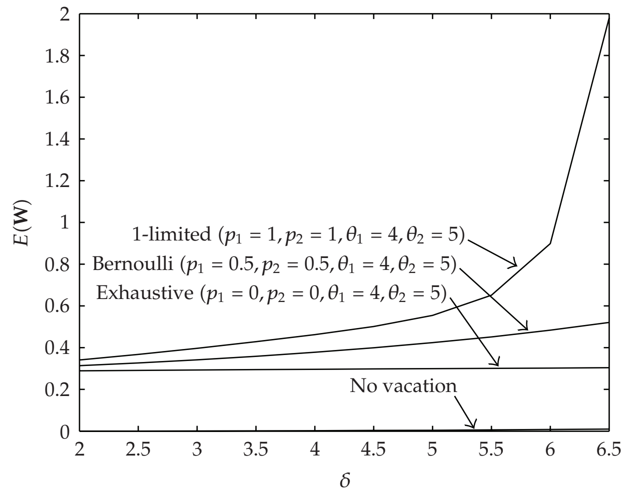

Figure 5: $E(\mathbf{W})$ versus δ for $\mu_1 = 31.94$, $\mu_2 = 10$.

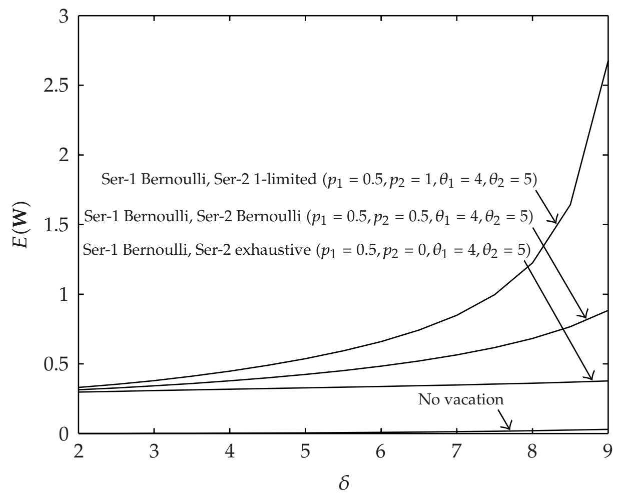

Figure 6: $E(\mathbf{W})$ versus δ for $\mu_1 = 31.94$, $\mu_2 = 10$.

For the arrival process, we consider the MAP input which is characterized by the matrices:

$$D_0 = \begin{bmatrix} -\lambda & \lambda \\ 0 & -\lambda \end{bmatrix}, \qquad D_1 = \begin{bmatrix} 0 & 0 \\ \lambda & 0 \end{bmatrix}. \tag{4.1}$$

In other words, the continuous-time Markov chain, which governs the input, has two states. For various values of λ ranging from 4 to 13, this MAP has the fundamental rate δ ranging from 2 to 6.5.

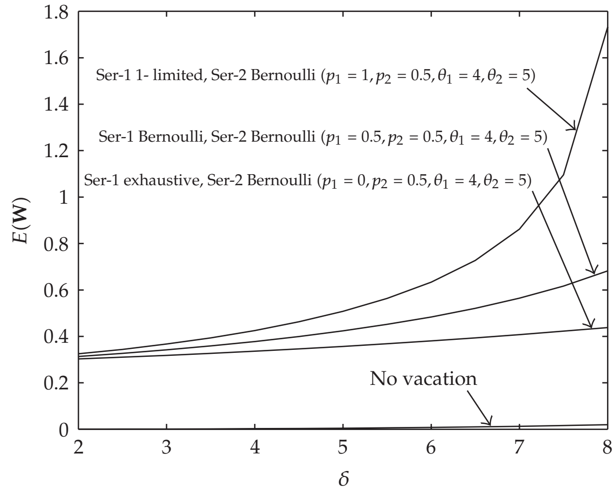

Figure 7: $E(\mathbf{W})$ versus δ for $\mu_1 = 31.94,\ \mu_2 = 10$.

Next, we consider the phase-type (PH) service time processes for server-1 and server-2 as

$$\alpha = (0.85,\ 0.14,\ 0.01), \qquad S = \begin{bmatrix} -200 & 0 & 0 \\ 0 & -20 & 0 \\ 0 & 0 & -0.5 \end{bmatrix}, \qquad S^0 = \begin{bmatrix} 200 \\ 20 \\ 0.5 \end{bmatrix},$$
$$\beta = (1,0), \qquad T = \begin{bmatrix} -20 & 20 \\ 0 & -20 \end{bmatrix}, \qquad T^0 = \begin{bmatrix} 0 \\ 20 \end{bmatrix}. \tag{4.2}$$

The average intensities of the services of server-1 and server-2 are given by $\mu_1 = 1/-\alpha S^{-1} e_3 = 31.94$ and $\mu_2 = 1/-\beta T^{-1} e_2 = 10$, respectively.

In Figures 1 and 2, the values of $E(N)$, the mean numbers of customers in the system and $E(N_B)$, the mean number of customers in the system when both servers being busy are plotted against the fundamental rate δ for chosen values of μ_1, μ_2, θ_1, θ_2, p_1, and p_2 satisfying stability condition (2.9).

By considering higher vacation rates (of order 10^5) in the vacation model under study, we obtain approximate results for corresponding nonvacation model. By fixing $\mu_1 = 31.94$, $\mu_2 = 10$, $\theta_1 = 4$, and $\theta_2 = 5$, the values of $E(N)$ versus the fundamental rate δ are plotted in Figure 1 for the following cases:

(1) 1-limited service policy ($p_1 = p_2 = 1$),

(2) Bernoulli vacation scheduling service ($p_1 = 0.4,\ p_2 = 0.5$),

(3) exhaustive service policy ($p_1 = p_2 = 0$),

(4) with no vacation system.

In all the cases, it is observed that $E(N)$ steadily increases as δ increases and decreases with decreasing values of p_1 and p_2. Figure 2 indicates that $E(N_B)$, the mean of the system size when both servers being busy for 1-limited service policy, Bernoulli vacation scheduling service, and exhaustive service with multiple vacation grow at a faster rate than nonvacation system for increasing values of the fundamental arrival rate δ.

In Figure 3, the values of P_0, the probability that there is no customer in the system, and both server-1 and server-2 on vacation are plotted against the fundamental arrival rate δ for the chosen parametric values of the system satisfying the stability condition (2.9). It is seen that the probability P_0 steadily decreases as the values of δ increase, and it decreases for increasing values of p_1 and p_2.

In Figure 4, we plot the probability P_{12} that both server-1 and server-2 are busy versus the fundamental arrival rate δ for the different values of p_1 and p_2. It is seen that the probability P_{12} exhibits the opposite trend to that of the probability P_0 in the sense that it increases with increasing δ as expected. However, P_{12} decreases for increasing values of p_1 and p_2. Moreover, for the case of nonvacation system, the value of P_{12} is seen to be more than that for any of the vacation policies under discussion.

The trends of the average waiting time $E(\mathbf{W})$ are depicted in Figures 5–7. The computation of the mean waiting time $E(\mathbf{W})$ of customer at the arrival epoch for the exhaustive service policy is carried out in accordance with procedure given in the work of Kao and Narayanan [11] and in the case of 1-limited service discipline, the approach due to the work of Tyagi et al. [14] is followed. It is observed from Figure 5 that $E(\mathbf{W})$ grows in an unbounded fashion for system with/without vacation for increasing values of δ. For nonvacation system and exhaustive service discipline ($p_1 = p_2 = 0$), the growth of $E(\mathbf{W})$ is not much faster whereas in the cases of Bernoulli vacation scheduling service ($p_1 = p_2 = 0.5$) and 1-limited service discipline ($p_1 = p_2 = 1$), there is a steep increase in $E(\mathbf{W})$ for larger values of δ.

Figure 6 illustrates the trend of expected waiting time $E(\mathbf{W})$ of the customer at arrival epoch versus the fundamental arrival rate δ if server-1 follows Bernoulli vacation scheduling service and server-2 follows either of the following: (i) 1-limited service policy, (ii) Bernoulli vacation scheduling, (iii) exhaustive service discipline, and (iv) no vacation system. In all the cases, the expected waiting time $E(\mathbf{W})$ increases for increasing values of δ. Finally, Figure 7 exhibits a similar trend wherein the service disciplines of server-1 and server-2 are interchanged corresponding to Figure 6.

5. Conclusions and further research

A queueing system with two heterogeneous servers and Bernoulli vacation has been presented. Customers arrival pattern is described by the MAP, and service times have PH distributions. Based on the matrix geometric method, the stationary queue length distribution, mean system size, and other system performance measure have been computed. This system subsumes the 1-limited service discipline and the exhaustive service discipline as special cases. Moreover, the expected waiting time of the customer at arrival epoch has been analyzed in detail. Results of numerical experiments giving insight into behavior of the systems are presented. We expect that the method of analysis adopted in this paper can be used to discuss other complex queueing systems such as multiserver retrial queue with Bernoulli vacation policy.

References

[1] B. T. Doshi, "Queueing systems with vacations—a survey," *Queueing Systems*, vol. 1, no. 1, pp. 29–66, 1986.

[2] Y. Levy and U. Yechiali, "Utilization of idle time in an $M/G/1$ queueing system," *Management Science*, vol. 22, no. 2, pp. 202–211, 1975.

[3] J. Keilson and L. D. Servi, "Oscillating random walk models for GI/G/1 vacation systems with Bernoulli schedules," *Journal of Applied Probability*, vol. 23, no. 3, pp. 790–802, 1986.

[4] L. Servi, "Average delay approximation of M/G/1 cyclic service queues with Bernoulli schedules," *IEEE Journal on Selected Areas in Communication*, vol. 4, no. 6, pp. 813–820, 1986.

[5] T. E. Tedijanto, "Exact results for the cyclic-service queue with a Bernoulli schedule," *Performance Evaluation*, vol. 11, no. 2, pp. 107–115, 1990.

[6] J. Teghem Jr., "Control of the service process in a queueing system," *European Journal of Operational Research*, vol. 23, no. 2, pp. 141–158, 1986.

[7] B. Doshi, "Single server queues with vacations," in *Stochastic Analysis of Computer and Communication Systems*, H. Takagi, Ed., pp. 217–265, North-Holland, Amsterdam, The Netherlands, 1990.

[8] H. Takagi, *Queueing Analysis. A Foundation of Performance Evaluation. Volume 1: Vacation and Priority Systems*, North-Holland, Amsterdam, The Netherlands, 1991.

[9] Y. Levy and U. Yechiali, "An M/M/S queue with servers' vacations," *Information Systems and Operational Research*, vol. 14, no. 2, pp. 153–163, 1976.

[10] B. Vinod, "Exponential queues with servers' vacations," *Journal of the Operational Research Society*, vol. 37, no. 10, pp. 1007–1014, 1986.

[11] E. P. C. Kao and K. S. Narayanan, "Analyses of an M/M/N queue with servers' vacations," *European Journal of Operational Research*, vol. 54, no. 2, pp. 256–266, 1991.

[12] X. Chao and Y. Zhao, "Analysis of multi-server queue with station and server vacation," *European Journal of Operational Research*, vol. 110, no. 2, pp. 392–406, 1998.

[13] N. Tian and Q. Li, "The M/M/C queue with PH synchronous vacations," *Systems Science and Mathematical Sciences*, vol. 13, no. 1, pp. 7–16, 2000.

[14] A. Tyagi, J. Harms, and A. Kamal, "Analysis of M/M/S queues with 1-limited service," *Journal of the Operational Research Society*, vol. 53, no. 7, pp. 752–767, 2002.

[15] B. Krishna Kumar and S. Pavai Madheswari, "Analysis of an M/M/N queue with Bernoulli service schedule," to appear in *International Journal of Operational Research*.

[16] M. F. Neuts and Y. Takahashi, "Asymptotic behavior of the stationary distributions in the GI/PH/C queue with heterogeneous servers," *Zeitschrift für Wahrscheinlichkeitstheorie*, vol. 57, no. 4, pp. 441–452, 1981.

[17] H. Heffes and D. M. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE Journal on Selected Areas in Communication*, vol. 4, no. 6, pp. 856–868, 1986.

[18] D. M. Lucantoni, K. S. Meier-Hellstern, and M. F. Neuts, "A single-server queue with server vacations and a class of nonrenewal arrival processes," *Advances in Applied Probability*, vol. 22, no. 3, pp. 676–705, 1990.

[19] C. Blondia, "Finite capacity vacation models with nonrenewal input," *Journal of Applied Probability*, vol. 28, no. 1, pp. 174–197, 1991.

[20] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, vol. 2 of *Johns Hopkins Series in the Mathematical Sciences*, Johns Hopkins University Press, Baltimore, Md, USA, 1981.

[21] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on Circuits and Systems*, vol. 25, no. 9, pp. 772–781, 1978.

[22] M. F. Neuts and D. M. Lucantoni, "A Markovian queue with N servers subject to breakdowns and repairs," *Management Science*, vol. 25, no. 9, pp. 849–861, 1979.

[23] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, The University Series in Undergraduate Mathematics, D. Van Nostrand, Princeton, NJ, USA, 1960.