Queueing Syst (2009) 61: 139–166
DOI 10.1007/s11134-008-9103-8

Analysis of the M/G/1 queue with exponentially working vacations—a matrix analytic approach

Ji-hong Li · Nai-shuo Tian · Zhe George Zhang · Hsing Paul Luh

Received: 13 March 2008 / Revised: 15 December 2008 / Published online: 13 January 2009

Abstract In this paper, an M/G/1 queue with exponentially working vacations is analyzed. This queueing system is modeled as a two-dimensional embedded Markov chain which has an M/G/1-type transition probability matrix. Using the matrix analytic method, we obtain the distribution for the stationary queue length at departure epochs. Then, based on the classical vacation decomposition in the M/G/1 queue, we derive a conditional stochastic decomposition result. The joint distribution for the stationary queue length and service status at the arbitrary epoch is also obtained by analyzing the semi-Markov process. Furthermore, we provide the stationary waiting time and busy period analysis. Finally, several special cases and numerical examples are presented.

Keywords Working vacations · Embedded Markov chain · M/G/1-type matrix · Stochastic decomposition · Conditional waiting time

J.-h. Li
College of Economics and Management, Yanshan University, Qinhuangdao, 066004, China
e-mail: jhli_two@sina.com

N.-s. Tian (✉)
College of Science, Yanshan University, Qinhuangdao, 066004, China
e-mail: tiannsh@ysu.edu.cn

Z.G. Zhang
Dept. of Decision Sciences, Western Washington University, Bellingham, WA 98225, USA
e-mail: george.zhang@wwu.edu

Z.G. Zhang
Faculty of Business Administration, Simon Fraser University, Burnaby, BC, Canada

H.P. Luh
Department of Mathematical Sciences, National ChengChi University, Taipei, Taiwan
e-mail: slu@nccu.edu.tw

Mathematics Subject Classification (2000) 60K25 · 68M20

1 Introduction

We consider a queueing system where the server may not be fully available for a period of time, called a vacation. During the vacation period, the server may also perform other supplementary tasks. Over the last two decades, the queueing systems with vacations have been well studied because of their applications in modeling the computer networks, communication, and manufacturing/service systems (see Fuhrmann and Cooper [6]). Readers are referred to the surveys of Doshi [4, 5], and the monographs of Takagi [17] and Tian and Zhang [16]. In these previous studies, usually it is assumed that the server stops service completely during the vacation. In 2002, Servi and Finn [14] analyzed an M/M/1 queue with working vacations, denoted by M/M/1/WV, where the server works at a lower service rate rather than completely stopping service during the vacation period. The motivation of studying M/M/1/WV queue is to model approximately a multi-queue system where each queue can be served at one of two service rates and the fast service rate mode cyclically moves from queue to queue with exhaustive service. The working vacation model can be applied to analyze a wavelength division multiplexing (WDM) optical access network. Subsequently, Kim, Choi and Chae [8], Wu and Takagi [18] generalized the model in [14] to an M/G/1 queue with general working vacations. Baba [1] provided a study on a GI/M/1 queue with working vacations by using the matrix analytic method. Banik et al. [2] analyzed the GI/M/1/N queue with working vacations. Liu et al. [9] established the stochastic decomposition in the M/M/1 queue with working vacations. Li and Tian [10, 11] considered two types of discrete-time GI/Geo/1 queues with working vacations. In fact, the matrix geometric solution method developed by Neuts [12] is very powerful in analyzing the GI/M/1-type (including M/M/1-type) working vacation queues. However, there was no attempt in analyzing the M/G/1-type working vacation queues via the matrix analytic method. The study on the M/G/1 with working vacations in Wu and Takagi [18] is based on the Laplace–Stieltjes transform (LST) for an embedded Markov chain. Kim, Choi and Chae [8] analyzed the M/G/1 with exponentially working vacations using the results in systems with disasters.

In this paper, we treat the M/G/1 queue with exponentially working vacations, which is a special case of that in Wu and Takagi [18] and is the same as that in Kim, Choi and Chae [8]. But we utilize the matrix analytic approach which is very different from the methods used in previous studies and derive more results and properties about the system performance.

The rest of this paper is organized as follows. In Sect. 2, the M/G/1/WV queue is formulated as the two-dimensional embedded Markov chain at the departure epochs. The M/G/1-type transition probability matrix for the model is developed. In Sect. 3, using the matrix analytic approach, we derive the probability generating function (PGF) of the stationary queue length at the departure epochs. In Sect. 4, we use the stochastic decomposition property for the standard M/G/1 queue with general (non-working) vacations in Shanthikumar [15] to develop another expression for the PGF. Section 5 gives the joint distribution of the stationary queue length and service rate status at an arbitrary epoch based on the semi-Markov process theory. The waiting time and busy period analysis are provided in Sects. 6 and 7. Finally, some special cases and numerical examples are presented in Sects. 8 and 9.

2 Model formulation and embedded Markov chain

Consider an M/G/1 queue with a Poisson arrival process of rate λ. Whenever the system becomes empty at a service completion instant, the server starts a working vacation during which the service is at a low rate and the service times are also i.i.d. The notations used in our model are as follows:

(1) The normal service time during the busy period S_b follows a general distribution with the mean of $1/\mu_b$ and

$$G_b(x) = \mathrm{P}\{S_b < x\}, \qquad \widetilde{G}_b(s) = \int_0^\infty e^{-sx}\, dG_b(x),$$

$$b^{(k)} = \int_0^\infty x^k\, dG_b(x), \quad k \geq 2.$$

(2) The service time during the vacation period S_v also follows a general distribution with the mean of $1/\mu_v$ and

$$G_v(x) = \mathrm{P}\{S_v < x\}, \qquad \widetilde{G}_v(s) = \int_0^\infty e^{-sx}\, dG_v(x),$$

$$g_v^{(k)} = \int_0^\infty x^k\, dG_v(x).$$

(3) The vacation time is exponentially distributed with rate θ. At a vacation completion instant, if there are customers in the system, the server will start a new busy period. Otherwise, he/she takes another working vacation. Inter-arrival times, service times, and working vacation times are mutually independent. The service discipline is First Come First Served (FCFS).

Let $L(t)$ be the number of customers in the system at time t and L_n the number of the customers at the nth service completion instant. Note that a service completion (or customer departure) may occur during a normal service period (busy period) or a working vacation period. Define

$$J_n = \begin{cases} 1, & \text{after the } n\text{th departure, the system stays in a busy period,} \\ 0, & \text{after the } n\text{th departure, the system stays in a working vacation period.} \end{cases}$$

Due to the exponential vacations, the process $\{(L_n, J_n), n \geq 1\}$ is a two-dimensional embedded Markov chain with the state space

$$\Omega = \{(0,0)\} \cup \{(k, j), k \geq 1, j = 0, 1\}.$$

Obviously, the server only stays in the vacation period when there are no customers in our model.

To develop the transition matrix of (L_n, J_n), we introduce a few quantities:
(i) Define

$$a_k = \int_0^\infty e^{-\lambda t}\frac{(\lambda t)^k}{k!}\,dG_b(t), \quad k \ge 0.$$

Then, a_k ($k \ge 0$) is the probability that there are k arrivals during S_b (regular service time). Multiplying a_k by z^k, and summing over $k = 0, 1, \ldots, \infty$, we have

$$A(z) = \sum_{k=0}^\infty a_k z^k = \int_0^\infty e^{-\lambda(1-z)x}\,dG_b(x) = \widetilde{G}_b\big(\lambda(1-z)\big), \qquad A'(1) = \frac{\lambda}{\mu_b} = \rho.$$

(ii) Define

$$b_k = \int_0^\infty e^{-\theta x}\frac{(\lambda x)^k}{k!}e^{-\lambda x}\,dG_v(x), \quad k \ge 0,$$

$$v_k = \int_0^\infty \int_0^x \theta e^{-\theta u}\frac{(\lambda u)^k}{k!}e^{-\lambda u}\,du\,dG_v(x), \quad k \ge 0.$$

Evidently,

$$\sum_{k=0}^\infty b_k = \int_0^\infty e^{-\theta x}\,dG_v(x) = \mathrm{P}\{S_v < V\} = \widetilde{G}_v(\theta),$$

$$\sum_{k=0}^\infty v_k = \int_0^\infty \int_0^x \theta e^{-\theta u}\,du\,dG_v(x) = \mathrm{P}\{V \le S_v\} = 1 - \widetilde{G}_v(\theta).$$

Thus, $\{b_k, k \ge 0\}$ and $\{v_k, k \ge 0\}$ are two non-complete probability distributions. In fact, denote by A_x the number of arrivals during a random length x. Then, we have

$$\mathrm{P}\{A_{S_v} = k, S_v < V\} = b_k, \qquad \mathrm{P}\{A_V = k, V \le S_v\} = v_k, \quad k \ge 0.$$

Hence, b_k is the probability that $V > S_v$ and k customers arrive during S_v, and v_k is the probability that $V \le S_v$ and k customers arrive during V. The z-transforms of $\{b_k, k \ge 0\}$ and $\{v_k, k \ge 0\}$ can be obtained as follows:

$$B(z) = \sum_{k=0}^\infty b_k z^k = \int_0^\infty e^{-[\theta+\lambda(1-z)]x}\,dG_v(x) = \widetilde{G}_v\big(\theta + \lambda(1-z)\big);$$

$$\begin{aligned} V(z) &= \sum_{k=0}^\infty v_k z^k = \int_0^\infty \int_0^x \theta e^{-[\theta+\lambda(1-z)]u}\,du\,dG_v(x) \\ &= \frac{\theta}{\theta+\lambda(1-z)}\int_0^\infty \big[1 - e^{-[\theta+\lambda(1-z)]x}\big]\,dG_v(x) \\ &= \frac{\theta}{\theta+\lambda(1-z)}\big[1 - \widetilde{G}_v\big(\theta+\lambda(1-z)\big)\big] = \frac{\theta}{\theta+\lambda(1-z)}\big[1 - B(z)\big]. \end{aligned}$$

Define $\beta = B'(1)$. Obviously,

$$B'(1) = \lambda \int_0^\infty x e^{-\theta x}\, dG_v(x) = \beta, \qquad V'(1) = \frac{\lambda}{\theta}\big(1 - \widetilde{G}_v(\theta)\big) - \beta.$$

(iii) Define

$$c_k = \sum_{j=0}^{k} v_j a_{k-j}, \quad k \geq 0,$$

which represents the probability that the vacation time V is not longer than S_v and k customers arrive during V plus S_b. Therefore,

$$\sum_{k=0}^{\infty} c_k = 1 - \widetilde{G}_v(\theta), \qquad C(z) = \sum_{k=0}^{\infty} c_k z^k = V(z)A(z),$$

and

$$C'(1) = \left(\rho + \frac{\lambda}{\theta}\right)\big(1 - \widetilde{G}_v(\theta)\big) - \beta.$$

Using a_k, b_k, v_k, and c_k, we can present the transition probability matrix of $X_n = (L_n, J_n)$ by considering three transition cases.

- Case 1: if $X_n = (m, 1), m \geq 1$:

$$X_{n+1} = \begin{cases} (m-1+j, 1) & \text{with probability } a_j,\ m \geq 2,\ j \geq 0; \\ (j, 1) & \text{with probability } a_j,\ m = 1,\ j \geq 1; \\ (0, 0) & \text{with probability } a_0,\ m = 1; \end{cases}$$

- Case 2: if $X_n = (m, 0), m \geq 2$:

$$X_{n+1} = \begin{cases} (m-1+j, 0) & \text{with probability } b_j,\ j \geq 0; \\ (m-1+j, 1) & \text{with probability } c_j,\ j \geq 0; \end{cases}$$

- Case 3: if $X_n = (m, 0), m = 1, 0$:

$$X_{n+1} = \begin{cases} (j, 0) & \text{with probability } b_j,\ j \geq 1; \\ (j, 1) & \text{with probability } c_j,\ j \geq 1; \\ (0, 0) & \text{with probability } b_0 + c_0. \end{cases}$$

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

$$\widetilde{\boldsymbol{P}} = \begin{bmatrix} \boldsymbol{B}_0 & \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{B}_3 & \cdots \\ \boldsymbol{C}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \boldsymbol{A}_3 & \cdots \\ & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots \\ & & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \cdots \\ & & & \vdots & \vdots \end{bmatrix}, \tag{1}$$

where

$$\boldsymbol{B}_0 = b_0 + c_0; \qquad \boldsymbol{B}_i = (b_i, c_i), \quad i \geq 1; \qquad \boldsymbol{C}_0 = (b_0 + c_0, a_0)^T;$$

$$\boldsymbol{A}_i = \begin{bmatrix} b_i & c_i \\ 0 & a_i \end{bmatrix}, \quad i \geq 0.$$

Evidently,

$$\boldsymbol{B}_0 + \sum_{i=1}^{\infty} \boldsymbol{B}_i \boldsymbol{e} = 1, \qquad \boldsymbol{C}_0 + \sum_{i=1}^{\infty} \boldsymbol{A}_i \boldsymbol{e} = \boldsymbol{e}, \qquad \sum_{i=0}^{\infty} \boldsymbol{A}_i \boldsymbol{e} = \boldsymbol{e},$$

where $\boldsymbol{e} = (1, 1)^T$ and T represents *matrix transpose operation*. Then, the stochastic matrix $\widetilde{\boldsymbol{P}}$ is an M/G/1-type matrix (see Neuts [13]). For such a model, the minimal nonnegative solution of the equation $\boldsymbol{G} = \sum_{i=0}^{\infty} \boldsymbol{A}_i \boldsymbol{G}^i$ is required and is obtained first.

Lemma 1 *If* $\rho = \lambda/\mu_b < 1$, *the equation* $z = \widetilde{G}_b(\lambda(1-z))$ *has the minimal nonnegative root* $z = 1$ *and the equation* $z = \widetilde{G}_v(\theta + \lambda(1-z))$ *also has a unique root in the range* $0 < z < 1$.

Proof First, we consider the equation $z = \widetilde{G}_b(\lambda(1-z))$. Let $\psi(z) = \widetilde{G}_b(\lambda(1-z))$ and evidently, $0 < \psi(0) = \widetilde{G}_b(\lambda) < \psi(1) = 1$. And, for $0 < z < 1$,

$$\psi'(z) = \lambda \int_0^{\infty} t e^{-\lambda(1-z)t}\, d\widetilde{G}_b(t) > 0; \qquad \psi''(z) = \lambda^2 \int_0^{\infty} t^2 e^{-\lambda(1-z)t}\, d\widetilde{G}_b(t) > 0.$$

Meanwhile, it follows from $\rho = \lambda/\mu_b < 1$ that $\psi'(1) = \rho < 1$. Thus, the equation $z = \psi(z)$ has the unique root $z = 1$. Similarly, we set $\varphi(z) = \widetilde{G}_v(\theta + \lambda(1-z))$. Then we have

$$0 < \varphi(0) = \widetilde{G}_v(\theta + \lambda) < \varphi(1) = \widetilde{G}_v(\theta) < 1.$$

And, for $0 < z < 1$,

$$\varphi'(z) = \lambda \int_0^{\infty} t e^{-(\theta+\lambda(1-z))t}\, d\widetilde{G}_v(t) > 0;$$

$$\varphi''(z) = \lambda^2 \int_0^{\infty} t^2 e^{-(\theta+\lambda(1-z))t}\, d\widetilde{G}_v(t) > 0.$$

Therefore, $z = \varphi(z)$ has a unique root in the range $0 < z < 1$. □

Lemma 2 *If* $\rho = \lambda/\mu_b < 1$ *and* $\theta > 0$, *the matrix equation* $\boldsymbol{G} = \sum_{i=0}^{\infty} \boldsymbol{A}_i \boldsymbol{G}^i$ *has the minimal nonnegative solution*

$$\boldsymbol{G} = \begin{bmatrix} \gamma & 1-\gamma \\ 0 & 1 \end{bmatrix},$$

where γ *is the unique root in the range* $0 < z < 1$ *for* $z = \widetilde{G}_v(\theta + \lambda(1-z))$.

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

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

Then, for $i \geq 1$, we have

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

Substituting $\boldsymbol{G}^i$ into the matrix equation, we obtain

$$\begin{cases} r_{11} = \sum_{i=0}^{\infty} b_i r_{11}^i = \widetilde{G}_v(\theta + \lambda(1 - r_{11})), \\ r_{12} = \sum_{i=0}^{\infty} r_{22}^i c_i + r_{12} \sum_{i=1}^{\infty} b_i \sum_{j=0}^{i-1} r_{11}^j r_{22}^{i-1-j}, \\ r_{22} = \sum_{i=0}^{\infty} a_i r_{22}^i = \widetilde{G}_b(\lambda(1 - r_{22})). \end{cases} \tag{2}$$

From Lemma 1, the first equation has the unique root $r_{11} = \gamma$ in the range $0 < r_{11} < 1$ and the third equation has $r_{22} = 1$. Taking r_{22} and r_{11} into the second equation, we have $r_{12} = 1 - \gamma$. □

The steady state analysis is justified by the following condition.

Theorem 1 *The Markov chain $\widetilde{\boldsymbol{P}}$ is positive recurrent if and only if $\sum_{i=0}^{\infty} i a_i = \rho = \lambda \mu_b^{-1} < 1$.*

Proof Because

$$\boldsymbol{A} = \sum_{i=0}^{\infty} \boldsymbol{A}_i = \begin{bmatrix} \widetilde{G}_v(\theta) & 1 - \widetilde{G}_v(\theta) \\ 0 & 1 \end{bmatrix}$$

is a reducible stochastic matrix, with the notation of (2.3.18) in Neuts [13], $\boldsymbol{A}(2) = 1$ is the degenerative stochastic matrix, and has the degenerative stationary distribution $\boldsymbol{\pi}(2) = 1$. On the other hand, $\boldsymbol{A}_i(2) = a_i$, $i \geq 0$, and $\boldsymbol{\beta}(2) = \sum_{i=0}^{\infty} i \boldsymbol{A}_i(2) = \rho$. Thus, with Theorem 2.3.3 in Neuts [13], the Markov chain $\widetilde{\boldsymbol{P}}$ is positive recurrent if and only if

$$\boldsymbol{\pi}(2)\boldsymbol{\beta}(2) = \rho < 1.$$ □

3 Stationary queue length at the departure epoch

Let (L, J) be the stationary limit of the process (L_n, J_n). Let

$$\boldsymbol{\pi}_0 = \pi_{00}; \qquad \boldsymbol{\pi}_k = (\pi_{k0}, \pi_{k1}), \quad k \geq 1,$$

$$\pi_{kj} = \mathrm{P}\{L = k, J = j\} = \lim_{n \to \infty} \mathrm{P}\{L_n = k, J_n = j\}, \quad (k, j) \in \Omega,$$

where π_{kj} represents the stationary probability that there are k customers in the system with server status in j at a service completion instant.

There are two possible transitions for the system to reach state $(0,0)$: (i) the empty system resulted from the last service completion in a busy period; and (ii) the empty system left by a customer departure during the working vacation period. Certainly, for a classic vacation system, the service is stopped during the vacation period, thus, case (ii) does not exist. Using $\boldsymbol{\pi}\widetilde{\boldsymbol{P}} = \boldsymbol{\pi}$, we have the balance equations:

$$\begin{aligned}
\pi_{00} &= \pi_{00}(b_0 + c_0) + \boldsymbol{\pi}_1 \boldsymbol{C} = (\pi_{00} + \pi_{10})b_0 + (\pi_{00} + \pi_{10})c_0 + \pi_{11}a_0; \\
\boldsymbol{\pi}_k &= \pi_{00}\boldsymbol{B}_k + \sum_{j=1}^{k+1} \boldsymbol{\pi}_j \boldsymbol{A}_{k+1-j}, \quad k \ge 1.
\end{aligned} \tag{3}$$

Theorem 2 *The PGF of the stationary queue length L at the departure epochs is given by*

$$\begin{aligned}
L(z) &= \frac{\theta(1 - \widetilde{G}_v(\theta))(1-\rho)}{(1 - \widetilde{G}_v(\theta))(\theta + \lambda(1-\gamma)) - \rho\theta(1-\gamma)\widetilde{G}_v(\theta)} \\
&\quad \times \frac{A(z)(1-z)(B(z)-z) + z(\gamma - z)(A(z) - B(z) - C(z))}{(z - B(z))(z - A(z))},
\end{aligned} \tag{4}$$

where $A(z)$, $B(z)$, and $C(z)$ are given in Sect. 2.

Proof Introduce the row-vector generating function

$$\boldsymbol{\Phi}(z) = \sum_{k=1}^{\infty} z^k \boldsymbol{\pi}_k, \quad |z| < 1.$$

From the second equation in (3), we obtain

$$\begin{aligned}
\boldsymbol{\Phi}(z) &= \pi_{00} \sum_{k=1}^{\infty} z^k \boldsymbol{B}_k + \sum_{k=1}^{\infty} z^k \sum_{j=1}^{k+1} \boldsymbol{\pi}_j \boldsymbol{A}_{k+1-j} \\
&= \pi_{00} \sum_{k=1}^{\infty} z^k \boldsymbol{B}_k + \frac{1}{z} \sum_{j=1}^{\infty} \boldsymbol{\pi}_j z^j \sum_{k=j-1}^{\infty} z^{k+1-j} \boldsymbol{A}_{k+1-j} - \boldsymbol{\pi}_1 \boldsymbol{A}_0 \\
&= \pi_{00} \sum_{k=1}^{\infty} z^k \boldsymbol{B}_k + \frac{1}{z} \boldsymbol{\Phi}(z) \boldsymbol{A}^*(z) - \boldsymbol{\pi}_1 \boldsymbol{A}_0,
\end{aligned}$$

where

$$\boldsymbol{A}^*(z) = \sum_{k=0}^{\infty} z^k \boldsymbol{A}_k = \begin{bmatrix} B(z) & C(z) \\ 0 & A(z) \end{bmatrix}.$$

Then, $\boldsymbol{\Phi}(z)$ can be written as

$$\boldsymbol{\Phi}(z) = z\left[\pi_{00} \sum_{k=1}^{\infty} z^k \boldsymbol{B}_k - \boldsymbol{\pi}_1 \boldsymbol{A}_0\right]\big(z\boldsymbol{I} - \boldsymbol{A}^*(z)\big)^{-1}.$$

It follows from

$$z\boldsymbol{I}-\boldsymbol{A}^*(z)=\begin{bmatrix} z-B(z) & -C(z) \\ 0 & z-A(z) \end{bmatrix},$$

that

$$\big(z\boldsymbol{I}-\boldsymbol{A}^*(z)\big)^{-1}=\begin{bmatrix} \frac{1}{z-B(z)} & \frac{C(z)}{[z-B(z)][z-A(z)]} \\ 0 & \frac{1}{z-A(z)} \end{bmatrix}.$$

Noting $u=b_0(\pi_{00}+\pi_{10})$, and

$$\begin{aligned}
&\pi_{00}\sum_{k=1}^{\infty}z^k\boldsymbol{B}_k-\boldsymbol{\pi}_1\boldsymbol{A}_0\\
&\quad=\pi_{00}\big(B(z)-b_0, C(z)-c_0\big)-(\pi_{10}b_0,\pi_{10}c_0+\pi_{11}a_0)\\
&\quad=\big(\pi_{00}B(z)-b_0(\pi_{00}+\pi_{10}),\pi_{00}C(z)-(\pi_{00}+\pi_{10})c_0-\pi_{11}a_0\big)\\
&\quad=\big(\pi_{00}B(z)-u,u-\pi_{00}\big(1-C(z)\big)\big),
\end{aligned}$$

we have

$$\begin{aligned}
\Phi(z)&=z\big(\pi_{00}B(z)-u,\ u-\pi_{00}\big(1-C(z)\big)\big)\big(z\boldsymbol{I}-\boldsymbol{A}^*(z)\big)^{-1}\\
&=z\left(\frac{\pi_{00}B(z)-u}{z-B(z)},\frac{C(z)(\pi_{00}B(z)-u)}{(z-B(z))(z-A(z))}+\frac{u-\pi_{00}(1-C(z))}{z-A(z)}\right). \qquad (5)
\end{aligned}$$

Note that the PGF of the stationary queue length at the departure epochs $L(z)=\pi_{00}+\Phi(z)\boldsymbol{e}$. Firstly, we compute

$$\begin{aligned}
&\Phi(z)\boldsymbol{e}\\
&\quad=z\frac{(\pi_{00}B(z)-u)(z-A(z))+C(z)(\pi_{00}B(z)-u)+(u-\pi_{00}(1-C(z)))(z-B(z))}{(z-B(z))(z-A(z))}\\
&\quad=z\frac{\pi_{00}B(z)(1+z-A(z))-\pi_{00}z(1-C(z))+u(A(z)-B(z)-C(z))}{(z-B(z))(z-A(z))}.
\end{aligned}$$

Then, $L(z)$ has the expression

$$\frac{\pi_{00}(z-B(z))(z-A(z))+z[\pi_{00}B(z)(1+z-A(z))-\pi_{00}z(1-C(z))+u(A(z)-B(z)-C(z))]}{(z-B(z))(z-A(z))}. \qquad (6)$$

From Lemma 1, γ is the root of the equation $z=\widetilde{G}_v(\theta+\lambda(1-z))=B(z)$, and the numerator of (6) equals 0 for $z=\gamma$. Thus, substituting $z=\gamma$ into the numerator of (6), and using $B(\gamma)=\gamma$, we get

$$\pi_{00}\gamma\big[1+\gamma-A(\gamma)-\big(1-C(\gamma)\big)\big]+u\big(A(\gamma)-\gamma-C(\gamma)\big)=0. \qquad (7)$$

Substituting

$$\begin{aligned}
&A(\gamma)-\gamma-C(\gamma)=A(\gamma)-\gamma-A(\gamma)V(\gamma)=A(\gamma)\big(1-V(\gamma)\big)-\gamma,\\
&\gamma\big[1+\gamma-A(\gamma)-\big(1-C(\gamma)\big)\big]=\gamma\big(\gamma-A(\gamma)\big(1-V(\gamma)\big)\big),
\end{aligned}$$

into (7), we obtain $u = \pi_{00}\gamma$. It follows from (6) that

$$\begin{aligned} L(z) &= \pi_{00} + \Phi(z)\boldsymbol{e} \\ &= \pi_{00} \frac{(z - B(z))(z - A(z)) + z[B(z)(1 + z - A(z)) - z(1 - C(z)) + \gamma(A(z) - B(z) - C(z))]}{z - B(z))(z - A(z))} \\ &= \pi_{00} \frac{A(z)(1 - z)(B(z) - z) + z(\gamma - z)(A(z) - B(z) - C(z))}{(z - B(z))(z - A(z))}. \end{aligned} \tag{8}$$

Using the normalizing condition $L(1) = 1$, we can obtain

$$\pi_{00} = \frac{\theta(1 - \widetilde{G}_v(\theta))(1 - \rho)}{(1 - \widetilde{G}_v(\theta))(\theta + \lambda(1 - \gamma)) - \rho\theta(1 - \gamma)\widetilde{G}_v(\theta)}. \tag{9}$$

This completes the proof. □

Let $P_v(P_b)$ be the probability that an arbitrary customer is served completely at the low service rate during a working vacation, denoted by $S = 0$ (at normal service rate during a busy period, denoted by $S = 1$), then

$$\begin{aligned} P_v &= \mathrm{P}\{S = 0\} = (\pi_{00} + \pi_{10})b_0 + \Phi(z)\boldsymbol{e_1}|_{z=1} \\ &= \pi_{00}\gamma + \pi_{00} \frac{\widetilde{G}_v(\theta) - \gamma}{1 - \widetilde{G}_v(\theta)} = \pi_{00} \frac{\widetilde{G}_v(\theta)(1 - \gamma)}{1 - \widetilde{G}_v(\theta)} \\ &= \frac{\theta(1 - \rho)\widetilde{G}_v(\theta)(1 - \gamma)}{(1 - \widetilde{G}_v(\theta))(\theta + \lambda(1 - \gamma)) - \rho\theta(1 - \gamma)\widetilde{G}_v(\theta)}, \end{aligned}$$

where $\boldsymbol{e_1} = (1, 0)^T$. Similarly,

$$\begin{aligned} P_b &= \mathrm{P}\{S = 1\} = (\pi_{00} + \pi_{10})c_0 + \pi_{11}a_0 + \Phi(z)\boldsymbol{e_2}|_{z=1} = 1 - P_v \\ &= \frac{(1 - \widetilde{G}_v(\theta))(\theta + \lambda(1 - \gamma)) - \theta(1 - \gamma)\widetilde{G}_v(\theta)}{(1 - \widetilde{G}_v(\theta))(\theta + \lambda(1 - \gamma)) - \rho\theta(1 - \gamma)\widetilde{G}_v(\theta)}, \end{aligned}$$

where $\boldsymbol{e_2} = (0, 1)^T$.

4 Conditional stochastic decomposition for $L(z)$

It is not convenient to use $L(z)$ in (4) to interpret the queue length distribution or to relate the working vacation model to the classical M/G/1 model. Now based on the stochastic decomposition method in Shanthikumar [15], we obtain an alternative expression for $L(z)$. First, we derive the distribution of the queue length at the beginning epoch of a busy period (the ending epoch of a vacation period).

4.1 The queue length at the beginning epoch of a busy period

Define the queue length at the beginning of a busy period Q_b and τ_k ($k \geq 1$) represents its probability distribution, i.e.,

$$\tau_k = \mathrm{P}\{Q_b = k\}, \quad k \geq 1.$$

Let the last slow rate (during working vacations) service completion before the start of a busy period be an embedded point, then $Q_b = k$ happens in one of two possible cases: Case 1, under the condition that the system stays in the vacation period after the last slow rate service and the current vacation time is not longer than S_v, i.e., $V \le S_v$, j ($j \ge 1$) customers are left by the last slow rate served customer and $k - j$ customers arrive during the vacation time V; Case 2, under the same condition, no customers are left by the last slow rate served customer and $k - 1$ customers arrive during the vacation time V. The probability that the system stays in the vacation period after the last service before the beginning of the busy period and $V \le S_v$ is

$$\mathrm{P}\{J = 0, V \le S_v\} = \sum_{i=0}^{\infty} \pi_{i0} \times \mathrm{P}\{V \le S_v\} = \pi_{00}(1 - \gamma),$$

where J represents the system state after one service completion. Thus, we have

$$\tau_k = \mathrm{P}\{Q_b = k\} = \frac{1}{\pi_{00}(1-\gamma)} \left(\sum_{j=1}^{k} \pi_{j0} v_{k-j} + \pi_{00} v_{k-1} \right), \quad k \ge 1.$$

Multiplying it by z^j and summing over $j = 1, 2, \ldots,$ we get

$$\begin{aligned} Q_b(z) &= \sum_{k=1}^{\infty} \tau_k z^k = \frac{1}{\pi_{00}(1-\gamma)} \left\{ \sum_{j=1}^{\infty} \pi_{j0} z^j \sum_{k=j}^{\infty} v_{k-j} z^{k-j} + \pi_{00} \sum_{k=1}^{\infty} v_{k-1} z^k \right\} \\ &= \frac{1}{\pi_{00}(1-\gamma)} \left\{ \Phi(z) \boldsymbol{e_1} V(z) + \pi_{00} z V(z) \right\} \\ &= \frac{1}{1-\gamma} \left\{ \frac{z(B(z)-\gamma)}{z - B(z)} + z \right\} V(z) = \frac{1}{1-\gamma} \frac{z(z-\gamma)V(z)}{z - B(z)}. \end{aligned} \tag{10}$$

Evidently, $Q_b(1) = 1$ and

$$E(Q_b) = \frac{(1 - \widetilde{G}_v(\theta))(\theta + \lambda(1-\gamma)) - \theta(1-\gamma)\widetilde{G}_v(\theta)}{\theta(1-\gamma)(1 - \widetilde{G}_v(\theta))} = \frac{P_b(1-\rho)}{\pi_{00}(1-\gamma)}.$$

4.2 Conditional stochastic decomposition structure for $L(z)$

As the system dynamics during the normal busy period in our model is the stochastically equivalent to the classical (or non-working) vacation model in Shanthikumar [15], the results in Shanthikumar [15] can be used to analyze the normal service period in the M/G/1 queue with working vacations. Note that we can obtain the stationary distribution of the number of customers at the beginning of the normal service period.

As in Shanthikumar [15], the service time of every customer is called an active period and the length of vacation is called an inactive period. Obviously, in our M/G/1/WV model, the server can work during an inactive period. The server may be viewed as alternating between active and inactive states if we allow the inactive period to have zero length. To keep coherence with results in Shanthikumar [15], let $L^s(L^T)$ be the number of the customers at the starting (ending) instant of an inactive period in the steady state, and $L^s(z)(L^T(z))$ is the corresponding PGF which will be used in the proof of the following theorem.

Theorem 3 *The PGF of the stationary queue length L at the departure epoch can be expressed as*

$$L(z) = P_v \frac{1-\widetilde{G}_v(\theta)}{(1-\gamma)\widetilde{G}_v(\theta)} \frac{B(z)(z-\gamma)}{z-B(z)} \\ + P_b \frac{(1-\rho)(1-z)A(z)}{A(z)-z} \frac{1-Q_b(z)}{E(Q_b)(1-z)}. \tag{11}$$

Proof Using the conditional argument, we have

$$L(z) = E\big(z^L|S=1\big)\mathrm{P}\{S=1\} + E\big(z^L|S=0\big)\mathrm{P}\{S=0\}. \tag{12}$$

In an M/G/1 queue with working vacations, we choose the customer departure instants during a normal service period (busy period) as regeneration points, and customer arrivals and departures during a working vacation period only affect the deviation $L^s - L^T$. Therefore, applying (2) in Shanthikumar [15], we obtain

$$E\big(z^L|S=1\big) = E\big(z^N\big)E\big(z^X\big),$$

where N is the number of customers in a classic M/G/1 queue without vacations. N and the additional variable X have the PGFs

$$E\big(z^N\big) = \frac{(1-\rho)(1-z)A(z)}{A(z)-z}, \qquad E\big(z^X\big) = \frac{L^s(z)-L^T(z)}{(1-\rho)(1-z)}, \tag{13}$$

which is based on Lemma 1 in Shanthikumar [15].

Here, as we explained above, under the condition $S=1$, one customer is served in the normal busy period and the queue length L left by this customer can be zero. Thus, the expression for $E(z^L|S=1)$ should not be $E(z^N|N>0)E(z^X)$.

Certainly, for this system,

$$E\big(z^L|S=0\big) = P_v^{-1}\big((\pi_{00}+\pi_{10})b_0 + \Phi(z)\boldsymbol{e_1}\big) \\ = P_v^{-1}\pi_{00}\frac{B(z)(z-\gamma)}{z-B(z)} = \frac{1-\widetilde{G}_v(\theta)}{(1-\gamma)\widetilde{G}_v(\theta)} \frac{B(z)(z-\gamma)}{z-B(z)}. \tag{14}$$

Note that $L^T = k(k \geq 1)$ includes two disjoint cases: (1) $L^s = k$, if there is a zero length inactive period between two successive active periods (two continuously performed normal services); and (2) $L^s = 0$ and there are k customers in the system when a working vacation ends. Therefore, we have

$$\mathrm{P}\big\{L^T = k\big\} = \mathrm{P}\big\{L^s = k\big\} + \mathrm{P}\big\{L^s = 0\big\}\tau_k, \quad k \geq 1, \tag{15}$$

and $\mathrm{P}\{L^T = 0\} = 0$. From (10), we can obtain the relationship

$$L^T(z) = L^s(z) - \mathrm{P}\{L^s = 0\}\big(1 - Q_b(z)\big).$$

It follows from (13) that

$$E\big(z^X\big) = \frac{\mathrm{P}\{L^s = 0\}(1 - Q_b(z))}{(1-\rho)(1-z)}.$$

Using $E(z^X)|_{z=1} = 1$, we get $\mathrm{P}\{L^s = 0\} = (1-\rho)(E(Q_b))^{-1}$, and

$$E\big(z^X\big) = \frac{1 - Q_b(z)}{E(Q_b)(1-z)}.$$

Thus, it follows from (12) and (13) that (11) holds. □

Remark 1 Equation (11) and (4) can be shown to be equivalent. Note that

$$L(z) = \pi_{00}\gamma + \Phi(z)\boldsymbol{e_1} + \pi_{00}(1-\gamma) + \Phi(z)\boldsymbol{e_2}.$$

From (14), we have

$$\pi_{00}\gamma + \Phi(z)\boldsymbol{e_1} = P_v \frac{1 - \widetilde{G}_v(\theta)}{(1-\gamma)\widetilde{G}_v(\theta)} \frac{B(z)(z-\gamma)}{z - B(z)}.$$

It is sufficient to verify the relation

$$P_b \frac{(1-\rho)(1-z)A(z)}{A(z) - z} \frac{1 - Q_b(z)}{E(Q_b)(1-z)} = \pi_{00}(1-\gamma) + \Phi(z)\boldsymbol{e_2}. \tag{16}$$

Substituting $Q_b(z)$ into the left hand side of (16) and using $E(Q_b) = P_b(1 - \rho)(\pi_{00}(1-\gamma))^{-1}$, we obtain

$$\begin{aligned} &P_b \frac{(1-\rho)(1-z)A(z)}{A(z) - z} \frac{1 - Q_b(z)}{E(Q_b)(1-z)} \\ &\quad = \pi_{00}A(z)\frac{z(z-\gamma)V(z) - (1-\gamma)(z - B(z))}{(z - B(z))(z - A(z))}. \end{aligned}$$

Similarly, computing the right hand side of (16) yields the same expression. Therefore, (11) and (4) are equivalent.

Accordingly, the expected queue length is shown to be

$$\begin{aligned} E(L) = {} & P_v\left\{\frac{\beta}{\widetilde{G}_v(\theta)} + \frac{1}{1-\gamma} - \frac{1-\beta}{1 - \widetilde{G}_v(\theta)}\right\} \\ & + P_b\left\{\rho + \frac{\lambda^2 b^{(2)}}{2(1-\rho)} + \frac{E(Q_b(Q_b - 1))}{2E(Q_b)}\right\}, \end{aligned}$$

where β is defined in Sect. 2.

5 Stationary queue length at the arbitrary epoch

Let $J(t)$ be the system state (or server status) at any time t, i.e., $J(t)$ equals 0 or 1, if the system is in a working vacation or a busy period, respectively. Then $X(t)=(L(t),J(t))$ forms a continuous-time Markov regeneration process. Define the limiting distribution of $X(t)$ by

$$p_{00}=\lim_{t\to\infty}\mathrm{P}\big\{L(t)=0,J(t)=0\big\},$$

$$p_{nj}=\lim_{t\to\infty}\mathrm{P}\big\{L(t)=n,J(t)=j\big\},\quad n\geq 1,\ j=0,1.$$

To obtain the expressions for p_{nj}, we consider another stochastic process $(\widetilde{L}(t),\widetilde{J}(t))$, where $\widetilde{L}(t)$ denotes the queue length at the most recent departure, and $\widetilde{J}(t)$ equals 0 or 1 if the system stays in a working vacation or a busy service period after the most recent departure, respectively. Clearly, $(\widetilde{L}(t),\widetilde{J}(t))$ is a semi-Markov process (SMP) having (L_n,J_n) for its embedded Markov chain. Let γ_{kj} be the sojourn time in state (k,j) in $(\widetilde{L}(t),\widetilde{J}(t))$ process. Then

$$\mathrm{P}\{\gamma_{kj}<t\}=\begin{cases}\mathrm{P}\{S_b<t\}=G_b(t),\\ \quad k\geq 1,\ j=1;\\ \mathrm{P}\{S_v<t,V>S_v\}+\mathrm{P}\{V+S_b<t,V<S_v\}=F(t),\\ \quad k\geq 1,\ j=0;\\ \mathrm{P}\{A+S_v<t,V>S_v\}+\mathrm{P}\{A+V+S_b<t,V<S_v\}=A(t)*F(t),\\ \quad k=0,\ j=0,\end{cases}$$

where $*$ represents “convolution operation”. Thus, we have

$$m_{kj}=E(\gamma_{kj})=\begin{cases}\frac{1}{\mu_b}, & k\geq 1,\ j=1;\\ (\frac{1}{\mu_b}+\frac{1}{\theta})(1-\widetilde{G}_v(\theta)), & k\geq 1,\ j=0;\\ \frac{1}{\lambda}+(\frac{1}{\mu_b}+\frac{1}{\theta})(1-\widetilde{G}_v(\theta)), & k=0,\ j=0.\end{cases}$$

Let υ_{kj} be the steady-state probability that the SMP $(\widetilde{L}(t),\widetilde{J}(t))$ is in state (k,j). From the SMP theory in Gross and Harris [7], we have

$$\upsilon_{kj}=\frac{\pi_{kj}m_{kj}}{\sum_{i=0}^{1}\sum_{h=i}^{\infty}\pi_{hi}m_{hi}},$$

where $\{\pi_{kj},\ j=0,1;\ k\geq 0\}$ satisfy the equations in (3). It follows from the results in Theorem 2 and expressions for m_{kj} that $\sum_{i=0}^{1}\sum_{h=i}^{\infty}\pi_{hi}m_{hi}=1/\lambda$. Therefore,

$$\upsilon_{kj}=\begin{cases}\lambda\pi_{k1}m_{k1}=\rho\pi_{k1}, & k\geq 1;\\ \lambda\pi_{k0}m_{k0}=(\rho+\frac{\lambda}{\theta})(1-\widetilde{G}_v(\theta))\pi_{k0}, & k\geq 1;\\ \lambda\pi_{00}m_{00}=(1+(\rho+\frac{\lambda}{\theta})(1-\widetilde{G}_v(\theta)))\pi_{00}, & k=0,\ j=0.\end{cases}$$

For $n\geq 0$, $j=0,1$, the limiting distribution of $(L(t),J(t))$ has the following expressions (see in [7]):

$$
\begin{aligned}
p_{nj} = \sum_{i=0}^{1}\sum_{k=1}^{\infty}\frac{\upsilon_{ki}}{m_{ki}} \\
&\times \int_0^{\infty} P\Big\{\text{required changes in } t \text{ to bring state} \\
&\qquad\qquad \text{from } (k,i) \text{ to } (n,j),\ \gamma_{ki} > t\Big\}\,dt.
\end{aligned}
$$

First, we find p_{00}. Because of the fact that no arrivals occur in t means $\gamma_{00} > t$, we have

$$
p_{00} = \frac{\upsilon_{00}}{m_{00}}\int_0^{\infty} P\big\{\text{no arrivals occur in } t\big\}\,dt = \frac{\upsilon_{00}}{m_{00}}\int_0^{\infty} e^{-\lambda t}\,dt = \pi_{00}. \tag{17}
$$

For $n \geq 1$, $j = 0$, we have

$$
\begin{aligned}
p_{n0} &= \frac{\upsilon_{00}}{m_{00}}\int_0^{\infty} P\{(n-1) \text{ arrivals occur in } t,\ V > t,\ \gamma_{00} > t\}\,dt \\
&\quad + \sum_{k=1}^{n}\frac{\upsilon_{k0}}{m_{k0}} P\{(n-k) \text{ arrivals occur in } t,\ V > t,\ \gamma_{k0} > t\}\,dt \\
&= \lambda\pi_{00}\int_0^{\infty}\frac{(\lambda t)^{n-1}}{(n-1)!}e^{-\lambda t}\big(\mathrm{P}\{V > S_v, S_v > t\} + \mathrm{P}\{V > t, V < S_v\}\big)\,dt \\
&\quad + \lambda\sum_{k=1}^{n}\pi_{k0}\int_0^{\infty}\frac{(\lambda t)^{n-k}}{(n-k)!}e^{-\lambda t}\big(\mathrm{P}\{V > S_v, S_v > t\} + \mathrm{P}\{V > t, V < S_v\}\big)\,dt \\
&= \lambda\pi_{00}\int_0^{\infty}\frac{(\lambda t)^{n-1}}{(n-1)!}e^{-\lambda t}\left(\int_t^{\infty} e^{-\theta u}\,dG_v(u) + \int_t^{\infty}\theta e^{-\theta u}\big(1 - G_v(u)\big)\,du\right)dt \\
&\quad + \lambda\sum_{k=1}^{n}\pi_{k0}\int_0^{\infty}\frac{(\lambda t)^{n-k}}{(n-k)!} \\
&\quad \times e^{-\lambda t}\left(\int_t^{\infty} e^{-\theta u}\,dG_v(u) + \int_t^{\infty}\theta e^{-\theta u}\big(1 - G_v(u)\big)\,du\right)dt \\
&= \lambda\int_0^{\infty}\left(\pi_{00}\frac{(\lambda t)^{n-1}}{(n-1)!}e^{-\lambda t} + \sum_{k=1}^{n}\pi_{k0}\frac{(\lambda t)^{n-k}}{(n-k)!}e^{-\lambda t}\right)\big(1 - G_v(t)\big)e^{-\theta t}\,dt.
\end{aligned} \tag{18}
$$

Similarly, for $n \geq 1$,

$$
\begin{aligned}
p_{n1} &= \lambda\sum_{k=1}^{n}\pi_{k1}\int_0^{\infty}\frac{(\lambda t)^{n-k}}{(n-k)!}e^{-\lambda t}\mathrm{P}\{S_b > t\}\,dt \\
&\quad + \lambda\sum_{k=1}^{n}\pi_{k0}\int_0^{\infty}\frac{(\lambda t)^{n-k}}{(n-k)!}e^{-\lambda t}\mathrm{P}\{V < t, V < S_v, V + S_b > t\}\,dt
\end{aligned}
$$

$$+\lambda\pi_{00}\int_0^\infty \frac{(\lambda t)^{n-1}}{(n-1)!}e^{-\lambda t}\mathrm{P}\{V<t, V<S_v, V+S_b>t\}\,dt$$

$$=\lambda\sum_{k=1}^{n}\pi_{k1}\int_0^\infty \frac{(\lambda t)^{n-k}}{(n-k)!}e^{-\lambda t}\big(1-G_b(t)\big)\,dt$$

$$+\lambda\int_0^\infty \pi_{00}\frac{(\lambda t)^{n-1}}{(n-1)!}e^{-\lambda t}\int_0^t \theta e^{-\theta u}\big(1-G_v(u)\big)\big(1-G_b(t-u)\big)\,du\,dt$$

$$+\lambda\int_0^\infty \sum_{k=1}^{n}\pi_{k0}\frac{(\lambda t)^{n-k}}{(n-k)!}e^{-\lambda t}\int_0^t \theta e^{-\theta u}\big(1-G_v(u)\big)\big(1-G_b(t-u)\big)\,du\,dt. \tag{19}$$

Then we get the limiting distribution p_{kj} of the system at an arbitrary epoch.

Define the z-transforms of p_{n1} and p_{n0} as

$$P_b(z)=\sum_{n=1}^{\infty}p_{n1}z^n,\qquad P_v(z)=\sum_{n=0}^{\infty}p_{n0}z^n,$$

respectively. Utilizing (18) and (19), we easily get

$$\begin{aligned}
P_v(z)&=\pi_{00}+\lambda\left(\pi_{00}z+\sum_{k=1}^{\infty}\pi_{k0}z^k\right)\int_0^\infty e^{-\lambda(1-z)t}\big(1-G_v(t)\big)e^{-\theta t}\,dt\\
&=\pi_{00}+\lambda\left(\pi_{00}z+\sum_{k=1}^{\infty}\pi_{k0}z^k\right)\frac{1-\widetilde{G}_v(\theta+\lambda(1-z))}{\theta+\lambda(1-z)},\\
P_b(z)&=\lambda\sum_{k=1}^{\infty}\pi_{k1}z^k\frac{1-\widetilde{G}_b(\lambda(1-z))}{\lambda(1-z)}\\
&\quad+\lambda\left(\pi_{00}z+\sum_{k=1}^{\infty}\pi_{k0}z^k\right)\frac{\theta(1-\widetilde{G}_v(\theta+\lambda(1-z)))}{\theta+\lambda(1-z)}\frac{1-\widetilde{G}_b(\lambda(1-z))}{\lambda(1-z)}.
\end{aligned} \tag{20}$$

Obviously, the stationary distribution of the queue length at an arbitrary epoch is given by

$$p_0=\lim_{t\to\infty}\mathrm{P}\big\{L(t)=0\big\}=p_{00},\qquad p_n=\lim_{t\to\infty}\mathrm{P}\big\{L(t)=n\big\}=p_{n1}+p_{n0},\quad n\ge 1.$$

Using the expressions for $A(z)$, $B(z)$ and $C(z)$, the PGF of $\{p_n, n\ge 0\}$ can be expressed as

$$\begin{aligned}
P(z)&=P_v(z)+P_b(z)\\
&=\pi_{00}+\sum_{k=1}^{\infty}\pi_{k1}z^k\frac{1-A(z)}{1-z}\\
&\quad+\left(\pi_{00}z+\sum_{k=1}^{\infty}\pi_{k0}z^k\right)\frac{1-B(z)-C(z)}{1-z}.
\end{aligned} \tag{21}$$

Substituting $\Phi(z)\boldsymbol{e}_1$, $\Phi(z)\boldsymbol{e}_2$ into (21), after routine algebraic manipulation, we obtain

$$P(z) = L(z),$$

which demonstrates the relationship $p_n = \pi_n$, $n \geq 0$, which indicates the fact that the stationary queue length distribution at arbitrary epochs is the same as that at departure epochs. In other words, PASTA (Poisson arrivals see time average) property also holds in the M/G/1 queue with exponentially working vacations.

Remark 2 As the special case (exponential working vacation) of the model in Wu and Takagi [18], (4) can be verified to be in agreement with (50) in Wu and Takagi [18].

6 Waiting time analysis

Let W and $\widetilde{W}(s)$ be the stationary waiting time and its LST, respectively. Because it is possible for each customer to be served during either the normal service or working vacation period, the service time is different for the two possible service periods. Define the conditional waiting times as

$$W_b = \{W|S = 1\}, \qquad W_v = \{W|S = 0\},$$

with $\widetilde{W}_b(s)$, $\widetilde{W}_v(s)$ representing their LSTs.

For a customer served during a busy period, the customers in the system left by this customer are those arrived during his/her waiting time W_b and normal service time S_b. Because, if the service of a customer is not complete at the end of the vacation, he/she will be served at the normal service rate as a new customer, we regard the elapsed service time of the interrupted customer as part of his/her waiting time. Thus, we have

$$E\big(z^L|S = 1\big) = \widetilde{W}_b\big(\lambda(1-z)\big)\widetilde{G}_b\big(\lambda(1-z)\big).$$

Substituting $s = \lambda(1-z)$ into the equation above, we have

$$\widetilde{W}_b(s) = \frac{(1-\rho)s}{s - \lambda(1-\widetilde{G}_b(s))} \frac{\lambda(1 - Q_b(1-\frac{s}{\lambda}))}{E(Q_b)s}.$$

Theorem 4 *The conditional waiting time W_b can be decomposed into the sum of two independent random variables: $W_b = W_0 + W_d$, where W_0 is the waiting time of a classic M/G/1 queue without vacation. W_0 and the additional delay W_d have the LSTs*

$$\widetilde{W}_0(s) = \frac{(1-\rho)s}{s - \lambda(1-\widetilde{G}_b(s))}, \qquad \widetilde{W}_d(s) = \frac{\lambda(1 - Q_b(1-\frac{s}{\lambda}))}{E(Q_b)s}.$$

Similarly, for W_v, we have

$$E\big(z^L|S = 0\big) = \widetilde{W}_v\big(\lambda(1-z)\big)\widetilde{G}_v\big(\lambda(1-z)\big).$$

Substituting $s=\lambda(1-z)$ into the equation above yields

$$\widetilde{W}_v(s)=\frac{1-\widetilde{G}_v(\theta)}{(1-\gamma)\widetilde{G}_v(\theta)}\frac{\widetilde{G}_v(\theta+s)[s-\lambda(1-\gamma)]}{s-\lambda(1-\widetilde{G}_v(\theta+s))}\frac{1}{\widetilde{G}_v(s)}.$$

Thus, the LST of the unconditional waiting time W of any customer is given by

$$\begin{aligned}\widetilde{W}(s)&=\mathrm{P}\{S=0\}\widetilde{W}_v(s)+\mathrm{P}\{S=1\}\widetilde{W}_b(s)\\&=P_v\frac{1-\widetilde{G}_v(\theta)}{(1-\gamma)\widetilde{G}_v(\theta)}\frac{\widetilde{G}_v(\theta+s)[s-\lambda(1-\gamma)]}{s-\lambda(1-\widetilde{G}_v(\theta+s))}\frac{1}{\widetilde{G}_v(s)}\\&\quad+P_b\frac{(1-\rho)s}{s-\lambda(1-\widetilde{G}_b(s))}\frac{\lambda(1-Q_b(1-\frac{s}{\lambda}))}{E(Q_b)s}.\end{aligned}\tag{22}$$

It follows from (22) that the mean $E(W)$ and second moment $E(W^2)$ are found to be

$$\begin{aligned}E(W)&=P_v\left\{\frac{1}{\lambda}\left[\frac{\beta}{\widetilde{G}_v(\theta)}+\frac{1}{1-\gamma}-\frac{1-\beta}{1-\widetilde{G}_v(\theta)}\right]-\frac{1}{\mu_v}\right\}\\&\quad+P_b\left\{\frac{\lambda b^{(2)}}{2(1-\rho)}+\frac{E(Q_b(Q_b-1))}{2\lambda E(Q_b)}\right\},\\E(W^2)&=P_v\frac{\Lambda}{\lambda^2(1-\gamma)\widetilde{G}_v(\theta)(1-\widetilde{G}_v(\theta))^2}\\&\quad+P_b\left\{\frac{b^{(2)}Q_b^{(2)}(1)}{2(1-\rho)E(Q_b)}+\frac{3\lambda^2(b^{(2)})^2-2\lambda(1-\rho)b^{(3)}}{6(1-\rho)^2}+\frac{Q_b^{(3)}(1)}{3\lambda^2E(Q_b)}\right\},\end{aligned}$$

where $Q_b^{(k)}(1)$ represents the kth order derivative value for $Q_b(z)$ at $z=1$,

$$\begin{aligned}\Lambda&=\widetilde{G}_v^{(2)}(\theta)\lambda(1-\gamma)\big(1-\widetilde{G}_v(\theta)\big)\left(\lambda+2(1-\beta)+2\frac{\lambda}{\mu_v}\big(1-\widetilde{G}_v(\theta)\big)\right)\\&\quad-\lambda^2(1-\gamma)\big(1-\widetilde{G}_v(\theta)\big)^2g_v^{(2)}-2\widetilde{G}_v(\theta)\big(1-\widetilde{G}_v(\theta)\big)^2\\&\quad\times\left(\lambda+\frac{\lambda}{\mu_v}-\left(\frac{\lambda}{\mu_v}\right)^2(1-\gamma)\right)\\&\quad+2(1-\beta)\left((1-\beta)\widetilde{G}_v(\theta)(1-\gamma)+\widetilde{G}_v(\theta)(1-\gamma)\frac{\lambda}{\mu_v}\big(1-\widetilde{G}_v(\theta)\big)\right),\end{aligned}$$

and

$$\widetilde{G}_v^{(2)}(\theta)=\int_0^\infty x^2e^{-\theta x}\,dG_v(x).$$

7 Busy period analysis

The duration in which the server works at the normal rate continuously is called a busy period, denoted by D_b. A busy period starts at a vacation completion instant with a non-empty system and ends at a regular service completion with an empty system. Let D_v be the total working vacation period which begins at the end of a busy period and ends at the start of the next busy period. Obviously, during D_v, if there are arrivals, the server will work at the low service rate. A busy cycle C is then defined as the sum of a working vacation period D_v, and a subsequent busy period D_b.

We first give the busy period and its distribution function in non-vacation classic M/G/1 queue, denoted by D and $D(x)$, respectively. It follows from the classic busy period analysis of M/G/1 queue in Cohen [3] that

$$D(x) = \mathrm{P}\{D < x\} = \sum_{j=0}^{\infty} \int_0^x \frac{(\lambda u)^{j-1}}{j!} e^{-\lambda u}\, d\mathrm{P}\{S_b^{(j)} < u\} \tag{23}$$

where $S_b^{(j)}$ is the j-fold convolution of S_b. And its LST $\widetilde{D}(s)$ satisfies the equation

$$\widetilde{D}(s) = E(e^{-sD}) = \widetilde{G}_b[s + \lambda(1 - \widetilde{D}(s))],$$

and

$$E(D) = \frac{E(S_b)}{1-\rho}.$$

7.1 Busy period distribution

In our model, if there are k customers at a vacation completion instant, i.e., $Q_b = k$, due to the memoryless property of the arrival process, the conditional busy period $\{D_b | Q_b = k\}$ is the k-fold convolution of D, i.e.,

$$\{D_b | Q_b = k\} = D^{(k)}.$$

Therefore, the distribution of the busy period length D_b is obtained as

$$\mathrm{P}\{D_b < x\} = \sum_{k=1}^{\infty} \mathrm{P}\{D_b < x | Q_b = k\}\mathrm{P}\{Q_b = k\} = \sum_{k=1}^{\infty} D^{(k)}(x)\tau_k.$$

Denoting the LST of D_b by $\widetilde{D}_b(s)$, we have

$$\begin{aligned}\widetilde{D}_b(s) = E(e^{-sD_b}) &= \sum_{k=1}^{\infty} E\{e^{-sD_b} | Q_b = k\}\mathrm{P}\{Q_b = k\} \\ &= \sum_{k=1}^{\infty} \big(E(e^{-sD})\big)^k \tau_k = Q_b(\widetilde{D}(s)),\end{aligned}$$

where $Q_b(z)$ is given as in (10). Furthermore, $E(D_b)$ is found to be

$$E(D_b) = E(Q_b)E(D) = \frac{(1-\widetilde{G}_v(\theta))(\theta+\lambda(1-\gamma)) - \theta(1-\gamma)\widetilde{G}_v(\theta)}{\mu_b(1-\rho)\theta(1-\gamma)(1-\widetilde{G}_v(\theta))}.$$

7.2 Working vacation period distribution

Let a random variable U be the number of vacations continuously taken after a busy period, then D_v must be the sum of U vacations. If the server takes k vacations during the working vacation period, i.e., $U = k$, it means that there are no customers in the system at the end of jth vacation where $j = 1, 2, \dots, k-1$, and there are customers in the system at the end of kth vacation. Letting p be the probability that there are no customers in the system at the end of a vacation, we have the probability distribution of U as

$$\mathrm{P}\{U = k\} = p^{k-1}(1-p), \quad k \ge 1.$$

The key is to find an expression for p. If one vacation length is fixed, then there may be several times to make the system empty due to possible customer departures during this vacation. The instant of the system becoming empty each time in this fixed time length is equivalent to the ending instant of a busy period in the classic M/G/1 queue with the service time S_v. Thus, assume the busy period with the service time S_v as D^v, and it follows from the results in Cohen [3] that

$$D^v(x) = \mathrm{P}\big\{D^v < x\big\} = \sum_{j=1}^{\infty} \mathrm{P}\big\{D^v < x, M = j\big\} = \sum_{j=1}^{\infty} D_j^v(x),$$

where M represents the number of customers served during the busy period D^v. Then, the LST of D^v satisfies

$$\widetilde{D}^v(s) = \widetilde{G}_v\big[s + \lambda\big(1 - \widetilde{D}^v(s)\big)\big],$$

where $\widetilde{D}_j^v(s)$ is the LST of $D_j^v(x)$.

Then, if we see j $(j \ge 1)$ customer arrivals during one vacation, k busy periods may be formed, for $k = 1, 2, \dots, j$, and all j customers are served completely during this vacation. Let $Y = Y_1 + Y_2 + \cdots + Y_k$ be the time interval from the beginning of this vacation to all k busy periods ending during this vacation, where $Y_i = A_i + D_i^v$ represents the sum of the interval from the end of the previous busy period to the ith arrival causing the ith busy period to start, and the ith busy period. Note that $Y_1, Y_2, \dots, Y_k$ form a series of i.i.d. random variables with $A + D^v$. Meanwhile, M_i is the number of customers served during the ith busy period. M_i's are i.i.d. random variables, denoted by M. Then for $i = 1, 2, \dots, k$, we have the joint distribution

$$H_l(x) = \mathrm{P}\{Y_i < x, M_i = l\} = A * D_l^v(x), \quad l \ge 1,$$

and its LST

$$\widetilde{H}_l(s) = \int_0^{\infty} e^{-sx}\, dH_l(x) = \frac{\lambda}{\lambda+s}\widetilde{D}_l^v(s).$$

Then we get

$$
\begin{aligned}
p &= \int_0^\infty \theta e^{-\theta t} e^{-\lambda t}\,dt \\
&\quad + \int_0^\infty \theta e^{-\theta t} \sum_{j=1}^{\infty}\sum_{k=1}^{j} \mathrm{P}\{Y < t < Y + A, M_1 + M_2 + \cdots + M_k = j\}\,dt \\
&= \frac{\theta}{\lambda+\theta} \\
&\quad + \frac{\theta}{\lambda+\theta} \sum_{j=1}^{\infty}\sum_{k=1}^{j} \sum_{\substack{n_1+n_2+\cdots+n_k=j,\\ n_1,n_2,\ldots,n_k\ge 1}} \int_0^\infty e^{-\theta u}\,dH_{n_1}(u) * H_{n_2}(u) * \cdots * H_{n_k}(u) \\
&= \frac{\theta}{\lambda+\theta} + \frac{\theta}{\lambda+\theta} \sum_{j=1}^{\infty}\sum_{k=1}^{j} \sum_{\substack{n_1+n_2+\cdots+n_k=j,\\ n_1,n_2,\ldots,n_k\ge 1}} \widetilde{H}_{n_1}(\theta)\widetilde{H}_{n_2}(\theta)\cdots\widetilde{H}_{n_k}(\theta) \\
&= \frac{\theta}{\lambda+\theta} + \frac{\theta}{\lambda+\theta} \sum_{k=1}^{\infty}\left(\frac{\lambda}{\lambda+\theta}\right)^k \sum_{j=1}^{k} \sum_{\substack{n_1+n_2+\cdots+n_k=j,\\ n_1,n_2,\ldots,n_k\ge 1}} \widetilde{D}^v_{n_1}(\theta)\widetilde{D}^v_{n_2}(\theta)\cdots\widetilde{D}^v_{n_k}(\theta) \\
&= \frac{\theta}{\lambda+\theta} + \frac{\theta}{\lambda+\theta} \sum_{k=1}^{\infty}\left(\frac{\lambda}{\lambda+\theta}\right)^k \left(\sum_{l=1}^{\infty} \widetilde{D}^v_l(\theta)\right)^k \\
&= \frac{\theta}{\lambda+\theta} + \frac{\theta}{\lambda+\theta} \sum_{k=1}^{\infty}\left(\frac{\lambda}{\lambda+\theta}\right)^k \left(\widetilde{D}^v(\theta)\right)^k = \frac{\theta}{\theta+\lambda(1-\gamma)},
\end{aligned}
\tag{24}
$$

where we use the relationship $\widetilde{D}^v(\theta) = \widetilde{G}_v[\theta + \lambda(1 - \widetilde{D}^v(\theta))]$ which shows $\widetilde{D}^v(\theta) = \gamma$.

Then the distribution of the vacation number U on D_v is obtained. Obviously, we get the distribution function of D_v

$$
\mathrm{P}\{D_v < x\} = \sum_{k=1}^{\infty} \mathrm{P}\{V^{(k)} < x\}\mathrm{P}\{U = k\} = \sum_{k=1}^{\infty} (1-p)p^{k-1}\mathrm{P}\{V^{(k)} < x\},
$$

where $V^{(k)}$ represents the k-fold convolution of the vacation time V. Furthermore, the mean $E(D_v)$ is shown to be

$$
E(D_v) = \frac{E(V)}{1-p} = \frac{1}{\theta} + \frac{1}{\lambda(1-\gamma)}.
$$

Certainly, for the busy cycle C, we have

$$
\mathrm{P}\{C < x\} = \mathrm{P}\{D_b + D_v < x\}
$$

and the mean $E(C)$ can be computed as

$$E(C) = E(D_b) + E(D_v).$$

In fact, denote J the system state at the arbitrary epoch, then from the results in Sect. 5, we have

$$\mathrm{P}\{J=0\} = \sum_{n=0}^{\infty} p_{n0} = \frac{(\theta+\lambda(1-\gamma))(1-\widetilde{G}_v(\theta))(1-\rho)}{(1-\widetilde{G}_v(\theta))(\theta+\lambda(1-\gamma)) - \rho\theta(1-\gamma)\widetilde{G}_v(\theta)},$$

$$\mathrm{P}\{J=1\} = \sum_{n=0}^{\infty} p_{n1} = \frac{\rho((1-\widetilde{G}_v(\theta))(\theta+\lambda(1-\gamma)) - \theta(1-\gamma)\widetilde{G}_v(\theta))}{(1-\widetilde{G}_v(\theta))(\theta+\lambda(1-\gamma)) - \rho\theta(1-\gamma)\widetilde{G}_v(\theta)}.$$

We can also compute these probabilities by using

$$\mathrm{P}\{J=1\} = \frac{E(D_b)}{E(C)} = \frac{E(D_b)}{E(D_b)+E(D_v)},$$

$$\mathrm{P}\{J=0\} = \frac{E(D_v)}{E(C)} = \frac{E(D_v)}{E(D_b)+E(D_v)}.$$

Remark 3 The mean results in the cycle analysis above are in agreement with those in Kim et al. [8]. Meanwhile, with the expressions for the state probabilities, we can further show the conditional queue-length distribution given the server is (or not) on the working vacation at the arbitrary epoch. By the definition and expression for $P_v(z)$, $P_b(z)$ in (20), after some computation, we can verify that

$$P_0(z) = \frac{P_v(z)}{\mathrm{P}\{J=0\}} = \frac{\theta}{\theta+\lambda(1-\gamma)} + \frac{\theta z}{\theta+\lambda(1-z)} \frac{\lambda(z-\gamma)}{\theta+\lambda(1-\gamma)} \frac{1-B(z)}{z-B(z)},$$

$$P_1(z) = \frac{P_b(z)}{\mathrm{P}\{J=1\}} = \frac{1-\psi(z)}{\psi'(1)(1-z)} \frac{1-\rho}{\rho} \frac{z(1-A(z))}{z-A(z)},$$

where

$$\psi(z) = \frac{\theta z}{\theta+\lambda(1-z)} \frac{z-\gamma}{1-\gamma} \frac{1-B(z)}{z-B(z)}.$$

The results for the conditional queue lengths at the arbitrary epoch are the same as those in Kim et al. [8].

8 Some special examples

In this section, we show that some vacation models in the literatures published previously are special cases of our model.

Example 1 The M/G/1 queue with classical (non-working) vacations.

If the server does not serve customers during the vacation period, then $\widetilde{G}_v(s)=0$. Thus some quantities in our model become zero, i.e.,

$$B(z)=0, \qquad \widetilde{G}_v(\theta)=0, \qquad \gamma=0, \qquad \beta=0, \qquad P_v=0.$$

Substituting these zeros into (11), we obtain

$$\begin{aligned} L(z) &= \frac{(1-\rho)(1-z)A(z)}{A(z)-z}\frac{\theta}{\theta+\lambda(1-z)} \\ &= \frac{(1-\rho)(1-z)\widetilde{G}_b(\lambda(1-z))}{\widetilde{G}_b(\lambda(1-z))-z}\frac{\theta}{\theta+\lambda(1-z)}, \end{aligned}$$

which is the PGF of the queue length L in the M/G/1 queue with exponential vacations. The mean queue length is

$$E(L)=\rho+\frac{\lambda^2 b^{(2)}}{2(1-\rho)}+\frac{\lambda}{\theta}.$$

The LST and mean of the waiting time are

$$\widetilde{W}(s)=\frac{(1-\rho)s}{s-\lambda(1-\widetilde{G}_b(s))}\frac{\theta}{\theta+s}, \quad E(W)=\frac{\lambda b^{(2)}}{2(1-\rho)}+\frac{1}{\theta}.$$

All these results are in agreement with the results of M/G/1 queue with exponential vacations reported in [15]. Furthermore, the expected busy period, busy cycle and vacation period are given by

$$E(D_b)=\frac{1}{\mu_b(1-\rho)}\frac{\lambda+\theta}{\theta}, \qquad E(D_v)=\frac{\lambda+\theta}{\lambda\theta}, \qquad E(C)=\frac{1}{\lambda(1-\rho)}\frac{\lambda+\theta}{\theta}.$$

Example 2 The M/M/1 queue with working vacations (M/M/1/WV).

If service times S_b and S_v all follow the exponential distributions with parameters μ_b and μ_v, respectively, the system becomes an M/M/1/WV studied in Servi and Finn [14] and Liu et al. [9]. Now we have

$$\widetilde{G}_v(s)=\frac{\mu_v}{\mu_v+s}, \qquad \widetilde{G}_b(s)=\frac{\mu_b}{\mu_b+s}.$$

Thus, γ is the root of the equation

$$z=\frac{\mu_v}{\mu_v+\theta+\lambda(1-z)}$$

in the range of $0<z<1$, and the other root $\gamma'>1$. They are obtained as

$$\begin{aligned} \gamma &= \frac{1}{2\lambda}\left(\lambda+\theta+\mu_v-\sqrt{(\lambda+\theta+\mu_v)^2-4\lambda\mu_v}\right)=\frac{\mu_v}{\lambda}\gamma^*, \\ \gamma' &= \frac{1}{2\lambda}\left(\lambda+\theta+\mu_v+\sqrt{(\lambda+\theta+\mu_v)^2-4\lambda\mu_v}\right)=(\gamma^*)^{-1}. \end{aligned}$$

To relate the results to those in Liu et al. [9], we introduce γ^*. First, we compute π_{00} as,

$$\pi_{00} = \frac{(1-\rho)\theta}{\theta - (1-\gamma)(\rho\mu_v - \lambda)} = \frac{1-\rho}{1 + \frac{1-\gamma}{\theta}\rho(\mu_b - \mu_v)}.$$

Because γ satisfies,

$$\lambda\gamma^2 - (\lambda + \theta + \mu_v)\gamma + \mu_v = 0,$$

after some algebraic manipulation, we have the relation,

$$\frac{\mu_v}{\gamma} = \theta + \mu_v + \lambda(1-\gamma) = \frac{\theta}{1-\gamma} + \lambda.$$

Thus,

$$\frac{\theta}{1-\gamma} = \frac{\mu_v}{\gamma} - \lambda = \frac{\mu_v - \lambda\gamma}{\gamma} = \frac{\lambda(1-\gamma^*)}{\gamma^*}.$$

Substituting it into the expression for π_{00},

$$\pi_{00} = \frac{(1-\rho)(1-\gamma^*)}{1 - \gamma^* \frac{\mu_v}{\mu_b}}.$$

Using (4), we have

$$\begin{aligned}
L(z) &= \pi_{00} \frac{A(z)(1-z)(B(z)-z) + z(\gamma - z)(A(z) - B(z) - C(z))}{(z - B(z))(z - A(z))} \\
&= \pi_{00} \frac{\mu_b(1-z)[\mu_v - (\theta + \mu_v + \lambda(1-z))z] + z(\gamma - z)\lambda(\mu_b - \mu_v)(1-z)}{[(\theta + \mu_v + \lambda(1-z))z - \mu_v][(\mu_b + \lambda(1-z))z - \mu_b]} \\
&= \pi_{00} \left(\frac{1}{1-\rho z} + \frac{\lambda(1 - \frac{\mu_v}{\mu_b})z(\gamma - z)}{[\mu_v - (\theta + \mu_v + \lambda(1-z))z](1-\rho z)} \right).
\end{aligned}$$

Because γ and γ' are the roots of the equation

$$\lambda z^2 - (\lambda + \theta + \mu_v)z + \mu_v = 0,$$

then,

$$\mu_v - \big(\theta + \mu_v + \lambda(1-z)\big)z = \lambda z^2 - (\lambda + \theta + \mu_v)z + \mu_v = \lambda(z-\gamma)(z-\gamma').$$

Using the relationship $(\gamma')^{-1} = \gamma^*$ and substituting π_{00} into $L(z)$, we get

$$\begin{aligned}
L(z) &= \pi_{00} \left(\frac{1}{1-\rho z} + \frac{(1 - \frac{\mu_v}{\mu_b})z(\gamma - z)}{(z-\gamma)(z-\gamma')(1-\rho z)} \right) \\
&= \pi_{00} \left(\frac{1}{1-\rho z} + \frac{1 - \frac{\mu_v}{\mu_b}}{1-\rho z} \frac{\gamma^* z}{1 - \gamma^* z} \right) \\
&= \frac{1-\rho}{1-\rho z} \left(\frac{1-\gamma^*}{1 - \gamma^* \frac{\mu_v}{\mu_b}} + \frac{\gamma^*(1 - \frac{\mu_v}{\mu_b})}{1 - \gamma^* \frac{\mu_v}{\mu_b}} \frac{(1-\gamma^*)z}{1-\gamma^* z} \right),
\end{aligned} \tag{25}$$

and

$$E(L) = \frac{\rho}{1-\rho} + \left(1 - \frac{\mu_v}{\mu_b}\right)\left(1 - \gamma^* \frac{\mu_v}{\mu_b}\right)^{-1} \frac{\gamma^*}{1-\gamma^*}. \tag{26}$$

These results are in agreement with those in Liu et al. [9]. Meanwhile, letting $\hat{z} = \gamma' = (\gamma^*)^{-1}$, we further derive

$$\begin{aligned} L(z) &= \frac{1-\rho}{1-\rho z} \frac{1-\gamma^*}{1-\gamma^* z}\left(\frac{1-\gamma^* z}{1-\gamma^*\mu_v/\mu_b} + \frac{\gamma^*(1-\mu_v/\mu_b)z}{1-\gamma^*\mu_v/\mu_b}\right) \\ &= \frac{1-\rho}{1-\rho z} \frac{1-\gamma^*}{1-\gamma^* z} \frac{1-\gamma^*\mu_v/\mu_b}{1-\gamma^*\mu_v/\mu_b z} \\ &= \frac{1-\rho}{1-\rho z} \frac{1-\hat{z}^{-1}}{1-\hat{z}^{-1} z} \frac{1-\mu_v/(\mu_b \hat{z})}{1-\mu_v/(\mu_b \hat{z})z}, \end{aligned}$$

which also is the same as (1.1) in Servi and Finn [14]. Then, we obtain

$$P_v = \frac{(1-\rho)\frac{\mu_v}{\lambda}\gamma^*}{1-\gamma^*\frac{\mu_v}{\mu_b}}, \qquad P_b = \frac{1-\frac{\mu_v}{\lambda}\gamma^*}{1-\gamma^*\frac{\mu_v}{\mu_b}}.$$

The LST of the stationary waiting time is

$$\begin{aligned} \widetilde{W}(s) &= P_v \widetilde{W}_v(s) + P_b \widetilde{W}_p(s) \\ &= \frac{\frac{\theta}{\lambda}(1-\rho)\gamma^*}{(1-\gamma^*\frac{\mu_v}{\mu_b})(1-\frac{\mu_v}{\lambda}\gamma^*)} \frac{[s-\lambda(1-\frac{\mu_v}{\lambda}\gamma^*)](\mu_v+s)}{\mu_v s - (\lambda - s)(\theta + s)} \\ &\quad + \frac{1-\gamma^*}{1-\gamma^*\frac{\mu_v}{\mu_b}} \frac{(1-\rho)(\mu_b+s)}{s(s-(\lambda-\mu_b))}\left[\frac{\theta(\lambda-s)(\lambda-s-\gamma^*\mu_v)}{\mu_v s-(\lambda-s)(\theta+s)} + \lambda\left(1-\frac{\mu_v}{\lambda}\gamma^*\right)\right]. \end{aligned}$$

The expected waiting time is given by

$$E(W) = \frac{1}{\mu_b(1-\rho)} + \frac{1}{\lambda}\left(\left(1-\frac{\mu_v}{\lambda}\right)\frac{(\gamma^*)^2}{1-\gamma^*} - \rho(1-\gamma^*)\right)\left(1-\frac{\mu_v}{\lambda}\gamma^*\right)^{-1}. \tag{27}$$

Meanwhile, the expected busy period, the expected vacation period, and the expected busy cycle are

$$\begin{aligned} E(D_b) &= \frac{\theta-(1-\gamma)(\mu_v-\lambda)}{\theta\mu_b(1-\rho)(1-\gamma)}, \qquad E(D_v) = \frac{1}{\theta} + \frac{1}{\lambda(1-\gamma)}, \\ E(C) &= \frac{\theta\rho^{-1} + (1-\gamma)(\mu_b-\mu_v)}{\theta\mu_b(1-\rho)(1-\gamma)}. \end{aligned}$$

Thus, taking different service time distributions, we can obtain the results of various special M/G/1 queues with working vacations or classic vacations.

9 Numerical results

In this section, we present some numerical examples for the M/G/1/WV queue with several service time distributions such as deterministic (D), exponential (M), and Erlang (E_2).

First we consider the model denoted as $M/(D_1, D_2)/1$, where the service times S_b, S_v are deterministic, and $S_b = \mu_b^{-1}, S_v = \mu_v^{-1}$ with parameters taken as $\mu_b = 2.0, \lambda = 1.25$.

Figure 1 shows how the expected queue length changes with the working-vacation service rate μ_v at different θ values and different ρ values, respectively. Similarly, Fig. 2 shows how the expected waiting time and the probability of a customer being served completely at the normal service rate change with μ_v. Clearly, these two figures show that in general the expected queue length, expected waiting time and P_b all decrease with μ_v. Figure 3 shows the relation between the expected queue length or expected waiting time and μ_v for an $M/(M, E_2)/1$ queue, where S_b follows the exponential distribution with the rate of μ_b and S_v follows an Erlang-2 distribution with the mean μ_v. In both Figs. 1 and 3, we can examine the effects of the vacation rate θ on the performance measures such as the expected queue length. It has been

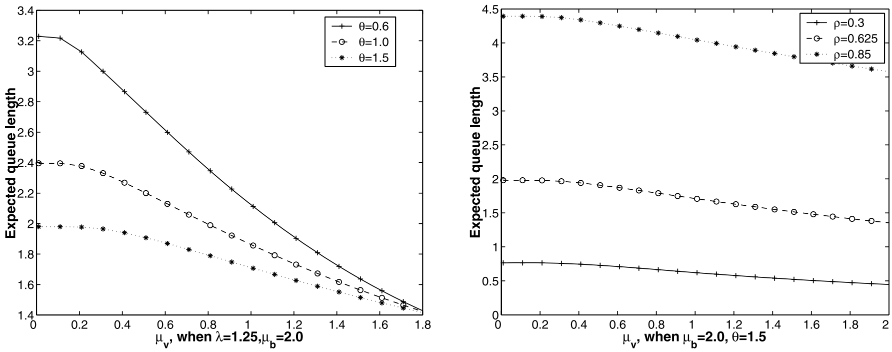


Fig. 1 Expected queue length against μ_v in $M/(D_1, D_2)/1$

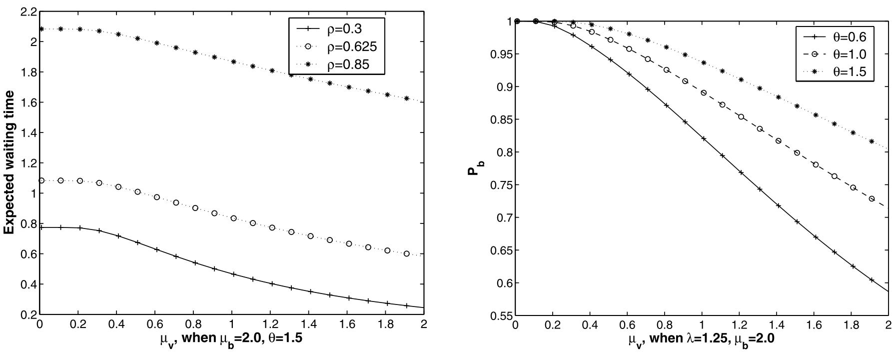


Fig. 2 $E(W)$ and P_b against μ_v, respectively

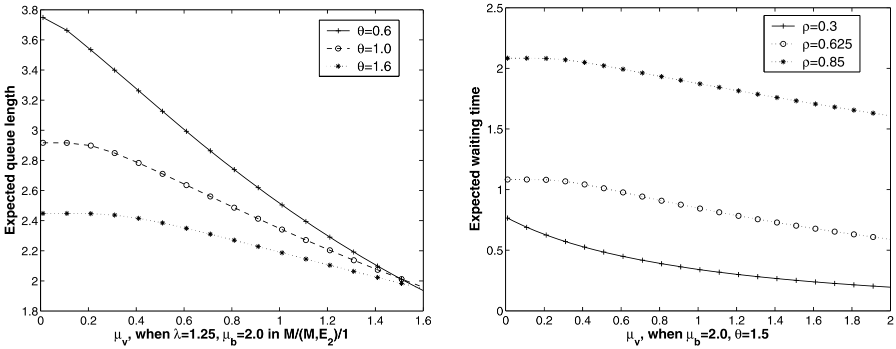
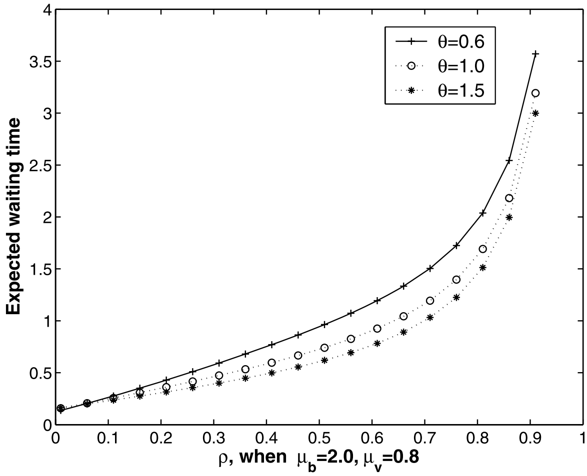


Fig. 3 $E(L)$ and $E(W)$ against μ_v in $M/(M, E_2)/1$, respectively

Fig. 4 $E(W)$ against ρ in $M/(M, E_2)/1$



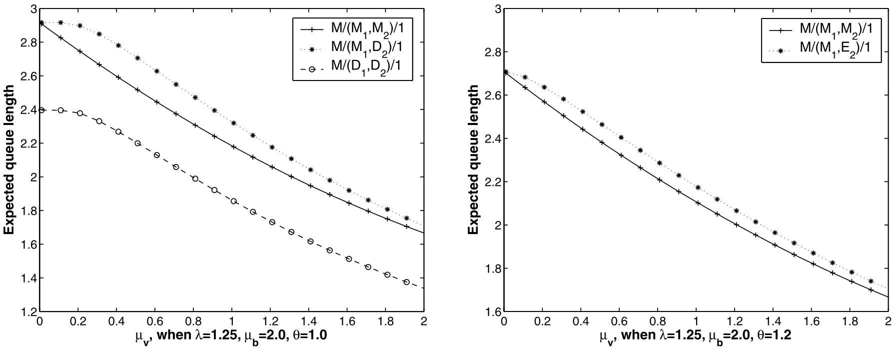


Fig. 5 Comparisons among different models

illustrated that the expected queue length is decreasing in θ and the effect of θ becomes smaller and turns to zero when $\mu_v = \mu_b = 2$. Another extreme case is when

$\mu_v = 0$ (no service during the vacation period or the classical vacation model), the effect of θ is largest.

Figure 4 illustrates that $E(W)$ increases with ρ. Note that the effect of θ on $E(W)$ is relatively small compared with that of ρ on $E(W)$.

Finally, in Fig. 5, we compare several models with different service distributions in terms of the expected queue lengths. The similar quantitative effect of the working-vacation service rate on the expected queue length has been demonstrated.

Acknowledgements The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the paper. And, the authors also thank the support from National Natural Science Foundation of China (No. 10671170), the support from the Science Foundation of Yanshan University for the Excellent Ph.D.Students, and the NSF of Hebei Province (F2008000864).

References

1. Baba, Y.: Analysis of a GI/M/1 queue with multiple working vacations. Oper. Res. Lett. **33**, 201–209 (2005)
2. Banik, A.D., Gupta, U.C., Pathak, S.S.: On the GI/M/1/N queue with multiple working vacations-analytic analysis and computation. Appl. Math. Model. **31**, 1701–1710 (2007)
3. Cohen, J.: The Single Server Queue. North-Holland, Amsterdam (1982), p. 250
4. Doshi, B.T.: Single server queues with vacations. In: Takagi, H. (ed.), Stochastic Analysis of Computer and Communication Systems. North-Holland/Elsevier, Amsterdam (1990), pp. 217–264
5. Doshi, B.T.: Queueing systems with vacations-a survey. Queueing Syst. **1**, 29–66 (1986)
6. Fuhrmann, S.W., Cooper, R.B.: Stochastic decomposition in the M/G/1 queue with generalized vacations. Oper. Res. **33**, 1117–1129 (1985)
7. Gross, D., Harris, C.: Fundamentals of Queueing Theory, 2nd edn. Wiley, New York (1985), pp. 347–354
8. Kim, J.D., Choi, D.W., Chae, K.C.: Analysis of queue-length distribution of the M/G/1 queue with working vacations. Hawaii International Conference on Statistics and Related Fields, 5–8 June 2003
9. Liu, W., Xu, X., Tian, N.: Stochastic decompositions in the M/M/1 queue with working vacations. Oper. Res. Lett. **35**, 595–600 (2007)
10. Li, J., Tian, N.: The discrete-time GI/Geo/1 queue with working vacations and vacation interruption. Appl. Math. Comput. **185**, 1–10 (2007)
11. Li, J., Tian, N., Liu, W.: Discrete-time GI/Geo/1 queue with working vacations. Queueing Syst. **56**, 53–63 (2007)
12. Neuts, M.F.: Matrix-Geometric Solutions in Stochastic Models. Hopkins University Press, Baltimore (1981)
13. Neuts, M.F.: Structured Stochastic Matrices of M/G/1 Type and Their Applications. Marcel Dekker, New York (1989)
14. Servi, L.D., Finn, S.G.: M/M/1 queue with working vacations (M/M/1/WV). Perform. Eval. **50**, 41–52 (2002)
15. Shanthikumar, J.G.: On stochastic decomposition in M/G/1 type queues with generalized server vacations. Oper. Res. **36**, 566–569 (1988)
16. Tian, N., Zhang, Z.G.: Vacation Queueing Models: Theory and Applications. Springer, New York (2006)
17. Takagi, H.: Vacation and Priority Systems, Part 1, Queueing Analysis: A Foundation of Performance Evaluation, vol 1. North-Holland/Elsevier, Amsterdam (1991)
18. Wu, D., Takagi, H.: M/G/1 queue with multiple working vacations. Perform. Eval. **63**, 654–681 (2006)