Computers & Industrial Engineering 65 (2013) 594–604

Contents lists available at SciVerse ScienceDirect

Computers & Industrial Engineering

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

Analysis of the discrete-time Geo/G/1 working vacation queue and its application to network scheduling

Ji-hong Li *

College of Management, Shanxi University, Taiyuan 030006, PR China

A R T I C L E   I N F O

Article history:
Received 10 October 2012
Received in revised form 20 February 2013
Accepted 17 April 2013
Available online 3 May 2013

Keywords:
Discrete-time queue
Working vacations
Embedded Markov chain
M/G/1-type matrix
Stochastic decomposition

A B S T R A C T

In this paper we present an exact steady-state analysis of a discrete-time Geo/G/1 queueing system with working vacations, where the server can keep on working, but at a slower speed during the vacation period. The transition probability matrix describing this queuing model can be seen as an M/G/1-type matrix form. This allows us to derive the probability generating function (PGF) of the stationary queue length at the departure epochs by the M/G/1-type matrix analytic approach. To understand the stationary queue length better, by applying the stochastic decomposition theory of the standard M/G/1 queue with general vacations, another equivalent expression for the PGF is derived. We also show the different cases of the customer waiting to obtain the PGF of the waiting time, and the normal busy period and busy cycle analysis is provided. Finally, we discuss various performance measures and numerical results, and an application to network scheduling in the wavelength division-multiplexed (WDM) system illustrates the benefit of this model in real problems.

1. Introduction

Discrete time queues with vacations have been studied in the past by many researchers due to their wide applications in the performance analysis of digital communication systems and telecommunication networks including the Broadband Integrated Services Digital Network (B-ISDN), Asynchronous transfer mode (ATM), and related technologies. In all these cases, continuous-time queues which can be found in the monographs of Takagi (1991) and Tian and Zhang (2006), cannot be applied to related fields from the fact that the arrival and departure occur only at certain fixed time epochs. Meanwhile, it is well known that all discrete-time queues have two variations due to the specific nature of the arrivals and departures, arrival-first (AF) and departure-first (DF), which both have special implications in practice. According to AF policy, arrivals take precedence over departures, while under DF policy the opposite effect is observed, seen in Hunter (1983) and Bruneel and Kim (1993). Therefore, discrete-time vacation queues have received a rapid increase of research attention both in queueing theorists and practitioners.

For discrete-time queues with different vacation policies, the analysis can be found in Takagi (1993), Tian and Zhang (2006), Alfa (2003) and references therein. In these papers, they all assumed that during the vacations, the servers stop original service completely. Servi and Finn (2002) considered an M/M/1 queue in which the server works at a different rate rather than completely stopping service during the vacation period. This policy is called working vacations and the vacation period becomes the slower speed operation period of the queueing system. Subsequently, the steady-state analysis of the M/G/1 queue with working vacations was presented by Kim, Choi, and Chae (2003), Wu and Takagi (2006), while Baba (2005) further extended to a GI/M/1 queue with working vacations by the matrix-analytic method developed by Neuts (1981). A finite buffer GI/M/1/ queue with working vacations was first studied by Banik, Gupta, and Pathak (2007). In recent years, by using the matrix geometric solution, Liu, Xu, and Tian (2007) obtained the steady-state results and the stochastic decomposition structures for the system indices in the M/M/1 queue with working vacations. By using the matrix-analytic method, the discrete-time GI/Geo/1 queues with working vacations have been studied by Li and Tian (2007), Li, Tian, and Liu (2007). Li, Tian, Zhang, and Luh (2009) and Li, Liu, and Tian (2009) studied the M/G/1 queue with working vacations where the server takes exponential multiple vacations.

In this paper, we consider a discrete-time Geo/G/1 queueing system with multiple working vacations where the server is working at the slower rate during the vacation period, noted as Geo/G/1(MWV). The analysis of this model is based on the use of the M/G/1-type matrix analytic approach developed by Neuts (1989) and the stochastic decomposition theory of the standard M/G/1 queue with general vacations in Shanthikumar (1988). The steady-state probability generating functions (PGFs) of the stationary performances are derived under the AF management policy, and the

* Tel.: +86 13485330831.
E-mail address: lijh1982@sxu.edu.cn

0360-8352/$ - see front matter
http://dx.doi.org/10.1016/j.cie.2013.04.009

conditional stochastic decomposition structures will be established. The queueing model discussed in the paper has potential applications in slotted digital communication systems such as ATM switching systems, circuit-switched TDMA systems and traffic concentrators. In such systems, messages or data arrive into the packet switching multiplexers that transmit packets using some service protocol. If there are few packets for transmission, the multiplexer can be assigned to execute both the current job and some other jobs, that is, some transmission ability is distributed to other jobs. A period in which the multiplexer executes both the current job and some other jobs can be treated as a vacation. By this mode, the utilization of the multiplexer can be improved in the real system.

The rest of this paper is organized as follows. In Section 2, the model is formulated as a two-dimensional embedded Markov chain at the departure epochs. In Section 3, by the Markov chain and matrix analytic methods, we obtain an expression for the PGF of the steady-state queue length. In Section 4, a more direct expression for the PGF of the queue length is derived by using the stochastic decomposition theory. By analyzing the conditional waiting time of a customer, the PGF of the waiting time is presented in Section 5. The busy period analysis is provided in Section 6. Finally, Section 7 contains some numerical results and an application in the WDM system to demonstrate the effectiveness of the model.

2. Model formulation and embedded Markov chain

We consider a discrete-time queueing system where the time axis is segmented into a sequence of equal time intervals (called slots). It is assumed that all queueing activities (arrivals and departures) occur at the slot boundaries, and therefore they may occur at the same time. For mathematical clarity, we define the arrival first (AF) system and suppose that the departures occur at the moment immediately before the slot boundaries which is similar to $t = n^-$, $n = 1, 2, \cdots$, and the arrivals occur at the moment immediately after the slot boundaries, i.e., $t = n^+$, $n = 0, 1, \cdots$.

Customers arrive at the system according to a geometric arrival process with rate λ, that is, λ is the probability that an arrival occurs in a slot. The arrival-interval time T follows a geometric distribution with rate λ.

$$P\{T = j\} = \lambda\bar{\lambda}^{j-1}, \quad \bar{\lambda} = 1 - \lambda, \quad 0 < \lambda < 1, \quad j \geqslant 1.$$

If, upon arrival, the server is idle, the service of the arriving customer commences immediately. Otherwise, the arriving customer joins the waiting line in order to be served. Service times of customers are synchronized on slot boundaries, that is, a customer can only start service at slot boundaries. Further, the service of a customer takes an integer number of slots, which implies that customers also leave the system at slot boundaries. After completion of one service, if the server finds the system becomes empty, he leaves for a vacation. During the vacation period, if there are new customers arriving, the server provides service for them at a slower rate by their arrival order.

At a vacation completion instant, or on return from a vacation, if the server finds customers in the system, he will come back to the normal working level and take the service with the normal service time. On the other hand, if he finds no customer, he immediately proceeds for another vacation and continues in this manner until he finds customers in the system at one vacation completion instant. If one customer is still being served when the vacation ends, we assume that this customer will begin to receive the service as a new one. This assumption can reasonably represent reality in some cases, because various service (vacation) time distributions arising in practical applications can be approximated by some special service distributions, such as geometric or phase type services, thus to some extent the elapsed service of this customer can be omitted. In other systems not revealing such service property, we should consider the case that the remaining service time of the customer can be completed by the normal speed. In this case, the analysis and results may be different and we will analyze future.

We assume that inter-arrival times, service times, and working vacation times are mutually independent. Although the slower service time commences during the vacation period, how long the service for one customer will take is not related to when the vacation occurs and how long this vacation will hold. This is possible since due to the memoryless arrival and vacation time distribution, at times of departure we need only to remember the number of users in the system in order to predict this number immediately after the next service completion. One whole vacation period, which begins at the instant that the system enters into vacation, and ends at the instant when a working vacation finishes, is directly related to the slower service time. One whole vacation period may be composed of several vacation times and is different from one vacation time. Here this system is denoted as Geo/G/1(MWV) model. The service discipline is First Come First Served (FCFS).

Some recommended notations of the model are given as follows:

- S_b: the customer's service time (in slots) during the busy period with the mean $E(S_b)$;
  $\left\{b_j^{(1)}, j \geqslant 1\right\}$: the general distribution of the normal service time S_b;
  $G_b(z)$: the probability generating function of the normal service time S_b;
  μ_b: the service rate during the normal period as $1/E(S_b)$.
- S_v: the customer's service times during the vacation period with the mean $E(S_v)$;
  $\left\{b_j^{(2)}, j \geqslant 1\right\}$: the general distribution of the vacation service time S_v;
  $G_v(z)$: the probability generating function of the vacation service time S_v;
  μ_v: the service rate during the vacation period as $1/E(S_v)$.
- V: one vacation time geometrically distributed with rate $\theta(0 < \theta < 1)$.

We make assumption $\mu_v \leqslant \mu_b$ to enable a slower service rate during the vacation period, so that the working vacation policy is an scheme under which the system can operate at two different rates alternatively.

Let $L(t)$ be the number of customers in the system at time t and L_n be the number of the customers at the instant of the nth service completion or customer departure. For the working vacation model, any service completion may occur during a service period or a working vacation period. Thus, we define a process

$$X_n = (L_n, J_n), \qquad n \geqslant 1,$$

where J_n symbolizes the system state at the nth departure (0 if the system stays in a working vacation period, or 1 if the system stays in a busy service period), then $\{X_n, n \geqslant 1\}$ is a two-dimensional embedded Markov chain of our queueing system and its state space is

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

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

We introduce the expressions below

$$a_j = \sum_{k=j}^{\infty} b_k^{(1)} \binom{k}{j} \lambda^j (1-\lambda)^{k-j}, \quad j \geqslant 0,$$

where $\{a_j, j \geqslant 0\}$ expresses the probability that there are j customer arrivals during S_b and has PGF and mean

$$\begin{aligned} A(z) = & \sum_{j=0}^{\infty} a_j z^j = \sum_{k=0}^{\infty} b_k^{(1)} \sum_{j=0}^{k} \binom{k}{j} (\lambda z)^j (1-\lambda)^{k-j} \\ = & \sum_{k=0}^{\infty} b_k^{(1)} [1-\lambda(1-z)]^k = G_b(1-\lambda(1-z)), \\ A'(1) = & \lambda G_b'(1) = \frac{\lambda}{\mu_b} = \rho. \end{aligned}$$

Denote

$$b_j = \sum_{k=max(1,j)}^{\infty} b_k^{(2)} \binom{k}{j} \lambda^j (1-\lambda)^{k-j} (1-\theta)^k, \quad j \geqslant 0;$$

$$v_j = \sum_{k=max(1,j)}^{\infty} \sum_{n=k}^{\infty} b_n^{(2)} \binom{k}{j} \lambda^j (1-\lambda)^k (1-\theta)^{k-1} \theta, \quad j \geqslant 0.$$

$\{b_j, j \geqslant 0\}$ has the probability explanation that $V > S_v$ and j customers arrive during the service time S_v. Meanwhile, $\{v_j, j \geqslant 0\}$ expresses the probability that $V \leqslant S_v$ and j customers arrive during the vacation time V. We have

$$\sum_{j=0}^{\infty} b_j = \sum_{k=0}^{\infty} b_k^{(2)} (1-\theta)^k = G_v(1-\theta),$$
$$\sum_{j=0}^{\infty} v_j = \sum_{n=1}^{\infty} b_n^{(2)} [1-(1-\theta)^n] = 1 - G_v(1-\theta).$$

$\{b_j, j \geqslant 0\}$ and $\{v_j, j \geqslant 0\}$ form two non-complete probability distributions. The corresponding partial probability generating functions are given as follows:

$$\begin{aligned} B(z) = \sum_{j=0}^{\infty} b_j z^j = & \sum_{j=0}^{\infty} \sum_{k=max(1,j)}^{\infty} b_k^{(2)} \binom{k}{j} \lambda^j (1-\lambda)^{k-j} (1-\theta)^k z^j \\ = & \sum_{k=0}^{\infty} b_k^{(2)} (1-\theta)^k \sum_{j=0}^{k} \binom{k}{j} (\lambda z)^j (1-\lambda)^{k-j} \\ = & \sum_{k=0}^{\infty} b_k^{(2)} (1-\theta)^k [1-\lambda(1-z)]^k = G_v((1-\theta)(1-\lambda(1-z))); \end{aligned}$$

$$\begin{aligned} V(z) = \sum_{j=0}^{\infty} v_j z^j = & \sum_{j=0}^{\infty} \sum_{k=max(1,j)}^{\infty} \sum_{n=k}^{\infty} b_n^{(2)} \binom{k}{j} \lambda^j (1-\lambda)^{k-j} (1-\theta)^{k-1} \theta z^j \\ = & \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} b_n^{(2)} (1-\theta)^{k-1} \theta [1-\lambda(1-z)]^{k-1} \\ = & \sum_{n=1}^{\infty} b_n^{(2)} \sum_{k=1}^{n} (1-\theta)^{k-1} \theta [1-\lambda(1-z)]^{k-1} \\ = & \frac{\theta}{1-(1-\theta)(1-\lambda(1-z))} \sum_{n=1}^{\infty} b_n^{(2)} [1-[(1-\theta)(1-\lambda(1-z))]^n] \\ = & \frac{\theta}{1-(1-\theta)(1-\lambda(1-z))} [1-G_v((1-\theta)(1-\lambda(1-z)))] \\ = & \frac{\theta}{1-(1-\theta)(1-\lambda(1-z))} [1-B(z)], \end{aligned}$$

where

$$B'(1) = \lambda(1-\theta) \sum_{j=1}^{\infty} j b_j^{(2)} (1-\theta)^j = \beta,$$
$$V'(1) = \frac{\lambda(1-\theta)}{\theta} (1 - G_v(1-\theta)) - \beta.$$

Let $c_j = \sum_{k=0}^{j} v_k a_{j-k}, j \geqslant 0$, then c_j represents that $V \leqslant S_v$ and j customers arrive during the sum of V and S_b, and we have

$$\sum_{j=0}^{\infty} c_j = 1 - G_v(1-\theta), \qquad C(z) = \sum_{j=0}^{\infty} c_j z^j = V(z)A(z)$$

and,

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

We consider the one-step transition probabilities of (L_n, J_n).

(1) If $X_n = (m,1)$, $m \geqslant 2$: at the next departure epoch, the remaining customers are those $m-1$ customers and new arriving ones during the service time S_b, thus,

$$X_{n+1} = (m-1+j, 1) \quad \text{with probability } a_j, \ m \geqslant 2, \ j \geqslant 0.$$

(2) If $X_n = (1,1)$, two cases happen. Firstly, at the $(n+1)$ th departure epoch, the remaining customers are those that have arrived during the normal service time S_b; on the other hand, no customers arrive during the normal service time S_b and the system becomes empty. Thus,

$$X_{n+1} = \begin{cases} (j,1) & \text{with probability } a_j, \quad j \geqslant 1; \\ (0,0) & \text{with probability } a_0. \end{cases}$$

(3) If $X_n = (m,0)$, $m \geqslant 2$: there also are two cases at the next departure epoch. Case i: if $V > S_v$, after this departure, the system stay in the vacation period and the remaining customers are those $m-1$ customers and new arriving customers during the service time S_v; Case ii: if $V \leqslant S_v$, next departure happens during the normal service period and the remaining customers are those $m-1$ customers and new arriving customers during the vacation time V and the service time S_b. From the definition of b_j and c_j, we have

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

(4) If $X_n = (m,0)$, $m = 0, 1$, by the similar analysis, we have

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

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

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

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

It follows from Eq. (1) that

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

where $\boldsymbol{e} = (1,1)^T$ and T represents 'transpose operation'. The stochastic matrix $\widetilde{\boldsymbol{P}}$ is an M/G/1-type matrix (see in Neuts, 1989). For such a model, the minimal nonnegative solution of the equation $\boldsymbol{G} = \sum_{i=0}^{\infty} \boldsymbol{A}_i \boldsymbol{G}^i$ is important. To obtain the minimal nonnegative solution, one lemma is needed.

Lemma 1. *If $\rho = \lambda/\mu_b < 1$, the equation $z = G_b(1-\lambda(1-z))$ has the minimal nonnegative root $z = 1$ and the equation $z = G_v((1-\theta)(1-\lambda(1-z)))$ also has the unique root in the range $0 < z < 1$.*

Proof. First, we consider the equation $z = G_b(1-\lambda(1-z))$. Let $\psi(z) = G_b(1-\lambda(1-z))$ and $0 < \psi(0) = G_b(1-\lambda) < \psi(1) = 1$. And, for $0 < z < 1$,

$$\psi'(z) = \lambda G_b'(1-\lambda(1-z)) > 0; \qquad \psi''(z) = \lambda^2 G_b^{(2)}(1-\lambda(1-z)) > 0.$$

Meanwhile, from $\rho = \lambda/\mu_b < 1$, $\psi'(1) = \rho < 1$. Thus, the equation $z = \psi(z)$ has the minimal nonnegative root $z = 1$. We set $\varphi(z) = z - G_v((1-\theta)(1-\lambda(1-z)))$, and have

$$\varphi(0) = -G_v((1-\theta)(1-\lambda)) < 0, \quad \varphi(1) = 1 - G_v(1-\theta) > 0.$$

For $0 < z < 1$, $y = G_v((1-\theta)(1-\lambda(1-z)))$ is an increasing function, thus the equation $z = G_v((1-\theta)(1-\lambda(1-z)))$ has a unique root in the range $0 < z < 1$. □

Lemma 2. *If $\rho = \lambda/\mu < 1$ 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 $0 < z < 1$ of the equation $z = G_v((1-\theta)(1-\lambda(1-z)))$.

Proof. Because all A_i are upper triangular, we can assume that the minimal nonnegative solution G has the same structure as

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

For $i \geqslant 1$, 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 G^i into the matrix equation, we obtain

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

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

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

Proof. Because

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

is a reducible stochastic matrix. With the notation of the equation (2.3.18) in Neuts (1989), $A(2) = 1$ is the degenerative stochastic matrix, and has the degenerative stationary distribution $\pi(2) = 1$. On the other hand, $A_i(2) = a_i, i \geqslant 0$, and $\boldsymbol{\beta}(2) = \sum_{i=0}^{\infty} i\boldsymbol{A}_i(2) = \rho$. Thus, with Theorem 2.3.3 in Neuts (1989), the Markov chain $\widetilde{\boldsymbol{P}}$ is positive recurrent if and only if $\pi(2)\beta(2) = \rho < 1$. □

3. Queue length analysis

Our next objective is to study the steady-state distribution of the Markov chain $\{X_n, n \geqslant 1\}$. Assume (L, J) be the stationary limit of $X_n = (L_n, J_n)$, and the stationary distribution is defined by

$$\pi_{kj} = \mathrm{P}\{L = k, J = j\} = \lim_{n\to\infty} \mathrm{P}\{L_n = k, J_n = j\}, \quad (k,j) \in \Omega,$$
$$\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \cdots), \quad \boldsymbol{\pi}_0 = \pi_{00}; \quad \boldsymbol{\pi}_k = (\pi_{k0}, \pi_{k1}), \quad k \geqslant 1.$$

The evolution of the chain is governed by the one-step transiton probabilites given by (1), then the Kolmogorov equations $\boldsymbol{\pi}\widetilde{\boldsymbol{P}} = \boldsymbol{\pi}$ together with the normalization condition for the stationay distributions are given by

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

3.1. Probability generating function

To resolve the system of Eq. (3), we introduce the row vector generating function

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

From the second equation in (3), it can be written as

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

where

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

Furthermore,

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

and

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

We easily obtain the expression for $(z\boldsymbol{I} - \boldsymbol{A}^*(z))^{-1}$ below:

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

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

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

After some computation,

$$\begin{aligned} \Phi(z) &= z(\pi_{00}B(z) - u, \quad u - \pi_{00}(1 - C(z)))(z\boldsymbol{I} - \boldsymbol{A}^*(z))^{-1} \\ &= z\left(\frac{\pi_{00}B(z) - u}{z - B(z)}, \frac{C(z)(\pi_{00}B(z) - u)}{(z - B(z))(z - A(z))} + \frac{u - \pi_{00}(1 - C(z))}{z - A(z)}\right). \end{aligned} \tag{4}$$

We get

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

The expression for the PGF $L(z) = \pi_{00} + \Phi(z)\boldsymbol{e}$ of the stationary queue length at the departure epoch can be derived by the followings:

$$\frac{\pi_{00}(z-B(z))(z-A(z))+z[\pi_{00}(B(z)(1+z-A(z))-z(1-C(z)))+u(A(z)-B(z)-C(z))]}{(z-B(z))(z-A(z))}. \tag{5}$$

From Lemma 1, γ is the root of the equation $z = G_v(1-\theta)(1-\lambda(1-z)) = B(z)$. Regularity demands that the numerator of (5) equals 0 for $z = \gamma$. Therefore, substituting $z = \gamma$ into the numerator of (5), and noting $B(\gamma) = \gamma$, one gets

$$\pi_{00}\gamma[1+\gamma-A(\gamma)-(1-C(\gamma))]+u(A(\gamma)-\gamma-C(\gamma))=0. \tag{6}$$

Compute

$$A(\gamma)-\gamma-C(\gamma)=A(\gamma)-\gamma-A(\gamma)V(\gamma)=A(\gamma)(1-V(\gamma))-\gamma,$$
$$\gamma[1+\gamma-A(\gamma)-(1-C(\gamma))]=\gamma(\gamma-A(\gamma)(1-V(\gamma))).$$

Substituting two equations into Eq. (6), we can obtain

$$u=\pi_{00}\gamma.$$

It follows from Eq. (5) that

$$\begin{aligned}L(z) &= \pi_{00}+\Phi(z)\boldsymbol{e}\\ &= \pi_{00}\frac{(z-B(z))(z-A(z))+z[B(z)(1+z-A(z))-z(1-C(z))+\gamma(A(z)-B(z)-C(z))]}{(z-B(z))(z-A(z))}\\ &= \pi_{00}\frac{A(z)(1-z)(B(z)-z)+z(r-z)(A(z)-B(z)-C(z))}{(z-B(z))(z-A(z))}.\end{aligned} \tag{7}$$

Using the normalizing condition $L(1) = 1$,

$$\pi_{00}=\frac{(1-\rho)(1-G_v(1-\theta))}{1-G_v(1-\theta)-(1-\gamma)\left(\rho G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\right)}. \tag{8}$$

Theorem 2. *The stationary queue length L at the departure epoch has the following generating function*

$$L(z)=(1-\rho)\delta\frac{A(z)(1-z)(B(z)-z)+z(r-z)(A(z)-B(z)-C(z))}{(z-B(z))(z-A(z))}, \tag{9}$$

where A(z), B(z) and C(z) are denoted in Section 2, *and*

$$\delta=\frac{1-G_v(1-\theta)}{1-G_v(1-\theta)-(1-\gamma)\left(\rho G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\right)}.$$

The steady-state queue length at the arbitrary epoch has the same distribution with that at the departure epoch, in other words, PASTA (Poisson arrivals see time average) property also exists in the discrete-time Geo/G/1 queue with working vacations. This result has been precisely explained in Li et al. (2009).

Moreover, we can derive the state probabilities of the system in the steady state.

$$\begin{aligned}P\{J=0\} &= \pi_{00}+\Phi(z)\boldsymbol{e_1}|_{z=1}=\pi_{00}+\pi_{00}\frac{G_v(1-\theta)-\gamma}{1-G_v(1-\theta)}\\ &= \frac{(1-\rho)(1-\gamma)}{1-G_v(1-\theta)-(1-\gamma)\left(\rho G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\right)},\\ P\{J=1\} &= \Phi(z)\boldsymbol{e_2}|_{z=1}=1-P\{J=0\}\\ &= \frac{1-G_v(1-\theta)-(1-\gamma)\left(1-\rho+\rho G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\right)}{1-G_v(1-\theta)-(1-\gamma)\left(\rho G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\right)}\end{aligned} \tag{10}$$

where $\boldsymbol{e}_1 = (1,0)^T$, and $\boldsymbol{e}_2 = (0,1)^T$.

3.2. Stationary probability vector $\boldsymbol{\pi}$

With the results in Neuts (1989), we can generate the recursive equations of $\boldsymbol{\pi}_k$, $k \geqslant 0$ by the following steps.

Step 1: Calculation of $\boldsymbol{\pi}_1$. From the relationship $b_0(\pi_{00} + \pi_{10}) = -\gamma\pi_{00}$, we have

$$\pi_{10}=\frac{\gamma-b_0}{b_0}\pi_{00}.$$

On the other hand, from the first equation in (3),

$$\pi_{00}=\pi_{00}(b_0+c_0)+\pi_{00}\frac{\gamma-b_0}{b_0}(b_0+c_0)+\pi_{11}a_0,$$

Compute

$$\begin{aligned}\pi_{11}a_0 &= \pi_{00}\left(1-b_0-c_0-\frac{\gamma-b_0}{b_0}(b_0+c_0)\right)\\ &= \pi_{00}\left(1-\frac{\gamma}{b_0}(b_0+c_0)\right),\end{aligned}$$

and π_{11} can be obtained by the following equation:

$$\pi_{11}=\frac{1}{a_0}\frac{b_0-\gamma b_0-\gamma c_0}{b_0}\pi_{00}=\frac{b_0(1-\gamma)-\gamma c_0}{a_0b_0}\pi_{00}.$$

The expression for $\boldsymbol{\pi}_1$ is given by

$$\boldsymbol{\pi}_1=(\pi_{10},\pi_{11})=\pi_{00}\left(\frac{\gamma-b_0}{b_0},\quad\frac{b_0(1-\gamma)-\gamma c_0}{a_0b_0}\right).$$

Remark 1. To give the calculation of π_0, π_1, Neuts (1989) introduced complicated stochastic matrices H and K. (For more details, see Neuts, 1989, pp.25). However, without using H and K, we can use the simple method to compute π_1 in the Geo/G/1 queue with working vacations.

Step 2. The recursive equations of $\boldsymbol{\pi}_k$, $k \geqslant 2$. From Theorem 3.2.5 in Neuts (1989),

$$\boldsymbol{\pi}_k=\left[\pi_{00}\overline{\boldsymbol{B}}_k+\sum_{j=1}^{k-1}\boldsymbol{\pi}_j\overline{\boldsymbol{A}}_{k+1-j}\right](\boldsymbol{I}-\overline{\boldsymbol{A}}_1)^{-1},\qquad k\geqslant 1, \tag{11}$$

where

$$\overline{\boldsymbol{B}}_r=\sum_{i=r}^{\infty}\boldsymbol{B}_i\boldsymbol{G}^{i-r},\qquad \overline{\boldsymbol{A}}_r=\sum_{i=r}^{\infty}\boldsymbol{A}_i\boldsymbol{G}^{i-r}.$$

$\{\overline{\boldsymbol{B}}_r, r\geqslant 1\}$ are a series of two-dimensional row vectors, and $\{\overline{\boldsymbol{A}}_r, r\geqslant 1\}$ two-dimensional matrices. For $k = 1$, Eq. (11) can be verified directly.

4. Conditional stochastic decomposition

In the previous section, we obtain the expression for the PGF of the queue length at the departure epoch. But, as in other papers, expression (9) has no certain implication and probability explanation, so that we cannot establish the relationship between the Geo/G/1(MWV) and the standard Geo/G/1 queues, especially for the stochastic decomposition structure which exists in the standard queue with vacations and can be seen in the key papers Fuhrmann and Cooper (1985), Shanthikumar (1988). In this section, we mainly use the stochastic decomposition method in Shanthikumar (1988) to obtain the concise and definite expression for $L(z)$.

Though Fuhrmanm and Cooper[8], Shanthikumar (1988) considered the M/G/1 queues with general (nonworking) vacations, in the model we analyze, under the condition that a customer is served by the normal rate, the operating process is stochastically equivalent to that of the M/G/1 queue with general (nonworking) vacations. Because we can obtain the stationary distribution of the number of customers at the busy period beginning instant, the results in Shanthikumar (1988) are used to analyze the normal service period in the M/G/1 queue with working vacations.

As in Shanthikumar (1988), the service time of every customer in a regular busy period is called an active period and the length of vacation or service interruptions is called an inactive period. The server can work during the inactive period and alternates between active and inactive states. We allow the inactive period to have zero length. Using the definition of conditional probability, we have

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

where $S = 0(1)$ represents the event that an arbitrary customer is served completely by the working vacation (normal) service rate, and:

L^s: the number of the customers at the starting instant of an inactive period in the steady state;
L^T: the number of the customers at the ending instant of an inactive period in the steady state;

$L^s(z)(L^T(z))$: the corresponding PGF.

From the results in Section 3, we derive these important probabilities:

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

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

where $P_v \neq \mathrm{P}\{J=0\}$ and $P_b \neq \mathrm{P}\{J=1\}$.

We first obtain the distribution of the number of the customers Q_b at the busy period beginning instant, in other words, the working vacation period ending instant. Introduce the last service completion before the busy period beginning instant as an embedded point, then $Q_b = k$ can be caused by two cases, case 1: under the condition the system stays in the vacation period after the last service and the vacation time is not longer than S_v, i.e., $V \leqslant S_v, j (j \geqslant 1)$ customers are left and $k-j$ customers arrive during the vacation time V; case 2: under the same condition, no customers are left 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 instant of beginning of the busy period and $V \leqslant S_v$ is related by

$$\begin{aligned}\mathrm{P}\{J=0, V\leqslant S_v\} &= \sum_{i=0}^{\infty}\pi_{i0}\times \mathrm{P}\{V\leqslant S_v\} = \mathrm{P}\{J=0\}(1-G_v(1-\theta))\\ &= \pi_{00}(1-\gamma).\end{aligned}$$

The distribution of Q_b is derived following:

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

The generating function for τ_k, $k \geqslant 1$ is

$$\begin{aligned}Q_b(z) = \sum_{k=1}^{\infty}\tau_k z^k &= \frac{1}{\pi_{00}(1-\gamma)}\left\{\sum_{j=1}^{\infty}\pi_{j0}z^j\sum_{k=j}^{\infty}v_{k-j}z^{k-j} + \pi_{00}\sum_{k=1}^{\infty}v_{k-1}z^k\right\}\\ &= \frac{1}{\pi_{00}(1-\gamma)}\{\Phi(z)\boldsymbol{e_1}V(z) + \pi_{00}zV(z)\}\\ &= \frac{1}{1-\gamma}\left\{\frac{z(B(z)-\gamma)}{z-B(z)} + z\right\}V(z)\\ &= \frac{1}{1-\gamma}\frac{z(z-r)V(z)}{z-B(z)}.\end{aligned} \tag{13}$$

We obtain

$$\begin{aligned}E(Q_b) &= \frac{1-G_v(1-\theta)-(1-\gamma)\big(G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\big)}{(1-\gamma)(1-G_v(1-\theta))}\\ &= \frac{P_b(1-\rho)}{\pi_{00}(1-\gamma)}.\end{aligned}$$

Theorem 3. *The PGF of the stationary queue length L at the departure epoch has the expression*

$$\begin{aligned}L(z) = P_v\frac{1-G(1-\theta)}{(1-\gamma)G(1-\theta)}\frac{B(z)(z-\gamma)}{z-B(z)} + P_b\frac{(1-\rho)(1-z)A(z)}{A(z)-z}\\ \times\frac{1-Q_b(z)}{E(Q_b)(1-z)}.\end{aligned} \tag{14}$$

Proof. From (12),

$$\begin{aligned}E\{z^L|S=0\} &= P_v^{-1}((\pi_{00}+\pi_{10})b_0 + \Phi(z)\boldsymbol{e_1})\\ &= P_v^{-1}\pi_{00}\frac{B(z)(z-r)}{z-B(z)} = \frac{1-G_v(1-\theta)}{(1-\gamma)G_v(1-\theta)}\frac{B(z)(z-\gamma)}{z-B(z)}.\end{aligned} \tag{15}$$

In a discrete-time Geo/G/1 queue with working vacations, we only choose the departure instants of customers in a regular busy period as embedded points, and the arrival and departure of customers in a working vacation period only affect the deviation $L^s - L^T$. Therefore, applying the results of Shanthikumar (1988) for the first term on the right-hand side in (12), we obtain

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

where N is the number of customers in a standard Geo/G/1 queue in the steady state under AF policy, and the PGFs of N and the additional variable X are

$$E(z^N) = \frac{(1-\rho)(1-z)A(z)}{A(z)-z}, \qquad E(z^X) = \frac{L^s(z)-L^T(z)}{(1-\rho)(1-z)}. \tag{16}$$

The important problem below is to solve $E(z^X)$. Firstly, consider $L^s(L^T)$. Note that $L^T = k (k \geqslant 1)$ includes two disjoint cases: (1) $L^s = k$, if there is an inactive period of zero length in two successive active periods. (2) $L^s = 0$ and there are k customers in the system when a working vacation ends. From the definition and results of Q_b above, we have

$$\mathrm{P}\{L^T=k\} = \mathrm{P}\{L^s=k\} + \mathrm{P}\{L^s=0\}\tau_k, \quad k\geqslant 1, \tag{17}$$

and $\mathrm{P}\{L^T = 0\} = 0$. Multiplying z^k for two sides of (17) and summing for all k, after some transform, we get

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

From (16),

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

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

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

The expression of (14) is derived from (12). □

Remark 2. We can easily prove that Eq. (14) is equivalent to (9). As stated above,

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

It can be demonstrated from Eq. (15) that

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

We only need to verify the equation

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

Substituting the expression for $Q_b(z)$ into the left side and with the relationship $E(Q_b) = P_b(1-\rho)(\pi_{00}(1-\gamma))^{-1}$, we obtain

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

By the similar computing, we can verify that Eqs. (14) and (9) are equivalent.

The average number of customers in the queue is given by

$$E(L) = P_v\left\{\frac{\beta}{G_v(1-\theta)} + \frac{1}{1-\gamma} - \frac{1-\beta}{1-G_v(1-\theta)}\right\} + P_b\left\{\rho + \frac{\lambda^2 E(S_b(S_b-1))}{2(1-\rho)} + \frac{E(Q_b(Q_b-1))}{2E(Q_b)}\right\}.$$

Denote L_w the number of the left customers in the system after a service completion, under the condition that such customer is served completely by the normal service rate, i.e., $S = 1$. For this conditional queue length L_w at the departure epoch, the conditional stochastic decomposition exists.

Theorem 4. *The conditional queue length L_w can be decomposed into two independent random variables: $L_w = L_0 + L_d$, where L_0 is the stationary queue length of a standard Geo/G/1 queue without vacations under AF policy, and the additional queue length L_d has the probability generating function below:*

$$L_d(z) = \frac{1-Q_b(z)}{E(Q_b)(1-z)}.$$

From the expression of (14), we have the results in some special cases:

- If we assume that there is no working during the vacation period, i.e., $\mu_v \to 0$, the model reduces to the Geo/G/1 queue with the classic vacation. In this case, $P_v = 0$, $P_b = 1$, and the PGF and mean the queue length at the departure epoch are agreement in the results in Tian and Zhang (2006).
- If the service time during the vacation period follows the same distribution with the time during the normal period, i.e., $\mu_v = \mu_b$, the model becomes the Geo/G/1 queue without vacations, which has been analyzed in Takagi (1993) and Tian and Zhang (2006).

5. Waiting time analysis

Because each customer is possible to be served during the normal period or the vacation period, the service times are different under such two periods. We need to introduce some notations:

W: the waiting time of any customer in the Geo/G/1(MWV) with its PGF $W(z)$;
W_0: the waiting time of a standard Geo/G/1 queue without vacations under AF policy;
W_p: the waiting time of one customer, under the condition that such customer is served by the normal service rate;
W_v: the waiting time of one customer, under the condition that such customer is served by the vacation service rate;
$W_0(z)$, $W_p(z)$, $W_v(z)$: the corresponding PGFs.

We have

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

For one customer who is served during the busy period, the remaining customers at the departure instant of this customer are those who arrive during his waiting time W_p and normal service time S_b. The relationship

$$E\{z^L|S=1\} = W_p(1-\lambda(1-z))G_b(1-\lambda(1-z))$$

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

$$W_p(z) = \frac{(1-\rho)(1-z)}{1-z-\lambda(1-G_b(z))} \frac{\lambda\left(1-Q_b\left(\frac{z-1}{\lambda}+1\right)\right)}{E(Q_b)(1-z)} = W_0(z)W_d(z).$$

Theorem 5. *The conditional waiting time W_p can be decomposed into the sum of two independent random variables: $W_p = W_0 + W_d$, where the additional delay W_d has the PGF*

$$W_d(z) = \frac{\lambda\left(1-Q_b\left(\frac{z-1}{\lambda}+1\right)\right)}{E(Q_b)(1-z)}.$$

By the similar reasoning as W_v, we get the following relationship

$$E\{z^L|S=0\} = W_v(1-\lambda(1-z))G_v(1-\lambda(1-z)).$$

Taking $s = 1-\lambda(1-z)$ gives

$$W_v(z) = \frac{1-G_v(1-\theta)}{(1-\gamma)G_v(1-\theta)} \frac{G_v((1-\theta)z)(1-z-\lambda(1-\gamma))}{1-z-\lambda(1-G_v((1-\theta)z))} \frac{1}{G_v(z)}.$$

The expression for the PGF of the waiting time W is

$$\begin{aligned} W(z) &= P_vW_v(z) + P_bW_p(z) \\ &= P_v\frac{1-G_v(1-\theta)}{(1-\gamma)G_v(1-\theta)} \frac{G_v((1-\theta)z)(1-z-\lambda(1-\gamma))}{1-z-\lambda(1-G_v((1-\theta)z))} \\ &\quad\times \frac{1}{G_v(z)} + P_b\frac{(1-\rho)(1-z)}{1-z-\lambda(1-G_b(z))} \\ &\quad\times \frac{\lambda\left(1-Q_b\left(\frac{z-1}{\lambda}+1\right)\right)}{E(Q_b)(1-z)}. \end{aligned} \tag{18}$$

The average waiting time in the queue is given by

$$E(W) = P_v\left\{\frac{1}{\lambda}\left[\frac{\beta}{G_v(1-\theta)} + \frac{1}{1-\gamma} - \frac{1-\beta}{1-G_v(1-\theta)}\right] - \frac{1}{\mu_v}\right\} + P_b\left\{\frac{\lambda E(S_b(S_b-1))}{2(1-\rho)} + \frac{E(Q_b(Q_b-1))}{2\lambda E(Q_b)}\right\}.$$

Remark 3. We can obtain the results of the steady-state waiting time in two special models presented in Section 4, which are agreement in those in Takagi (1993) and Tian and Zhang (2006).

6. Busy period analysis

In this section, we present the busy period analysis and denote some notation below:

D_v: a regular busy period which means the duration in which the server works at the normal rate continuously;
V_g: a whole vacation period which means the duration in which the server takes vacations continuously;
C: a busy cycle which is composed of a working vacation period V_g, and subsequent regular busy period D_v;
D: the regular busy period in the standard Geo/G/1 queue without vacations under the AF policy.

A regular busy period D_v starts at the instant when a working vacation finishes, and ends at the instant that the system becomes empty and enters into vacation. A whole vacation period V_g begins at the instant that the system enters into vacation, and ends at the instant when a working vacation finishes. During a vacation period V_g, if there are customer arrivals, the server will work at the vacation service rate.

We assume $D(z)$ the PGF of the regular busy period in the standard Geo/G/1 queue without vacations under the AF policy. As we all know, $D(z)$ satisfies the equation

$$D(z) = E\{Z^D\} = G_b[z(1-\lambda(1-D(z)))],$$

and

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

In our model, when there are k customers at the instant when a working vacation ends, i.e., $Q_b = k$, due to the memoryless property of the arrival process, this regular busy period D_v is the k-dimensional convolution of D, in other words, $\{D_v|Q_b = k\} = D^{(k)}$.

Denoting the PGF of the regular busy period D_v by $D_v(z)$, we have

$$D_v(z) = E\{z^{D_v}\} = \sum_{k=1}^{\infty} E\{z^{D_v}|Q_b = k\}\mathrm{P}\{Q_b = k\} = \sum_{k=1}^{\infty}(E\{Z^D\})^k \tau_k = Q_b(D(z)).$$

Further, the expected regular busy period $E(D_v)$ can be given by

$$E(D_v) = E(Q_b)E(D) = \frac{1-G_v(1-\theta)-(1-\gamma)\left(G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\right)}{\mu_b(1-\rho)(1-\gamma)(1-G_v(1-\theta))}.$$

The relationships

$$\mathrm{P}\{J=1\} = \frac{E(D_v)}{E(C)} = \frac{E(D_v)}{E(D_v)+E(V_g)}, \quad \mathrm{P}\{J=0\} = \frac{E(V_g)}{E(C)} = \frac{E(V_g)}{E(D_v)+E(V_g)}$$

exist. We can easily compute the expected busy cycle $E(C)$ and working vacation period $E(V_g)$:

$$E(V_g) = \frac{(1-\rho)(1-\gamma)}{1-G_v(1-\theta)-(1-\gamma)\left(1-\rho+\rho G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\right)} \times \frac{1-G_v(1-\theta)-(1-\gamma)\left(G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\right)}{\mu_b(1-\rho)(1-\gamma)(1-G_v(1-\theta))}$$

$$E(C) = \frac{1-G_v(1-\theta)-(1-\gamma)\left(\rho G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\right)}{1-G_v(1-\theta)-(1-\gamma)\left(1-\rho+\rho G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\right)} \times \frac{1-G_v(1-\theta)-(1-\gamma)\left(G_v(1-\theta)-\frac{\lambda}{\theta}(1-\theta)(1-G_v(1-\theta))\right)}{\mu_b(1-\rho)(1-\gamma)(1-G_v(1-\theta))}.$$

7. Numerical results

To demonstrate the applicability of the results obtained in the previous sections, some numerical results have been presented in the form of tables and graphs. Numerical results show the behavior of some performance measures against the variation of a critical model parameters. All the results are given in one type of special Geo/Geo/1 models with geometric service time S_b with parameter $\mu_b = 0.6$. We need two recommended parameters:

- $E(W^G)$: the stationary waiting time in the Geo/Geo/1 queue with classic (non-working) vacations.
- $E(W_0)$: the stationary waiting time in the Geo/Geo/1 queue without vacations.

Firstly, we consider a model with the arrival rate $\lambda = 0.3$, vacation rate $\theta = 0.55$, the geometric S_v with parameter $\mu_v(0 < \mu_v < 0.6)$ and this model is noted as $Geo/(Geo^1,Geo^2)/1$. Table 1 shows a series of steady-state performance measures for some special values of μ_v in the range of *(0,0.6)*. We have $E(W^G) = 1.4848$, and $E(L)$, $E(W)$ and $E(D_v)$ all decrease with corresponding increase of the vacation service rate μ_v. We notice the difference between $E(W)$ and $E(W^G)$, in that the waiting time in working vacation queue will be declined and the extent of decreasing will be larger.

Further, to present the effects of the variation of some critical parameters on performance measures, we consider a same $Geo/(Geo^1,Geo^2)/1$ model for the numerical results in Table 1, Figs. 1 and 2 to provide the continuity for the results analysis. Fig. 1 shows the expected queue length $E(L)$ and waiting time $E(W)$ against $\mu_v(0 < \mu_v < 0.6)$. $E(L)$ and $E(W)$ are almost linearly decreases to a fixed value in the given range of μ_v. Furthermore, for a fixed μ_v, $E(L)$ decreases as the vacation rate θ increases. When the system

Table 1
Steady-state performance measures in the system for $Geo/(Geo^1,Geo^2)/1/WV$ queue with parameters: $\lambda = 0.3$, $\mu_b = 0.6$, $\theta = 0.55$, and with $E(W^G) = 1.4848$ in $Geo/Geo/1/MV$.

μ_v	γ	$E(L)$	P_v	$E(W)$	$E(W) - E(W^G)$	$E(D_v)$
0.10	0.0072	0.9301	0.0081	1.3661	−0.1188	4.1083
0.15	0.0149	0.9156	0.0163	1.3040	−0.1809	4.0674
0.20	0.0248	0.8989	0.0261	1.2426	−0.2423	4.0204
0.25	0.0363	0.8809	0.0372	1.1827	−0.3022	3.9701
0.30	0.0490	0.8620	0.0492	1.1248	−0.3600	3.9181
0.35	0.0629	0.8428	0.0618	1.0692	−0.4157	3.8656
0.40	0.0776	0.8235	0.0750	1.0159	−0.4690	3.8133
0.45	0.0929	0.8043	0.0885	0.9650	−0.5199	3.7619
0.50	0.1087	0.7852	0.1024	0.9165	−0.5683	3.7116
0.55	0.1248	0.7664	0.1164	0.8705	−0.6144	3.6627

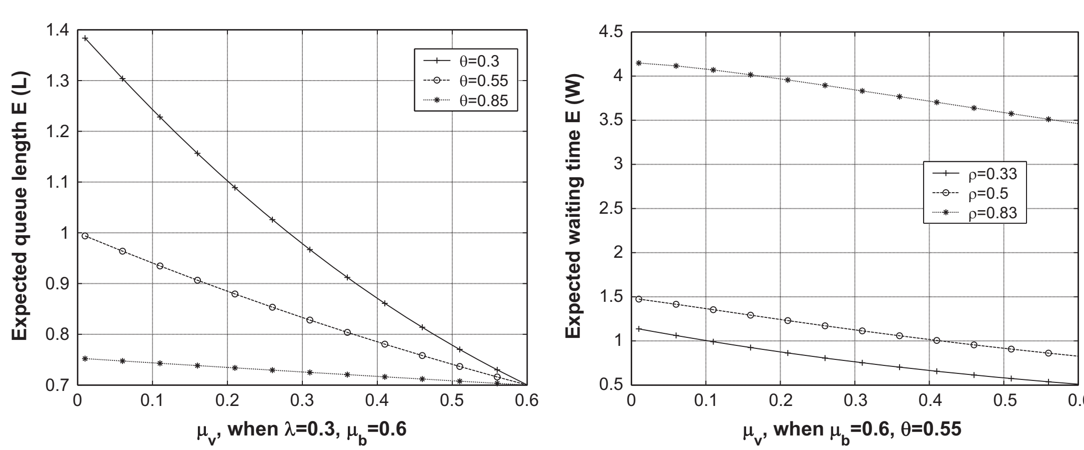

Fig. 1. $E(L)$ and $E(W)$ versus μ_v in $Geo/(Geo^1,Geo^2)/1$, respectively.

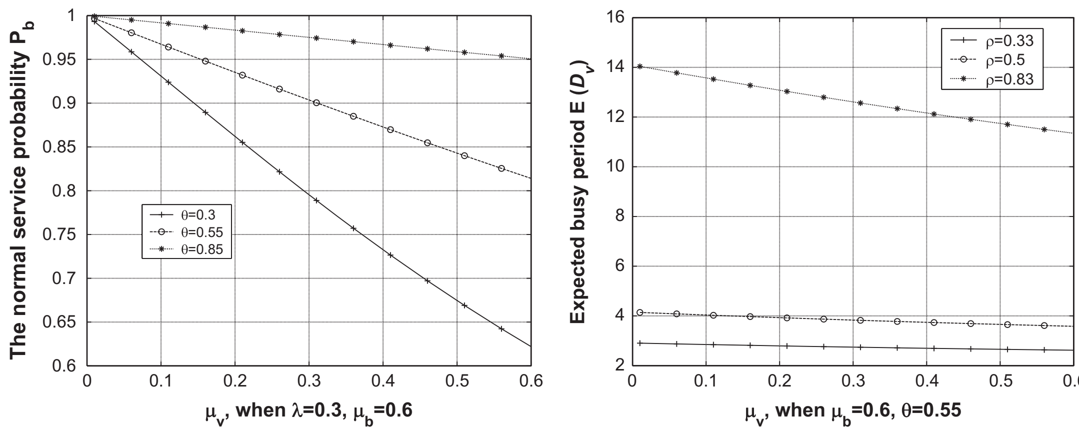

Fig. 2. P_b and $E(D_v)$ versus μ_v in $Geo/(Geo^1,Geo^2)/1$, respectively.

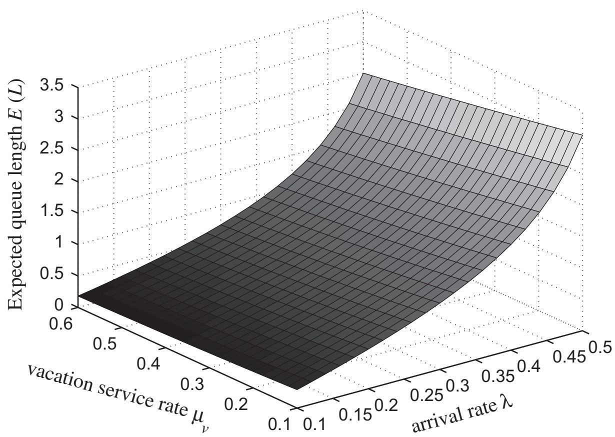

Fig. 3. The expected queue length $E(L)$ in Negative Binomial vacation service model.

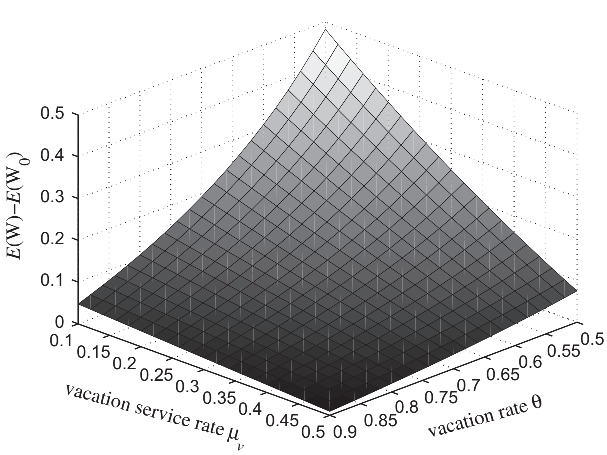

Fig. 4. The difference of $E(W)$ and $E(W_0)$ in Negative Binomial vacation service model.

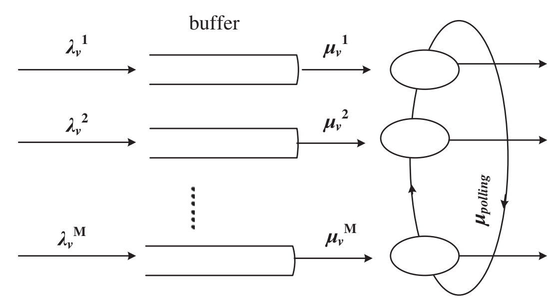

Fig. 5. The working vacation polling system.

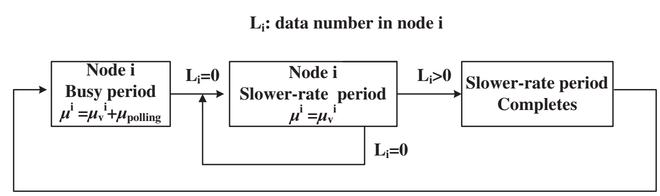

Fig. 6. Schematic representation of the queueing model.

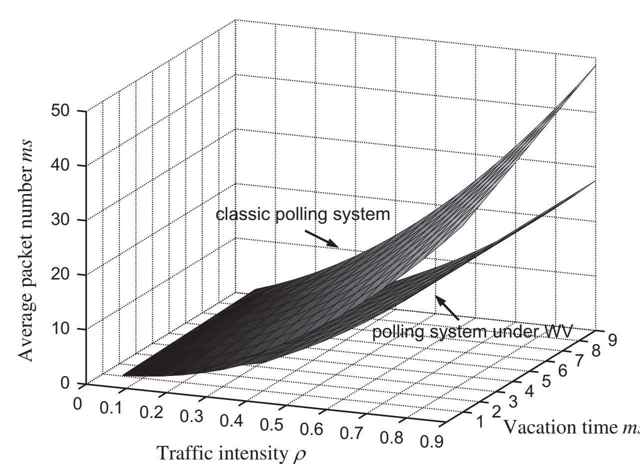

Fig. 7. Average packet number in the polling system.

has the loaded traffic (for example, $\rho = 0.83$), $E(W)$ is larger than that in unloaded traffic (for example, $\rho = 0.33$ or $\rho = 0.5$). Meanwhile, in Fig. 2, we have plotted P_b and the expected busy period $E(D_v)$ against μ_v. When $\mu_v = 0$, i.e., no service exists during the vacation, $P_b = 1$ means that each service is completed by the normal rate μ_b. With the increase of the μ_v, this probability decreases and part of those customers or jobs who arrive during the vacation period may be processed in the vacation period, thus P_b decreases. If keeping a fixed μ_v and θ in a loaded system, P_b will become

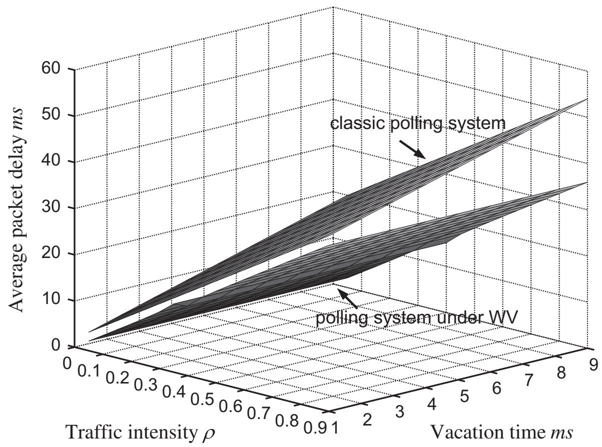

Fig. 8. Average packet delay in the polling system.

larger. In Figs. 1 and 2, the same θ (=0.55) is used in the expected waiting time and busy period charts to provide the continuity for the result analysis, which also show the coherence with the results in Table 1.

Finally, in Figs. 3 and 4, we have presented another models where S_v follows Negative Binomial distribution NB (r,p) where $r = 2$, $p = \mu_v (0 < \mu_v < 0.6)$. In Fig. 4, to present the effect of the working vacation on the waiting time, we show the difference of $E(W)$ and $E(W_0)$. We have run this experiment for $E(L)$ and $E(W)$ against two parameters and observe almost similar behavior as described for the above figures.

8. Application to network scheduling

Various applications based on traffic in the wavelength division-multiplexed (WDM) system need effective control protocols and efficient algorithms to support quality of service (QoS). Tremendous traffic demand in all optical networks needs an appropriate optical control layer supported with configurability, restorability, bandwidth utilization and node accessibility. Several approaches have been proposed to manage the optical traffic through WDM network involving optical system, where using a queue scheduling algorithm is one of the core mechanisms to achieve the network intermediate nodes (routers and switches) for the QoS control. By scheduling the transmission rates of the wavelength, QoS in each node will be achieved. One of the solution methods of intermediate nodes in multimedia service access network is the polling scheduling approach, where a token will operate in a cycle mode from one node to another node. This scheme is widely applied in computer and communication networks, but data in some nodes may not be transmitted timely and make them hungry. To overcome this disadvantage, based on the working vacation scheme, we propose a cyclic polling model in WDM by reserving wavelength transmission ability.

The cyclic polling system consists of M nodes (as queues) or terminals (all of the queues have infinite capacity). In the working vacation polling system, node i has permanent wavelengths assigned to it corresponding to the capability of being serviced at a reserved rate μ_v^i. The polling wavelengths serve at an additional rate of $\mu_{polling}$, so that the total service rate is $\mu_b^i = \mu_v^i + \mu_{polling}$. Inter-arrival times of data packets to node $i (i = 1,2,M)$ are an independently and identically distributed sequence and follow a geometrical distribution with the parameter λ_i. The reserved service of node i has geometrical distribution with the parameter μ_v^i. The polling time follows the geometrical distribution with the parameter θ and the service rate $\mu_v = \sum_{i=1}^M \mu_v^i = 0.85, \mu_{polling} = 0.15$. This polling scheme in the WDM system can be modeled as Fig. 5.

The multiple wavelengths serving a node can be engineered to indeed perform as if they are a single server. Using the decomposition approach, we decompose this polling service system into a single server working vacation model, as seen in Fig. 6.

In a symmetry polling system with 5 nodes, we consider a system where data packets have a fixed size of 1000 bits and the transmission speed is 1 Gb/s (hence, the transmission time of a packet is 1 μs). We have plotted the average packet number and delay in the classic system without polling wavelength (i.e., $\mu_{polling} = 0$) and the working vacation polling system with described parameters of Figs. 7 and 8. One may observe from Fig. 7 that average data number in each node under the working vacation polling scheme is smaller than that in the classic polling system whenever the values of ρ and θ are fixed. Fig. 8 presents almost similar behavior as described in the above Fig. 7. It is illustrated that the performance of the WDM is improved considerably under this polling service based on the working vacation scheme.

As a final remark, we observe that our numerical results have been illustrated for three specific choices of the service time distribution and those performance measures we have obtained in the previous sections can be used to reasonably represent some practical problems.

9. Conclusion

This paper suggests the system framework for a Geo/G/1 queue with working vacations. The results establish the evident relationship between the Geo/G/1 queue with regular vacations and the Geo/G/1 queue with working vacations. We have found two important methods, the matrix analytic approach and the stochastic decomposition theory, to analyze the working vacation M/G/1-type queues. The analysis method can be extended to other more general models, such as discrete-time Batch Markov Arrival Process (D-BMAP). The paper presents an application to network scheduling in slotted digital communication systems, which is called a cyclic polling model by reserving wavelength transmission ability. The performance analysis of a WDM system would suggest that the working vacation policy can represent various types of problems in practice efficiently. Naturally, the practical utility and further evolution of the models suggested in this paper need to be evaluated in light of actual applications. This points to a general tendency which deserves further study.

We have pointed out several directions in which our basic models can and should be generalized. Of the biggest interest to us is the incorporation of economic costs and customers' behavior. Under the working vacation policy, the server can economize the cost by transforming the service rates, but this may result in the dissatisfaction from the customers, who will choose "to be served" or "to leave". And the effect of intentionally supplied status information to customers is of great importance in practice, and appropriate methods for its incorporation and investigation within the rational model are yet to be explored. All in all, it is apparent that much remains to be done in this area.

Acknowledgements

The authors thank editors and referees for their valuable comments and suggestions, which are very helpful to improve the manuscript. This research is supported by Humanities and Social Science Fund in Ministry of Education in China (No. 10YJC630114).

References

Alfa, A. S. (2003). Vacation models in discrete time. *Queueing Systems, 44*, 5–30.
Baba, Y. (2005). Analysis of a GI/M/1 queue with multiple working vacations. *Operation Research Letters, 33*, 201–209.

Banik, A. D., Gupta, U. C., & Pathak, S. S. (2007). On the GI/M/1/N queue with multiple working vacationsanalytic analysis and computation. *Applied Mathematical Modelling, 31*, 1701–1710.

Bruneel, H., & Kim, B. G. (1993). *Discrete-time models for communication system including ATM*. Boston: Kluwer Academic Publishers..

Fuhrmann, S. W., & Cooper, R. B. (1985). Stochastic decomposition in the M/G/1 queue with generalized vacations. *Operations Research, 33*, 1117–1129.

Hunter, J. J. (1983). *Mathematical techniques of applied probability*. *Discrete time models: Techniques and applications* (Vol. 2). New York: Academic Press.

Kim, J. D., Choi, D. W. & Chae, K. C. (2003). Analysis of queue-length distribution of the M/G/1 queue with working vacations. In *International conference on statistics and related fields*. Hawaii.

Li, J., Liu, W., & Tian, N. (2009). Steady-state analysis of a discrete-time batch arrival queue with working vacations. *Performance Evaluation, 10*, 897–912.

Li, J., & Tian, N. (2007). The discrete-time GI/Geo/1 queue with working vacations and vacation interruption. *Applied Mathematics and Computation, 185*, 1–10.

Li, J., Tian, N., & Liu, W. (2007). Discrete-time GI/Geo/1 queue with working vacations. *Queueing Systems, 56*, 53–63.

Li, J., Tian, N., Zhang, Z. Z., & Luh, H. P. (2009). Analysis of the M/G/1 queue with exponentially working vacations – A matrix analytic approach. *Queueing Systems, 61*, 139–166.

Liu, W., Xu, X., & Tian, N. (2007). Stochastic decompositions in the M/M/1 queue with working vacations. *Operation Research Letters, 35*, 595–600.

Neuts, M. (1981). *Matrix-geometric solutions in stochastic models*. Baltimore: Johns Hopkins University Press.

Neuts, M. (1989). *Structured stochastic matrices of M/G/1 type and their applications*. New York: Marcel Dekker.

Servi, L. D., & Finn, S. G. (2002). M/M/1 queue with working vacations (M/M/1/WV). *Performance Evaluation, 50*, 41–52.

Shanthikumar, J. G. (1988). On stochastic decomposition in M/G/1 type queues with generalized server vacations. *Operations Research, 36*, 566–569.

Takagi, H. (1991). *Queueing analysis: A foundation of performance evaluation*. *Vacation and priority systems, Part 1* (Vol. 1). Amsterdam: North-Holland Elsevier.

Takagi, H. (1993). *Analysis: A foundation of performance evaluation*. *Discrete-time systems* (Vol. 3). Amsterdam: North-Holland Elsevier.

Tian, N., & Zhang, Z. G. (2006). *Vacation queueing models: theory and applications*. New York: Springer-Verlag.

Wu, D., & Takagi, H. (2006). M/G/1 queue with multiple working vacations. *Performance Evaluation, 63*, 654–681.