Applied Mathematical Modelling 33 (2009) 2967–2977

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Multi-server system with single working vacation

Chuen-Horng Lin [a], Jau-Chuan Ke [b,*]

[a] Graduate School of Computer Science and Information Technology, National Taichung Institute of Technology, Taichung 404, Taiwan, ROC
[b] Department of Applied Statistics, National Taichung Institute of Technology, No. 129, Sec. 3, Sanmin Rd., Taichung 404, Taiwan, ROC

ARTICLE INFO

Article history:
Received 14 March 2007
Received in revised form 26 September 2008
Accepted 13 October 2008
Available online 21 October 2008

Keywords:
Cost
Matrix–geometric approach
Optimization
Queue
Single working vacations

ABSTRACT

We consider an M/M/R queue with vacations, in which the server works with different service rates rather than completely terminates service during his vacation period. Service times during vacation period, service times during service period and vacation times are all exponentially distributed. Neuts' matrix–geometric approach is utilized to develop the computable explicit formula for the probability distributions of queue length and other system characteristics. A cost model is derived to determine the optimal values of the number of servers and the working vacation rate simultaneously, in order to minimize the total expected cost per unit time. Under the optimal operating conditions, numerical results are provided in which several system characteristics are calculated based on assumed numerical values given to the system parameters.

1. Introduction

We analyze an M/M/R queue with single working vacations and exhaustive service, in which the server works with variable service rates rather than completely stops service during his vacation period. Such a vacation is called a *working vacation* [1]. Each server starts a working vacation when the system is empty at his service completion instant. When a working vacation for some server ends, if there are no customers in the system, he stays idle and is ready for serving new arrivals. If the server returns from a working vacation to find the system not empty, he immediately switches to original service rate. The time interval between two successive vacations is called a service period [2, p. 107]. The service times during a service period, and service times during a working vacation are both exponentially distributed. The working vacation presented this paper is called a *single working vacation*.

It is assumed that customers arrive according to a Poisson process with parameter λ. The service times during a service period follow exponential distribution with mean $1/\mu_B$. The service times during a vacation period follow another exponential distribution with mean $1/\mu_v$. When there are no customers in the system at a service completion instant, the server begins a working vacation of mean duration $1/\eta$, where vacation times are exponentially distributed. We assume that arriving customers form a single waiting line based on the order of their arrivals; that is, the first-come, first-served discipline is followed. Suppose that the server can serve only one customer at a time. Customers entering into the service facility and finding that the server is busy have to wait in the queue until any server is available.

Our study is motivated by some practical systems. For providing network services, such as web service, file transfer service, and mail service, many network servers are located at computer center in a college. To keep the servers functioning well, virus scan is an important maintenance activity for the servers. It can be performed when the servers are idle. This type of maintenance could be programmed to perform on a regular basis. Although the operation of virus scan would consume

* Corresponding author. Tel.: +886 4 22196330; fax: +886 4 22196331.
E-mail address: jauchuan@ntit.edu.tw (J.-C. Ke).

0307-904X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.apm.2008.10.006

some system resources and reduce the processing speed, the servers could still provide his service with lower processing speed during the period of virus scan. When virus scan is done, the servers will enter the idle state again and wait the use requests arrive.

During the last two decades, the queueing systems with server vacations or working vacations have been investigated by many researchers. Past work may be divided into two categories: (i) the case of server vacation and (ii) the case of working vacation. In the case of server vacation, the readers are referred to the survey paper by Doshi [3] and the monograph of Takagi [2]. The works of Takagi [2] and Doshi [3] focus on a single server. As for multiple-server system with vacations, Zhang and Tian [4,5] recently gave a plenty analysis of M/M/c with synchronous multiple/single vacations of partial servers. In the case of working vacations, Servi and Finn [1] first examined an M/M/1 queue with multiple working vacations (Such model is denoted by M/M/1/WV queue), where inter-arrival times, service times during service period, service times during vacation period, and vacation times are all exponentially distributed. They developed the explicit formulae for the mean and variance number of customers in the system, and the mean and variance waiting time in the system. Later Wu and Takagi [6] extended Servi and Finn's M/M/1/WV queue to an M/G/1/WV queue. They assumed that service times during service period, service times during vacation period as well as vacation times are all generally distributed. Further, they assumed that when a working vacation ends, if there are customers in the system, the server changes to another service rate, where the service times follow a different distribution. Again, Baba [7] extended Servi and Finn's M/M/1/WV queue to a GI/M/1/WV queue. They not only assumed general independent arrival, they also assumed service times during service period, service times during vacation period as well as vacation times following exponential distribution. Furthermore, Baba [7] derived the steady-state system length distributions at arrival and arbitrary epochs. Recently, Banik et al. [8] studied a finite capacity GI/M/1 queue with multiple working vacations. They developed some important system performance measures such as, the probability of blocking and the expected waiting time in the system.

In the study of working vacations, the existing literature focuses mainly on a single server queue with multiple vacations. Also, this would motivate us to investigate the M/M/R queue with a single working vacation. The well-known matrix–geometric approach by Neuts [9] is used to analyze such multi-server system with a single working vacation as studied in [4,5]. This paper is organized as follows. In Section 2, we develop the quasi-birth–death chain for the single working vacation model. The steady-state equations are solved associated with the use of matrix–geometric approach given in Section 3. The system characteristics follow in Section 4. Cost analysis and sensitivity investigation are presented in Section 5. Finally, Section 6 consists of concluding remarks.

2. Quasi-birth–death chain model

The states of the system are described by the pair (k,n), $k = 0,1,2,\ldots,R$, $n = 0,1,2,\ldots$, where $k = 0$ denotes that all servers are on a working vacation, $k = 1$ denotes that there are $(R-1)$ servers on a working vacation, $k = j$ denotes that there are $(R–j)$ servers on a working vacation, and n represents the number of customers in the system.

The steady-state equations for $p_{k,n}$ $(k = 0,1,2,\ldots,R)$ relating to Fig. 1 are given by

(i) $k = 0$ and $n = 0$,

$$(\lambda + R\eta)p_{0,0} - \mu_v p_{0,1} - \mu_B p_{1,1} = 0. \tag{1}$$

(ii) $k = 0$ and $1 \leqslant n \leqslant R-1$,

$$-\lambda p_{0,n-1} + (\lambda + n\mu_v + R\eta)p_{0,n} - (n+1)\mu_v p_{0,n+1} = 0. \tag{2}$$

(iii) $k = 0$ and $n \geqslant R$,

$$-\lambda p_{0,n-1} + (\lambda + R\mu_v + R\eta)p_{0,n} - R\mu_v p_{0,n+1} = 0. \tag{3}$$

(iv) $1 \leqslant k \leqslant R-1$ and $n = 0$,

$$-(R-k+1)\eta p_{k-1,0} + [\lambda + (R-k)\eta]p_{k,0} - \mu_B p_{k+1,1} = 0. \tag{4}$$

(v) $2 \leqslant k \leqslant R-1$ and $1 \leqslant n \leqslant k-1$,

$$-(R-k+1)\eta p_{k-1,n} - \lambda p_{k,n-1} + [\lambda + (R-k)\eta + n\mu_B]p_{k,n} - (n+1)\mu_B p_{k+1,n+1} = 0. \tag{5}$$

(vi) $1 \leqslant k \leqslant R-1$ and $n = k$,

$$-(R-k+1)\eta p_{k-1,n} - \lambda p_{k,n-1} + [\lambda + (R-k)\eta + n\mu_B]p_{k,n} - (k\mu_B + \mu_v)p_{k,n+1} - (n+1)\mu_B p_{k+1,n+1} = 0. \tag{6}$$

(vii) $1 \leqslant k \leqslant R-1$ and $k+1 \leqslant n \leqslant R-1$,

$$-(R-k+1)\eta p_{k-1,n} - \lambda p_{k,n-1} + [\lambda + (R-k)\eta + k\mu_B + (n-k)\mu_v]\ p_{k,n} - \{k\mu_B + [(n+1)-k]\mu_v\}p_{k,n+1} = 0. \tag{7}$$

(viii) $1 \leqslant k \leqslant R-1$ and $n \geqslant R$,

$$-(R-k+1)\eta p_{k-1,n} - \lambda p_{k,n-1} + [\lambda + (R-k)\eta + k\mu_B + (n-k)\mu_v]p_{k,n} - [k\mu_B + (n-k)\mu_v]p_{k,n+1} = 0. \tag{8}$$