Available online at www.sciencedirect.com

ScienceDirect

Applied Mathematical Modelling 32 (2008) 2715–2730

APPLIED MATHEMATICAL MODELLING

www.elsevier.com/locate/apm

Performance analysis of GI/M/1 queue with working vacations and vacation interruption

Ji-Hong Li [a], Nai-Shuo Tian [b,*], Zhan-You Ma [b]

[a] *College of Economics and Management, Yanshan University, Qinhuangdao 066004, China*
[b] *College of Science, Yanshan University, Qinhuangdao 066004, China*

Received 15 November 2006; received in revised form 19 September 2007; accepted 20 September 2007
Available online 29 September 2007

Abstract

In this paper, we analyze a single-server vacation queue with a general arrival process. Two policies, working vacation and vacation interruption, are connected to model some practical problems. The GI/M/1 queue with such two policies is described and by the matrix analysis method, we obtain various performance measures such as mean queue length and waiting time. Finally, using some numerical examples, we present the parameter effect on the performance measures and establish the cost and profit functions to analyze the optimal service rate η during the vacation period.

Keywords: Vacation interruption; Working vacations; Embedded Markov chain; Matrix-analysis approach; Numerical analysis

1. Introduction

In this paper, we consider a single-server system where there is a class of main jobs or randomly arriving customers. In general, allowing the server to take the vacation when there are less customers can economize the cost, but also reduce the operational efficiency of the system and cause the loss and dissatisfaction from the customers. How can we solve this problem to make the system operate more efficiently? To answer this question, we develop a new quantitative model based on Markov process with the general arrival. Specially, we want to know (1) how the server uses the vacation time to serve main jobs or customers; and (2) when the system ends vacations to come to the normal working level. In this paper, we focus on working vacation and vacation interruption policies which provide the answers to the two questions, respectively. Under the working vacation policy, the server can also take service for the main jobs or customers during the vacation period with the lower rate rather than stopping completely. The vacation period becomes the lower speed operation period of the queueing system. Meanwhile, under the vacation interruption policy, the server can come back to the normal working level immediately once some indices of the system, such as the number

* Corresponding author. Tel./fax: +86 335 8074703.
E-mail address: tiannsh@ysu.edu.cn (N.-S. Tian).

0307-904X/$ - see front matter
doi:10.1016/j.apm.2007.09.017

of the main jobs, achieves the certain value. Thus, working vacation policy determines how the server use the vacation time and vacation interruption controls when the system ends the vacation period or the lower speed operation period.

The model presented in this paper belongs to the research area of single-server queues with vacations. General single-server vacation models have been well studied and surveyed by Doshi [1,2] and the monographs of Takagi [3] and Tian and Zhang [4]. For working vacation models, Servi and Finn [5] first studied an M/M/1 queue with working vacation. Their work is motivated and illustrated by the analysis of a WDM optical access network using multiple wavelengths which can be reconfigured. Subsequently, Kim et al. [6], Wu and Takagi [7] generalized results in [5] to an M/G/1 queue with working vacation. Baba [8] extended this study to a GI/M/1 queue with working vacation by the matrix-analysis method. Recently, Banik et al. [9] analyzed the GI/M/1/N queue with working vacations and presented a series of numerical results. Li and Tian [10] connected working vacation and vacation interruption and analyzed the discrete-time GI/Geo/1 queue with working vacations and vacation interruption.

In this paper, we will consider a GI/M/1 queue with working vacations and vacation interruption. For GI/M/1 queues with server vacations, Tian et al. [11] used the matrix geometric solution method to analyze. They obtained the expressions for the rate matrix and proved the stochastic decomposition properties for queue length and waiting time in a GI/M/1 vacation model with multiple exponential vacations. Independently, Chatterjee and Mukherjee [12] also researched the GI/M/1 with server vacations. Subsequently, Tian and Zhang [13,14] discussed the GI/M/1 queue with PH vacations or setup times and the discrete time GI/Geo/1 queue with server vacations. But, the server cannot take the original work during the vacation period as in the working vacation models. Meanwhile, in all models with vacations, he cannot come back to the normal working level until the vacation period ends.

In this paper, we will consider a single-server queue system which has the general arrival process. The working vacation and vacation interruption are connected and the server enters into vacation when there are no customers and he can take service at the lower rate during the vacation period. If there are customers in the system at the instant of a service completion during the vacation period, the server will come back to the normal working level no matter whether the vacation ends. Otherwise, he continues the vacation. The aim of our study is to develop a set of computable stationary performance measures such as the queue length and waiting time. With these performance measures, we demonstrate the parameter effect on the performance indices of the system. And, by some numerical results, we can also analyze economic performance measures.

The rest of this paper is organized as follows. In Section 2, the model of GI/M/1 with vacation interruption and multiple working vacations is formulated as an embedded Markov chain at the arrival epochs. In Sections 3 and 4, the main steady-state performance measures such as queue length and waiting time are obtained by the matrix analysis method. Finally, using some numerical examples, we also demonstrate the parameter effect on the performance indices of the system and establish the economic performance functions in Section 5.

2. Model formulation and embedded Markov chain

Consider a GI/M/1 queue such that the arrival process is a general distribution process. The server begins a vacation each time when the queue becomes empty and if there are customers arriving in a vacation period, the server continues to work at a lower rate. The working vacation period is an operation period in lower speed. At a service completion instant, if there are customers in the system in the vacation period, the server will come back to the normal working level, i.e. vacation interruption happens. Otherwise, he continues the vacation. Meanwhile, if there is no customer when a vacation ends, the server begins another vacation, otherwise, he switches to the normal working level.

Suppose τ_n be the arrival epoch of nth customers with $\tau_0 = 0$. The inter-arrival times $\{T_n, n \geqslant 1\}$ are independent and identically distributed with a general distribution function, denoted by $A(t)$ with a mean $1/\lambda$ and a Laplace Stieltjes transform (LST), denoted by $A^*(s)$. The service times during a service period, the service times during a working vacation and the working vacation times are exponentially distributed with rate μ, η and θ, respectively.

Let $L(t)$ be the number of customers in the system at time t and $L_n = L(\tau_n - 0)$ be the number of the customers before the nth arrival. Define

$$J_n = \begin{cases} 1, & \text{the } n\text{th arrival occurs during a service period,} \\ 0, & \text{the } n\text{th arrival occurs during a working vacation period.} \end{cases}$$

Then, the process $\{(L_n, J_n),\ n \geqslant 1\}$ is an embedded Markov chain with the state space

$$\Omega = \{(0,0)\} \bigcup \{(k,j),\ k \geqslant 1,\ j = 0,1\}.$$

Evidently, the server only stays in the vacation period when there are no customers.

In order to express the transition matrix of (L_n, J_n), let

$$p_{(i,j),(k,l)} = P(L_{n+1} = k, J_{n+1} = l | L_n = i, J_n = j).$$

Meanwhile, we introduce the expressions below

$$a_k = \int_0^\infty \mathrm{e}^{-\mu t} \frac{(\mu t)^k}{k!} \, \mathrm{d}A(t); \quad b_k = \int_0^\infty \int_0^t \eta \, \mathrm{e}^{-\eta x} \, \mathrm{e}^{-\theta x} \frac{(\mu(t-x))^{k-1}}{(k-1)!} \, \mathrm{e}^{-\mu(t-x)} \, \mathrm{d}x \, \mathrm{d}A(t);$$
$$c_k = \int_0^\infty \int_0^t \theta \, \mathrm{e}^{-\theta x} \, \mathrm{e}^{-\eta x} \frac{(\mu(t-x))^k}{k!} \, \mathrm{e}^{-\mu(t-x)} \, \mathrm{d}x \, \mathrm{d}A(t),$$

where $\{a_k,\ k \geqslant 0\}$ expresses the probability that k services complete during an inter-arrival time in the normal service period; and, $\{b_k,\ k \geqslant 1\}$ explains the probability that, during an inter-arrival time, the residual vacation time is longer than the service time during the vacation period, a customer is served completely, and there are other customers in the system, then at the instant of this service completion, the server ends the vacation and enters into the normal service period, i.e., vacation interruption happens, and k services complete during this inter-arrival time; certainly, $\{c_k,\ k \geqslant 0\}$ is the probability that, the residual vacation time is not longer than the service time during the vacation period, then no customer is served completely during the residual vacation time, and at the instant of vacation ending when there are customers in the system, the server enters into the normal service period and k services complete during an inter-arrival time.

Now, we consider the transition probabilities of (L_n, J_n). First, considering the case during a service period, the transition from $(i,1)$ to $(j,1)$ occurs if $i+1-j$ services complete during an inter-arrival time. Therefore, we have

$$p_{(i,1),(j,1)} = \begin{cases} \int_0^\infty \mathrm{e}^{-\mu t} \frac{(\mu t)^{i+1-j}}{(i+1-j)!} \, \mathrm{d}A(t) = a_{i+1-j}, & 1 \leqslant j \leqslant i+1. \\ 0, & j \geqslant i+2. \end{cases} \tag{1}$$

Second, the transition from $(i,0)$ to $(j,0)$ occurs only if $j = i+1$. So, we have

$$p_{(i,0),(j,0)} = \int_0^\infty \mathrm{e}^{-\theta t} \, \mathrm{e}^{-\eta t} \, \mathrm{d}A(t) = A^*(\theta + \eta), \qquad j = i+1. \tag{2}$$

Similarly, from the definition of b_k, c_k, we obtain

$$p_{(i,0),(j,1)} = \begin{cases} \int_0^\infty \int_0^t \eta \, \mathrm{e}^{-\eta x} \, \mathrm{e}^{-\theta x} \frac{(\mu(t-x))^{i-j}}{(i-j)!} \, \mathrm{e}^{-\mu(t-x)} \, \mathrm{d}x \, \mathrm{d}A(t) & \\ \quad + \int_0^\infty \int_0^t \theta \, \mathrm{e}^{-\theta x} \, \mathrm{e}^{-\eta x} \frac{(\mu(t-x))^{i+1-j}}{(i+1-j)!} \, \mathrm{e}^{-\mu(t-x)} \, \mathrm{d}x \, \mathrm{d}A(t) & \\ \quad = b_{i+1-j} + c_{i+1-j}, & 1 \leqslant j \leqslant i, \\ \int_0^\infty \int_0^t \theta \, \mathrm{e}^{-\theta x} \, \mathrm{e}^{-\eta x} \, \mathrm{e}^{-\mu(t-x)} \, \mathrm{d}x \, \mathrm{d}A(t) = c_0, & j = i+1. \end{cases} \tag{3}$$

And,

$$p_{(i,0),(0,0)} = \int_0^\infty \int_0^t \eta \mathrm{e}^{-\eta x} \mathrm{e}^{-\theta x} \left(1 - \sum_{k=0}^{i-1} \frac{(\mu(t-x))^k}{k!} \mathrm{e}^{-\mu(t-x)}\right) \mathrm{d}x \, \mathrm{d}A(t)$$
$$+ \int_0^\infty \int_0^t \theta \mathrm{e}^{-\theta x} \mathrm{e}^{-\eta x} \left(1 - \sum_{k=0}^{i} \frac{(\mu(t-x))^k}{k!} \mathrm{e}^{-\mu(t-x)}\right) \mathrm{d}x \, \mathrm{d}A(t)$$
$$= 1 - A^*(\theta+\eta) - \sum_{k=1}^{i} (c_k + b_k) - c_0, \quad i \geqslant 1. \tag{4}$$

From the above equations, we derive

$$p_{(0,0),(0,0)} = 1 - A^*(\theta+\eta) - c_0, \qquad p_{(i,1),(0,0)} = 1 - \sum_{k=0}^{i} a_k, \quad i \geqslant 1.$$

Using the lexicographical sequence for the states, the transition probability matrix of (L_n, J_n) can be written as the Block–Jacobi matrix

$$\widetilde{\boldsymbol{P}} = \begin{bmatrix} \boldsymbol{B}_{00} & \boldsymbol{A}_{01} & & & & \\ \boldsymbol{B}_1 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & & & \\ \boldsymbol{B}_2 & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & & \\ \boldsymbol{B}_3 & \boldsymbol{A}_3 & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{5}$$

where

$$\boldsymbol{B}_{00} = 1 - A^*(\theta+\eta) - c_0; \quad \boldsymbol{A}_{01} = (A^*(\theta+\eta), c_0);$$
$$\boldsymbol{A}_0 = \begin{bmatrix} A^*(\theta+\eta) & c_0 \\ 0 & a_0 \end{bmatrix}; \quad \boldsymbol{A}_k = \begin{bmatrix} 0 & b_k + c_k \\ 0 & a_k \end{bmatrix}, \quad k \geqslant 1;$$
$$\boldsymbol{B}_k = \begin{bmatrix} 1 - A^*(\theta+\eta) - \sum_{i=1}^{k} (c_i + b_i) - c_0 \\ 1 - \sum_{i=0}^{k} a_i. \end{bmatrix}, \quad k \geqslant 1.$$

The structure of $\widetilde{\boldsymbol{P}}$ indicates that (L_n, J_n) is irreducible and aperiodic.

3. Performance measure: queue length

First, we derive the steady-state distribution for (L_n, J_n) at arrival epochs using matrix–geometric approach. In order to derive the steady-state distribution, we need the lemma below.

Lemma 1. *If* $\rho = \lambda/\mu < 1$ *and* $\theta > 0$, *then the matrix equation* $\boldsymbol{R} = \sum_{k=0}^{\infty} \boldsymbol{R}^k \boldsymbol{A}_k$ *has the minimal nonnegative solution*

$$\boldsymbol{R} = \begin{bmatrix} A^*(\theta+\eta) & \beta(\xi - A^*(\theta+\eta)) \\ 0 & \xi \end{bmatrix},$$

where ξ *is the unique root in the range* $0 < z < 1$ *of the equation* $z = A^*(\mu(1-z))$, *and*

$$\beta = \frac{\theta + \eta A^*(\theta+\eta)}{\theta + \eta - \mu(1 - A^*(\theta+\eta))}.$$

Proof. Because all $\boldsymbol{A}_k$ are upper triangular, we can assume that $\boldsymbol{R}$ has the same structure as

$$\boldsymbol{R} = \begin{bmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{bmatrix}.$$

Then, for $k \geqslant 1$, we have

$$\boldsymbol{R}^k = \begin{bmatrix} r_{11}^k & r_{12}\sum_{j=0}^{k-1} r_{11}^j r_{22}^{k-1-j} \\ 0 & r_{22}^k \end{bmatrix}.$$

Substituting $\boldsymbol{R}^k$ into the matrix equation, we obtain

$$\begin{cases} r_{11} = A^*(\theta+\eta), \\ r_{12} = c_0 + \sum_{k=1}^{\infty} r_{11}^k (b_k + c_k) + r_{12}\sum_{k=1}^{\infty} a_k \sum_{j=0}^{k-1} r_{11}^j r_{22}^{k-1-j}, \\ r_{22} = \sum_{k=0}^{\infty} a_k r_{22}^k = A^*(\mu(1-r_{22})). \end{cases} \tag{6}$$

As is well known, if $\rho = \lambda/\mu < 1$, the third equation has the unique root $r_{22} = \xi$ in the range $0 < r_{22} < 1$ and $r_{11}{=}A^*(\theta+\eta)$. Taking r_{22} and r_{11} into the second equation, we have

$$r_{12}\left(1 - \sum_{k=1}^{\infty} a_k \sum_{j=0}^{k-1} r_{11}^j r_{22}^{k-1-j}\right) = \sum_{k=1}^{\infty} r_{11}^k b_k + \sum_{k=0}^{\infty} r_{11}^k c_k. \tag{7}$$

And, we easily compute

$$\sum_{k=1}^{\infty} r_{11}^k b_k + \sum_{k=0}^{\infty} r_{11}^k c_k = \frac{\theta + \eta A^*(\theta+\eta)}{\theta+\eta-\mu(1-A^*(\theta+\eta))}(A^*(\mu(1-r_{11})) - A^*(\theta+\eta)),$$

$$1 - \sum_{k=1}^{\infty} a_k \sum_{j=0}^{k-1} r_{11}^j r_{22}^{k-1-j} = 1 - \frac{r_{22} - A^*(\mu(1-r_{11}))}{r_{22}-r_{11}} = \frac{A^*(\mu(1-r_{11})) - A^*(\theta+\eta)}{r_{22}-r_{11}}.$$

Substituting the above equations into (7), we finally obtain the expression for $\boldsymbol{R}$. We have $\beta(\xi - A^*(\theta+\eta)) > 0$ for $\rho < 1$ and $\theta > 0$. Meanwhile, we easily verify that the Markov chain $\widetilde{\boldsymbol{P}}$ is positive recurrent if and only if $\rho < 1$ and $\theta > 0$. And, the matrix

$$B[\boldsymbol{R}] = \begin{bmatrix} \boldsymbol{B}_{00} & \boldsymbol{A}_{01} \\ \sum_{k=1}^{+\infty} \boldsymbol{R}^{k-1}\boldsymbol{B}_k & \sum_{k=1}^{+\infty} \boldsymbol{R}^{k-1}\boldsymbol{A}_k \end{bmatrix} = \begin{bmatrix} 1 - A^*(\theta+\eta) - c_0 & A^*(\theta+\eta) & c_0 \\ 1 - \psi(\theta+\eta) & 0 & \psi(\theta+\eta) \\ \frac{a_0}{\xi} & 0 & 1 - \frac{a_0}{\xi} \end{bmatrix},$$

where

$$\psi(\theta+\eta) = \frac{a_0}{\xi A^*(\theta+\eta)}\beta(\xi - A^*(\theta+\eta)) - \frac{c_0}{A^*(\theta+\eta)},$$

has a positive left invariant vector:

$$K(1, A^*(\theta+\eta), \beta(\xi - A^*(\theta+\eta))), \tag{8}$$

and K is a random positive real number. If $\rho < 1$, let (L, J) be the stationary limit of the process (L_n, J_n). Let

$$\pi_0 = \pi_{00}; \quad \boldsymbol{\pi}_k = (\pi_{k0}, \pi_{k1}), \quad k \geqslant 1,$$
$$\pi_{kj} = P\{L = k, J = j\} = \lim_{n\to\infty} P\{L_n = k, J_n = j\}, \quad (k,j) \in \Omega. \qquad \square$$

Theorem 1. *If* $\rho < 1$, *the stationary probability distribution of* (L, J) *is*

$$\begin{cases} \pi_{k0} = (1-\xi)\sigma(A^*(\theta+\eta))^k, & k \geqslant 0, \\ \pi_{k1} = (1-\xi)\sigma\beta(\xi - A^*(\theta+\eta))\sum_{j=0}^{k-1} \xi^j (A^*(\theta+\eta))^{k-1-j}, & k \geqslant 1, \end{cases} \tag{9}$$

where

$$\sigma = \frac{1 - A^*(\theta+\eta)}{1 - \xi + \beta(\xi - A^*(\theta+\eta))}.$$

Proof. With the Theorem 1.5.1 of Neuts (see in [15]), $(\pi_{00}, \pi_{10}, \pi_{11})$ is given by the positive left invariant vector (8) and satisfies the normalizing condition

$$\pi_{00} + (\pi_{10}, \pi_{11})(\boldsymbol{I} - \boldsymbol{R})^{-1}\mathbf{e} = 1,$$

where $\boldsymbol{I}$ expresses the identity matrix. And, substituting $\boldsymbol{R}$ into the above relationship, we easily get

$$K = (1-\xi)\frac{1 - A^*(\theta+\eta)}{1 - \xi + \beta(\xi - A^*(\theta+\eta))} = (1-\xi)\sigma.$$

Therefore, we obtain $(\pi_{10}, \pi_{11}) = (1-\xi)\sigma(A^*(\theta+\eta), \beta(\xi - A^*(\theta+\eta)))$. Using Theorem 1.5.1 of Neuts, we have

$$\boldsymbol{\pi}_k = (\pi_{k0}, \pi_{k1}) = (\pi_{10}, \pi_{11})\boldsymbol{R}^{k-1}, \quad k \geqslant 1. \tag{10}$$

Taking (π_{10}, π_{11}) and $\boldsymbol{R}^{k-1}$ into (10), we easily obtain the theorem. □

Then, we discuss the distribution of the queue length L at the arrival epochs. From Theorem 1, we easily obtain the distribution of L:

$$\begin{aligned}
\pi_0 &= P\{L=0\} = \pi_{00} = (1-\xi)\sigma, \\
\pi_k &= P\{L=k\} = \pi_{k0} + \pi_{k1} \qquad k \geqslant 1 \\
&= (1-\xi)\sigma\left((A^*(\theta+\eta))^k + \beta(\xi - A^*(\theta+\eta))\sum_{j=0}^{k-1}\xi^j(A^*(\theta+\eta))^{k-1-j}\right).
\end{aligned}$$

Theorem 2. *If $\rho < 1$ and $\mu > \eta$, the stationary queue length L can be decomposed into the sum of two independent random variables: $L = L_0 + L_d$, where L_0 is the stationary queue length of a classical GI/M/1 queue without vacation, and follows a geometric distribution with parameter ξ. Additional queue length L_d has a modified geometric distribution*

$$\begin{aligned}
&P\{L_d = 0\} = \sigma, \\
&P\{L_d = k\} = (1-\sigma)(1 - A^*(\theta+\eta))(A^*(\theta+\eta))^k, \quad k \geqslant 1,
\end{aligned} \tag{11}$$

where σ is defined as in Theorem 1.

Proof. From the probability expression for L above, the probability generating function of L is as follows:

$$\begin{aligned}
L(z) &= \sum_{k=0}^{\infty} z^k\pi_{k0} + \sum_{k=1}^{\infty} z^k\pi_{k1} = (1-\xi)\sigma\left[\frac{1}{1 - A^*(\theta+\eta)z} + \beta(\xi - A^*(\theta+\eta))\frac{1}{1-\xi z}\frac{z}{1 - A^*(\theta+\eta)z}\right] \\
&= \frac{1-\xi}{1-\xi z}\sigma\left[\frac{1}{1 - A^*(\theta+\eta)z}(1-\xi z) + \beta(\xi - A^*(\theta+\eta))\frac{z}{1 - A^*(\theta+\eta)z}\right] \\
&= \frac{1-\xi}{1-\xi z}\frac{1}{\delta}\left[1 - A^*(\theta+\eta) + (\beta-1)(\xi - A^*(\theta+\eta))\frac{(1 - A^*(\theta+\eta))z}{1 - A^*(\theta+\eta)z}\right] \\
&= \frac{1-\xi}{1-\xi z}\left[\sigma + (1-\sigma)\frac{(1 - A^*(\theta+\eta))z}{1 - A^*(\theta+\eta)z}\right] = L_0(z)L_d(z),
\end{aligned} \tag{12}$$

where $\delta = 1 - \xi + \beta(\xi - A^*(\theta+\eta))$. In the equation above, we easily verify

$$1 - \sigma = \frac{(\beta-1)(\xi - A^*(\theta+\eta))}{1 - \xi + \beta(\xi - A^*(\theta+\eta))} = \frac{\mu-\eta}{\theta + \eta A^*(\theta+\eta)}\frac{\beta(\xi - A^*(\theta+\eta))(1 - A^*(\theta+\eta))}{1 - \xi + \beta(\xi - A^*(\theta+\eta))}. \qquad □$$

Eq. (11) indicates that the additional delay L_d can be written as the mixture of two random variables: $L_d = \sigma X_0 + (1-\sigma)X_1$, where $X_0 \equiv 0$, X_1 follows a geometric distribution with parameter $A^*(\theta+\eta)$. Thus, we easily obtain a performance measure: mean queue length at the arrival epoch.

$$E(L) = \frac{\xi}{1-\xi} + \frac{\mu-\eta}{\theta+\eta A^*(\theta+\eta)}\frac{\beta(\xi - A^*(\theta+\eta))}{1-\xi+\beta(\xi - A^*(\theta+\eta))}.$$

Then we consider the queue length at arbitrary epoch. And, denote the limiting distribution of $L(t): p_k = \lim_{t\to\infty} P\{L(t)=k\}, \quad k \geqslant 0$.

Theorem 3. *If* $\rho < 1$ *and* $\theta > 0$, *the limiting distribution of* $L(t)$ *exists. And, we obtain*

$$\begin{cases} p_0 = 1 - \sigma\left\{\frac{\lambda}{\mu}\beta + \frac{\lambda}{\theta+\eta}(1-\beta)(1-\xi)\right\}, \\ p_k = (1-\xi)\sigma\left\{\frac{\lambda}{\mu}\beta\xi^{k-1} + \frac{\lambda}{\theta+\eta}(1-\beta)(1-A^*(\theta+\eta))[A^*(\theta+\eta)]^{k-1}\right\}, \quad k \geqslant 1. \end{cases}$$

(See Appendix A for the proof).

Let $\widetilde{L}$ denote the steady-state system size at an arbitrary epoch. Then, the expectation of $\widetilde{L}$ is given by

$$E(\widetilde{L}) = (1-\xi)\sigma\left\{\frac{\lambda}{\mu}\beta\frac{1}{(1-\xi)^2} + \frac{\lambda}{\theta+\eta}(1-\beta)\frac{1}{1-A^*(\theta+\eta)}\right\}.$$

Meanwhile, we easily obtain the state probabilities of a server in the steady-state.

$$\begin{aligned} P\{J=0\} &= \sum_{k=0}^{\infty}\pi_{k0} = \frac{1-\xi}{1-\xi+\beta(\xi-A^*(\theta+\eta))}, \\ P\{J=1\} &= \sum_{k=1}^{\infty}\pi_{k1} = \frac{\beta(\xi-A^*(\theta+\eta))}{1-\xi+\beta(\xi-A^*(\theta+\eta))}. \end{aligned} \tag{13}$$

Those performance measures such as queue length and the state probabilities of a server represent indices of the system and are important to analyze the operation in the last section.

4. Performance measure: waiting time

In this section, we assume that the service discipline is first-come first-served. Let W and $W^*(s)$ be the steady-state waiting time and its LST, respectively.

First, let H_0 be the probability that the server is in the service period when the new customer arrives, and H^1 be the probability that the server is in the vacation period and a new customer should wait when he arrives. We can easily compute

$$H_0 = \frac{\beta(\xi-A^*(\theta+\eta))}{1-\xi+\beta(\xi-A^*(\theta+\eta))}, \quad H_1 = \frac{(1-\xi)A^*(\theta+\eta)}{1-\xi+\beta(\xi-A^*(\theta+\eta))}.$$

Theorem 4. *If* $\rho < 1$ *and* $\theta > 0$, *the LST of stationary waiting time* W *is*

$$\begin{aligned} W^*(s) = {} & 1 - H_0 - H_1 + H_0\frac{\mu(1-\xi)}{s+\mu(1-\xi)}\frac{(\mu+s)(1-A^*(\theta+\eta))}{s+\mu(1-A^*(\theta+\eta))} + H_1\frac{\theta+\eta}{\theta+\eta+s} \\ & \times\frac{(\mu+s)(1-A^*(\theta+\eta))}{s+\mu(1-A^*(\theta+\eta))}\left[\frac{\eta}{\theta+\eta}+\frac{\theta}{\theta+\eta}\frac{\mu}{\mu+s}\right]. \end{aligned} \tag{14}$$

Proof. First, we obtain the probability that a new customer should not wait

$$P\{W=0\}=\pi_{00}=(1-\xi)\sigma=1-H_0-H_1.$$

When a new customer arrives, if there are k customers and the server is in the normal working period, the waiting time equals k service times by the rate μ. Then, we easily have

$$\sum_{k=1}^{\infty}\pi_{k1}W_{k1}^*(s)=\sum_{k=1}^{\infty}\pi_{k1}\left(\frac{\mu}{\mu+s}\right)^k=\frac{\beta(\xi-A^*(\theta+\eta))}{1-\xi+\beta(\xi-A^*(\theta+\eta))}\frac{\mu(1-\xi)}{s+\mu(1-\xi)}\frac{(\mu+s)(1-A^*(\theta+\eta))}{s+\mu(1-A^*(\theta+\eta))}$$
$$=H_0\frac{\mu(1-\xi)}{s+\mu(1-\xi)}\frac{(\mu+s)(1-A^*(\theta+\eta))}{s+\mu(1-A^*(\theta+\eta))}. \qquad (15)$$

Then, we consider the situation that there are k customers and the server is in the vacation period when the new customer arrives. For convenience, denote V and U^* be the vacation time and service time during the vacation period, respectively. There are two cases.

Case 1. if the residual vacation time is longer than one service time by the rate η, i.e., $V>U^*$, after a service completion in the vacation period, vacation interruption happens and the server comes back to the normal working level rather than keeping on the vacation. Thus, the waiting time is the sum of one service time by the rate η under the condition $V>U^*$ and $k-1$ service times by the rate μ. First, we compute the conditional probability of the service time by the rate η under the condition $V>U^*$. We have

$$P\{V>U^*\}=\int_0^{\infty}\eta\,\mathrm{e}^{-\eta x}\,\mathrm{e}^{-\theta x}\,\mathrm{d}x=\frac{\eta}{\theta+\eta},$$
$$P\{U^*<t|V>U^*\}=\frac{\theta+\eta}{\eta}\int_0^{t}\eta\,\mathrm{e}^{-\eta x}\,\mathrm{e}^{-\theta x}\,\mathrm{d}x=1-\mathrm{e}^{-(\eta+\theta)t}.$$

It can be easily explained that the service time by the rate η under the condition $V>U^*$ also follows the exponential distribution with parameter $\eta+\theta$, then we have

$$\sum_{k=1}^{\infty}\pi_{k0}P\{V>U^*\}W_{k0}^*(s)=\sum_{k=1}^{\infty}(1-\xi)\sigma(A^*(\theta+\eta))^k\frac{\eta}{\eta+\theta}\frac{\theta+\eta}{\theta+\eta+s}\left(\frac{\mu}{\mu+s}\right)^{k-1}$$
$$=\frac{\eta}{\theta+\eta}\frac{(1-\xi)A^*(\theta+\eta)}{1-\xi+\beta(\xi-A^*(\theta+\eta))}\frac{\theta+\eta}{\theta+\eta+s}\frac{(\mu+s)(1-A^*(\theta+\eta))}{s+\mu(1-A^*(\theta+\eta))}$$
$$=H_1\frac{\eta}{\theta+\eta}\frac{\theta+\eta}{\theta+\eta+s}\frac{(\mu+s)(1-A^*(\theta+\eta))}{s+\mu(1-A^*(\theta+\eta))}. \qquad (16)$$

Case 2. If the residual vacation time is not longer than the service time in the vacation period, i.e., $V\leqslant U^*$, when a vacation ends, the server has not completed a service during the vacation and comes back to the normal level. Therefore, the waiting time equals the sum of the residual vacation time under the condition $V\leqslant U^*$ and k service times by the rate μ. Similarly, we have the probabilities

$$P\{V\leqslant U^*\}=\int_0^{\infty}\theta\,\mathrm{e}^{-\eta x}\,\mathrm{e}^{-\theta x}\,\mathrm{d}x=\frac{\theta}{\theta+\eta},$$
$$P\{V<t|V\leqslant U^*\}=\frac{\theta+\eta}{\theta}\int_0^{t}\theta\,\mathrm{e}^{-\eta x}\,\mathrm{e}^{-\theta x}\,\mathrm{d}x=1-\mathrm{e}^{-(\eta+\theta)t}.$$

It can be verified that the residual vacation time under the condition $V\leqslant U^*$ also follows the exponential distribution with parameter $\eta+\theta$. Thus, we obtain

$$\sum_{k=1}^{\infty}\pi_{k0}P\{U^* \geqslant V\}\widetilde{W}^*_{k0}(s) = \sum_{k=1}^{\infty}(1-\xi)\sigma(A^*(\theta+\eta))^k\frac{\theta}{\eta+\theta}\frac{\theta+\eta}{\theta+\eta+s}\left(\frac{\mu}{\mu+s}\right)^k$$
$$= \frac{\theta}{\theta+\eta}\frac{(1-\xi)A^*(\theta+\eta)}{1-\xi+\beta(\xi-A^*(\theta+\eta))}\frac{\theta+\eta}{\theta+\eta+s}\frac{\mu(1-A^*(\theta+\eta))}{s+\mu(1-A^*(\theta+\eta))}$$
$$= H_1\frac{\theta}{\theta+\eta}\frac{\theta+\eta}{\theta+\eta+s}\frac{\mu(1-A^*(\theta+\eta))}{s+\mu(1-A^*(\theta+\eta))}. \quad (17)$$

And, we easily have

$$W^*(s) = \pi_{00} + \sum_{k=1}^{\infty}\pi_{k1}W^*_{k1}(s) + \sum_{k=1}^{\infty}\pi_{k0}\Big(P\{V > U^*\}W^*_{k0}(s) + P\{U^* \geqslant V\}\widetilde{W}^*_{k0}(s)\Big).$$

From (15)–(17), we have the result in Theorem 4 by some algebraic manipulation. With the structure in Theorem 4, we can easily get another performance measure: the expected waiting time of an arbitrary customer.

$$E(W) = H_0\left[\frac{1}{\mu(1-\xi)} + \frac{A^*(\theta+\eta)}{\mu(1-A^*(\theta+\eta))}\right] + H_1\left[\frac{1}{\theta+\eta} + \frac{A^*(\theta+\eta)}{\mu(1-A^*(\theta+\eta))} + \frac{\theta}{\theta+\eta}\frac{1}{\mu}\right]. \qquad \square$$

Remark 1. For Eq. (14), the steady-state waiting time of an arbitrary customer has the special probability explanation. The waiting time equals 0 with the probability $1 - H_0 - H_1$; with the probability H_0, it equals the sum of one exponential random variable with the rate $\mu(1-\xi)$ and one modified exponential random variable with the rate $\mu(1 - A^*(\theta+\eta))$; with the probability H_1, it equals the sum of three random variables: one exponential random variable with the rate $\theta+\eta$, and two modified exponential random variable with the rate $\mu(1 - A^*(\theta+\eta))$ and μ, respectively.

5. Numerical analysis

The model with two policies we consider can represent many problems in Management, Optical and Communication networks. For example, the banks establish the automatic teller machines (ATM) and the customers can fetch money from the employees or ATM in the day time. And, if the employees take vacations, the customers only can be served by ATM. This is the working vacation period or lower speed service period. Meanwhile, when some big operations arrive, the employees must also end vacation and come to work, then vacation interruption happens. Such model can also be used to analyze the priority queues and the changeable service rate systems, and model some internet systems, such as Optical nets, Electric nets and Communication nets according to the practical situations. According to the property of the data, one or some of data can be seen as the main task and the others as affiliated tasks. When there are few main data, the system can deal with the main data at the lower rate and affiliated data at the same time. But, when the number of the main data exceeds the certain value, the system must convert to the quick rate to deal with main data immediately. Such models can enable the data with higher priority to be processed firstly. Thus, to demonstrate the applicability of the results we obtain above, a variety of numerical results are presented below to analyze the operating and economic performance measures.

5.1. Operating performance measures

In Tables 1–3, we demonstrate the state probability in the system at pre-arrival and arbitrary epoches for three types of inter-arrival time distributions: (1) deterministic, (2) exponential, and (3) E_2. They all have equal mean ($=1/\lambda$) where $\lambda = 1.25$. Other parameters are taken as $\mu = 2.0$, $\theta = 1.0$, $\eta = 0.6$. In the titles of the tables, various performance measures are also given. Evidently, in D/M/1 and E_2/M/1 systems with working vacation

Table 1
Distribution of number of customers in the system at various epochs for D/M/1 queue with parameters: $\lambda = 1.25$, $\mu = 2.0$, $\theta = 1.0$, $\eta = 0.6$, and with $E(L)$=1.136291, $E(\widetilde{L}) = 1.618331$, $E(W)$=0.631126, $P\{J = 1\} = 0.482242$

k	π_{k0}	π_{k1}	π_k	p_k
0	0.373802	–	0.373802	0.170577
1	0.103931	0.223513	0.327444	0.381212
2	0.028897	0.142167	0.171063	0.245687
3	0.008034	0.068177	0.076211	0.118324
4	0.002234	0.029213	0.031447	0.050804
5	0.000621	0.011794	0.012415	0.020536
6	0.000173	0.004594	0.004767	0.008005
7	0.000048	0.001748	0.001796	0.003047
8	0.000013	0.000655	0.000668	0.001141
9	0.000004	0.000242	0.000246	0.000423
Sum	0.517757	0.482103	0.999860	0.999756

Table 2
Distribution of number of customers in the system at various epochs for M/M/1 queue with $E(L)$=2.296402, $E(\widetilde{L}) = 2.296402$, $E(W)$=1.231061, $P\{J = 1\} = 0.568181$

k	π_{k0}	π_{k1}	π_k	p_k
0	0.242424	–	0.242424	0.242424
1	0.106326	0.119617	0.225943	0.225943
2	0.046634	0.127224	0.173858	0.173858
3	0.020454	0.102526	0.122980	0.122980
4	0.008971	0.074171	0.083142	0.083142
5	0.003935	0.050783	0.054718	0.054718
6	0.001726	0.033681	0.035407	0.035407
7	0.000757	0.021902	0.022659	0.022659
8	0.000332	0.014062	0.014394	0.014394
9	0.000146	0.008953	0.009099	0.009099
Sum	0.431705	0.552919	0.984624	0.984624

Table 3
Distribution of number of customers in the system at various epochs for E_2/M/1 queue with $E(L)$=1.718788, $E(\widetilde{L}) = 1.953537$, $E(W)$=0.935032, $P\{J = 1\} = 0.535005$

k	π_{k0}	π_{k1}	π_k	p_k
0	0.292109	–	0.292109	0.215253
1	0.108607	0.159557	0.268162	0.282920
2	0.040380	0.143131	0.183511	0.204943
3	0.015013	0.097237	0.112250	0.128567
4	0.005582	0.059275	0.064857	0.075314
5	0.002075	0.034183	0.036258	0.042453
6	0.000772	0.019089	0.019861	0.023375
7	0.000287	0.010448	0.010735	0.012678
8	0.000107	0.005645	0.005752	0.006808
9	0.000040	0.003023	0.003063	0.003631
Sum	0.464972	0.531588	0.996558	0.996560

and vacation interruption, the pre-arrival state probability π_k does not equal the arbitrary state probability p_k, but in M/M/1 system, two indices are equal.

Fig. 1 provides the expected number of the main jobs with the change of vacation service rate η. In Fig. 2. we present the state probability of the server for the change of η and different vacation rate θ. In those figures, we present the performance measures for M/M/1 and E_2/M/1, respectively. Fig. 3 demonstrates the comparison of M/M/1 and E_2/M/1 with working vacations and vacation interruption. Finally, in Fig. 4 expected

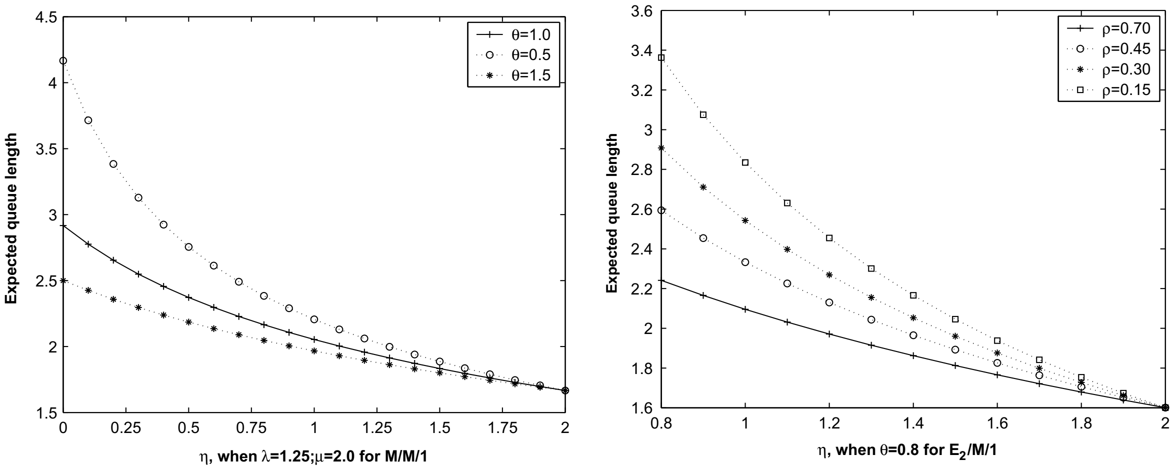


Fig. 1. The curve of expected queue length with the change of η.

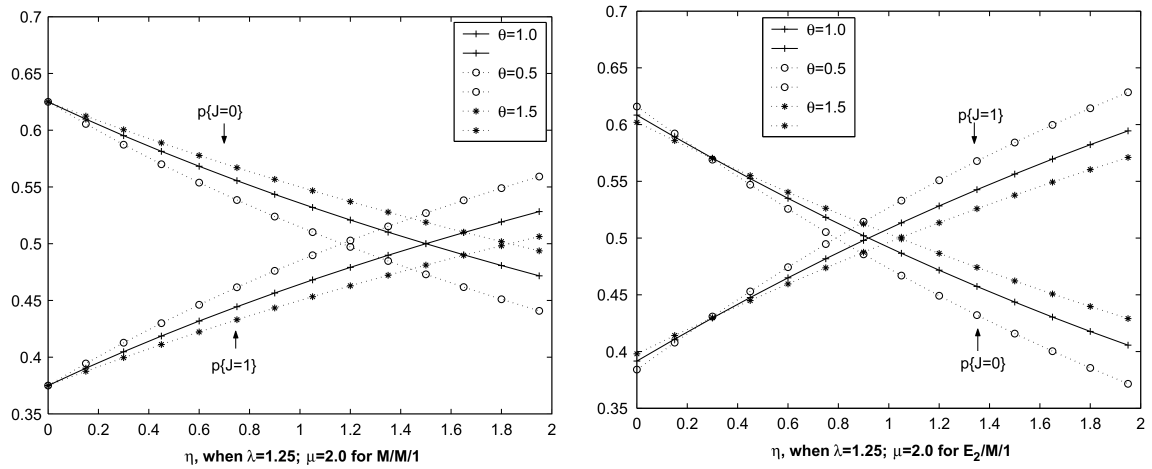


Fig. 2. The curve of state probability of the server with the change of η.

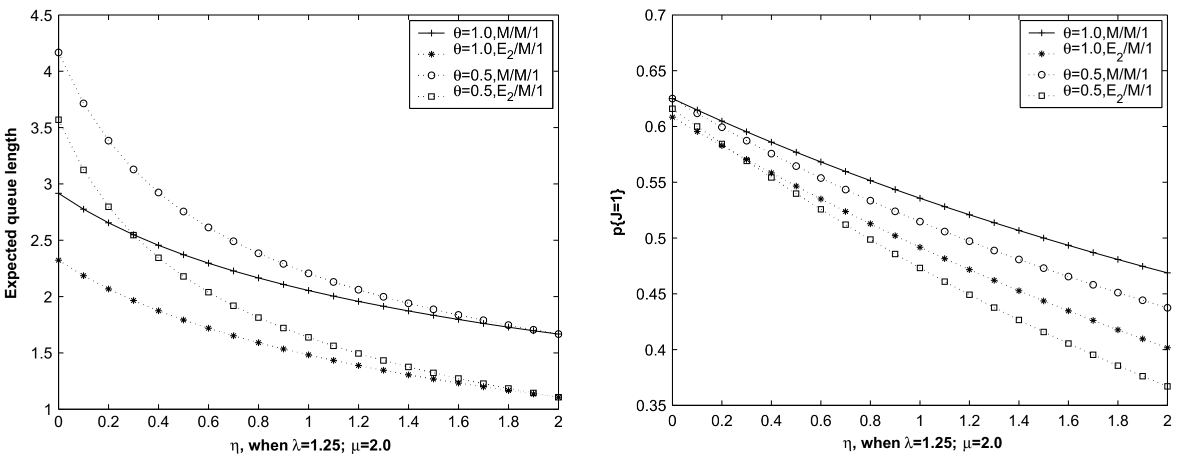


Fig. 3. The comparison of M/M/1 and E_2/M/1.

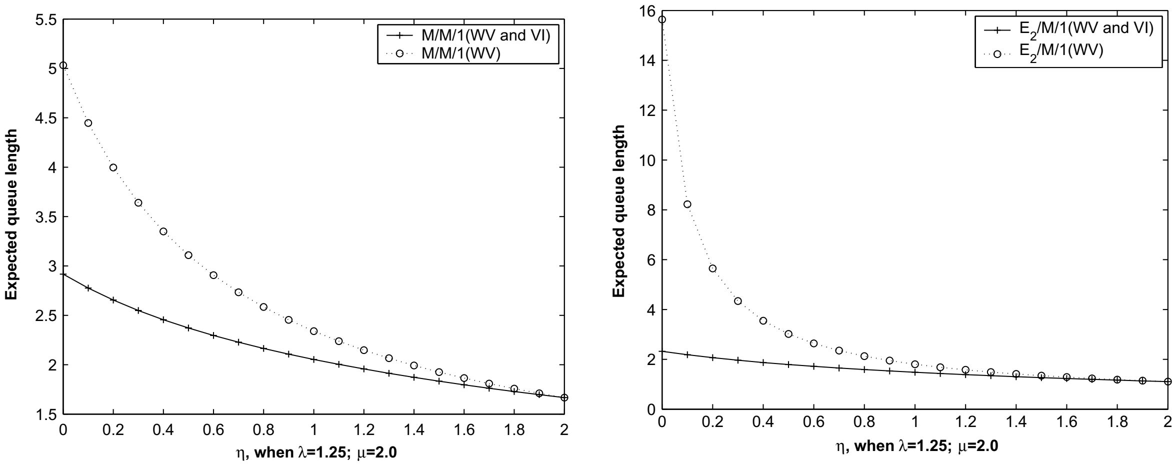


Fig. 4. The comparison of models with and without vacation interruption.

queue lengths are also compared in our model with working vacation (WV) and vacation interruption (VI) and that only with WV considered by Baba [8]. These computed performance measures illustrate the performance effects of assigning job ability during the vacation period.

Here are some findings:

(i) From Fig. 1, expected queue length can be measured with the change of η. With the vacation service rate increase, $E(L)$ decreases evidently and the utilization level of the system idle time becomes larger. Meanwhile, when service rate η approaches to μ, the vacation rate does not have the effect and the models become the corresponding queues without vacations. The effect of ρ on the expected queue length is also similar.

(ii) From Figs. 2 and 3, the state probability of the server can be also demonstrated and the probability that the server stays in working level, i.e., $P\{J=1\}$, evidently increases. The probability that the server stays in vacation decreases, thus, the utilization level of the system idle time also become larger. And, the vacation rate also has some effect on expected queue length and state probability of the server. For example, for E_2/M/1 queue, when $\theta = 0.5$, $E(L)$ and $P\{J=1\}$ are evidently larger than those when $\theta = 1.5$. It also shows that it is reasonable to establish the vacation period or lower speed operation period.

Thus, from (i) and (ii), working vacation policy is significantly better than general vacation policy without the vacation service in some situations.

(iii) From Fig. 4, in terms of $E(L)$, the vacation interruption policy is significantly better. For example, evidently, when the service rates η are the same in two situations, expected queue length in E_2/M/1 queue with WV is larger than that in E_2/M/1 queue with WV and VI and the former causes more jobs or customers to wait. Therefore, under the working vacation policy, if we want to develop a better service, we can consider vacation interruption policy that utilizes the server and decreases the waiting jobs effectively.

And, although we only show the results for some special systems, our model can also handle the other general and more complex arrivals. For example, for $E_{100}/M/1$ with large input parameters, using the results in sections above, we can also obtain various performance measures, but because of the complexity in computation, there are some skills to treat it in numerically, then it should not belong to our analysis scope in this paper and someone who are interested can do some work in this aspect.

5.2. Economic performance measures

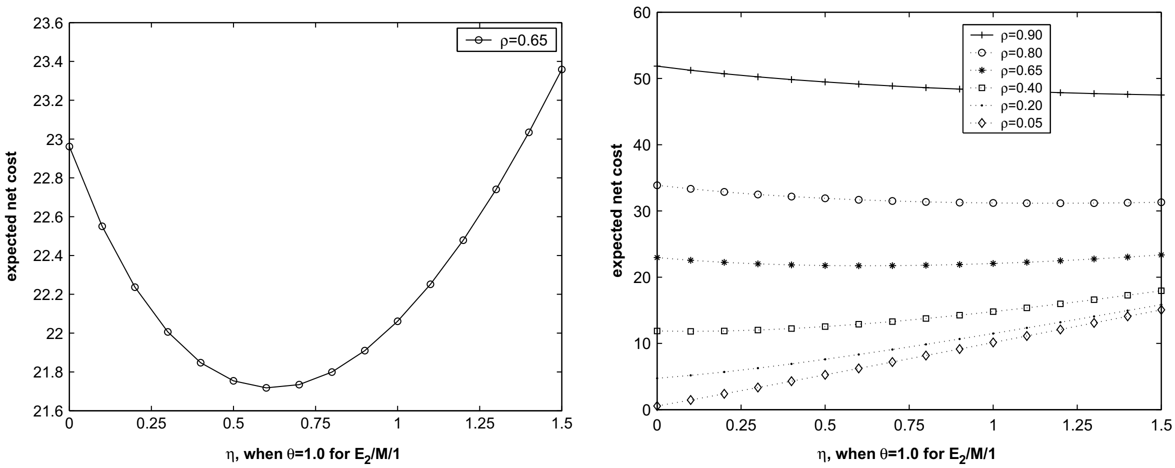
In practice, queueing managers always are interested in minimizing operating cost or maximizing business profits. Under a given cost/revenue structure, we can use the performance measures developed in this study to search for the cost minimization or profit maximization. In this model we consider, it is useful to optimize the service rate η during the vacation period. For $E_2/M/1$ queue with working vacation and vacation interruption, we establish the cost function to search for the optimal η policy. Assume c_w represent the unit time cost of every waiting customer, and c_1 and c_2 are the service costs every unit time during the normal working level and vacation period, respectively. Thus, we can establish the expected net cost function Z_c as:

$$\min : Z_c = c_w E(L) + c_1 \mu P\{J = 1\} + c_2 \eta P\{J = 0\},$$

where $E(L)$, $P\{J=1\}$ and $P\{J=0\}$ have been obtained in sections above. Fig. 5 shows this cost function for different traffic intensities and From Fig. 5a, evidently, there is the optimal vacation service rate η to make the cost minimize.

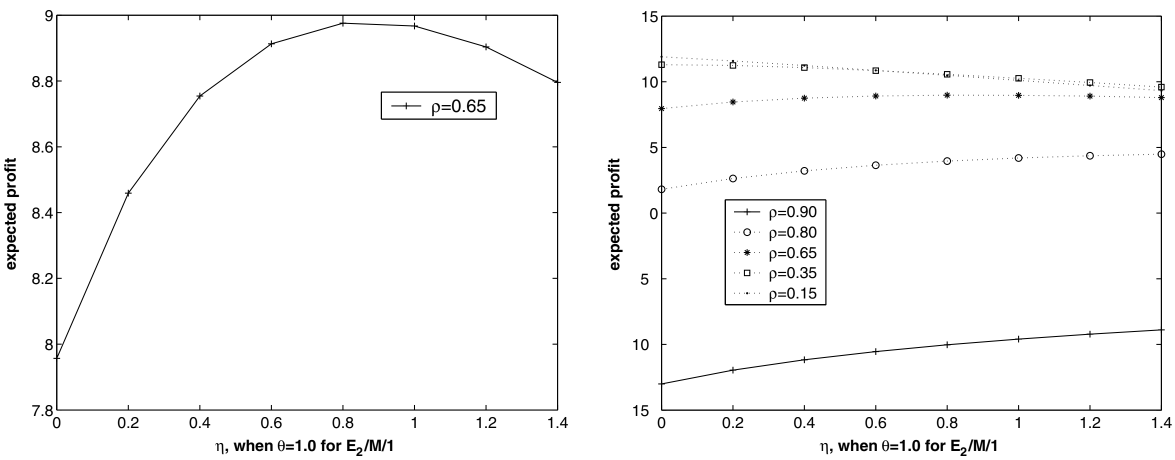
Meanwhile, we can also establish the cost and revenue structure which also consists of the saving rate of taking vacation, denoted by R_v, and the revenue rate of serving every unit time during the normal working level and vacation period, denoted by R_1 and R_2, respectively. The objective is to maximize the long-term average profit function as

$$\max : Z_p = R_v \theta + (R_1 - c_1)\mu P\{J = 1\} + (R_2 - c_2)\eta P\{J = 0\} - c_w E(L).$$



(a) Costs for $\rho = 0{\cdot}65$ (b) Costs for several cases

Fig. 5. Cost minimization chart for the case of $c_w = 4$, $c_1 = 15$, $c_2 = 10$. (a) Costs for $\rho = 0.65$. (b) Costs for several cases.



(a) Profits for $\rho = 0.65$ (b) Profits for several cases

Fig. 6. Profit chart for the case of $c_w = 4$, $c_1 = 15$, $R_1 = 20$, $R_2 = 8$, $c_2 = 10$, $R_v = 12$. Profits for (a) $\rho = 0.65$. (b) Profits for several cases.

In Fig. 6, we present this profit function for different traffic intensities and from Fig. 6a, evidently, there is the optimal vacation service rate η to make the profit maximize.

From Figs. 5b and 6b, some intuitive strategies have been found to the optimal η. In a heavily loaded system such as $\rho = 0.9$, the $(\eta = \mu)$ policy without vacation should be optimal to minimize the cost and maximize the profit of the system. On the other hand, in a very lightly loaded system such as $\rho = 0.05$ or 0.15, the $(\eta = 0)$ policy without working during the vacation period becomes the optimal. Meanwhile, in a low to moderate loaded system such as $\rho = 0.40$–0.80, the optimal level of service rate during the vacation period should be between the two extreme cases. That is, an $(\eta < \mu)$ policy should be optimal in terms of minimizing the cost or maximizing the profit as illustrated in Figs. 5 and 6.

It should be pointed out that other types of cost and profit structures can be studied similarly. The systems with the other general arrival can be also analyzed similarly. The corresponding optimal η policies can be obtained by a finite search to minimize cost functions (or maximize profit functions) based on these performance measures developed in this study. From the analysis above, managers or network researchers can establish the appropriate service rate during working vacation and apply the vacation interruption policy to enable the companies or networks operate flexibly.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions, which were helpful to improve the paper. Meanwhile, the authors also thank the support from National Natural Science Foundation of China #10671170, and the work is also supported by the Science Foundation of Yanshan University for the Excellent Ph.D. students.

Appendix A

Proof for Theorem 3. We consider another stochastic process $\{\widetilde{L}(t), \widetilde{J}(t)\}$. $\widetilde{L}(t)$ denotes the system size at the most recent arrival and $\widetilde{J}(t)$ equals 0 or 1 if the most recent arrival occurs during a working vacation or during a service period, respectively. Clearly, $\{\widetilde{L}(t), \widetilde{J}(t)\}$ is a SMP having $\{L_n, J_n\}$ for its embedded Markov chain. γ_{kj} denotes the sojourn time in state (k,j), $j = 0,1$; $k \geqslant j$, then $P\{\gamma_{kj} \leqslant t\} = A(t)$, $E(\gamma_{kj}) = \lambda^{-1}$.

From the theory of SMP, if the Markov chain (L_n, J_n) is irreducible, aperiodic and positive recurrent, and all $E(\gamma_{kj})$ are finite, then the limiting distribution of $\{\widetilde{L}(t), \widetilde{J}(t)\}$ exists. Let υ_{kj} be the steady-state probability that the SMP is in state (k,j). Thus, we have

$$\upsilon_{kj} = \frac{\pi_{kj} E(\gamma_{kj})}{\sum_{i=0}^{1} \sum_{h=i}^{\infty} \pi_{hi} E(\gamma_{hi})},$$

where $\{\pi_{kj}, j = 0, 1;\ k \geqslant j\}$ is given by (9). Thus, we have $\upsilon_{kj} = \pi_{kj}$. The limiting distribution of $\widetilde{L}(t)$ has following expressions (see in [16]):

$$p_k = \sum_{i=0}^{1} \sum_{j=i}^{\infty} \upsilon_{ji} \int_0^{\infty} P\{\text{required changes in } t \text{ to bring state from } (j,i) \text{ to } (k,i)\} \lambda(1 - A(t))\, dt.$$

For $k \geqslant 1$, we obtain

$$\begin{aligned} p_k = {} & \sum_{i=k-1}^{\infty} \pi_{i1} \int_0^{\infty} \frac{(\mu t)^{i+1-k}}{(i+1-k)!} e^{-\mu t} \lambda(1 - A(t))\, dt \\ & + \sum_{i=k}^{\infty} \pi_{i0} \int_0^{\infty} \int_0^{t} \eta e^{-\eta x} e^{-\theta x} \frac{(\mu(t-x))^{i-k}}{(i-k)!} e^{-\mu(t-x)}\, dx \lambda(1 - A(t))\, dt \\ & + \sum_{i=k-1}^{\infty} \pi_{i0} \int_0^{\infty} \int_0^{t} \theta e^{-\theta x} e^{-\eta x} \frac{(\mu(t-x))^{i-k+1}}{(i-k+1)!} e^{-\mu(t-x)}\, dx \lambda(1 - A(t))\, dt \\ & + \pi_{k-1,0} \int_0^{\infty} e^{-\eta t} e^{-\theta t} \lambda(1 - A(t))\, dt. \end{aligned} \tag{18}$$

Substituting the expressions for π_{kj} in Theorem 1into the above equation, we compute each part of the equation and easily have

$$\sum_{i=k-1}^{\infty}\pi_{i1}\int_0^{\infty}\frac{(\mu t)^{i+1-k}}{(i+1-k)!}\,\mathrm{e}^{-\mu t}\lambda(1-A(t))\,\mathrm{d}t = (1-\xi)\sigma\beta\left\{\frac{\lambda}{\mu}\beta\xi^{k-1}-\frac{\lambda}{\mu}[A^*(\theta+\eta)]^{k-1}\frac{1-A^*(\mu(1-A^*(\theta+\eta)))}{1-A^*(\theta+\eta)}\right\}. \quad (19)$$

Second, we also have

$$\sum_{i=k}^{\infty}\pi_{i0}\int_0^{\infty}\int_0^t\eta\,\mathrm{e}^{-\eta x}\,\mathrm{e}^{-\theta x}\frac{(\mu(t-x))^{i-k}}{(i-k)!}\,\mathrm{e}^{-\mu(t-x)}\,\mathrm{d}x\lambda(1-A(t))\,\mathrm{d}t = (1-\xi)\sigma[A^*(\theta+\eta)]^k\frac{\eta}{\theta+\eta-\mu(1-A^*(\theta+\eta))} \times\left\{\frac{\lambda[1-A^*(\mu(1-A^*(\theta+\eta)))]}{\mu(1-A^*(\theta+\eta))}-\frac{\lambda}{\theta+\eta}[1-A^*(\theta+\eta)]\right\}. \quad (20)$$

Similarly,

$$\sum_{i=k-1}^{\infty}\pi_{i0}\int_0^{\infty}\int_0^t\theta\,\mathrm{e}^{-\theta x}\,\mathrm{e}^{-\eta x}\frac{(\mu(t-x))^{i-k+1}}{(i-k+1)!}\,\mathrm{e}^{-\mu(t-x)}\,\mathrm{d}x\lambda(1-A(t))\,\mathrm{d}t = (1-\xi)\sigma[A^*(\theta+\eta)]^{k-1}\frac{\theta}{\theta+\eta-\mu(1-A^*(\theta+\eta))} \times\left\{\frac{\lambda[1-A^*(\mu(1-A^*(\theta+\eta)))]}{\mu(1-A^*(\theta+\eta))}-\frac{\lambda}{\theta+\eta}[1-A^*(\theta+\eta)]\right\}, \quad (21)$$

and,

$$\pi_{k-1,0}\int_0^{\infty}\mathrm{e}^{-(\eta+\theta)t}\lambda(1-A(t))\,\mathrm{d}t = (1-\xi)\sigma[A^*(\theta+\eta)]^{k-1}\frac{\lambda}{\theta+\eta}[1-A^*(\theta+\eta)]. \quad (22)$$

Substituting (19)–(22) in (18), we obtain the expression for p_k. For $k=0$, using the normal condition, we obtain p_0. □

References

[1] B.T. Doshi, Single server queues with vacations, in: H. Takagi (Ed.), Stochastic Analysis of the Computer and Communication Systems, North-Holland/Elsevier, Amsterdam, 1990, pp. 217–264.
[2] B.T. Doshi, Queueing systems with vacations – a survey, Queueing Syst. 1 (1986) 29–66.
[3] H. Takagi, Queueing analysis: a foundation of performance evaluation, Vacation and Priority Systems Part 1, vol. 1, North-Holland/Elsevier, Amsterdam, 1991.
[4] N. Tian, Z.G. Zhang, Vacation Queueing Models: Theory and Applications, Springer, New York, 2006.
[5] L.D. Servi, S.G. Finn, M/M/1 queue with working vacations (M/M/1/WV), Perform. Evaluation 50 (2002) 41–52.
[6] J.D. Kim, D.W. Choi, K.C. Chae, Analysis of queue-length distribution of the M/G/1 queue with working vacations, in: Hawaii International Conference on Statistics and Related Fields, June 5–8, 2003.
[7] D. Wu, H. Takagi, M/G/1 queue with multiple working vacations, Perform. Evaluation 63 (7) (2006) 654–681.
[8] Y. Baba, Analysis of a GI/M/1 queue with multiple working vacations, Oper. Res. Lett. 33 (2005) 201–209.
[9] A.D. Banik, U.C. Gupta, S.S. Pathak, On the GI/M/1/N queue with multiple working vacations-analytic analysis and computation, Appl. Math. Model. 31 (2007) 1701–1710.
[10] J. Li, N. Tian, The discrete-time GI/Geo/1 queue with working vacations and vacation interruption, Appl. Math. Comput. 185 (1) (2007) 1–10.
[11] N. Tian, D. Zhang, C. Cao, The GI/M/1 queue with exponential vacations, Queueing Syst. 5 (1989) 331–344.
[12] U. Chatterjee, S. Mukherjee, GI/M/1 queue with server vacations, J. Oper. Res. Soc. 41 (1990) 83–87.

[13] N. Tian, Z.G. Zhang, A note on GI/M/1 queues with phase-type setup times or server vacations, INFOR 41 (2003) 341–351.
[14] N. Tian, Z.G. Zhang, The discrete time GI/Geo/1 queue with multiple vacations, Queueing Syst. 40 (2002) 283–294.
[15] M. Neuts, Matrix–Geometric Solutions in Stochastic Models, Johns Hopkins University Press, Baltimore, MD, 1981.
[16] D. Gross, C.M. Harris, Fundamentals of Queueing Theory, third ed., Wiley, New York, 1998.