MATHEMATICA APPLICANDA
Vol. 41(2) 2013, p. 145–169
doi: 10.14708/ma.v41i2.389

Srinivas R. Chakravarthy (Flint)

Analysis of a multi-server queuing model with vacations and optional secondary services

Abstract In this paper we study a multi-server queueing model in which the customer arrive according to a Markovian arrival process. The customers may require, with a certain probability, an optional secondary service upon completion of a primary service. The secondary services are offered (in batches of varying size) when any of the following conditions holds good: (a) upon completion of a service a free server finds no primary customer waiting in the queue and there is at least one secondary customer (including possibly the primary customer becoming a secondary customer) waiting for service; (b) upon completion of a primary service, the customer requires a secondary service and at that time the number of customers needing a secondary service hits a pre-determined threshold value; (c) a server returning from a vacation finds no primary customer but at least one secondary customer waiting. The servers take vacation when there are no customers (either primary or secondary) waiting to receive service. The model is studied as a QBD-process using matrix-analytic methods and some illustrative examples are discussed.

2010 Mathematics Subject Classification: 60K25; 90B22.

Key words and phrases: Markovian arrival process, queueing, waiting time, vacation, secondary batch services, multi-service, phase type distribution, algorithmic probability.

1. Introduction In the literature queueing models with vacations have been studied extensively (see e.g. [22,32,58]). Such models occur naturally in many practical situations. For example, the servers may take a break from their routine work to tend to items such as clearing overhead work, training, helping others, and personal reasons. The breaks may also occur due to interruptions caused by emergencies or breakdowns in the machines/systems needed for servicing the customers. Such models are studied under the umbrella of queueing models with server interruptions or queueing models with vacations (see e.g. [36]).

Queuing systems in which customers may request an additional service or leave behind some secondary (overhead/administrative type) work arise commonly in production and manufacturing systems, computer and communications engineering, and service industries. Some of the earlier works on

such models include [6, 25, 34]. Later, several authors [12–20, 25, 28, 29, 31, 33, 35, 41–43, 46, 59, 60] studied a variety of queuing models wherein such additional services are referred to as optional services and under different contexts including retrials, breakdowns and repairs of service facility, and vacationing servers. The model studied in this paper has a number of applications in practice. For example, customers receiving services in industries such as health care normally leave behind some additional work (such as billing, referral, missing information, etc). Also, in computer and communications engineering, messages or packets often leave behind some secondary (overhead) work, usually of an *administrative* nature. In production line systems, jobs may require additional work, usually of an "inspection" nature. This secondary service has a lower priority than the primary one, but should not accumulate unduly and hence a need for attending to those (in batches) once a threshold is met.

The additional services are offered by the same server by taking a break from the primary services (also referred to as *essential* services by some authors) using a variety of criteria such as when the server is blocked due to not having enough buffer space to place a customer requiring an additional service; when the server becomes idle at the primary service node, etc.

Recently, Chakravarthy [9] studied a single server queueing model in which the customers arrive according to a Markovian arrival process (MAP). With a certain probability a customer may opt for a secondary service. The secondary services are provided by the same server either immediately (if no one is waiting to receive a primary service) or waits until the number of customers requiring secondary services hits a pre-determined threshold. The server goes on a vacation whenever the system becomes empty. Assuming the services to be of phase type, the model is studied as a QBD-process using matrix-analytic methods and some illustrative examples are discussed.

In this paper we consider a multi-server queueing system with MAP arrivals. These customers will be referred to as primary customers. Upon completion of a service, the primary customer may request an additional service (with a certain probability) or leave the system (with the complement probability). The customers requiring additional services, referred henceforth as secondary customers, are served in batches of varying size but not to exceed a pre-determined threshold. A server initiates a secondary service to a waiting batch of one or more customers soon after finishing a primary service whenever the system has no primary customers or when the number of secondary customers hits the threshold value. It is also possible for a server returning from a vacation to offer secondary services provided there are no primary customers waiting for service. It should be pointed out that there cannot be any customers waiting for secondary services without at least one server busy in the system. We assume that service times are exponential with parameter depending on the type of service (primary or secondary). A server will go on

a vacation when at a service completion there are no (primary or secondary) customers waiting for service. The servers can take multiple vacations at a time due to the system has no (primary or secondary) customers waiting for service at the time of returning from a vacattion. The vacation times are assumed to be exponentially distributed. The motivation for this paper comes from a need to (a) study a multi-server system wherein the customers require optional secondary services, which to our knowledge has not been studied in the literature; (b) highlight the fact that for the special case when a customer's secondary service has to be provided immediately for every such request the model reduces to an appropriate classical queueing model. This fact has been overlooked by a number of authors (see e.g., [29, 41, 46]) until recently (see e.g. [9]); (c) see the impact of the probability (of opting for a secondary service) on some system performance measures as compared to the classical $MAP/M/c$ queue with vacation; and (d) study the effective service time (defined as the time spent in getting a primary service and possibly a secondary service) that has been ignored in the literature until recently (see e.g. [9]).

The paper is organized as follows. In Section 2 the model under study is described. The steady-state analysis of the model is performed in Section 3 and in this section we also display some key system performance measures. Illustrative numerical examples are discussed in Section 4. Some concluding remarks are given in Section 5.

2. Model description We consider a multi-server queueing system in which customers (henceforth referred to as *primary customers*) arrive according to a Markovian arrival process (MAP) with representation (D_0, D_1) of order m. The MAP, first introduced by Neuts [47] as a versatile Markovian point process, is a rich class of point processes that includes many well-known processes such as Poisson, PH-renewal processes, and Markov-modulated Poisson process. The generator Q^*, defined by $Q^* = D_0 + D_1$, governs the underlying Markov chain of the MAP such that D_0 accounts for the transitions corresponding to no arrival and D_1 governs those corresponding to an arrival. For further details on MAP and their usefulness in stochastic modelling, we refer to [39, 49, 50] and for a review and recent work on MAP we refer the reader to [2, 7, 8].

Upon completion of a service the primary customer may require a secondary service with probability $p, 0 \leqslant p \leqslant 1$, and with probability $q = 1 - p$ the customer will leave the system. We refer the customers requiring secondary services as *secondary customers* henceforth. We assume that primary and secondary services follow exponential distribution with parameters μ_1 and μ_2, respectively. A free server finding no primary or secondary customer waiting for service will go on a vacation, and we assume that the vacation times are exponentially distributed with parameter θ. It is possible for a server to go on multiple vacations without serving any customers between vacations.

The secondary services are offered (in batches of varying size) when any of the following conditions holds good: (a) upon completion of a service the free server finds no primary customer waiting in the queue and there is at least one secondary customer (including possibly the primary customer who just finished a service needing a secondary service) waiting for service; (b) upon completion of a primary service, the customer requires a secondary service and at that time the number of customers needing secondary service hits $b, 1 \leqslant b < \infty$, a pre-determined threshold value; (c) a server returning from a vacation finds no primary customer but at least one secondary customer waiting.

A server returning from a vacation will always serve a primary customer, if any waiting, before offering services to a batch of secondary customers, if any. If no customers (primary or secondary) are waiting the server will go back for another vacation.

In the sequel we need the following notations. By $\boldsymbol{e}$ we will denote a column vector (of appropriate dimension) of 1's; $\boldsymbol{e}_i$ we will denote a unit column vector (of appropriate dimension) with 1 in the ith position and 0 elsewhere; and I an identity matrix (of appropriate dimension). We will display the dimension should there be a need to emphasize it. For example, if there is a need to display the dimension of an identity matrix of order m, we will do so by writing I_m rather than I; a unit vector of dimension m will be denoted as $\boldsymbol{e}(m)$ rather than $\boldsymbol{e}$. The notation "$\prime$" will stand for the transpose of a matrix and the symbol $\otimes$ denotes the Kronecker product of matrices. Thus, if A is a matrix of order $m \times n$ and if B is a matrix of order $p \times q$, then $A \otimes B$ will denote a matrix of order $mp \times nq$ whose $(i,j)^{th}$ block matrix is given by $a_{ij}B$. For details and properties on Kronecker products we refer the reader to [27, 44].

Let $\boldsymbol{\delta}$ be the stationary probability vector of the Markov process with generator Q^*. That is, $\boldsymbol{\delta}$ is the unique (positive) probability vector satisfying.

$$\boldsymbol{\delta} Q^* = \mathbf{0}, \boldsymbol{\delta}\boldsymbol{e} = 1. \tag{1}$$

The constant $\lambda = \boldsymbol{\delta} D_1 \boldsymbol{e}$, referred to as the **fundamental rate**, gives the expected number of arrivals per unit of time in the stationary version of the MAP. Often, in model comparisons, it is convenient to select the time scale of the MAP so that λ has a certain value. That is accomplished, in the continuous MAP case, by multiplying the coefficient matrices D_0 and D_1, by the appropriate common constant.

3. The steady-state analysis Let $N_1(t), N_2(t), N_3(t), N_4(t)$, and $J(t)$ denote, respectively, the number of primary customers in the queue, the number of secondary customers in the queue, the number of servers busy with primary customers, the number of servers busy with secondary customers, and the phase of the arrival process at time t. Note that the number of

servers on vacation at time t is $c - N_3(t) - N_4(t)$. The process

$$\{(N_1(t), N_2(t), N_3(t), N_4(t), J(t))\}_{t \geqslant 0}$$

is a continuous-time Markov chain with state space given by

$$\begin{aligned}\Omega &= \{(i_1, 0, i_3, i_4, k) : i_1 \geqslant 0, 0 \leqslant i_3, i_4 \leqslant c, 0 \leqslant i_3 + i_4 \leqslant c, 1 \leqslant k \leqslant m\} \\ &\quad \cup \{(i_1, i_2, i_3, i_4, k) : i_1 \geqslant 0, 1 \leqslant i_2 \leqslant b-1, 1 \leqslant i_3 \leqslant c, \\ &\qquad 0 \leqslant i_4 \leqslant c, 1 \leqslant i_3 + i_4 \leqslant c, 1 \leqslant k \leqslant m\}.\end{aligned}$$

For $i_1 \geqslant 0$, we now define the set of states

$$\begin{aligned}\underline{\boldsymbol{i}}_1 &= \{(i_1, 0, i_3, i_4, k), 0 \leqslant i_3, i_4 \leqslant c, 0 \leqslant i_3 + i_4 \leqslant c, 1 \leqslant k \leqslant m\} \\ &\quad \cup \{(i_1, i_2, i_3, i_4, k) : 1 \leqslant i_2 \leqslant b-1, 1 \leqslant i_3 \leqslant c, 0 \leqslant i_4 \leqslant c, \\ &\qquad 1 \leqslant i_3 + i_4 \leqslant c, 1 \leqslant k \leqslant m\}.\end{aligned}$$

Note that the level $\underline{\boldsymbol{i}}_1, i_1 \geqslant 0$, is of dimension $0.5(c+1)(2+bc)m$. The infinitesimal generator of the Markov chain governing the system is given by

$$Q = \begin{pmatrix} B_1 & A_0 & & & \\ A_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}, \tag{2}$$

where

$$B_1 = \begin{pmatrix} B_{11} & & & & \\ B_{12} & B_{13} & & & \\ B_{12} & & B_{13} & & \\ \vdots & & & \ddots & \\ B_{12} & & & & B_{13} \end{pmatrix}, \; A_0 = I \otimes D_1, \tag{3}$$

$$A_1 = \begin{pmatrix} A_{10} & & & \\ & A_{11} & & \\ & & \ddots & \\ A_{12} & & & A_{11} \end{pmatrix}, \tag{4}$$

$$A_2 = \begin{pmatrix} A_{20} & A_{22} & & & & \\ & A_{21} & A_{23} & & & \\ & & A_{21} & A_{23} & & \\ & & \ddots & \ddots & & \\ & & & & A_{21} & A_{23} \\ & & & & & A_{21} \end{pmatrix}, \tag{5}$$

and before we display the matrices $B_{11}, B_{12}, B_{13}, A_{10}, A_{11}, A_{12}, A_{20}, A_{21}, A_{22}$, and A_{23} appearing in (3), (4), and (5), we need to set up some auxiliary matrices.

Define $\boldsymbol{\alpha}_r = \theta(r, r-1, \cdots, 1, 0)$, $\boldsymbol{\beta}_r = \mu_2(0, 1, \cdots, r)$, $\boldsymbol{\gamma}_r = r\mu_1 \boldsymbol{e}'(c+1-r)$ for $0 \leqslant r \leqslant c$,

$$E_r = \begin{pmatrix} r\theta & & & & \\ \mu_2 & (r-1)\theta & & & \\ & 2\mu_2 & (r-2)\theta & & \\ & \ddots & \ddots & & \\ & & & (r-1)\mu_2 & \theta \\ & & & & r\mu_2 \end{pmatrix}, \quad 1 \leqslant r \leqslant c,$$

$$\tilde{E}_r = \begin{pmatrix} \mathbf{0} & E_r \end{pmatrix}, \quad \tilde{I}_r = (c+1-r)\mu_1 \begin{pmatrix} \mathbf{0} & I_r \end{pmatrix}, \quad 1 \leqslant r \leqslant c,$$

$$\tilde{B}_r = (c+1-r)\mu_1 \begin{pmatrix} q & p & & & \\ & q & p & & \\ & & \ddots & \ddots & \\ & & & q & p \end{pmatrix}, \quad 1 \leqslant r \leqslant c.$$

If $\boldsymbol{a_i} = (a_{i1}, \cdots, a_{in_i})$ is a vector of dimension $n_i, 1 \leqslant i \leqslant r$, then by $\Delta(\boldsymbol{a_1}, \cdots, \boldsymbol{a_r})$ we denote a diagonal matrix of dimension $n_1 + \cdots + n_r$ with diagonal entries given by the components of the vectors $\boldsymbol{a_i}, 1 \leqslant i \leqslant r$. That is,

$$\Delta(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r) = \begin{pmatrix} a_{11} & & & \\ & a_{12} & & \\ & & \ddots & \\ & & & a_{rn_r} \end{pmatrix}.$$

Now we are ready to define the matrices appearing in (3), (4) and (5).

$$\begin{aligned} B_{11} &= I \otimes D_0 - \begin{pmatrix} \Delta(\boldsymbol{\beta}_c) & & & \\ \tilde{B}_c & \Delta(\boldsymbol{\beta}_{c-1} + \boldsymbol{\gamma}_1) & & \\ & \ddots & \ddots & \\ & & \tilde{B}_1 & \Delta(\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_c) \end{pmatrix} \otimes I, \\ B_{12} &= \begin{pmatrix} \tilde{I}_c & \tilde{E}_{c-1} & & & \\ & \tilde{I}_{c-1} & \tilde{E}_{c-2} & & \\ & \ddots & \ddots & & \\ & & & \tilde{I}_2 & \tilde{E}_1 \\ & & & & \tilde{I}_1 \quad 0 \end{pmatrix} \otimes I, \\ B_{13} &= I \otimes D_0 - \Delta(\boldsymbol{\beta}_{c-1} + \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\beta}_0 + \boldsymbol{\gamma}_c) \otimes I, \end{aligned}$$

$$
\begin{aligned}
A_{10} &= I \otimes D_0 - \Delta(\boldsymbol{\alpha}_c + \boldsymbol{\beta}_c + \boldsymbol{\gamma}_0, \cdots, \boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0 + \boldsymbol{\gamma}_c) \otimes I, \\
A_{11} &= I \otimes D_0 - \Delta(\boldsymbol{\alpha}_{c-1} + \boldsymbol{\beta}_{c-1} + \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0 + \boldsymbol{\gamma}_c) \otimes I, \\
A_{12} &= p \begin{pmatrix} \tilde{I}_c & & & & \\ & \tilde{I}_{c-1} & & & \\ & & \ddots & & \\ & & & \tilde{I}_2 & \\ & & & & \tilde{I}_1 \quad 0 \end{pmatrix} \otimes I, \\
A_{20} &= \begin{pmatrix} 0 & E_c & & & & \\ & q\mu_1 I_c & E_{c-1} & & & \\ & & 2q\mu_1 I_{c-1} & E_{c-2} & & \\ & & & \ddots & & \\ & & & & (c-1)q\mu_1 & E_1 \\ & & & & & cq\mu_1 \end{pmatrix} \otimes I, \\
A_{21} &= \begin{pmatrix} q\mu_1 I_c & E_{c-1} & & & \\ & 2q\mu_1 I_{c-1} & E_{c-2} & & \\ & \ddots & \ddots & & \\ & & & (c-1)q\mu_1 & E_1 \\ & & & & cq\mu_1 \end{pmatrix} \otimes I, \\
A_{22} &= \begin{pmatrix} 0 & & & & \\ p\mu_1 I_c & & & & \\ & 2p\mu_1 I_{c-1} & & & \\ & & \ddots & & \\ & & & (c-1)p\mu_1 & \\ & & & & cp\mu_1 \end{pmatrix} \otimes I, \\
A_{23} &= \begin{pmatrix} p\mu_1 I_c & & & & \\ & 2p\mu_1 I_{c-1} & & & \\ & & \ddots & & \\ & & & (c-1)p\mu_1 & \\ & & & & cp\mu_1 \end{pmatrix} \otimes I,
\end{aligned}
$$

3.1. The stability condition Let $\boldsymbol{\pi} = (\boldsymbol{\pi}(0), \boldsymbol{\pi}(1), \cdots, \boldsymbol{\pi}(b-1))$ be the steady state probability vector of the generator $A = A_0 + A_1 + A_2$. That is, $\boldsymbol{\pi}$ satisfies

$$\boldsymbol{\pi} A = \mathbf{0}, \boldsymbol{\pi}\, \boldsymbol{e} = 1. \tag{6}$$

We further partition $\boldsymbol{\pi}(i), 0 \leqslant i \leqslant b-1$, as

$$\boldsymbol{\pi}(i) = (\boldsymbol{\pi}_{00}(i), \cdots, \boldsymbol{\pi}_{0c}(i), \boldsymbol{\pi}_{10}(i), \cdots, \boldsymbol{\pi}_{1c-1}(i), \cdots, \boldsymbol{\pi}_{c-10}(i), \boldsymbol{\pi}_{c-11}(i), \boldsymbol{\pi}_{c0}(i)).$$

The following theorem establishes the stability condition of the queueing system under study.

THEOREM 3.1 *The queuing system under study is stable if and only if the following condition is satisfied.*

$$\lambda < \frac{bc\mu_1\mu_2}{b\mu_2 + p\mu_1}. \tag{7}$$

PROOF The queuing system under study with the QBD type generator given in (2) is stable (see, e.g., [48]) if and only if $\boldsymbol{\pi} A_0 \boldsymbol{e} < \boldsymbol{\pi} A_2 \boldsymbol{e}$. First note that the equations in (6) reduce to

$$\begin{gathered}
\boldsymbol{\pi}(0)[A_{10} + A_{20} + I \otimes D_1] + \boldsymbol{\pi}(b-1)A_{12} = \mathbf{0}, \\
\boldsymbol{\pi}(0)A_{22} + \boldsymbol{\pi}(1)[A_{11} + A_{21} + I \otimes D_1] = \mathbf{0}, \\
\boldsymbol{\pi}(i-1)A_{23} + \boldsymbol{\pi}(i)[A_{11} + A_{21} + I \otimes D_1] = \mathbf{0},\ 2 \leqslant i \leqslant b-1,
\end{gathered} \tag{8}$$

subject to

$$\sum_{i=0}^{b-1} \boldsymbol{\pi}(i)\mathbf{e} = 1.$$

It is easy to verify from (8) that

$$\boldsymbol{\pi}_{00}(0)[Q^* - c\theta I] = \mathbf{0},$$

$$\left[\boldsymbol{\pi}_{01}(0) + \boldsymbol{\pi}_{10}(0) + \textstyle\sum_{i=1}^{b-1} \boldsymbol{\pi}_{10}(i)\right] [Q^* - (c-1)\theta I] = \mathbf{0}, \quad 1 \leqslant j \leqslant b-1,$$

$$\left[\textstyle\sum_{k=0}^{j} \boldsymbol{\pi}_{kj-k}(0) + \sum_{i=1}^{b-1} \sum_{k=1}^{j} \boldsymbol{\pi}_{kj-k}(i)\right] [Q^* - (c-1)\theta I] = \mathbf{0}, \quad 1 \leqslant j \leqslant c-1$$

which imply that

$$\boldsymbol{\pi}_{00}(0) = \mathbf{0} \text{ and } \boldsymbol{\pi}_{kj-k}(i) = \mathbf{0}, \quad 1 \leqslant i \leqslant b-1,\ 1 \leqslant k \leqslant j,\ 1 \leqslant j \leqslant c-1. \tag{9}$$

In view of (9) the only nonzero components of $\boldsymbol{\pi}$ are $\boldsymbol{\pi}_{jc-j}(0),\ 0 \leqslant j \leqslant c$, and $\boldsymbol{\pi}_{jc-j}(i),\ 1 \leqslant i \leqslant b-1,\ 1 \leqslant j \leqslant c$. Using this fact and defining

$$a_j(0) = \boldsymbol{\pi}_{jc-j}(0)\mathbf{e},\ 0 \leqslant j \leqslant c, \quad a_j(i) = \boldsymbol{\pi}_{jc-j}(i)\mathbf{e},\ 1 \leqslant j \leqslant c, \tag{10}$$

it is easy to verify that

$$\boldsymbol{\pi} A_2 \mathbf{e} = \sum_{j=0}^{c} [j\mu_1 + (c-j)\mu_2] a_j(0) + \sum_{i=1}^{b-1} \sum_{j=1}^{c} [j\mu_1 + (c-j)\mu_2] a_j(i) - p\mu_1 \sum_{j=1}^{c} a_j(b-1). \tag{11}$$

Using equations (10), (11), and the fact that $\boldsymbol{\pi} A_0 \mathbf{e} = \lambda$, the steady-state equations in (8) after post-multiplying by $\mathbf{e}$ can be rewritten as

$$\begin{aligned}
-c\mu_2 a_0(0) + p\mu_1 a_1(b-1) &= 0, && (12)\\
(c+1-j)\mu_2 a_{j-1}(0) - [jp\mu_1 + (c-j)\mu_2]a_j(0) & && (13)\\
+(j+1)\mu_1 a_{j+1}(b-1) &= 0,\ 1 \leqslant j \leqslant c-1, &&\\
\mu_2 a_{c-1}(0) - cp\mu_1 a_c(0) &= 0, && (14)\\
p\mu_1 a_1(i-1) - [p\mu_1 + (c-1)\mu_2]a_1(i) &= 0,\ 1 \leqslant i \leqslant b-1, && (15)\\
jp\mu_1 a_j(i-1) - [jp\mu_1 + (c-j)\mu_2]a_j(i) & && (16)\\
+(c+1-j)\mu_2 a_{j-1}(i) &= 0,\ 2 \leqslant j \leqslant c,\ 1 \leqslant i \leqslant b-1, &&
\end{aligned}$$

subject to the normalizing condition

$$\sum_{j=0}^{c} a_j(0) + \sum_{i=1}^{b-1}\sum_{j=1}^{c} a_j(i) = 1.$$

From equations (12) through (16), it is easy to verify that

$$\sum_{j=1}^{c} j a_j(i) = \sum_{j=1}^{c} j a_j(0),\ 1 \leqslant i \leqslant b-1. \tag{17}$$

For use in sequel we define

$$\hat{a} = \sum_{j=1}^{c} j a_j(0).$$

From (12) through (14) we establish that

$$(c-j)\mu_2 a_j(0) = p\mu_1 \left[\sum_{k=1}^{j+1} k a_k(b-1) - \sum_{k=1}^{j} k a_k(0)\right],\ 0 \leqslant j \leqslant c-1. \tag{18}$$

Similarly from equations (14) and (16) we get

$$(c-j)\mu_2 a_j(i) = p\mu_1 \sum_{k=1}^{j} k[a_k(i-1) - a_k(i)],\ 1 \leqslant j \leqslant c,\ 1 \leqslant i \leqslant b-1. \tag{19}$$

Adding the equations given in (18) over j we get

$$c\mu_2 \sum_{j=0}^{c} a_j(0) - \mu_2 \hat{a} = p\mu_1 \left[\hat{a} - \sum_{k=1}^{c} k^2 a_k(b-1) + \sum_{k=1}^{c} k^2 a_k(0)\right]. \tag{20}$$

Now adding the equations given in (19) over j we get

$$c\mu_2 \sum_{j=1}^{c} a_j(i) - \mu_2 \hat{a} = p\mu_1 \left[\sum_{k=1}^{c} k^2 a_k(i) - \sum_{k=1}^{c} k^2 a_k(i-1)\right],\ 1 \leqslant i \leqslant b-1. \tag{21}$$

From (20) and (21) it can readily be verified that

$$\hat{a} = \frac{c\mu_2}{b\mu_2 + p\mu_1}. \tag{22}$$

The stated result follows immediately from (11), (17) and (22) on noting that the right-hand side of (11) can be simplified as $c\mu_2 + \hat{a}[b\mu_1 - (b\mu_2 + p\mu_1)]$.■

Note: Intuitively one can try to explain the stability condition. Suppose we rewrite equation (7) as

$$\lambda < c\left[\frac{1}{\frac{1}{\mu_1} + \frac{p}{b\mu_2}}\right].$$

The quantity $\frac{1}{\mu_1} + \frac{p}{b\mu_2}$ appearing within the parentheses on the right can be thought of as the mean time to process a customer. First note that the quantity $\frac{1}{\mu_1}$ is the mean service time for a primary customer and the quantity $\frac{p}{b\mu_2}$ can be thought as the mean service time for that primary customer receiving a service as a secondary customer (requesting probability is p) and is obtained by being in a group of b customers. See Theorem 3.5 below where this intuitive reasoning is supported for a special case.

3.2. The steady-state probability vector

Let $\boldsymbol{x} = (\boldsymbol{x}(0), \boldsymbol{x}(1), \cdots)$ denote the steady-state probability vector of Q. That is, $\boldsymbol{x}$ satisfies

$$\boldsymbol{x}\, Q = \mathbf{0}, \boldsymbol{x}\ \boldsymbol{e} = 1. \tag{23}$$

We further partition $\boldsymbol{x}(i)$ of dimension $0.5(c+1)(2+bc)m$ as

$$\boldsymbol{x}(i) = (\boldsymbol{x}_0(i), \cdots, \boldsymbol{x}_{b-1}(i)),\ i \geqslant 0,$$

with $\boldsymbol{x}_0(i) = (\boldsymbol{x}_{000}(i), \cdots, \boldsymbol{x}_{0c0}(i))$, $\boldsymbol{x}_j(i) = (\boldsymbol{x}_{j10}(i), \cdots, \boldsymbol{x}_{jc0}(i))$, $1 \leqslant j \leqslant b-1$, $i \geqslant 0$. Note that the vector $\boldsymbol{x}_0(i)$ is of dimension $0.5(c+1)(c+2)m$ while $\boldsymbol{x}_j(i)$ is of dimension $0.5c(c+1)m$. The steady-state probability that there are i primary customers waiting in the queue, no secondary customers waiting in the queue with $r, 0 \leqslant r \leqslant c$, servers busy with primary customers and $k, 0 \leqslant k \leqslant c-r$, servers busy serving batches of secondary customers with the arrival process in one of m phases is given by the components of $\boldsymbol{x}_{0rk}(i)$. Note that in this case exactly $c-r-k$ servers are on vacation. A similar interpretation is given for the components of $\boldsymbol{x}_{jrk}(i)$ but with j secondary customers waiting in the queue.

Under the stability condition given in (7) the steady-state probability vector $\boldsymbol{x}$ is obtained (see, e.g., [48]) as follows

$$\boldsymbol{x}(i) = \boldsymbol{x}(0)R^i, i \geqslant 0,$$

where the matrix R is the minimal nonnegative solution to the matrix quadratic equation:

$$R^2 A_2 + RA_1 + A_0 = 0,$$

and the vector $\boldsymbol{x}(0)$ is obtained by solving

$$\boldsymbol{x}(0)[B_1 + RA_2] = \mathbf{0},$$

subject to the normalizing condition

$$\boldsymbol{x}(0)(I - R)^{-1}\boldsymbol{e} = 1.$$

The computation of the R matrix can be carried out using a number of well-known methods such as logarithmic reduction and (block) Gauss-Seidel iterative by exploiting the special structure of the coefficient matrices A_0, A_1, and A_2, which are of dimension $0.5(c+1)(2+bc)m$. This is very important especially when one is dealing with large values of b, c, and m. The details of such exploitation will be omitted; however, the key steps in the logarithmic reduction are given below and for full details on this we refer the reader to [37].

Logarithmic Reduction Algorithm for R:

Step 0: $H \leftarrow (-A_1)^{-1}A_0$, $L \leftarrow (-A_1)^{-1}A_2$, $G = L$, and $T = H$.

Step 1:

$$\begin{gathered} U = HL + LH \\ M = H^2 \\ H \leftarrow (I - U)^{-1}M \\ M \leftarrow L^2 \\ L \leftarrow (I - U)^{-1}M \\ G \leftarrow G + TL \\ T \leftarrow TH \end{gathered}$$

Continue Step 1 until $||\boldsymbol{e} - G\boldsymbol{e}||_\infty < \epsilon$.

Step 2: $R = -A_0(A_1 + A_0 G)^{-1}$.

We now establish three useful results. These are intuitively obvious

Lemma 3.2 *We have*

$$\sum_{i=0}^{\infty} \boldsymbol{x}(i)(\boldsymbol{e} \otimes I) = \boldsymbol{\delta}.$$

where $\boldsymbol{\delta}$ is as given in (1).

PROOF First observe that the steady-state equations in (23) can be rewritten as

$$\begin{aligned} \boldsymbol{x}(0)B_1 + \boldsymbol{x}(1)A_2 &= \boldsymbol{0}, \\ \boldsymbol{x}(i-1)A_0 + \boldsymbol{x}(i)A_1 + \boldsymbol{x}(i+1)A_2 &= \boldsymbol{0}, \ i \geqslant 1. \end{aligned} \tag{24}$$

Post-multiplying equations given in (24) by $\boldsymbol{e} \otimes I$ and adding the resulting equations over i and using the uniqueness of the steady-state vector $\boldsymbol{\delta}$, the stated result follows immediately. ■

LEMMA 3.3 *We have*

$$\sum_{i=0}^{\infty}\sum_{j=0}^{b-1}\sum_{r=1}^{c}\sum_{k=0}^{c-r} r\boldsymbol{x}_{jrk}(i)\boldsymbol{e} = \frac{\lambda}{\mu_1}. \tag{25}$$

PROOF Note that in steady-state the rate at which customers enter into the system is equal to the rate at which the customers leave the primary facility (either as secondary customers to get additional service or leave the system). This leads to

$$\mu_1 \sum_{i=0}^{\infty}\sum_{j=0}^{b-1}\sum_{r=1}^{c}\sum_{k=0}^{c-r} r\boldsymbol{x}_{jrk}(i)\boldsymbol{e} = \lambda,$$

from which the stated equation follows. ■

Note: The equation (25) says that the mean number of servers busy with primary customers is given by $\frac{\lambda}{\mu_1}$.

The following lemma gives an expression for the mean number of servers busy with secondary customers.

LEMMA 3.4 *We have*

$$\sum_{i=0}^{\infty}\left[\sum_{k=1}^{c}\sum_{r=0}^{c-k} k\boldsymbol{x}_{0rk}(i)\boldsymbol{e} + \sum_{j=0}^{b-1}\sum_{k=1}^{c-1}\sum_{r=1}^{c-k} k\boldsymbol{x}_{jrk}(i)\boldsymbol{e}\right] = \frac{\lambda p}{\mu_2\mu_{BSS}}, \tag{26}$$

where μ_{BSS} is the mean number of secondary customers served in a batch.

PROOF On noting that in steady-state the rate of customers entering into secondary facility is equal to the rate at which they leave the system, we obtain

$$\mu_2\mu_{BSS}\sum_{i=0}^{\infty}\left[\sum_{k=1}^{c}\sum_{r=0}^{c-k} k\boldsymbol{x}_{0rk}(i)\boldsymbol{e} + \sum_{j=1}^{b-1}\sum_{k=1}^{c-1}\sum_{r=1}^{c-k} k\boldsymbol{x}_{jrk}(i)\boldsymbol{e}\right] = \lambda p,$$

from which the stated result follows from immediately. ■

3.3. Special cases In this section we will look at some special cases of the model under study.

3.3.1. Case $p = 0$ In this case all primary customers leave the system without requiring any secondary service. Thus, the model under study for this case, reduces to the classical $MAP/M/c$ queue with vacation where the service time distribution is exponentially distributed with parameter μ_1.

3.3.2. Case $b = 1$ In this case any primary customer who requires a secondary service (which occurs with probability p) has to be served immediately irrespective of whether there is any primary customer waiting or not. Thus, in this case, the current model reduces to a special case of $MAP/PH/c$ queue with vacation. This fact was first pointed out recently in [9] even though many papers dealing with such systems in the context of single-server queue overlooked this. Recall that a PH-distribution is obtained as the time until absorption in a finite state Markov chain with one absorption state. It is characterized by an initial probability vector and a square matrix governing the transitions to various transient states. PH-distributions are defined for both discrete and continuous time. For details on PH-distributions and their properties, we refer the reader to [48, 49, 51].

THEOREM 3.5 *In the case when $b = 1$, the current model reduces to the classical $MAP/PH/c$ queue with vacation where the $PH-$representation of service time is of the form $(\boldsymbol{\zeta}, T)$ of dimension 2 given by*

$$\boldsymbol{\zeta} = (1, 0), \quad T = \begin{pmatrix} -\mu_1 & p\mu_1 \\ 0 & -\mu_2 \end{pmatrix}.$$

PROOF In the case when $b = 1$ any primary customer requesting a secondary service (which occurs with probability p) has to be served soon after the primary service. Hence, the server will initiate a secondary service to this customer immediately. The stated result follows immediately. ■

Note: (a) In the case when $b = 1$, the mean service time, μ'_{ST}, is given by

$$\mu'_{ST} = \boldsymbol{\zeta}(-T)^{-1}\boldsymbol{e} = \frac{1}{\mu_1} + \frac{p}{\mu_2}.$$

(b) The stability condition for this special case is $\lambda\mu'_{ST} < c$ which, as it should, agrees with (7) by setting $b = 1$.

3.4. The System Performance Measures In this section we will list a number of system performance measures of interest along with their expressions.

1. Probability that the system is idle. The probability, P_{idle}, that the system is idle (i.e., there are no customers in the system) is given by

$$P_{idle} = \boldsymbol{x}_{000}(0)\boldsymbol{e}.$$

2. Mean number of servers busy with primary customers. The mean, μ_{BPC}, number of servers busy with primary customers is as given in (25).

3. Mean number of servers busy with secondary customers. The mean, μ_{BSC}, number of servers busy with batches of secondary customers is as given in (26).

4. Mean number of servers on vacation. The mean, μ_{NV}, number of servers on vacation customers in the queue is given by $c - \mu_{BPC} - \mu_{BSC}$.

5. Mean number of primary customers waiting in the queue. The mean, μ_{NPQ}, number of primary customers waiting in the queue is given by

$$\mu_{NPQ} = \boldsymbol{x}(0)R(I-R)^{-2}\boldsymbol{e}.$$

6. Mean number of secondary customers waiting in the queue. The mean, μ_{NSQ}, number of secondary customers waiting in the queue is given by

$$\mu_{NSQ} = \sum_{j=1}^{b-1} j \sum_{i=0}^{\infty} \boldsymbol{x}_j(i)\boldsymbol{e}.$$

7. Probability mass function of the number of secondary customers served in a batch. The probability mass function, $\{\alpha_j\}$, of the number of secondary customers served in a batch is given by

$$\xi_j = \begin{cases} d[p\mu_1\ \hat{\boldsymbol{x}}_0(0)\boldsymbol{a}_1 + q\mu_1\ \boldsymbol{x}_1(0)\boldsymbol{a}_1 + \mu_2\ \boldsymbol{x}_1(0)\boldsymbol{a}_2 + \theta\ \boldsymbol{x}_1(0)\boldsymbol{a}_3], & j = 1, \\ d[p\mu_1\ \boldsymbol{x}_{j-1}(0)\boldsymbol{a}_1 + \boldsymbol{x}_j(0)(q\mu_1\boldsymbol{a}_1 + \mu_2\ \boldsymbol{a}_2 + \theta\ \boldsymbol{a}_3)], & 2 \leqslant j \leqslant b-1, \\ dp\mu_1\ \sum_{i=0}^{\infty} \boldsymbol{x}_{b-1}(i)\boldsymbol{a}_1, & j = b, \end{cases}$$

where

$$\hat{\boldsymbol{x}}_0(0) = (\boldsymbol{x}_{010}(i), \cdots, \boldsymbol{x}_{0c0}(i)), \quad \boldsymbol{a}_1 = (1, \cdots, 1, 2, \cdots, 2, \cdots, c-1, c-1, c)'$$

$$\boldsymbol{a}_2 = (0, 1, \cdots, c-1, 0, 1, \cdots, c-2, \cdots, 0, 1, 0)',$$

$$\boldsymbol{a}_3 = (c-1, c-2, \cdots, 1, c-2, c-3, \cdots, 1, 0, \cdots, 1, 0, 0)',$$

and d is the normalizing constant. From the probability mass functions, $\{\xi_j\}$, one can compute the mean (μ_{BS}) and standard deviation (σ_{BS}) of the number of secondary customers served in a batch.

8. Mean number of customers in the system. The mean, μ_{NS}, number of customers in the system is given by

$$\mu_{NS} = \mu_{NPQ} + \mu_{NSQ} + \frac{\lambda}{\mu_1} + \frac{\lambda p}{\mu_2}.$$

9. Mean effective service time. The mean, μ'_{EFS}, effective service time can be obtained using Little's law as

$$\mu'_{EFS} = \frac{\mu_{NS} - \mu_{NPQ}}{\lambda}.$$

4. Illustrative numerical examples In this section we discuss the qualitative aspects of the queueing system under consideration through illustrative numerical examples. First we note that it is very important to verify the correctness and the accuracy of the code written to compute various system performance measures. This is accomplished through verifying a number of results. For example, one can use the results of Lemmas 3.2 and 3.3. Also, one can use the special case when $p = 0$ to verify the correctness of the code. We also obtained the numerical solution for the Poisson arrivals in its simple form. Next, we implemented the general algorithm, but using the following MAP representation: Let D_0 be an irreducible, stable matrix with eigenvalue of maximum real part -$\eta < 0$. Let $\boldsymbol{\kappa}$ denote the corresponding left eigenvector, normalized by $\boldsymbol{\kappa e}$=1. Taking $D_1 = -D_0 \boldsymbol{e}\ \boldsymbol{\kappa}$ the MAP representation reduces to the Poisson arrival process with intensity rate η [51]. The general algorithm does not utilize this fact in any manner, but the numerical results were identical.

For the arrival process, we consider the following five sets of values for D_0 and D_1.

1. Erlang (ERL):

$$D_0 = \begin{pmatrix} -5 & 5 \\ & -5 \end{pmatrix}, D = \begin{pmatrix} & \\ 5 & \end{pmatrix}$$

2. Exponential (EXP):

$$D_0 = \begin{pmatrix} -1 \end{pmatrix}, D = \begin{pmatrix} 1 \end{pmatrix}$$

3. Hyperexponential (HEX):

$$D_0 = \begin{pmatrix} -1.90 & 0 \\ 0 & -0.19 \end{pmatrix}, D = \begin{pmatrix} 1.710 & 0.190 \\ 0.171 & 0.019 \end{pmatrix}$$

4. MAP with negative correlation (MNC):

$$D_0 = \begin{pmatrix} -1.00222 & 1.00222 & 0 \\ 0 & -1.00222 & 0 \\ 0 & 0 & -225.75 \end{pmatrix}, D = \begin{pmatrix} 0 & 0 & 0 \\ 0.01002 & 0 & 0.9922 \\ 223.4925 & 0 & 2.2575 \end{pmatrix}$$

5. MAP with positive correlation (MPC):

$$D_0 = \begin{pmatrix} -1.00222 & 1.00222 & 0 \\ 0 & -1.00222 & 0 \\ 0 & 0 & -225.75 \end{pmatrix}, D = \begin{pmatrix} 0 & 0 & 0 \\ 0.9922 & 0 & 0.01002 \\ 2.2575 & 0 & 223.4925 \end{pmatrix}.$$

The above MAP processes will be normalized so as to have a specific arrival rate. However, these are qualitatively different in that they have different variance and correlation structure. The first three arrival processes, namely ERL, EXP, and HEX, correspond to renewal processes and so the correlation is 0. The arrival process labeled MNC has correlated arrivals with correlation between two successive inter-arrival times given by -0.4889 and and the arrivals corresponding to the processes labeled MPC has a positive correlation with values 0.4889. The ratio of the standard deviations of the inter-arrival times of these five arrival processes with respect to ERL are, respectively, 1, 1.41421, 3.17451, 1.99336, and 1.99336.

In our illustrations below, we fix $\lambda = 0.9$, $\frac{\lambda}{c\mu_1} = 0.9$, $\mu_2 = 10$, and $\theta = 1.0$. All other parameters are varied, and the details are spelled out in appropriate places. First, we look at the effect of p and b by comparing the current model, $MAP/M, M/c$, to that of the clasical $MAP/M/c$ queueing model with vacation. In the classical queueing model with vacation, the mean number of servers busy serving (primary) customers is $\frac{\lambda}{\mu_1}$ and hence the mean number on vacation is given by $c - \frac{\lambda}{\mu_1}$. For the current model, the mean number of customers busy with primary customers is also $\frac{\lambda}{\mu_1}$(see Lemma 3.3). We now look at the reduction in the mean number of servers going on vacation due to the (optional) secondary services as functions of p and b. This reduction is quantified by looking at the ratio, $\frac{\mu_{NV}}{c-\frac{\lambda}{\mu_1}}$, of the mean number of servers on vacation for these two models. In Figure 1 we display this ratio for various scenarios and for c varying from 1 to 5. A quick look at this figure reveals the following observations. Note that the observations for the case $c = 1$ are (as it should be) as summarized in [9].

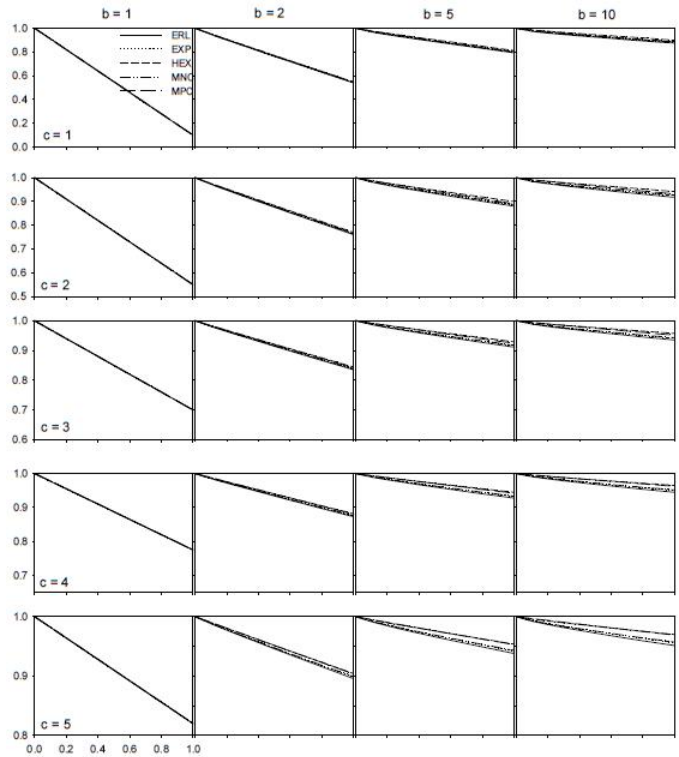


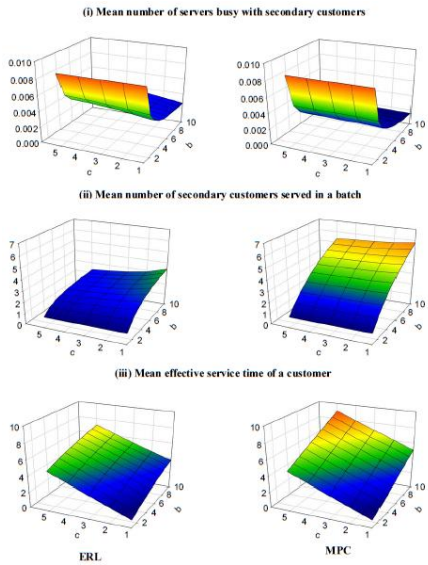
Figure 1: Ratio of P(server is on vacation) for $MAP/ERS, ERS/1$

- As p increases the ratio decreases as it should be. This is the case for all cases considered. However, the rate of decrease (as a function p) is higher for smaller values of b. This is to be expected since a smaller b will cause the server to visit the secondary customers more frequently resulting in less vacation.

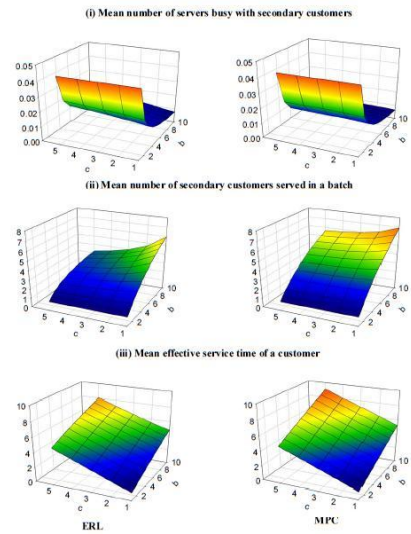
- As c increases we notice that this ratio appears to increase when all other parameters are fixed. Further, the differences in this ratio when the arrival process is varied are noticeably seen for large values of b. Also, the range for the values of this ratio appears to decrease when c is increased. This is true for all values of b.

- In the case when $b = 1$ even for small values of p there is a significant change in the ratio. This appears to be the case for all scenarios.

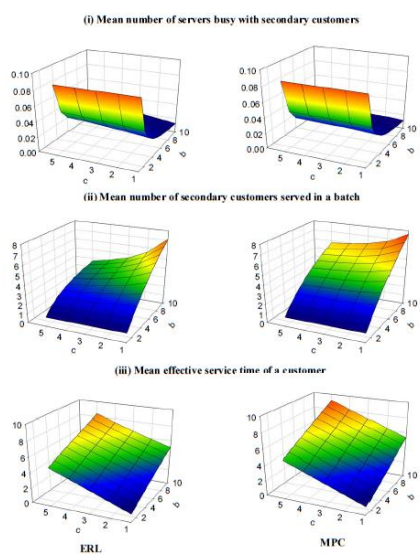
- In the case of large b, even when every customer requires an additional service (i.e., $p = 1.0$) the maximum reduction in the mean number of servers on vacation ranges from about 12% (when $c = 1$) to 5% (when $c = 5$). This indicates that it is better to postpone serving secondary



(a) $p = 0.1$



(b) $p = 0.5$



(c) $p = 1.0$

Figure 2: Comparison of selected measures for two arrival processes

customers to a later point in time by housing them in a secondary buffer.

With respect to the system performance measures, μ_{BSC}, μ_{BS}, and μ'_{EFS}, we noticed that the arrival processes EXP and MNC behave similar to that of ERL, and the arrival process HEX yield similar results to that of MPC process. Hence, we will display only the graphs of these measures (in Figure 2) for the two arrival processes, ERL and MPC, for various scenarios. In Figure 3 we display the ratios (MPC over ERL) of these measures. From these two figures we notice the following observations.

- With regard to the measure μ_{BSC} we notice an increasing trend as p increases; an increasing trend as b increases, and decreasing trend as b increases. These are as expected. Further this measure for ERL arrivals is much higher as compared to MPC arrivals (only when $b = 1$ these two arrivals have the same value for this measure) for all combinations.

- With regard to the measure μ_{BS} we notice an increasing trend as p increases; an increasing trend as b increases; a decreasing trend as c increases. These are as expected. For example, in the case when c is increased there is a higher probability for a server to be free (either from finishing a service or by returning back from vacation) to offer services to waiting secondary customers. Further this measure for MPC arrivals is much higher as compared to ERL arrivals (only when $b = 1$ these two arrivals have the same value for this measure) for all combinations. The ratio of this measure (MPC to ERL) appears to widen as c is increased as well as when p is decreased.

- With regard to the measure μ'_{EFS} we notice an increasing trend as p increases; an increasing trend as b increases; an increasing trend as c increases. These are as expected. Also this measure for MPC arrivals is much higher as compared to ERL arrivals (only when $b = 1$ these two arrivals have the same value for this measure) for all combinations. However, we see an interesting trend when $p = 0.1$ as compared to $p = 0.5$ and $p = 1.0$. In the case when $p = 0.1$, we see that the ratio appears to decrease as c increases; but for the other two cases, the ratio appears to increase with increasing c.

5. Concluding remarks In this paper we studied a multi-server queueing model with non-renewal arrivals in which servers take vacation when becoming idle. After receiving (primary) services the customers may require an optional service with a certain probability. Such models have applications in practice notably in service industries. Assuming the service times to be exponentially distributed, we employed a threshold for offering secondary services. The model was analyzed in steady state using matrix-analytic methods.

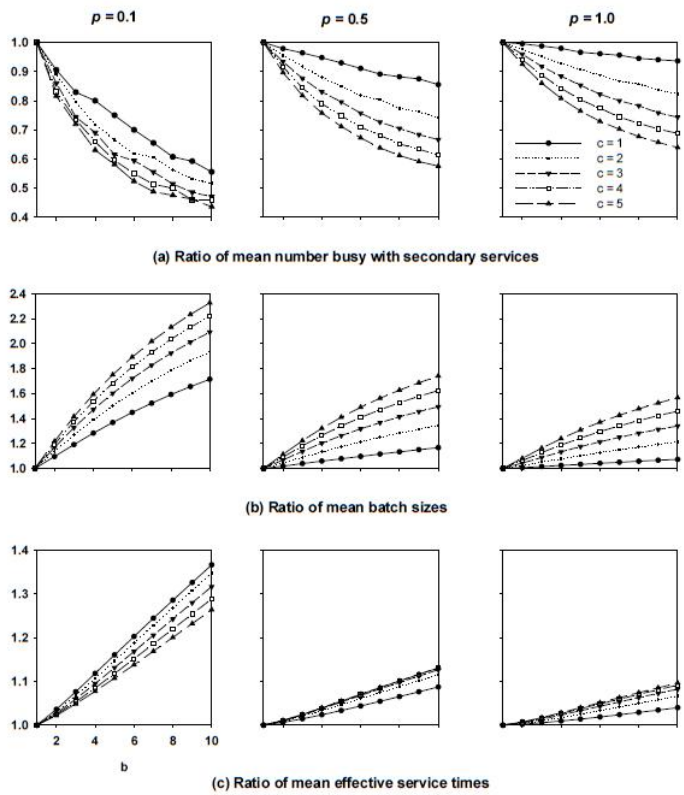


Figure 3: Ratios of MPC to ERL for selected measures

Through some illustrative examples, we presented the effects of the probability and the threshold on selected system performance measures. Specifically, we showed that in the case when $b = 1$ even for small values of p there is a significant change in the the mean number of servers going on vacation (comparing the current model to the classical $MAP/M/c$ queue with vacations). In the case of large b, even when every customer requires an additional service (i.e., $p = 1$) we noticed that it is better to postpone serving secondary customers to a later point in time by housing them in a secondary buffer. The significant role played by the correlated arrivals was highlighted.

Acknowledgment: The author thanks the reviewers and the editor for their valuable suggestions that improved the presentation.

References

[1] Alfa, A.S. Vacation models in discrete time. Queueing Systems, 44, 5-30, 2003. MR 1989864, Zbl 1046.90016, doi: 10.1023/A:1025653200299

[2] Artalejo, J.R., Gomez-Correl, A., and He, Q.M. Markovian arrivals in stochastic modelling: a survey and some new results. SORT, 34 (2), 101-144, 2010. MR 2797472, Zbl 1230.90059, doi: 10.1007/s10626-009-0075-6

[3] Banik, A.D., Gupta, U.C., and Pathak, S.S. $BMAP/G/1/N$ queue with vacations and limited service discipline. Applied Mathematics and Computation, 180, 707-721, 2006. MR 2270048, Zbl 1105.60069, doi: 10.1016/j.amc.2006.01.006

[4] Blondia, C. Finite capacity vacation models with non-renewal input. Journal of Applied Probability, 28, 174-197, 1991. MR 1090456, Zbl 0721.60107, doi: 10.2307/3214749

[5] Chae, K.C., Lee, H.W., and Ahn, C.W. An arrival time approach to $M/G/1-$type queues with generalized vacations. Queueing Systems, 38, 91-100, 2001. MR 1839241, Zbl 1017.90015, doi: 10.1023/A:1010876229827

[6] Chakravarthy, S., and Parthasarathy, R. A finite capacity queueing model with primary and secondary services. Computers and Industrial Engineering, 16, 97-108, 1989. doi: 10.1016/0360-8352(89)90012-0

[7] Chakravarthy, S.R. The batch Markovian arrival process: A review and future work. Advances in Probability Theory and Stochastic Processes. Krishnamoorthy, A., Raju, N., and Ramaswami, V (Eds.), Notable Publications Inc., NJ, 21-39, 2001.

[8] Chakravarthy, S.R. Markovian arrival processes. Wiley Encyclopedia of Operations Research and Management Science. Published Online: 15 JUN 2010.

[9] Chakravarthy, S.R. Analysis of $MAP/PH_1/PH_2/1$ queue with vacations and optional secondary services. To appear in **Applied Mathematical Modeling**.

[10] Chang, S.H., Takine, T., Chae, K.C., and Lee, H.W. A unified queue length formula for $BMAP/G/1$ queue with generalized vacations. Stochastic Models, 18, 369-386, 2002. MR 1928493, Zbl 1013.60063, doi: 10.1081/STM-120014218

[11] Chang, S.H., and Takine, T. Factorization and stochastic decomposition properties in bulk queues with generalized vacations. Queueing Systems, 50, 165-183, 2005. MR 2143119

[12] Choi, D.I., and Kim, T. Analysis of a two-phase queueing system with vacations and Bernoulli feedback. Stochastic Analysis and Applications, 21, 1009-1019, 2003. MR 2000410

[13] Choudhury, G. Some aspects of an $M/G/1$ queueing system with optional second service. TOP, 11, 141-150, 2003. MR 1991213

[14] Choudhury, G., and Paul, M. A batch arrival queue with an additional service channel under N- policy. Applied Mathematics and Computation, 156, 115-130, 2004. MR 2087256

[15] Choudhury, G., and Madan, K. C. (2004). A two-phase batch arrival queueing system with a vacation time under Bernoulli schedule. Applied Mathematics and Computation, 149, 337-349, 2004. MR 2033072

[16] Choudhury, G., and Paul, M. Analysis of a two phase batch arrival queueing model with Bernoulli vacation schedule. Revista Investigation Operacional, 25, 3, 217-228, 2004. MR 2957676

[17] Choudhury, G., and Madan, K.C. A two-stage batch arrival queueing system with a modified Bernoulli schedule vacation under N-policy. Mathematical and Computer Modelling, 42, 71-85, 2005. MR 2162388

[18] Choudhury, G. Steady state analysis of a $M/G/1$ queue with linear retrial policy and two-phase service under Bernoulli vacation schedule. Applied Mathematical Modelling, 32, 2480-2489, 2008. MR 2456307

[19] Dimitriou, I., and Langaris, C. Analysis of a retrial queue with two-phase service and server vacations. Queueing Systems, 60, 111-129, 2008. MR 2452752

[20] Dimitriou, I., and Langaris, C. A repairable queueing model with two-phase service, start-up times and retrial customers. Computers and Operations Research, 37, 1181-1190, 2010. MR 2577279

[21] Doshi, B.T. A note on stochastic decomposition in a $GI/G/1$ queue with vacations or set-up times. Journal of Applied Probability, 22, 419-428, 1985. MR 0789364

[22] Doshi, B.T. Queueing systems with vacations - a survey. Queueing Systems, 1, 29-66, 1986. MR 0896237

[23] Doshi, B.T. Generalization of the stochastic decopmosition results for the single-server queues with vacations, Stochastic Models, 6, 307-333, 1990. MR 1047974

[24] Doshi, B.T. Single server queues with vacations, in: Stochastic Analysis of Computer and Communications Systems, ed. H. Takagi, 217-265, 1990. MR 1150292

[25] Doshi, B.T. Analysis of a two-phase queueing system with general service times. Operations Research Letters, 10, 265-272, 1991. MR 1122328

[26] Ferrandiz, J.M. The $BMAP/GI/1$ queue with server set-up times and server vacations. Advances in Applied Probability, 25, 235-254, 1993. MR 1206541

[27] Graham, A. Kronecker Products and Matrix Calculus with Applications. Ellis Horwood, Chichester, UK, 1981. MR 0640865

[28] Jeyakumar, S., and Arumuganathan. A non-Markovian bulk queue with multiple vacations and control policy on request for re-service. Quality Technology and Quantitative Management, 8, 253-269, 2011.

[29] Kalyanaraman, R., and Bala Murugan, S.P. A single server queue with additional optional service in batches and server vacation. Applied Mathematical Sciences, 2, 2765-2776, 2008. MR 2480408

[30] Kasahara, S., Takine, T., Takahashi, Y., and Hasegawa, T. $MAP/G/1$ queues under N-policy with and without vacations. Journal of Operations Research Society of Japan, 39, 188-212, 1996. MR 1401017

[31] Katayama, T., and Kobayashi, K. Sojourn time analysis of a queueing system with two-phase service and server vacations. Naval Research Logistics, 54, 59-65, 2007. MR 2285485

[32] Ke, J-C., C-H. Wu., and Zhang, Z.G. Recent developments in vacation queueing models: A short survey. International Journal of Operations Research, 7, 3-8, 2010.

[33] Kim, T.S., and Park, A.Q. Cycle analysis of a two-phase queueing model with threshold. European Journal of Operational Research, 144, 157-165, 2003. MR 1940081

[34] Krishna, C.M., and Lee, Y.H. A study of a two-phase service. Operations Research Letters, 9, 91-97, 1990. MR 1049817

[35] Krishna Kumar, B., Vijayakumar, A., and Arivudainambi, D. An $M/G/1$ retrial queueing system with two-phase service and preemptive resume. Annals of Operations Research, 113, 61-79, 2002. MR 1960682

[36] Krishnamoorthy, A., Pramod, P., and Chakravarthy, S.R. Queues with interruptions: A survey. TOP, doi: 10.1007/s11750-012-0256-6. Published online: 16 March 2012.

[37] Latouche G. and Ramaswami V. Introduction to matrix analytic methods in stochastic modeling. SIAM, 1999. MR 1674122, Zbl 0922.60001

[38] Lee, H.W., Ahn, B.Y., and Park, N.I. Decompositions of the queue length distributions in the $MAP/G/1$ queue under multiple and single vacations with N-policy. Stochastic Models, 17, 157-190, 2001. MR 1853439

[39] Lucantoni, D.M. New results on the single server queue with a batch Markovian arrival process. Stochastic Models, 7, 1-46, 1991. MR 1102528

[40] Lucantoni, D., Meier-Hellstern, K.S., and Neuts, M.F. A single-server queue with server vacations and a class of nonrenewal arrival processes. Advances in Applied Probability, 22, 676-705, 1990. MR 1066970

[41] Madan, K.C. An $M/G/1$ queue with second optional service. Queueing Systems, 34, 36-46, 2000. MR 1769764

[42] Madan, K.C. On a single server queue with two-stage heterogeneous service and deterministic server vacations. International Journal of System Science, 32, 837-844, 2001. MR 1959614

[43] Madan, K. C., Al-Nasser, A. D., and Al-Masri, A. Q. On $M^X/G(G_1, G_2)/1$ queue with optional re-service. Applied Mathematics and Computation, 152, 71-88, 2004. MR 2050054

[44] Marcus, M., and Minc, H. A survey of matrix theory and matrix inequalities. Allyn and Bacon, Boston, MA, 1964. MR 0162808

[45] Matendo, S.K. A single-server queue with server vacations and a batch Markovian arrival process. Cahiers C.E.R.O. 35, 87-114, 1993. MR 1268965

[46] Medhi, J. A single server Poisson input queue with a second optional channel. Queueing Systems, 42, 239-242, 2002. MR 1935140

[47] Neuts, M.F. A versatile Markovian point process. Journal of Applied Probability, 16, 764-779, 1979. MR 0549556

[48] Neuts, M.F. Matrix-geometric solutions in stochastic models: An algorithmic approach. The Johns Hopkins University Press, Baltimore, MD, 1981. [1994 version is Dover Edition]. MR 1313503

[49] Neuts, M.F. Structured stochastic matrices of $M/G/1$ type and their applications. Marcel Dekker, NY, 1989. MR 1010040

[50] Neuts, M.F. Models based on the Markovian arrival process. IEICE Transactions on Communications, E75B, 1255-1265, 1992.

[51] Neuts, M.F. Algorithmic Probability: A collection of problems. Chapman and Hall, NY, 1995. MR 1369776

[52] Nishimura, S. A spectral method for a nonpreemptive priority BMAP/G/1/N queue. Stochastic Models, 21, 579-597, 2005. MR 2148774

[53] Saffer, Z., and Telek, M. Unified analysis of $BMAP/G/1$ cyclic polling models. Queueing Systems, 64, 69-102, 2010. MR 2584755

[54] Schellhaas, H. Single server queues with a batch Markovian arrival process and server vacations. OR Spektrum, 15, 189-196, 1994. MR 1263420

[55] Stewart, W.J. Introduction to the Numerical Solution of Markov Chains. Princeton University Press, Princeton, NJ, 1994. MR 1312831

[56] Takine, T., and Takahashi, Y. On the relationship between queue lengths at a random instant and at a departure in the stationary queue with BMAP arrivals. Stochastic Models, 14, 601-610, 1998. MR 1621362

[57] Takine, T. The nonpreemptive priority $MAP/G/1$ queue. Operations Research, 47, 917-927, 1999. MR 1755190

[58] Tian, N.S., and Zhang, Z.G. Vacation queueing models: Theory and applications. Springer Publishers, New York. 2006. MR 2248264

[59] Wang, J. An $M/G/1$ queue optional second service and server breakdowns. Computers and Mathematics with Applications, 47, 1713-1723, 2004. MR 2079878

[60] Wu, J.; Liu., Z. and Peng. Y. On the $BMAP/G/1$ $G-$queues with second optional service and multiple vacations. Applied Mathematical Modelling, 33, 4314-4325, 2009. MR 2566982, Zbl 1171.90374, doi: 10.1016/j.apm.2009.03.013

Analiza systemu kolejkowego z wieloma stanowiskami obsługi dostępnymi na żądanie i fakultatywnym spektrum czynności

Srinivas R. Chakravarthy

Streszczenie Ten artykuł poświęcony jest modelom kolejkowym dla systemów z wieloma serwerami z markowskim strumieniem zgłoszeń. Klienci żądają, aby obsługa świadczyła również pewne opcjonalne usługi po zakończeniu podstawowego procesu. Te usługi dodatkowe (o różnym zakresie) mają być dostępne i oferowane z pewnym prawdopodobieństwem, gdy którykolwiek z następujących warunków jest spełniony: (a) po zakończeniu obsługi na darmowy, podstwowy, serwis nie czeka klient w kolejce i jest co najmniej jeden chętny klient na serwis wtórny (tym chętnym prawdopodobnie jest klientem, który własnie otrzymał podstawową usługe), (b) po zakończeniu podstawowego serwisu, klient wymaga dodatkowego serwisu i w tym czasie liczba klientów, którzy reflektują na tę dodatkową usługę przekroczy wcześniej ustaloną wartość progową; (c) serwer który wznawia obsługę po przerwie nie ma klientów na podstawową usługę, ale przynajmniej jeden klient czeka na dodatkowy serwis. Serwery mogą zostać wyłączone na pewien czas, gdy nie ma klientów (podstawowych

lub chętnych na serwis dodatkowy) czekających na obsługę. Model jest badane jako uogólniony proces urodzin i śmierci (quasi-birth-death-matrix-process) analizowany analitycznie. Podane są przykłady ilustrujące zastosowane podejście.

2010 Klasyfikacja tematyczna AMS (2010): 60K25; secondary: 90B22, 60K37, 60K30, 60H30, 90B05.

Słowa kluczowe: Markowski proces zgłoszeń, kolejka, czas oczekiwania, wyłączenia, drugoplanowe zestawy usług, obsługa wielostanowiskowa, mieszanki rozkładów wykładniczego typu, prawdopodobieństwo algorytmiczne.

Srinivas R. Chakravarthy is Professor and Interim Head of the Department of Industrial and Manufacturing Engineering, Kettering University, Flint, MI 48504. He has a PhD in Operations Research, MS in Statistics and BS in Mathematics. His research interests are in applied stochastic models, algorithmic probability, queuing, reliability and inventory. He has published more than 95 papers in leading journals, and presented papers at national and international conferences. He has received NSF awards and organized the First and Second International Conferences on Matrix-analytic Methods in Stochastic Models. He received Kettering University Alumni's Outstanding Teacher Awards and Kettering University's Outstanding Researcher Awards. He is a member of INFORMS, ASA, IIE and Sigma Xi Research Society. He is currently an area editor for Simulation Modeling Practice and Theory, associate editor of IAPQR Transactions, and a member of the advisory boards of many journals.

Srinivas R. Chakravarthy
Kettering University
Department of Industrial and Manufacturing Engineering
Flint, MI-48504, USA
E-mail: schakrav@kettering.edu
URL: http://paws.kettering.edu/ schakrav/
Communicated by: Krzysztof Szajowski

(Received: 5th December 2012)
