Performance Evaluation 67 (2010) 897–912

Contents lists available at ScienceDirect

Performance Evaluation

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

Steady-state analysis of a discrete-time batch arrival queue with working vacations

Ji-hong Li [a], Wei-qi Liu [a], Nai-shuo Tian [b,*]

[a] College of Management, Shanxi University, Taiyuan, 030006, China

[b] College of Sciences, Yanshan University, Qinhuangdao, 066004, China

ARTICLE INFO

Article history:
Received 22 March 2008
Received in revised form 21 October 2009
Accepted 22 March 2010
Available online 20 April 2010

Keywords:
Discrete-time queue
Batch arrival
Working vacations
M/G/1-type matrix
Stochastic decomposition
Ethernet Passive Optical Network (EPON)

ABSTRACT

This paper analyzes a discrete-time batch arrival queue with working vacations. In a $\mathrm{Geo}^X/G/1$ system, the server works at a lower speed during the vacation period which becomes a lower speed operation period. This model is more appropriate for the communication systems with the transmit units arrived in batches. We formulate the system as an embedded Markov chain at the departure epoch and by the M/G/1-type matrix analytic approach, we derive the probability generating function (PGF) of the stationary queue length. Then, we obtain the distribution for the number of the customers at the busy period initiation epoch, and use the stochastic decomposition technique to present another equivalent PGF of the queue length. We also develop a variety of stationary performance measures for this system. Some special models and numerical results are presented. Finally, a real-world example in an Ethernet Passive Optical Network (EPON) is provided.

1. Introduction

Discrete time queues with vacations have been widely studied in recent years by many researchers because these systems are more appropriate than their continuous-time counterparts for modeling computer and telecommunication systems since the basic units in these systems are digital such as a machine cycle time, bits and packets, etc. Some references are given by Doshi [1,2], and an excellent and complete study on discrete-time queueing systems with vacations has been presented by Takagi [3] who gave the analysis of Geo/G/1-type queues (including batch arrivals) with different vacation policies. In recent years, Tian and Zhang [4] and Zhang and Tian [20] studied the Geo/G/1 queues with multiple adaptive vacations. The batch arrival discrete-time $\mathrm{Geo}^X/G/1$ queue under multiple vacations governed by a geometrically distributed timer was analyzed by Fiems and Bruneel [5]. Recently, Chang and Choi [6] analyzed a single-server batch arrival bulk-service queue where customers are served in batches of random size and the server takes multiple vacations whenever the queue is empty. Samanta et al. [7] presented the discrete-time $\mathrm{Geo}^X/\mathrm{Geo}^{(a,b)}/1/N$ queues with batch arrival and bulk-service under single and multiple vacation policies. In all these references, it is assumed that the server cannot take service during the vacations.

In this paper, we will consider a batch arrival Geo/G/1 system with working vacations. Working vacation (WV) policy was introduced by Servi and Finn [8] and the server works at a different rate rather than completely stopping service during the vacation period. In this case, the vacation period becomes the lower speed operation period of the queueing system. Servi and Finn [8] analyzed an M/M/1/ queue with working vacations, denoted as M/M/1/WV, and modeled a wavelength division multiplexing (WDM) optical access network using multiple wavelengths which can be reconfigured. Subsequently, Wu and Takagi [9] generalized results in [8] to an M/G/1 queue with working vacations. Baba [10] and Banik et al. [11] presented two

* Corresponding author. Tel.: +86 335 8074703; fax: +86 335 8074703.
E-mail addresses: lijh1982@sxu.edu.cn (J.-h. Li), tiannsh@ysu.edu.cn (N.-s. Tian).

0166-5316/$ – see front matter
doi:10.1016/j.peva.2010.03.001

types of GI/M/1 queues with working vacations. Liu et al. [12] gave the stochastic decomposition structure in the M/M/1/WV, which also exists in the queues with general vacations. Li and Tian [13–15] analyzed a discrete-time Geo/Geo/1 and two types of GI/Geo/1 queues with working vacations, respectively. So far, only three literatures concentrated on the discrete-time working vacation queues, and especially, there was no attempt on the Geo/G/1-type working vacation queues and no uniform methods and results are established.

Thus, we will generalize the analysis of classic and working vacation queues in other literatures to a discrete-time batch arrival Geo/G/1 queue with working vacations. In fact, the working vacation policy is more general than the classical one in the sense that, if the service rate during the vacation period becomes zero, the classic vacation policies are the special cases of the working vacation policies. In practice, there are many real-world systems that fit this model. Under the assumption, messages which are transmitted could consist of a random number of packets. If, upon arrival, the channel is free, one packet is randomly chosen to be transmitted and the rest of them are waiting for transmission; otherwise, if the channel is occupied, all the packets wait for transmission and will try the retransmission after a random period of time. If there are less messages in the channel, part of the transmission ability is stored, and the packet will be transmitted slowly and it saves the cost and operation ability in the system. On the other hand, if more messages arrive, the transmission ability can be released completely and the packet can be transmitted quickly.

In this paper, we mainly apply the M/G/1-type matrix analytic approach developed by Neuts [16]. By this method, a matrix $\boldsymbol{G}$, as a minimal nonnegative solution of a matrix equation, plays an important role. In this paper, we can get the precise analytic expression for $\boldsymbol{G}$ since the coefficient matrices of the equation are upper triangular. Furthermore, based on this result and the stochastic decomposition theory in [17], we can obtain two equivalent probability generating functions(PGFs) of the stationary queue length at the departure epoch, one of which has the evident relationship with the result in the Geo^X/G/1 queue with regular(non-working) vacations. Meanwhile, we show that the probabilities of the server's state and the probabilities that an arbitrary customer is served by the normal or working vacation service rate can be computed. We carry on the complete analysis of the stationary indices in the system and derive the PGF of the number of the customers at the busy period initiation epoch and the computation formulas of the expected sojourn time.

Compared to the other literatures in this field, especially Refs. [12–15], the contributions of the present paper, in our opinion, are both theoretical and practical. First, we carry on the research on the model by a method which is different from those used in other literatures. As we all know, the different-type queueing models always have the different analytic methods. For the continuous-time and discrete-time GI/M/1-type models, including the M/M/1 ones, the Matrix Geometric Solution (MGS) method is very powerful in deriving the stationary distributions of the queue length and waiting time, and the details can be seen in Refs. [12–15]. But this method is not helpful in analyzing the M/G/1-type models and other approaches must be found. We innovate to apply the connection of the M/G/1-type matrix analytic approach and the stochastic decomposition method to solve the batch-arrival M/G/1-type model in this paper. Thus, our work is very different from the other publications both in the models and in the analysis method. In a view, this analyzing process can be extended to the other relevant queueing models, such as the Markov arrival process (MAP) models and Phase-type (PH) vacation models. Secondly, we revisit and extend the M/G/1-type matrix analytic approach in [16]. It follows from the results in [16] that there is no direct numerical expression for $\boldsymbol{G}$ in general M/G/1-type matrices, which can only be approximated by some numerical approximations. But, the fact that an evident expression for the matrix $\boldsymbol{G}$ exists in this queue will make the analysis in this paper more specific and interesting. The third contribution in this paper comes from the analysis of the sojourn time and stochastic decomposition property, because as far as we know, until now, no literatures were involved in these system indices in discrete-time Geo/G/1 working vacation queues. Finally, our theoretical results have the potential to add the support and necessary ground for performance evaluation in practical problems. A real example in Section 7 will verify our opinion.

The remainder of the paper is structured as follows. In Section 2, the model of Geo^X/G/1/WV is formulated, and we establish an embedded Markov chain and obtain the matrix $\boldsymbol{G}$. In Section 3, by Markov stationary equations, we obtain a some complicated expression for the PGF of the queue length. In Section 4, using the stochastic decomposition theory, we get a more concise expression for the PGF and the analysis of the sojourn time is also processed. The corresponding indices in two special models are presented in Section 5. The numerical discussion of the stationary probabilities is postponed to Section 6. The performance analysis of an Ethernet Passive Optical Network (EPON) is presented in Section 7.

2. Model formulation and embedded Markov chain

We consider a single-server discrete-time queue where the time axis is divided into equal intervals (called slots). It is assumed that all queueing activities (arrivals, departures and vacations) occur at the slot boundaries and therefore they may occur at the same time. For mathematical convenience, we will suppose that the departures occur at the moment immediately before the slot boundaries, whereas arrivals occur at the moment immediately after the slot boundaries. Thus, we will discuss the model for the early arrival system (EAS) policy. Details on the EAS discipline and related concepts can be found in [18].

Batches of customers arrive at the system according to a geometrical process with probability p $(0 < p < 1)$. The number of individual external customers arriving in each batch is $k \geq 1$ with probability χ_k, and we denote by $E(X)$ and $X(z)$ the mean and PGF of the sequences $\{\chi_k\}_{k=1}^{\infty}$, respectively. The service time during the normal busy period is an independent and identically distributed random variable S_b which follows a general distribution $\{b_k^{(1)}\}_{k=1}^{\infty}$ with the PGF $F(z)$ and mean $E(S_b)$.

If at any time any customer arrives, he goes to the service facility for service. Arriving customers are queued according to the first-come, first-served (FCFS) discipline. The server can serve only one customer at a time. The server begins a working vacation each time whenever the queue becomes empty and if there are customers arriving during a vacation period, the server continues to work at a lower rate. The service time S_v during the vacation period also follows a general distribution $\{b_k^{(2)}\}_{k=1}^{\infty}$ with the PGF $G(z)$ and mean $E(S_v)$.

The working vacation period is an operation period in a lower speed, and the vacation time V is geometrically distributed with rate θ $(0 < \theta < 1)$, i.e.,

$$\mathrm{P}\{V = k\} = \theta(1-\theta)^{k-1}, \quad k \geq 1.$$

If no customers are found in the queue when the server returns from the vacation, he again leaves for another vacation with the same length. This pattern continues until he returns from a vacation to find at least one customer waiting in the queue. We denote this model by $\mathrm{Geo}^{X}/\mathrm{G}/1/\mathrm{WV}$.

Let L_n be the number of the customers at the nth service completion or customer departure instant. In the working vacation model, any service completion may occur during a service period or a working vacation period. Define

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

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

$$\Omega = \Big\{(k, 0), k \geq 0\Big\} \bigcup \Big\{(k, 1), k \geq 1\Big\}.$$

A_x denotes the number of the batches arriving during the random length x; then we introduce three probability notations:

$$a_k = \mathrm{P}\{A_{S_b} = k\} = \sum_{j=max(1,k)}^{\infty} b_j^{(1)} \binom{j}{k} p^k \overline{p}^{j-k}, \quad \overline{p} = 1 - p,\ k \geq 0;$$

$$b_k = \mathrm{P}\{A_{S_v} = k, V > S_v\} = \sum_{j=max(1,k)}^{\infty} b_j^{(2)} \binom{j}{k} p^k \overline{p}^{j-k} \overline{\theta}^j, \quad \overline{\theta} = 1 - \theta,\ k \geq 0;$$

$$v_k = \mathrm{P}\{A_V = k, V \leq S_v\} = \sum_{j=k+1}^{\infty} \sum_{n=j}^{\infty} b_n^{(2)} \binom{j-1}{k} p^k \overline{p}^{j-1-k} \overline{\theta}^{j-1} \theta, \quad k \geq 0,$$

where a_k represents the probability that there are k batches arriving during S_b (regular service time), b_k represents the probability that $V > S_v$ and k batches arrive during S_v (vacation service time), and v_k represents the probability that $V \leq S_v$ and k batches arrive during V (vacation time).

Assume that $\alpha_j, j = 0, 1, \ldots$, is the probability that j customers arrive during S_b. Because one batch includes n customers with probability χ_n, $n = 1, 2, \ldots$, we have the following relationship:

$$\alpha_j = \sum_{k=0}^{j} a_k \chi_j^{(k)}, \quad j = 0, 1, \ldots, \tag{1}$$

where $\chi_j^{(k)}$ is the probability that j $(j \geq k)$ customers arrive in k batches and is the k-fold convolution of χ_j, and $\chi_0^{(0)} = 1$. Now define z as the function variable, and then derive the PGF of $\{\alpha_j\}_{j=0}^{\infty}$,

$$\begin{aligned} \alpha(z) &= \sum_{j=0}^{\infty} \alpha_j z^j = \sum_{k=0}^{\infty} a_k \sum_{j=k}^{\infty} \chi_j^{(k)} z^j = \sum_{k=0}^{\infty} a_k [X(z)]^k \\ &= A(X(z)) = F[1 - p(1 - X(z))], \end{aligned}$$

where $A(z) = \sum_{k=0}^{\infty} a_k z^k = F[1 - p(1 - z)]$.

Furthermore, assume that $\beta_j, j = 0, 1, \ldots$ is the probability that the vacation time V is longer than S_v, i.e., $V > S_v$, and j customers arrive during S_v, and

$$\beta_j = \sum_{k=0}^{j} b_k \chi_j^{(k)}, \quad j = 0, 1, \ldots. \tag{2}$$

Now multiplying (2) by appropriate powers of z and then taking summation over all possible values of j, we obtain

$$\begin{aligned} \beta(z) &= \sum_{j=0}^{\infty} \beta_j z^j = \sum_{k=0}^{\infty} b_k [X(z)]^k = B(X(z)) \\ &= G[(1-\theta)(1 - p(1 - X(z)))], \end{aligned}$$

where $B(z) = \sum_{k=0}^{\infty} b_k z^k = G[(1-\theta)(1 - p(1 - z))]$.

Similarly, $\nu_j, j = 0, 1 \ldots$ is the probability that $V \le S_v$ and j customers arrive during the vacation time V, and

$$\nu_j = \sum_{k=0}^{j} \upsilon_k \chi_j^{(k)}, \quad j = 0, 1, \ldots. \tag{3}$$

Multiplying (3) by appropriate powers of z and then taking summation over all possible values of j, we obtain

$$\begin{aligned} \nu(z) &= \sum_{j=0}^{\infty} \nu_j z^j = \sum_{k=0}^{\infty} \upsilon_k [X(z)]^k = V(X(z)) \\ &= \frac{\theta(1-\beta(z))}{1-(1-\theta)(1-p(1-X(z)))}, \end{aligned}$$

where

$$V(z) = \sum_{k=0}^{\infty} \upsilon_k z^k = \frac{\theta(1-B(z))}{1-(1-\theta)(1-\lambda(1-z))}.$$

Evidently,

$$\sum_{j=0}^{\infty} \beta_j = G(1-\theta), \qquad \sum_{j=0}^{\infty} \nu_j = 1 - G(1-\theta).$$

Thus, $\{\beta_j, j \ge 0\}$ and $\{\nu_j, j \ge 0\}$ are two non-complete probability distributions.

Let $\zeta_k = \sum_{j=0}^{k} \nu_j \alpha_{k-j}, k \ge 0$; then ζ_k represents the probability that the vacation time V is not longer than the vacation service time S_v and k customers arrive during V plus S_b. Then,

$$\sum_{k=0}^{\infty} \zeta_k = 1 - G(1-\theta), \qquad \zeta(z) = \sum_{k=0}^{\infty} \zeta_k z^k = \nu(z)\alpha(z).$$

With the assumptions above, we consider the one-step transition probabilities of (L_n, J_n).

Case 1: If $X_n = (m, 1), m \ge 1$:

$$X_{n+1} = \begin{cases} (m-1+j, 1) & \text{with probability } \alpha_j, \; m \ge 2, \; j \ge 0; \\ (j, 1) & \text{with probability } \alpha_j, \; m = 1, \; j \ge 1; \\ (0, 0) & \text{with probability } \alpha_0, \; m = 1. \end{cases}$$

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

$$X_{n+1} = \begin{cases} (m-1+j, 0) & \text{with probability } \beta_j, \; j \ge 0; \\ (m-1+j, 1) & \text{with probability } \zeta_j, \; j \ge 0. \end{cases}$$

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

$$X_{n+1} = \begin{cases} (j, 0) & \text{with probability } \beta_j, \; j \ge 1; \\ (j, 1) & \text{with probability } \zeta_j, \; j \ge 1; \\ (0, 0) & \text{with probability } \beta_0 + \zeta_0. \end{cases}$$

Based on one-step transition situation analysis, using the lexicographical sequence for the states, the one-step transition probability matrix of (L_n, J_n) can be written as the Block–Jacobi matrix

$$\widetilde{\boldsymbol{P}} = \begin{bmatrix} \boldsymbol{B}_0 & \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{B}_3 & \cdots \\ \boldsymbol{C}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \boldsymbol{A}_3 & \cdots \\ & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots \\ & & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \cdots \\ & & & \vdots & \vdots \end{bmatrix}, \tag{4}$$

where

$$\boldsymbol{B}_0 = \beta_0 + \zeta_0; \qquad \boldsymbol{B}_i = (\beta_i, \zeta_i), \quad i \ge 1; \qquad \boldsymbol{C}_0 = (\beta_0 + \zeta_0, \alpha_0)^T$$

$$\boldsymbol{A}_i = \begin{bmatrix} \beta_i & \zeta_i \\ 0 & \alpha_i \end{bmatrix}, \quad i \ge 0,$$

where T represents 'matrix transpose operation'. The stochastic matrix $\widetilde{\boldsymbol{P}}$ is an M/G/1-type matrix (see [16]). For such a model, to demonstrate whether this chain is positive recurrent, the minimal nonnegative solution of the equation $\boldsymbol{G} = \sum_{i=0}^{\infty} \boldsymbol{A}_i \boldsymbol{G}^i$ is needed. To obtain the matrix $\boldsymbol{G}$ and the positive recurrent result, we present some preliminaries based on

Neuts [16] that we will use in our analysis. Evidently, $A = \sum_{i=0}^{\infty} \boldsymbol{A}_i$ is reducible in this Markov chain, and from the equation (2.3.18) in [16], $\boldsymbol{A}$ can be rewritten as the structure below:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}(1) & & & & 0 \\ & \boldsymbol{A}(2) & & & 0 \\ & & \ddots & & \vdots \\ & & & \boldsymbol{A}(c) & 0 \\ \boldsymbol{T}(1) & \boldsymbol{T}(2) & \cdots & \boldsymbol{T}(c) & \boldsymbol{T}(0) \end{pmatrix} \tag{5}$$

where $\boldsymbol{A}(1), \ldots, \boldsymbol{A}(c)$ are irreducible stochastic matrices, and matrices in the bottom line may not exist, where $\boldsymbol{T}(1), \ldots, \boldsymbol{T}(c)$ are general matrices, and $\boldsymbol{T}(0)$ are reducible. In this case, all $\boldsymbol{A}_i$ $(i \geq 0)$ also have the same structure; then its diagonal part $\boldsymbol{A}_i(j)$ $(1 \leq j \leq c)$ is also irreducible. Introduce

$$\boldsymbol{\pi}(j)\boldsymbol{A}(j) = \boldsymbol{\pi}(j), \qquad \boldsymbol{\pi}(j)\boldsymbol{e} = 1, \quad 1 \leq j \leq c,$$
$$\boldsymbol{\vartheta}(j) = \sum_{i=1}^{\infty} i\boldsymbol{A}_i(j)\boldsymbol{e}, \quad 1 \leq j \leq c.$$

Then Theorem 2.3.3 in [16] can be shown in Preliminary 1 here.

Preliminary 1. *If $A = \sum_{i=0}^{\infty} \boldsymbol{A}_i$ is reducible, $\boldsymbol{G}$ is a stochastic matrix if and only if*

$$\boldsymbol{\pi}(j)\boldsymbol{\vartheta}(j) \leq 1, \quad 1 \leq j \leq c,$$

and $\widetilde{\boldsymbol{P}}$ is positive recurrent if and only if

$$\boldsymbol{\pi}(j)\boldsymbol{\vartheta}(j) \leq 1, \quad 1 \leq j \leq c; \qquad \sum_{k=1}^{\infty} k\boldsymbol{B}_k\boldsymbol{e} < \infty,$$

Having the results in Preliminary 1 as preparation, we come back to our model below. Because all $\boldsymbol{A}_i$ are upper triangular, we can assume that $\boldsymbol{G}$ has the same structure as

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

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

$$\begin{cases} r_{11} = \sum_{i=0}^{\infty} \beta_i r_{11}^i = G[(1-\theta)(1-p(1-X(r_{11})))], \\ r_{12} = \sum_{i=0}^{\infty} r_{22}^i \zeta_i + r_{12} \sum_{i=0}^{\infty} \beta_i \sum_{j=0}^{i-1} r_{11}^j r_{22}^{i-1-j}, \\ r_{22} = \sum_{i=0}^{\infty} \alpha_i r_{22}^i = F[1-p(1-X(r_{22}))]. \end{cases} \tag{6}$$

To obtain the minimal nonnegative solution, one lemma is first provided. First we introduce $\rho = pE(X)E(S_b)$, which presents the system intensity.

Lemma 1. *If $\rho < 1$, the equation $z = F[1-p(1-X(z))]$ has the minimal nonnegative root $z = 1$ and the equation $z = G[(1-\theta)(1-p(1-X(z)))]$ has the unique root in the range $0 < z < 1$.*

Proof. First, we consider the equation $z = F[1-p(1-X(z))]$. Let $\psi(z) = F[1-p(1-X(z))]$ and evidently, $0 < \psi(0) = F(1-p) < \psi(1) = 1$. For $0 < z < 1$, $\psi'(z) > 0$, $\psi''(z) > 0$. Meanwhile, from $\rho = pE(X)E(S_b) < 1$, $\psi'(1) = \rho < 1$. Thus, the equation $z = \psi(z)$ has the minimal nonnegative root $z = 1$. Similarly, we set $\varphi(z) = G[(1-\theta)(1-p(1-X(z)))]$. Then we have $0 < \varphi(0) < \varphi(1) < 1$. For $0 < z < 1$, $\varphi'(z) > 0, \varphi''(z) > 0$. It is shown that the equation $z = G[(1-\theta)(1-p(1-X(z)))]$ has a unique root in the range $0 < z < 1$. □

After some computation, we have

$$\boldsymbol{G} = \begin{bmatrix} \gamma & 1-\gamma \\ 0 & 1 \end{bmatrix}, \tag{7}$$

where γ is the unique root in the range $0 < z < 1$ of the equation $z = G[(1-\theta)(1-p(1-X(z)))]$. Evidently, $\boldsymbol{G}$ is a reducible stochastic matrix.

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

Proof. Because

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

is a reducible stochastic matrix. With the notation of the Eq. (5), $\boldsymbol{A}(2) = 1$ is the degenerative stochastic matrix and has the degenerative stationary distribution $\boldsymbol{\pi}(2) = 1$. On the other hand, $\boldsymbol{A}_i(2) = \alpha_i$, $i \geq 0$, and $\boldsymbol{\vartheta}(2) = \sum_{i=0}^{\infty} i\boldsymbol{A}_i(2) = \rho$. It is obvious that $\sum_{k=1}^{\infty} k\boldsymbol{B}_k\boldsymbol{e} < \infty$. Thus, by Preliminary 1, the Markov chain $\widetilde{\boldsymbol{P}}$ is positive recurrent if and only if

$$\boldsymbol{\pi}(2)\boldsymbol{\vartheta}(2) = \rho < 1. \quad \square$$

3. Queue length analysis

If $\rho < 1$, let (L, J) be a set of random variables which follows the stationary distribution of the (L_n, J_n). Denote

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

$$\boldsymbol{\pi}_k = (\pi_{k0}, \pi_{k1}), \qquad \boldsymbol{\pi} = (\pi_{00}, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_k, \ldots).$$

Here, we explain that the state $(0, 0)$ is reached in two cases. Case 1: in the busy period, the service completion for the last customer leaves the system with no customer; case 2: a customer is served completely in the vacation period, and no customer is left. In the classic vacation policy, it is impossible to serve customers during the vacation period and case 2 does not exist.

We solve for the stationary distribution π_{kj} by noting that the vector $\boldsymbol{\pi}$ satisfies the equation $\boldsymbol{\pi}\widetilde{\boldsymbol{P}} = \boldsymbol{\pi}$ and have the following steady-state equations:

$$\begin{cases} \pi_{00} = \pi_{00}(\beta_0 + \zeta_0) + \pi_1\boldsymbol{C} = (\pi_{00} + \pi_{10})\beta_0 + (\pi_{00} + \pi_{10})\zeta_0 + \pi_{11}\alpha_0; \\ \pi_{k0} = \pi_{00}\beta_k + \displaystyle\sum_{j=1}^{k+1} \pi_{j0}\beta_{k+1-j}, \quad k \geq 1; \\ \pi_{k1} = \pi_{00}\zeta_k + \displaystyle\sum_{j=1}^{k+1} \pi_{j0}\zeta_{k+1-j} + \sum_{j=1}^{k+1} \pi_{j1}\alpha_{k+1-j}, \quad k \geq 1. \end{cases} \tag{8}$$

Evidently, based on the analysis of the transition, for state $(0, 0)$, with the probability $\zeta_0\pi_{00} + \zeta_0\pi_{10} + \alpha_0\pi_{11}$, the last service completion in the busy period leaves the system with no customer; with the probability $\beta_0(\pi_{00} + \pi_{10})$, a customer is served completely in the vacation period, and no customer is left. Let $u = \beta_0(\pi_{00} + \pi_{10})$, and under the classic vacation policy, it is impossible to serve customers in the vacation period and $u = 0$.

Now we introduce the following generating functions:

$$Q(z) = \beta_0(\pi_{00} + \pi_{10}) + \sum_{k=1}^{\infty} \pi_{k0} z^k, \qquad P(z) = \zeta_0\pi_{00} + \zeta_0\pi_{10} + \alpha_0\pi_{11} + \sum_{k=1}^{\infty} \pi_{k1} z^k,$$

where $Q(z)$ and $P(z)$ represent the generating functions of the number of the left customers at the instant when one customer is served completely during the working vacation period and normal service period, respectively.

First, we consider $Q(z)$ and multiplying the second equation in (8) by z^k and summing over k yield

$$\begin{aligned} Q(z) &= \beta_0(\pi_{00} + \pi_{10}) + \pi_{00}\sum_{k=1}^{\infty} \beta_k z^k + \sum_{k=1}^{\infty}\sum_{j=1}^{k+1} \pi_{j0}\beta_{k+1-j} z^k \\ &= \pi_{00}\beta(z) + \sum_{j=1}^{\infty} \pi_{j0} z^{j-1} \sum_{k=j-1}^{\infty} \beta_{k+1-j} z^{k+1-j} \\ &= \pi_{00}\beta(z) + \frac{\beta(z)}{z}\big(Q(z) - \beta_0(\pi_{00} + \pi_{10})\big). \end{aligned}$$

Noting $u = \beta_0(\pi_{00} + \pi_{10})$, we obtain

$$Q(z) = \frac{\beta(z)(\pi_{00}z - u)}{z - \beta(z)}. \tag{9}$$

From Lemma 1, γ is the root of the equation $z = \beta(z)$; thus it is the zero point of denominator of (9). Regularity demands that the numerator of (9) equals 0 for $z = \gamma$. Therefore, we have

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

Then,

$$Q(z) = \frac{\pi_{00}\beta(z)(z - \gamma)}{z - \beta(z)}. \tag{10}$$

Furthermore, from the first equation in (8),

$$\zeta_0\pi_{00} + \zeta_0\pi_{10} + \alpha_0\pi_{11} = \pi_{00} - \beta_0(\pi_{00} + \pi_{10}) = \pi_{00}(1-\gamma).$$

Using the expression for $P(z)$, multiplying the third equation in (8) by z^k and summing over k, we have

$$\begin{aligned} P(z) &= \pi_{00}(1-\gamma) + \pi_{00}\sum_{k=1}^{\infty}\zeta_k z^k + \sum_{k=1}^{\infty}\sum_{j=1}^{k+1}\pi_{j0}\zeta_{k+1-j}z^k + \sum_{k=1}^{\infty}\sum_{j=1}^{k+1}\pi_{j1}\alpha_{k+1-j}z^k \\ &= \pi_{00}\zeta(z) + \frac{Q(z) - \pi_{00}\gamma}{z}\zeta(z) + \frac{P(z) - \pi_{00}(1-\gamma)}{z}\alpha(z). \end{aligned} \tag{11}$$

After some computation, we obtain

$$P(z) = \pi_{00}\frac{z\zeta(z)(z-\gamma) - (1-\gamma)\alpha(z)(z-\beta(z))}{(z-\alpha(z))(z-\beta(z))}.$$

The PGF of the queue length at the departure epoch has the expression

$$\begin{aligned} L(z) &= \pi_{00} + \sum_{k=1}^{\infty}(\pi_{k0} + \pi_{k1})z^k = Q(z) + P(z) \\ &= \pi_{00}\frac{\alpha(z)(1-z)(\beta(z)-z) + z(r-z)(\alpha(z) - \beta(z) - \zeta(z))}{(z-\beta(z))(z-\alpha(z))}. \end{aligned} \tag{12}$$

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

$$\alpha'(1) = \rho, \qquad \beta'(1) = \beta, \qquad \zeta'(1) = \rho\nu'(1) + \nu(1) = \left[\rho + \frac{p\overline{\theta}}{\theta}E(X)\right](1 - G(1-\theta)) - \beta,$$

we can obtain

$$\pi_{00} = \frac{(1-\rho)(1-G(1-\theta))}{1 - G(1-\theta) - (1-\gamma)\left[\rho G(1-\theta) - \frac{p\overline{\theta}}{\theta}E(X)(1-G(1-\theta))\right]}. \tag{13}$$

$P_v(P_b)$ denotes the probability that an arbitrary customer is served completely by the working vacation service rate (normal service rate), noted by $S = 0$ ($S = 1$); then we have

$$\begin{aligned} P_v &= \mathrm{P}\{S=0\} = (\pi_{00} + \pi_{10})\beta_0 + \sum_{k=1}^{\infty}\pi_{k0} = Q(z)|_{z=1} \\ &= \frac{(1-\rho)G(1-\theta)(1-\gamma)}{1 - G(1-\theta) - (1-\gamma)\left[\rho G(1-\theta) - \frac{p\overline{\theta}}{\theta}E(X)(1-G(1-\theta))\right]}, \\ P_b &= \mathrm{P}\{S=1\} = \zeta_0\pi_{00} + \zeta_0\pi_{10} + \alpha_0\pi_{11} + \sum_{k=1}^{\infty}\pi_{k1} = P(z)|_{z=1} = 1 - P_v \\ &= \frac{1 - G(1-\theta) - (1-\gamma)\left[G(1-\theta) - \frac{p\overline{\theta}}{\theta}E(X)(1-G(1-\theta))\right]}{1 - G(1-\theta) - (1-\gamma)\left[\rho G(1-\theta) - \frac{p\overline{\theta}}{\theta}E(X)(1-G(1-\theta))\right]}. \end{aligned}$$

Meanwhile, we obtain the state probability of the server at the departure epoch, and

$$\begin{aligned} \mathrm{P}\{J=0\} &= \pi_{00} + \sum_{k=1}^{\infty}\pi_{k0} = \pi_{00}(1-\gamma) + Q(z)|_{z=1} \\ &= \frac{(1-\rho)(1-\gamma)}{1 - G(1-\theta) - (1-\gamma)\left[\rho G(1-\theta) - \frac{p\overline{\theta}}{\theta}E(X)(1-G(1-\theta))\right]}, \\ \mathrm{P}\{J=1\} &= \sum_{k=1}^{\infty}\pi_{k1} = P(z)|_{z=1} - \pi_{00}(1-\gamma) = 1 - \mathrm{P}\{J=0\} \\ &= \frac{1 - G(1-\theta) - (1-\gamma)\left[1 - \rho + \rho G(1-\theta) - \frac{p\overline{\theta}}{\theta}E(X)(1-G(1-\theta))\right]}{1 - G(1-\theta) - (1-\gamma)\left[\rho G(1-\theta) - \frac{p\overline{\theta}}{\theta}E(X)(1-G(1-\theta))\right]}. \end{aligned}$$

It is obvious that P_v and P_b are not $\mathrm{P}\{J=0\}$ and $\mathrm{P}\{J=1\}$, respectively. The differences are caused by what happens in the state $(0, 0)$. If there is no service during the vacation period, i.e., $G(1-\theta) = 0$, we derive $P_b = 1$; in other words, all customers are served completely by the normal service rate. Then P_v becomes $\mathrm{P}\{J=0\}$ and $P_b = \mathrm{P}\{J=1\}$.

4. Another expression for $L(z)$: stochastic decomposition structure

In last section, we obtain the Eq. (12) for the PGF of the queue length at the departure epochs. Now we obtain another equivalent expression for $L(z)$ to establish the relationship between $\mathrm{Geo}^X/G/1/WV$ and the classic $\mathrm{Geo}^X/G/1$ queue, including the stochastic decomposition structure. We find that, in the model we analyze, under the condition that a customer is served by the normal rate, the operation process is stochastically equivalent to that of the $\mathrm{Geo}^X/G/1$ queue with general (nonworking) vacations, and then the results in [19,17] can be used to analyze the normal service period in the working vacation queues. Now, we mainly use the stochastic decomposition result in [17] to obtain another equivalent expression for $L(z)$.

4.1. The queue length at the beginning of the busy period

First, we obtain the distribution of the number of the customers Q_b at the beginning instant of the busy period. Introduce the last service completion before the beginning instant of the busy period as a regenerating point; then $Q_b = k$ happens under two cases: in case 1, under the condition that the system stays in the vacation period after the last service and the vacation time V is not larger than the vacation service time S_v, j ($j \geq 1$), customers are left and $k - j$ customers arrive during the vacation time V; in case 2, under the same condition, no customers are left and $k - 1$ customers arrive during the vacation time V. We first can compute

$$\mathrm{P}\{J = 0, V \leq S_v\} = \sum_{k=0}^{\infty} \pi_{k0} \times \mathrm{P}\{V \leq S_v\} = \mathrm{P}\{J = 0\}(1 - G(1 - \theta)) = \pi_{00}(1 - \gamma),$$

which represents the probability that the system stays in the vacation period after the last service and $V \leq S_v$. Then, the distribution of Q_b at the beginning instant of the busy period is

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

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

$$Q_b(z) = \sum_{k=1}^{\infty} \tau_k z^k = \frac{1}{1-\gamma} \frac{z(z-r)\nu(z)}{z - \beta(z)}. \tag{14}$$

Evidently, $Q_b(1) = 1$ and

$$E(Q_b) = \frac{1 - G(1-\theta) - (1-\gamma)\left[G(1-\theta) - \frac{p\bar{\theta}}{\theta} E(X)(1 - G(1-\theta)) \right]}{(1-\gamma)(1 - G(1-\theta))}.$$

From the expression for P_b, we have the relation: $E(Q_b) = P_b(1-\rho)(\pi_{00}(1-\gamma))^{-1}$.

4.2. The conditional stochastic decomposition of $L(z)$

As in [17], the service time of every customer in a regular busy period is called an active period and the length of a vacation or service interruption is called an inactive period. Certainly, the server can work during the inactive period and alternates between active and inactive states. Note that for this view to be general we allow the inactive period to have zero length. Let $L^s(L^T)$ be the number of the customers at the starting (ending) instant of an inactive period in the steady state with the PGF $L^s(z)(L^T(z))$. With the same notation, we first give the known stochastic decomposition result in [17] as Preliminary 2, which will be used in the analysis below.

Preliminary 2. *If $L^s \leq_{st} L^T$, the stationary distribution of the number of customers in the system at service completion epochs is the convolution of the distribution functions of two independent random variables, one of which is the stationary distribution of the number of customers in the system at service completion epochs in an ordinary M/G/1-type queue without server vacations, and the number of customers in the system at service completion epochs has the PGF*

$$P(z) = E(z^N)E(z^Y) = E(z^N) \times \frac{L^s(z) - L^T(z)}{(1-\rho)(1-z)},$$

where N represents the number of customers in the system at service completion epochs in an ordinary M/G/1-type queue without server vacations and Y is an additional variable caused by the vacation.

The detailed results and proofs for this Preliminary can also be seen from Lemma 1 and Theorem 1 in [17].

Below, we conduct the analysis of our working vacation model by the result in Preliminary 2. First, based on the conditional probability, we have the relation

$$L(z) = E\{z^L\} = E\{z^L | S = 1\}\mathrm{P}\{S = 1\} + E\{z^L | S = 0\}\mathrm{P}\{S = 0\}. \tag{15}$$

In a batch arrival 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 $L^s - L^T$. From the fact that the operation process under the condition that a customer is served by the normal rate is stochastically equivalent to that of the Geo^X/G/1 queue with general (nonworking) vacations, applying Preliminary 2 for the first term on the right-hand side in (15), we obtain

$$E\{z^L|S=1\} = E(z^N)E(z^Y) = \frac{(1-\rho)(1-z)\alpha(z)}{\alpha(z)-z} \times \frac{L^s(z)-L^T(z)}{(1-\rho)(1-z)}, \tag{16}$$

where N is the number of customers in a classic Geo^X/G/1 queue in the steady state. Here, as we explain above, under the condition $S = 1$, the left queue length L can be zero. Thus, the expression for $E\{z^L|S = 1\}$ should not be

$$E\{z^L|S=1\} = E\{z^N|N>0\}E(z^Y).$$

Based on the definition of $Q(z)$, we have

$$E\{z^L|S=0\} = P_v^{-1}Q(z) = \frac{1-G(1-\theta)}{(1-\gamma)G(1-\theta)} \frac{\beta(z)(z-\gamma)}{z-\beta(z)}. \tag{17}$$

Then, consider $L^s(L^T)$. Note that $L^T = k$ $(k \geq 1)$ 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. Therefore, we have

$$\mathrm{P}\{L^T = k\} = \mathrm{P}\{L^s = k\} + \mathrm{P}\{L^s = 0\}\tau_k, \quad k \geq 1, \tag{18}$$

and $\mathrm{P}\{L^T = 0\} = 0$. From (14), we obtain the relation

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

From (16),

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

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

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

Thus, from (15)–(17), we obtain

$$L(z) = P_v \frac{1-G(1-\theta)}{(1-\gamma)G(1-\theta)} \frac{\beta(z)(z-\gamma)}{z-\beta(z)} + P_b \frac{(1-\rho)(1-z)\alpha(z)}{\alpha(z)-z} \frac{1-Q_b(z)}{E(Q_b)(1-z)}. \tag{19}$$

It is easy to verify that Eqs. (19) and (12) are equivalent. The expression in (19) has the certain probability explanation. With the probability P_v, the queue length L is one random variable with the generating function

$$\frac{1-G(1-\theta)}{(1-\gamma)G(1-\theta)} \frac{\beta(z)(z-\gamma)}{z-\beta(z)},$$

and with the probability P_b, it is the sum of two random variables, one of which is the classic queue length in the Geo^X/G/1 queue without vacations.

From (19), the expected queue length is

$$\begin{aligned} E(L) = {} & P_v\left\{\frac{\beta}{G(1-\theta)} + \frac{1}{1-\gamma} - \frac{1-\beta}{1-G(1-\theta)}\right\} \\ & + P_b\left\{\rho + \frac{p^2(E(X))^2E(S_b(S_b-1))}{2(1-\rho)} + \frac{pE(S_b)E(X(X-1))}{2(1-\rho)} + \frac{E(Q_b(Q_b-1))}{2E(Q_b)}\right\}, \end{aligned} \tag{20}$$

where

$$E(S_b(S_b-1)) = F''(1), \qquad E(X(X-1)) = X''(1), \qquad E(Q_b(Q_b-1)) = Q_b''(1).$$

Evidently, no complete stochastic decomposition structure exists, but based on the Eq. (19), we can derive a conditional stochastic decomposition structure directly. Denote by L_q 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., $S = 1$. Then,

$$L_q = \{L|S=1\}.$$

A theorem can be given to demonstrate conditional stochastic decomposition.

Theorem 2. *The conditional queue length L_q can be decomposed into two independent random variables: $L_q = L_0 + L_d$, where L_0 is the stationary queue length of a classic* Geo^X/G/1 *queue without vacations under AF policy, and L_0 and additional queue* *length L_d have PGFs,*

$$L_0(z) = \frac{(1-\rho)(1-z)\alpha(z)}{\alpha(z)-z}, \qquad L_d(z) = \frac{1-Q_b(z)}{E(Q_b)(1-z)},$$

respectively.

4.3. System time of an arbitrary customer

We consider the period when an arbitrary customer spends in the queue and in the service, called the sojourn time S. Denote the PGF of S by $S(z)$. The customers left behind a departing customer include those who arrive during its sojourn time S and the residual customers of this batch to which it belongs. Let

$$p_j = \frac{1}{E(X)} \sum_{n=j+1}^{\infty} \chi_n, \quad j = 0, 1, \ldots .$$

Then, $\{p_j, j \geq 0\}$ is the probability that there are j left customers of one batch after an arbitrary customer, and is the 'residual life' of $\{\chi_j, j \geq 1\}$ with the PGF

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

Thus, we have

$$L(z) = S[1-p(1-X(z))]P(z).$$

Differentiating it with respect to z, we finally get the expected sojourn time in the queue

$$E(S) = \frac{E(L)}{pE(X)} - \frac{E(X(X-1))}{2p(E(X))^2}, \tag{21}$$

where $E(L)$ is given in Eq. (20).

Remark 1. Until now, only Wu and Takagi [9] and Yi et al. [20] have presented the research on the M/G/1-type working vacation queues. The former analyzed a continuous-time M/G/1 queue with general working vacation using the methods of embedded Markov chain and plural function, and the latter applied the supplementary variable method to provide the analysis of the discrete-time Geo/G/1 working vacation queue. Evidently, the plural function method is not applicable in analyzing the discrete-time cases. In Yi et al. [20], they first considered a Geo/G/1 queue with disasters and used its results to analyze the working vacation queue, where the analysis of the Geo/G/1 queue was based on the supplementary variable method. Under this method, we need to analyze the operation processes of two queueing models and establish the connection between two models. The solving process of the working vacation queue would become more difficult and complex. In our paper, applying the matrix analytic method can study the Geo/G/1 queue directly, and the matrix $\boldsymbol{G}$ makes the research process more easy. Thus, this method is more direct and simple compared to that used in [20]. But, if $X = 1$, the results are the same with those in [20], and details can be seen below.

5. Special models

Case 1: Geo/G/1 queue with working vacations. If $X = 1$, then $\rho = pE(S_b)$. The batch-arrival model becomes the Geo/G/1 queue with working vacations which was considered in [20]. From the results in sections above, a series of the stationary indices in the system can be obtained.

- The vacation service and normal service probabilities:

$$P_v = \frac{(1-\rho)G(1-\theta)(1-\gamma)}{1-G(1-\theta)-(1-\gamma)\left[\rho G(1-\theta) - \frac{p\bar{\theta}}{\theta}(1-G(1-\theta))\right]},$$

$$P_b = \frac{1-G(1-\theta)-(1-\gamma)\left[G(1-\theta) - \frac{p\bar{\theta}}{\theta}(1-G(1-\theta))\right]}{1-G(1-\theta)-(1-\gamma)\left[\rho G(1-\theta) - \frac{p\bar{\theta}}{\theta}(1-G(1-\theta))\right]}.$$

- The expected queue length:

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

- The expected sojourn time: $E(S) = E(L)p^{-1}$.

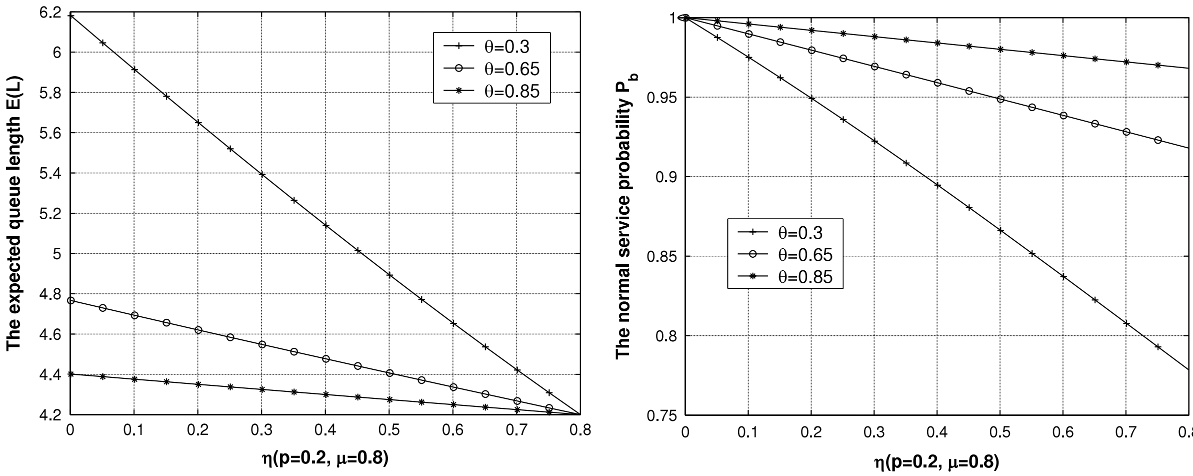

Fig. 1. $E(L)$ and P_b versus η in $\text{Geo}^D/(\text{Geo}^1, \text{Geo}^2)/1$.

These results are in agreement with the results of Geo/G/1 queue with working vacations reported in [20].

Case 2: $\text{Geo}^X/G/1$ queue with classic vacations. If the server cannot work during the vacation period, the working vacation policy becomes the classic vacation policy. In this case, evidently, $P_v = 0$, $P_b = 1$, and as follows.

- The PGF of the queue length at the departure epoch:

$$L(Z) = \frac{(1-\rho)(1-z)F[1-p(1-X(z))]}{F[1-p(1-X(z))]} \times \frac{1-Q_b(z)}{E(Q_b)(1-z)},$$

where

$$Q_b(z) = \frac{\theta z}{1-\bar{\theta}(1-p(1-X(z)))}, \qquad E(Q_b) = Q_b'(1).$$

- The expected queue length:

$$E(L) = \rho + \frac{p^2(E(X))^2E(S_b(S_b-1))}{2(1-\rho)} + \frac{pE(S_b)E(X(X-1))}{2(1-\rho)} + \frac{E(Q_b(Q_b-1))}{2E(Q_b)}.$$

- The expected sojourn time:

$$E(S) = \frac{E(L)}{pE(X)} - \frac{E(X(X-1))}{2p(E(X))^2}.$$

Thus, we show that some vacation models in the literatures are special cases of our model.

6. Numerical results

In this section, based on the results obtained, we show some numerical cases. Without loss of generality, we assume the service times during the normal service and working vacation periods follow the geometrical distributions with parameters μ and η, respectively, where $\mu = 0.8$, and η can change from 0 to 0.8. Thus, in this system, higher and lower speeds alternate according to the number of the customers (jobs). The customers (jobs) arrive in a batch X, and we consider two cases for X, (1) deterministic ($X = 3$); (2) geometric ($E(X) = 3$). Two models are denoted by $\text{Geo}^D/(\text{Geo}^1, \text{Geo}^2)/1$, $\text{Geo}^G/(\text{Geo}^1, \text{Geo}^2)/1$, respectively.

Figs. 1 and 2 show some trends of the system indices in $\text{Geo}^D/(\text{Geo}^1, \text{Geo}^2)/1$. In Fig. 1, we pay attention to the curves of the stationary queue length $E(L)$ and the normal service probability P_b with the change of the lower service rate η. Evidently, the queue length decreases with the increase of η, and the longer the vacation time, the larger the queue length $E(L)$ is; then when η approaches to $\mu(= 0.8)$, $E(L)$ will achieve a fixed value, i.e., the queue length without vacations, no matter how long the vacation time is. The vacation will have no impact on $E(L)$. It is reasonable and demonstrates that keeping some service ability during the vacation period decreases the system waiting customers, and when the service rate during the vacation period is equal to the normal service rate, the system becomes the classic Geo/Geo/1 queue without vacations. The normal service probability P_b has the similar change trend and P_b decreases with the increase of η, and when θ is smaller, for example, $\theta = 0.3$, the vacation time is larger and the probability that a customer is served by the normal service rate becomes smaller. Meanwhile, when $\eta = 0$, in other words, the server cannot work during the vacation period, all the customers will be served by the normal rate μ and P_b is equal to 1.

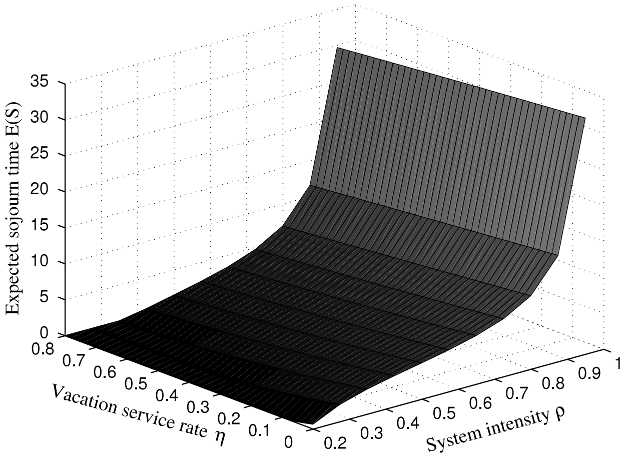

Fig. 2. Sojourn time $E(S)$ versus η and ρ in $\mathrm{Geo}^D/(\mathrm{Geo}^1, \mathrm{Geo}^2)/1$.

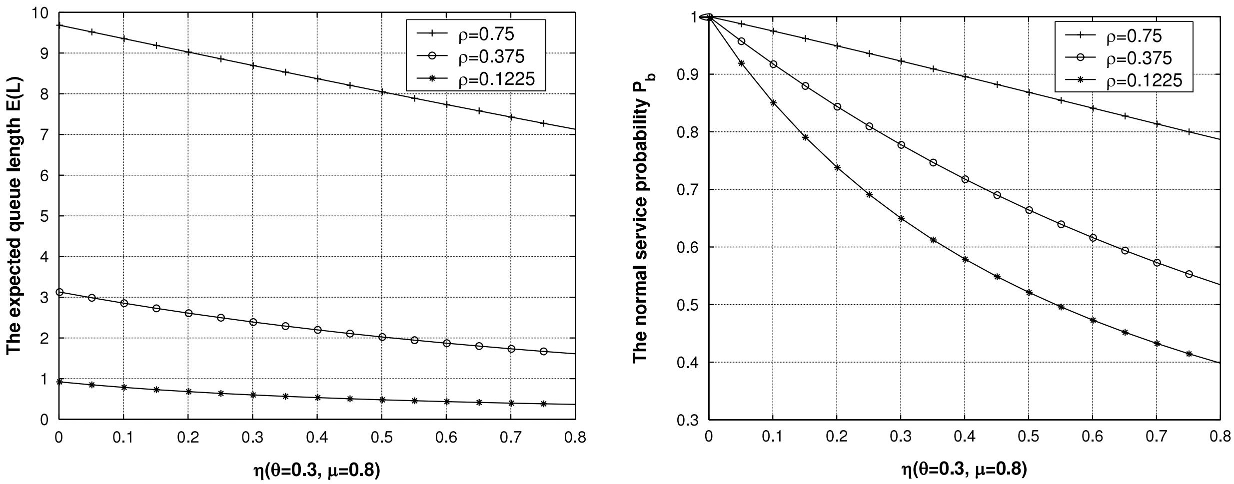

Fig. 3. $E(L)$ and P_b versus η in $\mathrm{Geo}^G/(\mathrm{Geo}^1, \mathrm{Geo}^2)/1$.

In Fig. 2, the trend of the sojourn time $E(S)$ of an arbitrary customer is presented with the changes of two parameters, the vacation service rate η and system intensity ρ. Certainly, the sojourn time will decrease with the increase of η. But from the figure, we find that, compared with the effect of the system intensity ρ, the effect of η is smaller and with the increase of the ρ, the sojourn time increases significantly. Thus, it demonstrates that, in an unloaded system which has the smaller system intensity, the sojourn time is smaller and it is not necessary to establish the working vacation service rate.

Similarly, in Figs. 3 and 4, we illustrate the trends of the system indices in $\mathrm{Geo}^G/(\mathrm{Geo}^1, \mathrm{Geo}^2)/1$ model. Fig. 3 shows the trends observed for the expected queue length $E(L)$ and the normal service probability P_b under three systems with $\rho = 0.1225$, $\rho = 0.375$ and $\rho = 0.75$, respectively. That is, $E(L)$ decreases as η increases and in a loaded system ($\rho = 0.75$), $E(L)$ is larger evidently. On the other hand, in an unloaded system ($\rho = 0.1225$), the number of vacations will be larger and the probability that a customer is served by the normal service rate is smaller than that in a loaded system ($\rho = 0.75$ or $\rho = 0.375$).

Furthermore, we have investigated the expected sojourn time $E(S)$ regarding the combinations of the values of the vacation rate θ and vacation service rate η in Fig. 4. Certainly, $E(S)$ decreases as η and θ increase. Meanwhile, with the increase of θ, in other words, the vacation time decreases, the decreasing degree of $E(S)$ regarding η declines and when $\theta = 1$, $E(S)$ achieves a fixed value, i.e., the sojourn time in the system without vacations. On the other hand, with the increase of η, the decreasing degree of $E(S)$ regarding θ declines and when $\eta = \mu$, the system reduces to the model without vacations and $E(S)$ achieves a fixed value.

7. Application to an Ethernet Passive Optical Network

The following is a possible scenario in which our results in working vacation queues can be used. Consider an EPON, which consists of one optical line terminal (OLT) situated at the central office (CO) and multiple optical network units (ONUs)

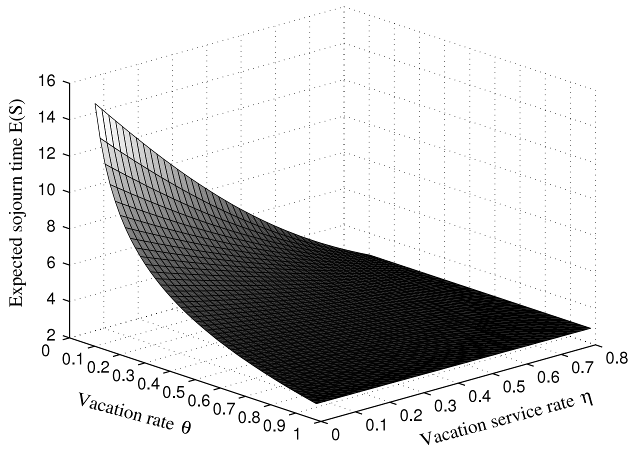

Fig. 4. Sojourn time $E(S)$ versus η and θ in $\mathrm{Geo}^G/(\mathrm{Geo}^1, \mathrm{Geo}^2)/1$.

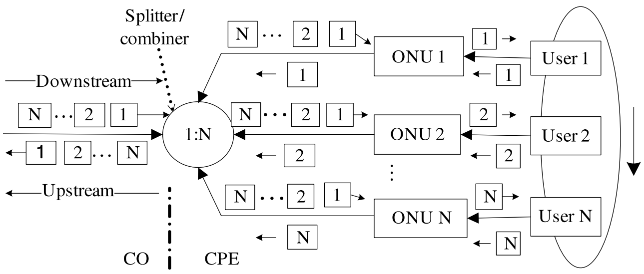

Fig. 5. Typical structure of an EPON.

located at customer premises equipment (CPE), and a passive splitter/combiner. As illustrated in Fig. 5, EPON provides bi-directional transmissions, where in the downstream direction (from OLT to the ONUs), the OLT broadcasts to all ONUs. The frames are sent to their destination ONUs by using media access control (MAC) layer. In the upstream direction (from ONUs to the OLT), because it is a multipoint-to-point network, the fiber channel is shared by all ONUs. Therefore, scheduling is needed to prevent data packet collision from different ONUs. A robust mechanism is needed for allocating time slots and the upstream bandwidth for each ONU to transmit data. In EPON, the mechanism is called MPCP involving both GATE message and REPORT message. The ONUs may send REPORT messages about the queue state of each ONU to the OLT. The OLT allocates upstream bandwidth to each ONU by sending GATE messages with the form of a 64-byte MAC control frame. GATE message contains a time-stamp and grants time slots which represent the periods that ONU can transmit data, so that the OLT can allocate the upstream bandwidth and time slots to each ONU accordingly. Other schemes include the interleaved polling scheme with an adaptive cycle time (IPACT), which allocates time slots in the buffer of the ONUs. These two strategies are efficient in the EPON performance analysis, but they may not match the immediate needs of each ONU and cause the congestion in the EPON.

If the working vacation scheme is applied to the EPON, we can imagine that each ONU can alternate to transmit the data/message at a high or low rate according to the number of the data, so that the immediate needs and avoidance of congestion are satisfied. Below we propose an adaptive scheme which combines the gated and polling service schemes together based on the working vacation for allocating time slots and the upstream bandwidth for each ONU to transmit data. We consider each ONU in the EPON system separately and treat it as a single server queueing model. First, the gated service and working vacation service queues are presented respectively.

Under the gated service scheme, ONU i requires the right of sending data, and it closes the gate at the beginning of every transmission period; then ONU i only transmits data packet which have been present in this transmission period, and the new arrival data packets will be waiting outside of the gate according to arrival ordering. When the data packets inside gate of ONU i have been transmitted, the system will turn to next ONU $i+1$ and transmit data packet inside the gate of ONU $i+1$. For simplicity but without loss of generality, it is assumed that each ONU generates a single packet during a slot, i.e., one data each batch, and the service time follows the geometric distribution. Then this single ONU with the gated service can be modeled as a $\mathrm{Geo}/(\mathrm{Geo}^1, \mathrm{Geo}^2)/1$ gated service multiple vacation queueing model.

Under the working vacation scheme, ONU i has permanent wavelengths assigned to it with the capability of transmitting data packet at a nominal rate η^i, and there are additional wavelengths in the upstream channel with the capacity to service at an additional average rate of η^*. Such additional wavelengths operate in a cycle mode from one ONU to another ONU, that is, if ONU i is currently operating at a total service rate of $\eta^i + \eta^*$, when the buffer of ONU i becomes empty at a service

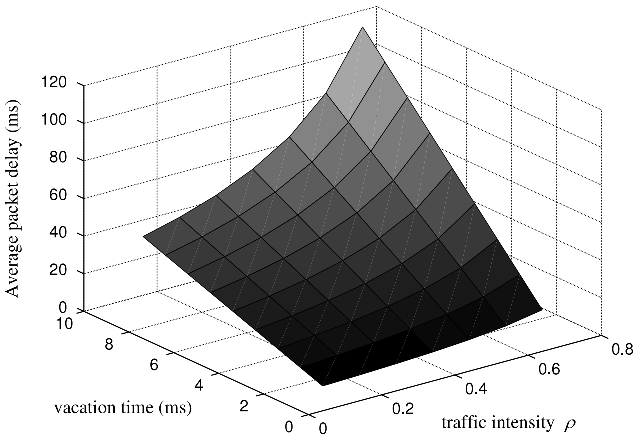

Fig. 6. Average packet delay time with gated service.

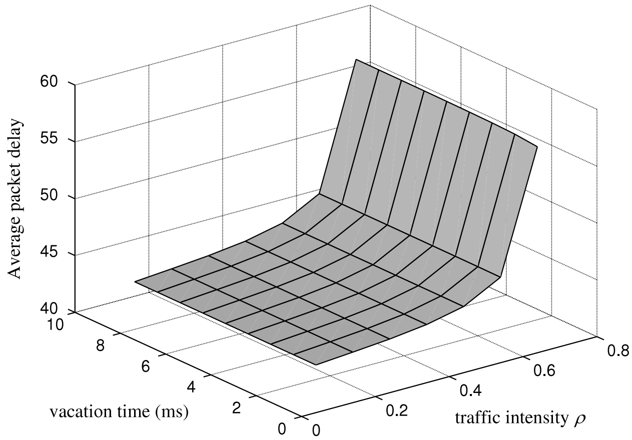

Fig. 7. Average packet delay time with working vacation scheme.

completion, the polling wavelengths are reconfigured to ONU $i+1$, then ONU i instantaneously reduces its service rate to η^i and only after a reconfiguration delay of Δ, ONU $i+1$ increases its service rate to $\eta^{i+1}+\eta^*$. This cycle continually repeats itself. This operation process is also similar to that in [8]. Then each ONU can be modeled as a Geo/(Geo^1, Geo^2)/1 working vacation queueing model.

To specify a detailed model, it is assumed that this EPON is a five router symmetric system. If we design data packets with fixed size as 1000 bits and the total transmission speeds of the EPON is 1 Gb/s, the transmission time of a data packet equals to 1 μs (i.e. the time length of a slot). We assume that the polling time follow a geometric distribution and the service time $\mu = 0.95$, $\eta = 0.05$. For working vacation service scheme, the polling time should have the additional time which is assumed as $10 \times$ traffic intensity (ρ). In Figs. 6 and 7, we plot the average packet delays from the effect of traffic intensity ρ and vacation time, under gated service and working vacation scheme, respectively. Evidently, along with the increase of traffic intensity ρ, not only under gated service scheme but also under working vacation service scheme, the average delay time increases. But compared to two figures, when $\rho > 0.6$ and $E(V) > 8$ ms, the average packet delay in ONUs of the gated service increases more obviously than working vacation service scheme. In other situation, the performance of the gated service excels the working vacation service. Therefore, to compare the gated service and working vacation service, we cannot find one service scheme to surpass another in all situations.

To exploit the advantages of two schemes, the adaptive service scheme is proposed, according to a criterion, i.e. the delay time of data packets in ONUs is minimal. The ONUs will decide which service scheme they will take between gated service and working vacation service scheme. The algorithm is

If $E(W_{working}) > E(W_{gate})$,

The adaptive service scheme adopt the working vacation service;

Else

The adaptive service scheme adopt the gated service.

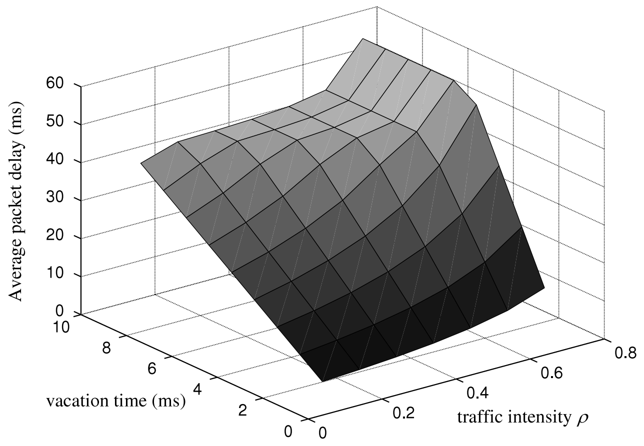

Fig. 8. Average packet delay time with adaptive service.

In Fig. 8, we adopt the adaptive service and depict the curve of the average packet delay. From the figure, the effects of the traffic intensity ρ and vacation time have become smoother than the effects under the single scheme. It is illustrated that the performance of EPON is improved considerably under this adaptive service based on the working vacation scheme.

Similar scheduling also happen in other real fields, such as wireless ATM, Call Center, and Digital Commerce. From this practical example, it is also shown that the results in this paper would be beneficial not only to queueing theorists but also to many practitioners who try to evaluate the performance of their real systems.

8. Conclusions

In this paper, we have studied the system indices in a batch arrival Geo/G/1 queue with working vacations. We have formulated two methods to obtain the PGFs of the queue length at the departure epoch and shown the formulas of some other performance measures, such as the queue length at the busy period initiation epoch, the normal (vacation) service probability and the sojourn time. 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. Meanwhile, our special models and numerical examples conclude that our working vacation model is justified and system parameters affect the steady-state mean queue length and mean delay. Lastly, the performance analysis of an EPON would suggest that the working vacation policy can represent various types of problems in practice efficiently.

Acknowledgements

The authors would like to thank editors and three referees for their valuable comments and suggestions, which were very helpful to improve the paper. This research is supported by National Natural Science Foundation of China #10671170.

References

[1] B.T. Doshi, Single server queues with vacations, in: H. Takagi (Ed.), Stochastic Analysis of the computer and Communication Systems, North-Holland Elsevier, Amsterdam, 1990, pp. 217–264.
[2] B.T. Doshi, Queueing systems with vacations-a survey, Queueing Syst. 1 (1986) 29–66.
[3] H. Takagi, Queueing Analysis: Discrete-Time Systems, vol. 3, North-Holland Elsevier, Amsterdam, 1993.
[4] N. Tian, Z.G. Zhang, Vacation Queueing Models: Theory and Applications, Springer-Verlag, New York, 2006.
[5] D. Fiems, H. Bruneel, Analysis of a discrete-time queueing system with timed vacations, Queueing Systems 42 (2002) 243–254.
[6] S.H. Chang, D.W. Choi, Performance analysis of a finite-buffer discrete-time queue with bulk arrival, bulk service and vacations, Comput. Oper. Res. 32 (2005) 2213–2234.
[7] S.K. Samanta, M.L. Chaudhry, U.C. Gupta, Discrete-time $\mathrm{Geo}^{\mathrm{X}}/\mathrm{Geo}^{(a,b)}/1/\mathrm{N}$ queues with single and multiple vacations, Math. Comput. Model. 45 (2007) 93–108.
[8] L.D. Servi, S.G. Finn, M/M/1 queue with working vacations (M/M/1/WV), Perform. Eval. 50 (2002) 41–52.
[9] D. Wu, H. Takagi, M/G/1 queue with multiple working vacations, Perform. Eval. 63 (2006) 654–681.
[10] Y. Baba, Analysis of a GI/M/1 queue with multiple working vacations, Oper. Res. Lett. 33 (2005) 201–209.
[11] A.D. Banik, U.C. Gupta, S.S. Pathak, On the GI/M/1/N queue with multiple working vacations-analytic analysis and computation, Appl. Math. Model. 31 (2007) 1701–1710.
[12] W. Liu, X. Xu, N. Tian, Stochastic decompositions in the M/M/1 queue with working vacations, Operations Research letters 35 (2007) 595–600.
[13] J. Li, N. Tian, Analysis of the discrete time Geo/Geo/1 queue with single working vacation, Qual. Technol. Quantit. Manag. 5 (1) (2008) 77–89.
[14] J. Li, N. Tian, The discrete-time GI/Geo/1 queue with working vacations and vacation interruption, Appl. Math. Comput. 185 (2007) 1–10.
[15] J. Li, N. Tian, W. Liu, Discrete-time GI/Geo/1 queue with working vacations, Queueing Systems 56 (2007) 53–63.
[16] M. Neuts, Structured Stochastic Matrices of M/G/1 Type and their Applications, Marcel Dekker, New York, 1989.
[17] J.G. Shanthikumar, On stochastic decomposition in M/G/1 type queues with generalized server vacations, Operations Research 36 (1988) 566–569.
[18] J.J. Hunter, Mathematical Techniques of Applied Probability, in: Discrete-Time Models: Techniques and Applications, vol. 2, Academic Press, New York, 1983.
[19] S.W. Fuhrmann, R.B. Cooper, Stochastic decomposition in the M/G/1 queue with generalized vacations, Operations Research 33 (1985) 1117–1129.
[20] X.W. Yi, et al., The Geo/G/1 queue with disasters and multiple working vacations, Stochastic Models 23 (2007) 537–549.

Further reading

[1] Z.G. Zhang, N. Tian, Discrete time Geo/G/1 queue with multiple adaptive vacations, Queueing Systems 38 (4) (2001) 419–430.

Ji-hong Li is currently a lector in the College of Management, Shanxi University in China. She received her Ph.D. degree in Management Science and Engineering at Yanshan Universit, China in 2009. Her research interests are queueing models with vacations, production management and performance analysis of Communication Networks. She has published several papers in international journals.

Wei-qi Liu is a professor in the College of Management, Shanxi University in China. His research interests include Finance Engineering and Time Series analysis. He has published several research monographs and many important papers.

Nai-shuo Tian is a professor in the College of Sciences, Yanshan University in China. His research interests are stochastic process, queueing systems, and performance evaluation models in computer networks and supply chain systems. He has published many important papers in several international journals including Queueing Systems, Performance Evaluation, Operations Research Letters, EJOR, INFOR and several other journals in China.