Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 11, No. 1, pp. 57–78, January 2015

ANALYSIS AND PERFORMANCE OPTIMIZATION OF A GEOM/G/1 QUEUE WITH GENERAL LIMITED SERVICE AND MULTIPLE ADAPTIVE VACATIONS*

Zhanyou Ma, Wuyi Yue and Li Chen

Abstract: In this paper, we first present an analysis for a general limited service Geom/G/1 queueing model with multiple adaptive vacations (MAVs). Then we give the Probability Generating Function (P.G.F.) of the queue length by using the embedded Markov chain method and the regeneration cycle approach. The P.G.F. of the waiting time is derived based on the independence between the arrival process and the waiting time. The probabilities of various states (such as "busy", "idle" and "vacation") are also derived in the system. We can explicitly calculate their averages and higher moments by differentiating these P.G.Fs. to numerically compare different system cases. Furthermore, we present an analysis of the cost model for this general limited service Geom/G/1 queueing model with multiple adaptive vacations to determine the optimal management policy by minimum cost. We can obtain the optimal values of service rate and the optimal expected traffic intensity in the system. Finally, based on numerical results, the relations of performance measures and the optimal expected traffic intensity are discussed.

Key words: *multiple adaptive vacations, general limited service, embedded Markov chain method, regeneration cycle approach, coefficient, cost function*

Mathematics Subject Classification: *60K25*

---

1 Introduction

Many Geom/G/1 queues with various vacation policies have been well investigated. Early research has focused on exhaustive service policy, where the server takes a vacation only if the system becomes idle presented in [10], [7]. However, multifarious non-exhaustive service policies also have important practical values in computer systems and communication networks presented in [8], [2].

Several non-exhaustive service policies have been introduced into the discussion of performance analysis of polling systems presented in [3], [12], where the server may take a vacation when there are customers waiting in the system. Stochastic decomposition results of vacation queues with general vacation policies were presented in [6], [11]. Geom/G/1 queues with a variety of vacation policies were systemically analyzed in [13], and research results of multi-server vacation queues were studied in [14].

In the adaptation of different application backgrounds, some new vacation policies were introduced into queueing systems. A class of Geom/G/1 queues with exhaustive service

*This work was supported by the National Natural Science Foundation of China (Nos. 71071134, 61472341), Hebei Province National Science Foundation (Nos. A2012203124, F2012203093, G2013203169), China, and was supported in part by MEXT, Japan.

© 2015 Yokohama Publishers

and multiple adaptive vacations were studied in [14]. A discrete time Geom/G/1 queue with multiple adaptive vacations was investigated in [20]. Banik [1] combined batch Markov arrival process (BMAP) and MAVs, and analyzed the BMAP/G/1/∞ queue with multiple adaptive vacation policy. They obtained the queue length distribution at a post-departure epoch by using matrix-analytic procedure. Some important performance measures were studied too. This is an important literature about MAVs, but this paper studied the exhaustive service discipline queue system. Therefore, we are inspired by [1], we will combine the general limited service and MAVs for the first time, and study non-exhaustive queue system by using the embedded Markov chain method and the regeneration cycle approach. The multiple adaptive vacation policy is a synthetic policy which generalizes several simple vacation policies.

Chang et al. considered an infinite capacity N-policy M/G/1 queueing system with a single removable server in [5]. In this system, the server is controllable that may be turned on at arrival epochs or off at service completion epochs. They applied a differential technique to study system sensitivity. And a cost model for infinite capacity queueing system under steady-state condition is developed, to determine the optimal management policy at minimum cost. So that this model could be used in real applications to obtain the optimal management policy.

Wang et al. studied the cost optimization and sensitivity analysis of the machine repair problem with variable servers and balking in [16]. They obtained various system performance measures, such as the expected number of failed machines, the expected number of operating machines, the average balking rate, the machine availability, the operative utilization, and so on. A cost model is constructed to determine the optimal values of the number of busy servers, and maintain the balking rate at a certain level.

Ke et al. investigated the cost optimization of queue with different vacation policy in [4], [19], such as (p, N)-policy M/G/1 queue with an unreliable server and single vacation, unreliable multi-server queue with a controllable repair policy, server breakdown, and so on. A quasi-birth-and-death process is used to model the complex system and the stability condition is examined. The rate matrix is calculated approximately and steady-state stationary distributions were obtained by a matrix-analytic approach. And they constructed a total expected cost function per unit time and applied the Tabu search method to find the minimum cost, then determined the optimal repair policy, the optimal value of service rate and the optimal value of repair rate.

In order to model more practical problems, and analyze the performance of these problems better, we studied a Geom/G/1 queue with a non-exhaustive service and multiple adaptive vacations in [9]. Using the embedded Markov chain method and the regeneration cycle approach, we obtain the transformation formulae of the stationary queue length and the waiting time, and give stochastic decomposition structures for the stationary performance indices. The probabilities for the system being in various states are also derived. In this paper, we can explicitly calculate their averages and higher moments of the stationary queue length L_v and the stationary waiting time W_v by differentiating these P.G.Fs. to numerically compare different system cases. We present an analysis of the cost function to determine the optimal expected traffic intensity in the system and show the effect of system parameters on performance measures. The optimal analysis provides the theoretical basis for the system performance optimization. Therefore, this paper is the further promotion of [9]. Moreover, general limited service Geom/G/1 queues with multiple vacations or single vacations presented in [13] are the special cases in the model presented in this paper.

The rest of this paper is organized as follows. In Section 2, we describe the queueing system model, as well as suppose system parameters and a service policy. In Section 3, we derive the performance analysis to give the number of customers at the beginning instant, the stationary queue length and waiting time, the service cycle, and so on. In Section 4, we develop an expected cost function for the Geom/G/1 queue with the general limited service and MAVs to determine the optimal expected traffic intensity and the best service rate μ. In Section 5, by numerical examples, we show the effect of the parameters on the mean state probability and the performance optimization with the cost function. In Section 6, we give conclusions.

2 System Model

Considering a classical Geom/G/1 queue, we introduce a general limited service and multiple adaptive vacation policy presented in [14], [20]: A service period is defined as the interval between a service starting instant and the next adjacent vacation starting instant. The difference between a service period and a busy period is that the former may end when some customers are still waiting in the system, but the latter ends only when no customers remain in the system. A general limited service policy means that the number of customers which is served in every service period does not exceed a determinate upper limit M, M is a positive integer.

Let $Q_b^{(n)}$ be the number of customers in the system at the beginning instant of the nth service period, then the number of customers who will be served in the next service period is given by

$$\Phi = \min\left\{Q_b^{(n)}, M\right\}.$$

Let H denote the number of the vacations. The server will take H times of vacation that have random lengths. The probability distribution h_k and the Probability Generating Function (P.G.F.) $H(z)$ of H are given as follows:

$$P(H=k) = h_k, \quad k \geq 1, \quad H(z) = \sum_{k=1}^{\infty} h_k z^k.$$

Let V_ξ $(\xi = 1, 2, ..., H)$ denote the length of vacation time for the ξth vacation. The vacation time lengths V_ξ are independent identically distributed (i.i.d.) random variables. Therefore, the state space of the queue model should be

$$\Omega = \{0, 1, 2, ...\}$$

where state 0 represents that the system may be in the vacation state or the idle state, state i $(i \geq 1)$ represents that the system may be in the busy state or the vacation state.

There are three cases in this model as follows:

(1) If there are residual customers in the system when a service period ends, and there are customers arriving during the ξth vacation, $1 \leq \xi \leq H$, the vacation period will stop in advance at the completion instant of the ξth vacation. The server will begin a new service period, and the server will take vacations again when Φ customers have been served.

(2) If there are no residual customers in the system when a service period ends, and there are customers arriving during the ξth vacation, $1 \leq \xi \leq H$, the vacation period will

stop in advance at the completion instant of the ξth vacation. The server will begin a new service period. In the new service period, the server will serve Φ customers, then takes vacations.

(3) If there are no residual customers in the system when a service period ends, and there are customers arriving during the Hth vacation, the vacation period will stop in advance at the completion instant of the Hth vacation, and the system enters an idle period. When a customer arrives in the idle period, the server emerges from an idle state to service state, and only begins to serve this customer. Then the server begins to take vacations, the new arrival should be served in the next service period.

The system will continually repeat the above process.

The basic assumptions of the system model are given as follows:

(1) Suppose that customer arrivals can only occur at the discrete time instant $t = n^-, n = 0, 1, ...$ The service starts and ends at the discrete time instant $t = n^+$ only, $n = 1, 2, ...$ The model is called a late arrival system. The inter-arrival time T is supposed to be an independent identically distributed (i.i.d.) discrete random variable following a geometric distribution with parameter λ $(0 < \lambda < 1)$. We can write the probability distribution of T as follows:

$$P(T = j) = \lambda\bar{\lambda}^{j-1}, \quad j = 1, 2, ...$$

where $\bar{\lambda} = 1 - \lambda$. We denote by N_n the number of customers arriving during the interval $[0, n]$, therefore N_n follows a Binomial distribution as follows:

$$P(N_n = j) = \binom{n}{j}\lambda^j\bar{\lambda}^{n-j}, \quad j = 0, 1, ..., n, \quad n = 1, 2, ...$$

(2) The service time S is supposed to be an i.i.d. discrete random variable with a general distribution s_j. The P.G.F. $S(z)$, the mean $E[S]$ and the second factorial moment $E[S(S-1)]$ of S are given as follows:

$$P(S = j) = s_j, \quad j \geq 1, \quad S(z) = \sum_{j=1}^{\infty} s_j z^j,$$

$$E[S] = \sum_{i=0}^{\infty} i s_i, \quad E[S(S-1)] = \frac{d^2 S(z)}{dz^2}\Big|_{z=1}.$$

Let μ be the reciprocal value of the mean $E[S]$. This means that we have $1/\mu = E[S]$.

(3) The time length V of an arbitrary vacation in H vacations is a non-negative i.i.d. discrete random variable with general probability distribution p_v and the P.G.F. $V(z)$ given by

$$P(V = v) = p_v, \quad v \geq 1, \quad V(z) = \sum_{v=1}^{\infty} p_v z^v$$

and the mean $E[V]$ and the second factorial moment $E[V(V-1)]$ of V are given as follows:

$$E[V] = \sum_{v=0}^{\infty} v p_v, \quad E[V(V-1)] = \frac{d^2 V(z)}{dz^2}\Big|_{z=1}.$$

Let $\rho = \lambda/\mu$ be the traffic intensity of the system.

Suppose that there is a single server in this system, and its capability is infinite. The inter-arrival time, the service time and the time length of a vacation are mutually independent. The service order is First-Come First-Served (FCFS). The model mentioned above can be denoted by Geom/G/1 (GL, MAVs), where GL and MAVs represent the General Limited service and the Multiple Adaptive Vacations, respectively. The schematic diagram for Geom/G/1 (GL, MAVs) queue is given in Fig. 1, where C_s is defined as a service cycle which is the interval between two adjacent points where the service ends and the vacation starts at the same time.

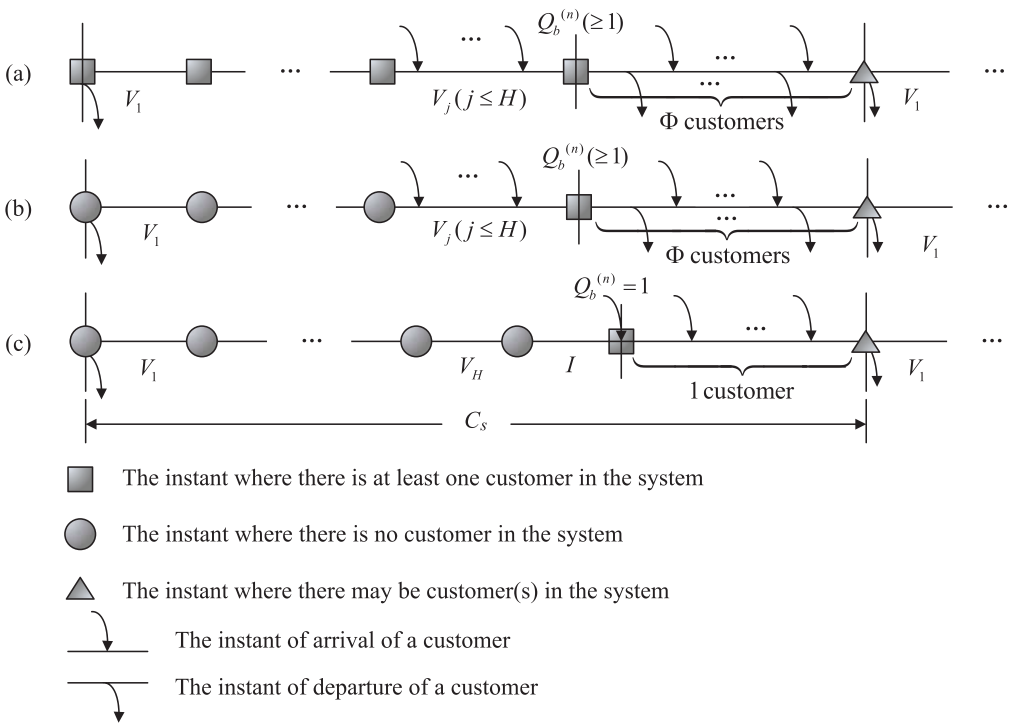

Figure 1: The schematic diagram for the Geom/G/1 (GL, MAVs) queue.

3 Performance Analysis

3.1 Preliminaries

According to a non-exhaustive service policy presented in this paper, the transition probability matrix of the queue length $\{L_n, n \geq 1\}$ at the departure instant of a customer is different from the transition probability matrix of a classical Geom/G/1 queue at not only boundary states, but all states. Therefore we cannot simply apply the results of a Geom/G/1 boundary state variation model to study the Geom/G/1 with the non-exhaustive service policy presented in this paper. The regeneration cycle approach is the most effective tool to apply to the stationary queue length of a system with a non-exhaustive service policy.

Let $L_v(t)$ represent the queue length process of a Geom/G/1 queue with a non-exhaustive service policy. The beginning instants of the regeneration cycle are chosen as regeneration points when the number of customers is zero. The process $L_v(t)$ can be assumed to restart at these instants.

If the system is positive recurrent, $L_v(t)$ will transit the zero state with an infinite amount of times, therefore the system has infinite regeneration points. The interval between two adjacent regeneration points is defined as a regeneration cycle. A regeneration cycle may include several service cycles. The length of a regeneration cycle is an i.i.d. random variable. In order to obtain the stationary queue length of Geom/G/1 (GL, MAVs) queue, we give the regeneration cycle approach as Lemma 3.1.

Lemma 3.1. *If a stationary distribution exists, its P.G.F. can be given as*

$$L_v(z) = \frac{E\left[\sum_{n=1}^{\Phi} z^{L_n}\right]}{E[\Phi]} \tag{3.1}$$

where L_n denotes the number of customers in the system at the nth departure instant of customers (see [13] and [14]).

3.2 Number of Customers at the Beginning Instant

Transition probabilities of Markov chain $\{L_n, n \geq 1\}$ are given as follows:

$$P_{jk} = \begin{cases} \sum_{r=k-j+M}^{\infty} \binom{r}{k-j+M} \lambda^{k-j+M} \bar{\lambda}^{r-k+j-M} P\left(B^{(M)} + V = r\right), & k \geq j - M > 0 \\ (1 - H(V(\bar{\lambda}))) \sum_{r=k}^{\infty} \binom{r}{k} \lambda^k \bar{\lambda}^{r-k} P\left(B^{(j)} + V = r\right), & j \leq M, k \neq 1 \\ (1 - H(V(\bar{\lambda}))) \sum_{r=k}^{\infty} \binom{r}{k} \lambda^k \bar{\lambda}^{r-k} P\left(B^{(j)} + V = r\right) + H(V(\bar{\lambda})), & j \leq M, k = 1 \\ 0, & j > M, k < j - M \end{cases} \tag{3.2}$$

where $B^{(j)}$ is the sum of j service times, $0 \leq j \leq M$, and V is the time length of an arbitrary vacation. $V(\bar{\lambda})$ represents the value of the P.G.F. $V(z)$ at $z = \bar{\lambda}$, $H(V(\bar{\lambda}))$ represents the value of the P.G.F. $H(z)$ at $z = V(\bar{\lambda})$. We define that

$$q_k = \lim_{n \to \infty} P\left(Q_b^{(n)} = k\right), \quad k \geq 0.$$

$\{q_k, k \geq 0\}$ is the distribution of Q_b customers in the system at the beginning instant of the service period. From the equilibrium equation of the Markov chain, we have

$$q_0 = V(\bar{\lambda})(1 - H(V(\bar{\lambda}))) \sum_{j=0}^{M} q_j (S(\bar{\lambda}))^j, \tag{3.3}$$

$$\begin{aligned} q_1 &= \sum_{j=0}^{M} q_j \left((1 - H(V(\bar{\lambda}))) \sum_{r=j}^{\infty} \binom{r}{1} \lambda \bar{\lambda}^{r-1} P\left(B^{(j)} + V = r\right) + H(V(\bar{\lambda})) \right) \\ &\quad + q_{M+1} \sum_{r=j}^{\infty} \binom{r}{0} \lambda^0 \bar{\lambda}^r P\left(B^{(M)} + V = r\right), \end{aligned} \tag{3.4}$$

$$\begin{aligned} q_k &= \sum_{j=0}^{M} q_j (1 - H(V(\bar{\lambda}))) \sum_{r=j}^{\infty} \binom{r}{k} \lambda^k \bar{\lambda}^{r-k} P\left(B^{(j)} + V = r\right) \\ &\quad + \sum_{j=M+1}^{k+M} \sum_{r=M}^{\infty} \binom{r}{k-j+M} \lambda^{k-j+M} \bar{\lambda}^{r-k+j-M} P\left(B^{(M)} + V = r\right), \\ &\qquad k \geq 2. \end{aligned} \tag{3.5}$$

We define the partial probability generating function $Q_M(z)$ as follows:

$$Q_M(z) = \sum_{k=0}^{M} q_k z^k.$$

Multiplying both sides of Eqs. (3.3)-(3.5) by z^0, z, z^k, respectively, and taking the summation with respect to k, we obtain the P.G.F. $Q_b(z)$ of $\{q_k, k \geq 0\}$ as follows:

$$\begin{aligned} Q_b(z) = &\frac{1}{z^M - (S(1-\lambda(1-z)))^M V(1-\lambda(1-z))} \\ &\times \Big((1 - H(V(\bar{\lambda}))) z^M Q_M(S(1-\lambda(1-z))) V(1-\lambda(1-z)) \\ &- (S(1-\lambda(1-z)))^M Q_M(z) V(1-\lambda(1-z)) \\ &+ H(V(\bar{\lambda})) Q_M(1) z^{M+1} \Big) \end{aligned} \tag{3.6}$$

where $V(1-\lambda(1-z))$ represents the composite function of $V(z)$ and $1-\lambda(1-z)$, $S(1-\lambda(1-z))$ represents the composite function of $S(z)$ and $1-\lambda(1-z)$.

To determine $Q_b(z)$, we should obtain the coefficients $q_0, q_1, ..., q_M$ by using the Rouche theorem [15] and the Lagrange theorem [17]. In the denominator of Eq. (3.6), we define that

$$f(z) = z^M, \quad g(z) = -V(1-\lambda(1-z))(S(1-\lambda(1-z)))^M.$$

For the probability distribution $\{x_k, k \geq 0\}$ of any non-negative random variable integer X and for a sufficiently small $\varepsilon > 0$, in $|z| = 1 + \varepsilon$, we have

$$|X(z)| = \left| \sum_{k=0}^{\infty} x_k z^k \right| \leq \sum_{k=0}^{\infty} x_k (1+\varepsilon)^k = \sum_{k=0}^{\infty} x_k (1 + k\varepsilon) + o(\varepsilon) = 1 + \varepsilon E[X] + o(\varepsilon)$$

where $E[X]$ is the mean of X and $o(\varepsilon)$ is the higher order infinitesimal of ε.

Applying the above inequality to $g(z)$, we obtain

$$|g(z)| \leq 1 + (M\rho + \lambda E[V])\varepsilon + o(\varepsilon), \quad |f(z)| = (1+\varepsilon)^M = 1 + M\varepsilon + o(\varepsilon).$$

If $\rho+\lambda E[V]M^{-1}<1$, then $|f(z)|>|g(z)|$ in $|z|=1+\varepsilon$. According to the Rouche theorem, $f(z)$ and $f(z)+g(z)$ have the same number of roots in $|z|=1+\varepsilon$. Therefore, the denominator of Eq. (3.6) has M roots in $|z|=1+\varepsilon$, where one root is $z=1$, and the other $M-1$ roots are given by applying the Lagrange theorem as

$$z_m=\sum_{n=1}^{\infty}\frac{e^{\frac{2\pi mn}{M}i}}{n!}\frac{d^{n-1}}{dz^{n-1}}\left(V(1-\lambda(1-z))(S(1-\lambda(1-z)))^M\right)^{\frac{n}{M}}\bigg|_{z=0} \tag{3.7}$$

where $m=1,2,...,M-1$. Because $Q_b(z)$ is analytic in $|z|<1$, the numerator of Eq. (3.6) has the same roots. q_k satisfies the set of equations comprised of the following $M-1$ linear equations:

$$\begin{aligned}\sum_{k=0}^{M}q_k\Big((1-H(V(\bar{\lambda})))z_m^M(S(1-\lambda(1-z_m)))^kV(1-\lambda(1-z_m))+H(V(\bar{\lambda}))z_m^{M+1}\\ -(S(1-\lambda(1-z_m)))^Mz_m^kV(1-\lambda(1-z_m))\Big)=0,\quad m=1,...,M-1.\end{aligned} \tag{3.8}$$

Based on the normalization condition $Q_b(1)=1$ and by applying the L'Hospital rule in Eq. (3.6), we have

$$\frac{(M(1-\rho)+H(V(\bar{\lambda}))(1-\lambda E[V]))Q_M(1)-((1-\rho)+\rho H(V(\bar{\lambda})))Q_M'(1)}{M(1-\rho)-\lambda E[V]}=1 \tag{3.9}$$

where $Q_M'(1)$ represents the derivative values of $Q_M(z)$ at $z=1$. From Eq. (3.9), we obtain the relation between $Q_M'(1)$ and $Q_M(1)$ as follows:

$$Q_M'(1)=\frac{\lambda E[V]-M(1-\rho)}{(1-\rho)+\rho H(V(\bar{\lambda}))}+\frac{M(1-\rho)+H(V(\bar{\lambda}))(1-\lambda E[V])}{(1-\rho)+\rho H(V(\bar{\lambda}))}Q_M(1). \tag{3.10}$$

According to Eq. (3.10), we obtain the Mth equation for q_k as follows:

$$\begin{aligned}&\sum_{k=0}^{M}q_k\left(M(1-\rho)+H(V(\bar{\lambda}))(1-\lambda E[V])+k((1-\rho)+\rho H(V(\bar{\lambda})))\right)\\ =\;&M(1-\rho)-\lambda E[V].\end{aligned} \tag{3.11}$$

From Eqs. (3.3), (3.8) and (3.11), we can obtain a set of equations comprised by $M+1$ linear equations, thus we can resolve $q_k, k=0,1,...,M$ and $Q_M(z)$. Taking derivatives for both sides of Eq. (3.6) with respect to z and by applying the L'Hospital rule, the mean number $E[Q_b]$ of customers at the beginning instant of a service cycle is given by

$$\begin{aligned}E[Q_b]\;=\;&\frac{1}{2(M(1-\rho)-\lambda E[V])}\Big\{\lambda^2ME[S(S-1)]-\Big(M(M-1)(1-\rho^2)+2\lambda\rho ME[V]\\ &+\lambda^2E[V(V-1)]\Big)+Q_M''(1)((1-H(V(\bar{\lambda})))\rho^2-1)\\ &+Q_M'(1)\Big((1-H(V(\bar{\lambda})))\Big(\lambda^2E[S(S-1)]+2\lambda\rho E[V]\Big)\\ &-2\rho MH(V(\bar{\lambda}))M-2\lambda E[V]\Big)+Q_M(1)\Big((1-H(V(\bar{\lambda})))\\ &\times\Big(M(M-1)+2\lambda ME[V]+\lambda^2E[V(V-1)]\Big)\\ &-M(M-1)\rho^2-\lambda^2ME[S(S-1)]-2\lambda\rho ME[V]\\ &-\lambda^2E[V(V-1)]+M(M+1)H(V(\bar{\lambda}))\Big)\Big\}.\end{aligned} \tag{3.12}$$

Combining $\Phi = \min\{Q_b, M\}$ and Eq. (3.10), we obtain that

$$\begin{aligned} E[\Phi] &= \sum_{k=1}^{M} kq_k + M \sum_{k=M+1}^{\infty} q_k \\ &= Q'_M(1) + M(1 - Q_M(1)) \\ &= \frac{\lambda E[V] + \rho M H(V(\bar{\lambda})) + H(V(\bar{\lambda}))(1 - \rho M - \lambda E[V]) Q_M(1)}{1 - \rho(1 - H(V(\bar{\lambda})))}. \end{aligned} \tag{3.13}$$

The equilibrium condition of the system requires that the mean number of customers arriving in a service cycle is less than M. Therefore, the equilibrium condition of the system is given by

$$M - \rho E[\Phi] - \lambda E[V] - H(V(\bar{\lambda}))(1 - \lambda E[V]) > 0. \tag{3.14}$$

Based on the regeneration cycle approach and the expression of $Q_b(z)$, we obtain the stochastic decomposition structure of stationary performance measures for a general limited service Geom/G/1 queue with multiple adaptive vacations.

3.3 Stationary Queue Length and Waiting Time

Let L_v be the stationary queue length of the Geom/G/1 (GL, MAVs), i.e. $L_v = \lim_{n\to\infty} L_n$. Therefore, we can obtain the following theorem.

Theorem 3.2. *If $M - \rho E[\Phi] - \lambda E[V] - H(V(\bar{\lambda}))(1 - \lambda E[V]) > 0$ and $\rho + \lambda E[V] M^{-1} < 1$, the stationary queue length L_v in a Geom/G/1 (GL, MAVs) queue can be decomposed into three independent random variables:*

$$L_v = L + L_d + L_r$$

where L is the stationary queue length of a classical Geom/G/1 queue presented in [13], [14], its P.G.F. is

$$L(z) = \frac{(1-\rho)(1-z)S(1-\lambda(1-z))}{S(1-\lambda(1-z)) - z}. \tag{3.15}$$

L_d is an additional queue length of the Geom/G/1 queue with multiple adaptive vacations. L_r is another additional queue length resulting from the general limited service policy. P.G.Fs. $L_d(z)$ and $L_r(z)$ of the additional queue lengths L_d and L_r are given by

$$L_d(z) = \frac{1 - zH(V(\bar{\lambda})) - \dfrac{1 - H(V(\bar{\lambda}))}{1 - V(\bar{\lambda})}(V(1-\lambda(1-z)) - V(\bar{\lambda}))}{\left(H(V(\bar{\lambda})) + \dfrac{1 - H(V(\bar{\lambda}))}{1 - V(\bar{\lambda})}\lambda E[V]\right)(1-z)}, \tag{3.16}$$

$$\begin{aligned} L_r(z) = {} & \frac{\beta}{z^M - (S(1-\lambda(1-z)))^M V(1-\lambda(1-z))} \\ & \times \frac{1}{1 - V(1-\lambda(1-z)) + H(V(\bar{\lambda}))\Big(V(1-\lambda(1-z)) - (1-z)V(\bar{\lambda})) - z\Big)} \\ & \times \Big(Q_M(S(1-\lambda(1-z)))\left(z^M(1 - V(1-\lambda(1-z)))\right. \\ & + H(V(\bar{\lambda}))V(1-\lambda(1-z))(z^M - (S(1-\lambda(1-z)))^M)) \\ & - Q_M(z)(S(1-\lambda(1-z)))^M(1 - V(1-\lambda(1-z))) \\ & + H(V(\bar{\lambda}))Q_M(1)z\left((S(1-\lambda(1-z)))^M - z^M\right)\Big). \end{aligned} \tag{3.17}$$

Proof. Because L_n is the number of customers at the departure instant of the nth customer in a service cycle, and A_k represents the number of customers arriving in the kth service period, then we have

$$L_n = Q_b - n + \sum_{k=1}^{n} A_k, \quad n = 1, 2, ..., \Phi.$$

Substituting Eq. (3.13) and $E\left[\sum_{n=1}^{\Phi} z^{L_n}\right]$ into Eq. (3.1) of Lemma 3.1 (regeneration cycle approach), we obtain

$$\begin{aligned}
L_v(z) &= \frac{E\left[\sum_{n=1}^{\Phi} z^{L_n}\right]}{E[\Phi]} \\
&= \frac{(1-\rho)(1-z)S(1-\lambda(1-z))}{S(1-\lambda(1-z))-z} \\
&\quad \times \frac{1 - zH(V(\bar{\lambda})) - \frac{1-H(V(\bar{\lambda}))}{1-V(\bar{\lambda})}(V(1-\lambda(1-z)) - V(\bar{\lambda}))}{\left(H(V(\bar{\lambda})) + \frac{1-H(V(\bar{\lambda}))}{1-V(\bar{\lambda})}\lambda E[V]\right)(1-z)} \\
&\quad \times \frac{\beta}{z^M - (S(1-\lambda(1-z)))^M V(1-\lambda(1-z))} \\
&\quad \times \frac{1}{1 - V(1-\lambda(1-z)) + H(V(\bar{\lambda}))} \\
&\quad \times \frac{1}{1 - V(1-\lambda(1-z)) + H(V(\bar{\lambda}))\Big(V(1-\lambda(1-z)) - (1-V(\bar{\lambda}))z - V(\bar{\lambda})\Big)} \\
&\quad \times \Big(Q_M(S(1-\lambda(1-z)))\left(z^M(1-V(1-\lambda(1-z)))\right. \\
&\quad \left. + H(V(\bar{\lambda}))V(1-\lambda(1-z))(z^M - (S(1-\lambda(1-z)))^M)\right) \\
&\quad - Q_M(z)(S(1-\lambda(1-z)))^M(1-V(1-\lambda(1-z))) \\
&\quad + H(V(\bar{\lambda}))Q_M(1)z\left((S(1-\lambda(1-z)))^M - z^M\right)\Big) \\
&= L(z)L_d(z)L_r(z) \qquad (3.18)
\end{aligned}$$

where $\beta = \dfrac{\left(1-\rho(1-H(V(\bar{\lambda})))\right)\left[\lambda E[V] + H(V(\bar{\lambda}))(1-V(\bar{\lambda})-\lambda E[V])\right]}{(1-\rho)\left[\lambda E[V] + \rho M H(V(\bar{\lambda})) + H(V(\bar{\lambda}))Q_M(1)(1-\rho M - \lambda E[V])\right]}$. □

Taking a derivative of $L_v(z)$ with respect to z, then applying the L'Hospital rule repeatedly twice and letting $z = 1$, we obtain the mean additional queue length $E[L_v]$ as follows:

$$\begin{aligned}
E[L_v] &= E[L] + E[L_d] + E[L_r] \\
&= \rho + \frac{\lambda^2 E[S(S-1)]}{2(1-\rho)} + \frac{1-H(V(\bar{\lambda}))}{1-V(\bar{\lambda})} \frac{\lambda^2 E[V(V-1)]}{2\left(H(V(\bar{\lambda})) + \frac{1-H(V(\bar{\lambda}))}{1-V(\bar{\lambda})}\lambda E[V]\right)} + E[L_r]
\end{aligned} \qquad (3.19)$$

where

$$\begin{aligned} E[L_r] &= \frac{A}{2[M(1-\rho)+\lambda E[V]][\lambda E[V](H(V(\bar{\lambda}))-1)+V(\bar{\lambda})-1]}, \\ A &= \rho Q_m'(1)\Big[-2\lambda E[V]+2MH(V(\bar{\lambda}))(1-\rho)-2\lambda ME[V]-\lambda^2 E[V(V-1)]-2\frac{E[V]}{E[S]} \\ &\quad +2\lambda MH(V(\bar{\lambda}))E[V](1-\rho)+MH(V(\bar{\lambda}))[(M-1)(1-\rho^2)-\lambda E[S(S-1)]]\Big] \\ &\quad +Q_M(1)\Big[MH(V(\bar{\lambda}))(\rho^2 M-M+1)-2MH(V(\bar{\lambda}))(1-\rho) \\ &\quad -2M\rho E[V]-\lambda^2 E[V(V-1)]\Big]. \end{aligned} \tag{3.20}$$

Based on the stochastic decomposition result of the queue length L_v and a classical relation, we can prove the stochastic decomposition result of the waiting time. The classical relation is that the number of customers in the system at the departure instant of a customer is equal to the number of customers arriving in the sojourn time of that customer.

Theorem 3.3. *If $M-\rho E[\Phi]-\lambda E[V]-H(V(\bar{\lambda}))(1-\lambda E[V])>0$ and $\rho+\lambda E[V]M^{-1}<1$, the stationary waiting time W_v in a Geom/G/1 (GL, MAVs) queue can be decomposed into three independent random variables:*

$$W_v = W + W_d + W_r$$

where W is the stationary waiting time of a classical Geom/G/1 queue presented in [13], [14].

W_d is an additional delay of the Geom/G/1 queue with multiple adaptive vacations and W_r is another additional delay resulting from the general limited service policy. P.G.Fs. $W_d(z)$ and $W_r(z)$ of the additional delays W_d and W_r are given by

$$W_d(z) = \frac{\lambda - H(V(\bar{\lambda}))(z-\bar{\lambda})-\lambda\frac{1-H(V(\bar{\lambda}))}{1-V(\bar{\lambda})}(V(z)-V(\bar{\lambda}))}{\left(H(V(\bar{\lambda}))+\frac{1-H(V(\bar{\lambda}))}{1-V(\bar{\lambda})}\lambda E[V]\right)(1-z)}, \tag{3.21}$$

$$\begin{aligned} W_r(z) &= \frac{\beta}{(z-\bar{\lambda})^M-(\lambda S(z))^M V(z)} \\ &\quad \times\frac{1}{\lambda+H(V(\bar{\lambda}))(\bar{\lambda}-V(\bar{\lambda}))-\lambda V(z)(1-H(V(\bar{\lambda})))+H(V(\bar{\lambda}))(1-V(\bar{\lambda}))z} \\ &\quad \times\Big(zQ_M(S(z))\left((z-\bar{\lambda})^M(1-V(z))+H(V(\bar{\lambda}))V(z)((z-\bar{\lambda})^M\right. \\ &\quad \left.-(\lambda S(z))^M)\right)-\lambda^{M+1}Q_M\left(\frac{z-\bar{\lambda}}{\lambda}\right)(S(z))^M(1-V(z)) \\ &\quad +H(V(\bar{\lambda}))Q_M(1)(z-\bar{\lambda})((\lambda S(z))^M-(z-\bar{\lambda})^M)\Big). \end{aligned} \tag{3.22}$$

Proof. Because the waiting time is independent of the input process after the arrival instant, the number of residual customers in the system after the departure instant is equal to the sum of the number of customers arriving in the waiting time W_v and the service time S for the Geom/G/1 (GL, MAVs) model. Also, because of the independent increment property of the input process which follows a binomial distribution, the number of customers arriving in the waiting time and the service time are mutually independent. We then have

$$L_v(z) = W_v(1-\lambda(1-z))S(1-\lambda(1-z)). \tag{3.23}$$

Substituting $L_v(z)$ in Theorem 3.2 into Eq. (3.23), and letting $z' = 1-\lambda(1-z)$, then displacing z' with z, we can obtain the P.G.Fs. $W_d(z)$ and $W_r(z)$ of the additional delays W_d and W_r given by Eqs. (3.21) and (3.22), respectively. □

Because the inter-arrival time follows Geometric distribution, by using Little Law $E[W_v] = \frac{1}{\lambda}E[L_q]$, where $E[L_q]$ is the mean waiting queue length, i.e. $E[L_q] = E[L_v]-\rho$, we can obtain the mean waiting time $E[W_v]$ as follows:

$$\begin{aligned} E[W_v] &= \frac{1}{\lambda}(E[L_v]-\rho) \\ &= \frac{\lambda E[S(S-1)]}{2(1-\rho)} + \frac{1-H(V(\bar{\lambda}))}{1-V(\bar{\lambda})}\frac{\lambda E[V(V-1)]}{2\left(H(V(\bar{\lambda}))+\dfrac{1-H(V(\bar{\lambda}))}{1-V(\bar{\lambda})}\lambda E[V]\right)} + E[W_r] \end{aligned} \tag{3.24}$$

where $E[W_r] = \frac{1}{\lambda}E[L_r]$ and $E[L_r]$ is given in Eq. (3.20).

3.4 Analysis of the Service Cycle

According to the definition of the consecutive vacations (denoted by $V_k, k = 1, 2, ..., H$ are i.i.d. random variables with the distribution function, the finite first and second moments) and the number J of consecutive vacations presented in [14] and [20], then the P.G.F. $J(z)$ of J can be obtained as follows:

$$\begin{aligned} &P(J \geq 1) = 1, \quad P(J \geq j) = (V(\bar{\lambda}))^{j-1}\sum_{k=j}^{\infty} h_k, \quad j \geq 2, \\ &J(z) = 1 - \frac{1-z}{1-V(\bar{\lambda})z}(1-H(V(\bar{\lambda})z)). \end{aligned} \tag{3.25}$$

The P.G.F. $V_G(z)$ and the mean vacation time $E[V_G]$ of V_G, which is the whole time length for every two consecutive vacations, are given as follows:

$$\begin{aligned} V_G(z) &= 1 - \frac{1-V(z)}{1-V(\bar{\lambda})V(z)}(1-H(V(\bar{\lambda})V(z))), \\ E[V_G] &= \frac{1-H(V(\bar{\lambda}))}{1-V(\bar{\lambda})}E[V]. \end{aligned} \tag{3.26}$$

In a general limited service Geom/G/1 queue with multiple adaptive vacations, the server may stay in an idle period. The idle period is equal to zero at the completion instant of J vacations when one of the following three cases holds: (i) there are customers arriving in the service period; (ii) there are no customers arriving in the service period, but there are customers arriving during the vacation period; (iii) there are no customers arriving in the service period or the vacation period, but there are residual customers that have resulted from a general limited policy.

The idle period is equal to an inter-arrival time if there are no customers arriving at the completion instant of J vacations and there are no residual customers resulting from a general limited policy. Let I_v be the time length of the idle period, the mean $E[I_v]$ of I_v is given as follows:

$$E[I_v] = \frac{(1-\rho)\beta H(V(\bar{\lambda}))\left(H(V(\bar{\lambda}))V(\bar{\lambda})Q_M(S(\bar{\lambda})) + Q_M(0)(1-V(\bar{\lambda}))\right)}{\lambda V(\bar{\lambda})\left(H(V(\bar{\lambda}))(1-V(\bar{\lambda})) + \lambda E[V](1-H(V(\bar{\lambda})))\right)}. \tag{3.27}$$

The mean busy period is given as follows:

$$E[S_\lambda] = \frac{\lambda E[V] + \rho M H(V(\bar{\lambda})) + H(V(\bar{\lambda}))Q_M(1)(1-M\rho-\lambda E[V])}{\mu(1-\rho(1-H(V(\bar{\lambda}))))}. \tag{3.28}$$

The mean service cycle $E[C_s]$ of C_s is given as follows:

$$\begin{aligned} E[C_s] &= E[V_G] + E[I_v] + E[S_\lambda] \\ &= \frac{\lambda E[V] + \rho M H(V(\bar{\lambda})) + H(V(\bar{\lambda}))Q_M(1)(1-M\rho-\lambda E[V])}{\mu(1-\rho(1-H(V(\bar{\lambda}))))} \\ &\quad + \frac{(1-\rho)\beta H(V(\bar{\lambda}))\left(H(V(\bar{\lambda}))V(\bar{\lambda})Q_M(S(\bar{\lambda})) + Q_M(0)(1-V(\bar{\lambda}))\right)}{\lambda V(\bar{\lambda})\left(H(V(\bar{\lambda}))(1-V(\bar{\lambda})) + \lambda E[V](1-H(V(\bar{\lambda})))\right)} \\ &\quad + \frac{1-H(V(\bar{\lambda}))}{1-V(\bar{\lambda})}E[V]. \end{aligned} \tag{3.29}$$

Let p_B, p_V and p_I be the state probabilities that the server is in a busy, vacation or idle state, respectively. We can give that

$$\begin{aligned} p_B &= \frac{\lambda E[V] + \rho M H(V(\bar{\lambda})) + H(V(\bar{\lambda}))Q_M(1)(1-M\rho-\lambda E[V])}{\mu(1-\rho(1-H(V(\bar{\lambda}))))E[C_s]}, \\ p_V &= \frac{E[V](1-H(V(\bar{\lambda})))}{E[C_s](1-V(\bar{\lambda}))}, \\ p_I &= \frac{(1-\rho)\beta H(V(\bar{\lambda}))\left(H(V(\bar{\lambda}))V(\bar{\lambda})Q_M(S(\bar{\lambda})) + Q_M(0)(1-V(\bar{\lambda}))\right)}{\lambda V(\bar{\lambda})E[C_s]\left[H(V(\bar{\lambda}))(1-V(\bar{\lambda})) + \lambda E[V](1-H(V(\bar{\lambda})))\right]}. \end{aligned} \tag{3.30}$$

If the random variables H and M are supposed to have different probability distributions, we can derive some vacation queueing systems as special cases of the model presented in this paper as follows.

3.5 Special Cases

Since $L_v(z)$ and $W_v(z)$ are very complicated, the variances $D[L_v]$ and $D[W_v]$ of L_v and W_v will be more complicated. In order to simplify, we assume that the random variables H and M to have the special probability distribution.

(1) Let $H \to \infty$ and $M \to \infty$, the queue turns into a Geom/G/1 queue with multiple vacations. There is no idle state in the system, and $H(z) = 0$. Then the P.G.Fs. of the additional queue length and the additional delay correspond with the results given in [13], [14]. Taking a derivative of $L_v(z)$ with respect to z, then applying the L'Hospital rule repeatedly twice and letting $z = 1$, we can obtain the mean additional queue length $E[L_{mv}]$. Taking the two order derivative of $L_v(z)$ with respect to z, then applying the L'Hospital

rule repeatedly three times and letting $z = 1$, we can obtain the variance $D[L_{mv}]$. They are as follows:

$$
\begin{aligned}
E[L_{mv}] &= \rho + \frac{\lambda^2 E[S(S-1)]}{2(1-\rho)} + \frac{\lambda E[V(V-1)]}{2E[V]}, \qquad (3.31)\\
D[L_{mv}] &= L''(z)|_{z=1} - (E[L])^2 + E[L] + L''_d(z)|_{z=1} - (E[L_d])^2 + E[L_d]\\
&= \rho + \frac{3\lambda^2 E[S(S-1)]}{2(1-\rho)} + \frac{\lambda^2 E[S(S-1)(S-2)]}{3(1-\rho)^2} - \left(\rho + \frac{\lambda^2 E[S(S-1)]}{2(1-\rho)}\right)^2\\
&\quad + \frac{1}{2}\left(\frac{\lambda E[S(S-1)]}{1-\rho}\right)^2 + \frac{\lambda^2 E[V(V-1)(V-2)]}{3E[V]}\\
&\quad - \left(\frac{\lambda E[V(V-1)]}{2E[V]}\right)^2 + \frac{\lambda E[V(V-1)]}{2E[V]} \qquad (3.32)
\end{aligned}
$$

where

$$
E[S(S-1)(S-2)] = \frac{d^3 S(z)}{dz^3}\Big|_{z=1}, \quad E[V(V-1)(V-2)] = \frac{d^3 V(z)}{dz^3}\Big|_{z=1}.
$$

Coefficient $CV(L_{mv})$ of Variation of L_{mv} is given by

$$
CV(L_{mv}) = \frac{\sqrt{D[L_{mv}]}}{E[L_{mv}]} = \frac{\sqrt{D[L_{mv}]}}{\rho + \dfrac{\lambda^2 E[S(S-1)]}{2(1-\rho)} + \dfrac{\lambda E[V(V-1)]}{2E[V]}}. \qquad (3.33)
$$

Taking a derivative of $W_v(z)$ with respect to z, then applying the L'Hospital rule repeatedly twice and letting $z = 1$, we can obtain the mean waiting time $E[W_{mv}]$. Taking the two order derivative of $W_v(z)$ with respect to z, then applying the L'Hospital rule repeatedly three times and letting $z = 1$, we can obtain the variance $D[W_{mv}]$. They are as follows:

$$
\begin{aligned}
E[W_{mv}] &= \frac{\lambda E[S(S-1)]}{2(1-\rho)} + \frac{E[V(V-1)]}{2E[V]}, \qquad (3.34)\\
D[W_{mv}] &= W''(z)|_{z=1} - (E[W])^2 + E[W] + W''_d(z)|_{z=1} - (E[W_d])^2 + E[W_d]\\
&= \frac{\lambda E[S(S-1)(S-2)]}{3(1-\rho)} + \frac{1}{4}\left(\frac{\lambda E[S(S-1)]}{(1-\rho)}\right)^2 + \frac{\lambda E[S(S-1)]}{2(1-\rho)}\\
&\quad + \frac{E[V(V-1)(V-2)]}{3E[V]} - \left(\frac{E[V(V-1)]}{2E[V]}\right)^2 + \frac{E[V(V-1)]}{2E[V]}. \qquad (3.35)
\end{aligned}
$$

Coefficient $CV(W_{mv})$ of Variation of W_{mv} is given by

$$
CV(W_{mv}) = \frac{\sqrt{D[W_{mv}]}}{E[W_{mv}]} = \frac{\sqrt{D[W_{mv}]}}{\dfrac{\lambda E[S(S-1)]}{2(1-\rho)} + \dfrac{E[V(V-1)]}{2E[V]}}. \qquad (3.36)
$$

(2) Let $H \equiv 1$ and $M \to \infty$, the queue turns into a Geom/G/1 queue with a single vacation. There is an idle state in the system, and $H(z) = z$. Then the P.G.Fs. of the additional queue length and the additional delay correspond with the results given in [13], [14]. Taking a derivative of $L_v(z)$ with respect to z, then applying the L'Hospital rule repeatedly twice and letting $z = 1$, we can obtain the mean additional queue length

$E[L_{sv}]$. Taking the two order derivative of $L_v(z)$ with respect to z, then applying the L'Hospital rule repeatedly three times and letting $z = 1$, we can obtain the variance $D[L_{sv}]$. They are as follows:

$$
\begin{aligned}
E[L_{sv}] &= \rho + \frac{\lambda^2 E[S(S-1)]}{2(1-\rho)} + \frac{\lambda^2 E[V(V-1)]}{2[V(\bar{\lambda}) + \lambda E[V]]}, \qquad (3.37)\\
D[L_{sv}] &= L''(z)|_{z=1} - (E[L])^2 + E[L] + L''_d(z)|_{z=1} - (E[L_d])^2 + E[L_d]\\
&= \rho + \frac{3\lambda^2 E[S(S-1)]}{2(1-\rho)} + \frac{\lambda^2 E[S(S-1)(S-2)]}{3(1-\rho)^2} - \left(\rho + \frac{\lambda^2 E[S(S-1)]}{2(1-\rho)}\right)^2\\
&\quad + \frac{1}{2}\left(\frac{\lambda E[S(S-1)]}{1-\rho}\right)^2 + \frac{\lambda^3 E[V(V-1)(V-2)]}{3[V(\bar{\lambda}) + \lambda E[V]]}\\
&\quad - \left(\frac{\lambda^2 E[V(V-1)]}{2[V(\bar{\lambda}) + \lambda E[V]]}\right)^2 + \frac{\lambda^2 E[V(V-1)]}{2[V(\bar{\lambda}) + \lambda E[V]]}. \qquad (3.38)
\end{aligned}
$$

Coefficient $CV(L_{sv})$ of Variation of L_{sv} is given by

$$
CV(L_{sv}) = \frac{\sqrt{D[L_{sv}]}}{E[L_{sv}]} = \frac{\sqrt{D[L_{mv}]}}{\rho + \frac{\lambda^2 E[S(S-1)]}{2(1-\rho)} + \frac{\lambda^2 E[V(V-1)]}{2[V(\bar{\lambda}) + \lambda E[V]]}}. \qquad (3.39)
$$

Taking a derivative of $W_v(z)$ with respect to z, then applying the L'Hospital rule repeatedly twice and letting $z = 1$, we can obtain the mean waiting time $E[W_{sv}]$. Taking the two order derivative of $W_v(z)$ with respect to z, then applying the L'Hospital rule repeatedly three times and letting $z = 1$, we can obtain the variance $D[W_{sv}]$. They are as follows:

$$
\begin{aligned}
E[W_{sv}] &= \frac{\lambda E[S(S-1)]}{2(1-\rho)} + \frac{\lambda E[V(V-1)]}{2[V(\bar{\lambda}) + \lambda E[V]]}, \qquad (3.40)\\
D[W_{sv}] &= W''(z)|_{z=1} - (E[W])^2 + E[W] + W''_d(z)|_{z=1} - (E[W_d])^2 + E[W_d]\\
&= \frac{\lambda E[S(S-1)(S-2)]}{3(1-\rho)} + \frac{1}{4}\left(\frac{\lambda E[S(S-1)]}{(1-\rho)}\right)^2 + \frac{\lambda E[S(S-1)]}{2(1-\rho)}\\
&\quad + \frac{\lambda E[V(V-1)(V-2)]}{3[V(\bar{\lambda}) + \lambda E[V]]} - \left(\frac{\lambda E[V(V-1)]}{2[V(\bar{\lambda}) + \lambda E[V]]}\right)^2 + \frac{\lambda E[V(V-1)]}{2[V(\bar{\lambda}) + \lambda E[V]]}. \qquad (3.41)
\end{aligned}
$$

Coefficient $CV(W_{sv})$ of Variation of W_{sv} is given by

$$
CV(W_{sv}) = \frac{\sqrt{D[W_{sv}]}}{E[W_{sv}]} = \frac{\sqrt{D[W_{sv}]}}{\frac{\lambda E[S(S-1)]}{2(1-\rho)} + \frac{E[V(V-1)]}{2E[V]}}. \qquad (3.42)
$$

4 Performance Optimization

In this section, we develop an expected cost function per unit time for Geom/G/1 queue with general limited service and MAVs. Our objective is to determine the optimal expected traffic intensity, and to minimize the total expected cost. Firstly, we define the following parameters:

C_1 = Cost of unit time when the system is on the busy period,

C_2 = Cost of unit time when the system is on the vacation period,
C_3 = Cost of unit time when the system is on the idle period.
Therefore, the total expected cost is given as follows:

$$\begin{aligned} F(E[V],H,\rho) &= C_1E[S_\lambda]+C_2E[V_G]+C_3E[I_v] \\ &= C_1\frac{\lambda E[V]+\rho MH(V(\bar{\lambda}))+H(V(\bar{\lambda}))Q_M(1)(1-M\rho-\lambda E[V])}{\mu(1-\rho(1-H(V(\bar{\lambda}))))} \\ &\quad +C_3\frac{(1-\rho)\beta H(V(\bar{\lambda}))\left(H(V(\bar{\lambda}))V(\bar{\lambda})Q_M(S(\bar{\lambda}))+Q_M(0)(1-V(\bar{\lambda}))\right)}{\lambda V(\bar{\lambda})\left(H(V(\bar{\lambda}))(1-V(\bar{\lambda}))+\lambda E[V](1-H(V(\bar{\lambda})))\right)} \\ &\quad +C_2\frac{1-H(V(\bar{\lambda}))}{1-V(\bar{\lambda})}E[V] \end{aligned} \tag{4.1}$$

where $E[V]$, λ, μ, H and M are decision variables that we can give them as parameters of the system. The values of the parameters λ, H and M in a queueing service system are not usually to be controlled due to the outside factors of the system. For some purposes, $E[V]$ and ρ can be viewed as controllable variables. Therefore, we think that $F(E[V],\rho)$ is the function of the mean vacation time length $E[V]$ and ρ, then we search the optimal cost function about $E[V]$ and ρ.

Since $E[V]$ and ρ are controllable variables, and they can be changed by the management of the system, $F(E[V],\rho)$ follows changes when the mean vacation time length $E[V]$ and ρ changes. We will find the minimum cost $F(E[V],\rho)$ about ρ, so that we can find the optimal value of service rate, and provide a useful performance evaluation tool for more general situations arising in practical applications such as the manufacturing system and other related systems.

We use a direct search method to obtain the optimal value ρ such that the expected cost function $F(E[V],\rho)$ attains a minimum value, say $F(E[V]^*,\rho^*)$. The procedure for finding the optimal solution is described in the following.

Step 1: For fixed values λ, M and H, find the minimum value $F(E[V]^*,\rho)$ for threshold value ρ, i.e.,

$$\min_{E[V]} F(E[V],\rho) = F(E[V]^*,\rho).$$

Step 2: Find the set of all minimum cost solutions for $\rho=\rho_{min},...,\rho_{max}$,

$$\Theta = \{F(E[V]^*,\rho) : \rho=\rho_{min},...,\rho_{max}\}.$$

Step 3: Find the minimum cost solution for $\rho=\rho_{min},...,\rho_{max}$,

$$\min_{\rho} \Theta = F(E[V]^*,\rho^*).$$

Due to the random variable H can be supposed to have different probability distributions, $F(E[V],\rho)$ will have different forms. Therefore, we can derive the cost function $F(E[V],\rho)$ to obtain the best traffic intensity ρ^* and the minimum cost by the analysis presented above.

5 Numerical Results

In this section, we present some numerical results that provide insight into the system behavior. Using the equations presented in Section 3, we can numerically compare the performance measures of the systems for two different Geom/G/1 (GL, MAVs) queueing models: the Geom/G/1 queue with multiple vacations and the Geom/G/1 queue with single vacation. Here we assume that the service time S and the time length V of an arbitrary vacation follow geometric distributions, i.e., S follows a geometric distribution with parameter $\mu = 1/5$. V follows another geometric distribution with parameters 1/3 and 1/6. Suppose that the range of the traffic intensity ρ is from 0.1 to 0.9.

In all figures in this section, "MV" represents that the numerical results are for the Geom/G/1 queueing model with multiple vacations and "SV" represents that the numerical results are for the Geom/G/1 queueing model with single vacation.

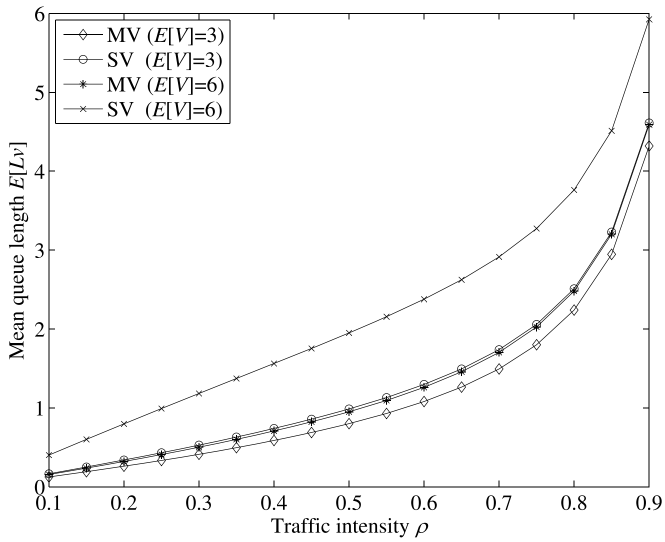

Figure 2: Mean queue length versus traffic intensity ρ.

Figure 2 shows how the mean queue length $E[L_v]$ changes with the traffic intensity ρ with $E[V] = 3$ and 6 for two cases of vacation policy, i.e., multiple vacations and single vacation. We can find that when ρ increases, the mean queue length $E[L_v]$ increases. This is because the greater the traffic intensity ρ is, the higher the possibility is that there will be customers arriving during the service cycle. We note that the mean queue length $E[L_v]$ of the Geom/G/1 (E, MV) queue is smaller than $E[L_v]$ of the Geom/G/1 (E, SV) queue. And we also find that the mean queue lengths of two queue models with $E[V] = 6$ are larger than that of queue models with $E[V] = 3$.

Figure 3 shows how the mean waiting time $E[W_v]$ changes with the traffic intensity ρ with $E[V] = 3$ and 6 for two cases of vacation policy, i.e., multiple vacations and single vacation. We can find that when ρ increases, the mean waiting time $E[W_v]$ decreases slowly, then increases fastly for the single vacation queue with $E[V] = 6$. However, the mean waiting time $E[W_v]$ increases for other three cases. This is because the greater the traffic intensity ρ is, the higher the possibility is that there will be customers arriving during the service cycle. We note that the mean waiting time $E[W_v]$ of the Geom/G/1 (E, MV) queue is smaller than $E[W_v]$ of the Geom/G/1 (E, SV) queue. We also find that the mean waiting times of two queue models with $E[V] = 6$ are larger than that of queue models with $E[V] = 3$.

Figure 4 shows how coefficient of variation $CV[L_v]$ changes with the traffic intensity ρ with $E[V] = 3$ and 6 for two cases of vacation policy, i.e., multiple vacations and single

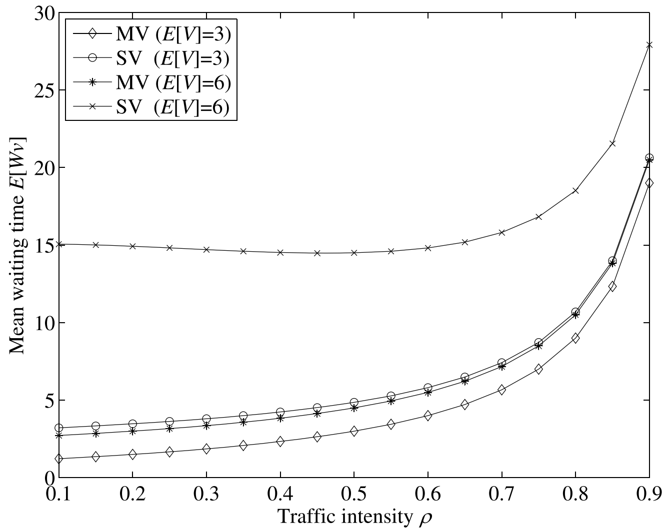

Figure 3: Mean waiting time versus traffic intensity ρ.

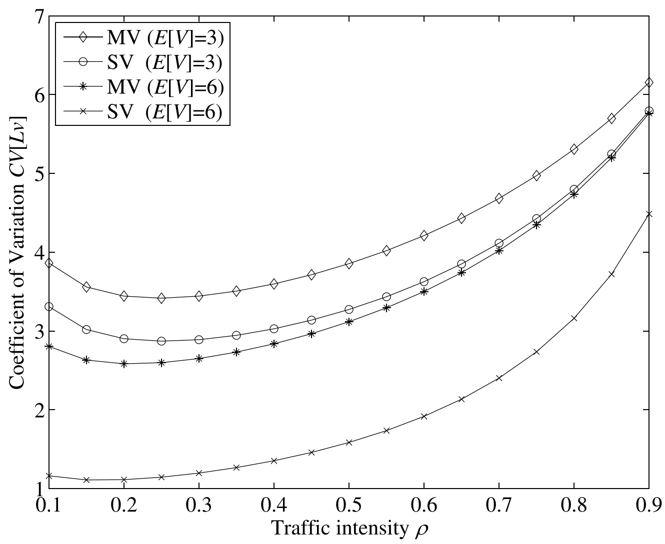

Figure 4: Coefficient of variation $CV[L_v]$ versus traffic intensity ρ.

vacation. We can find that when ρ increases, coefficient of variation $CV[L_v]$ decreases slowly firstly, then increases fast.

Figure 5 shows how coefficient of variation $CV[W_v]$ changes with the traffic intensity ρ with $E[V] = 3$ and 6 for two cases of vacation policy, i.e., multiple vacations and single vacation. We can find that when ρ increases, coefficient of variation $CV[W_v]$ decreases slowly firstly, then increases fast. We note that the coefficient of variation $CV[W_v]$ of the Geom/G/1 (E, MV) queue is larger than $CV[W_v]$ of the Geom/G/1 (E, SV) queue. We also find that the coefficient of variation $CV[W_v]$ of two queue models with $E[V] = 3$ is larger than that of queue models with $E[V] = 6$.

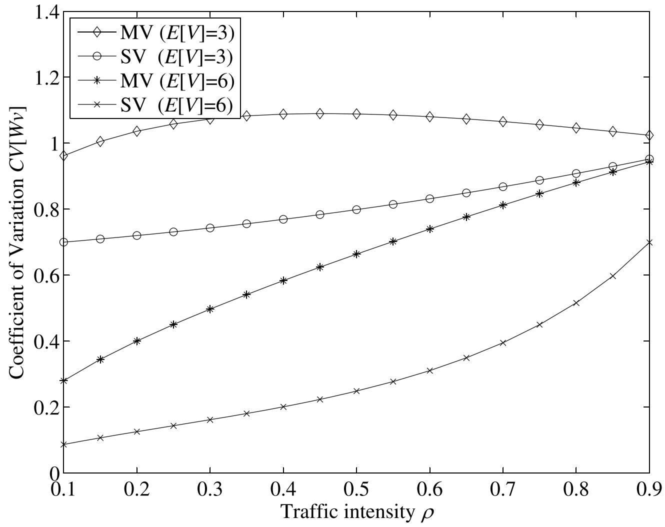

Figure 5: Coefficient of variation $CV[W_v]$ versus traffic intensity ρ.

Figure 6 shows how the probabilities p_B, p_V and p_I change with the traffic intensity ρ for two cases of vacation policy, i.e., multiple vacations and single vacation. We can

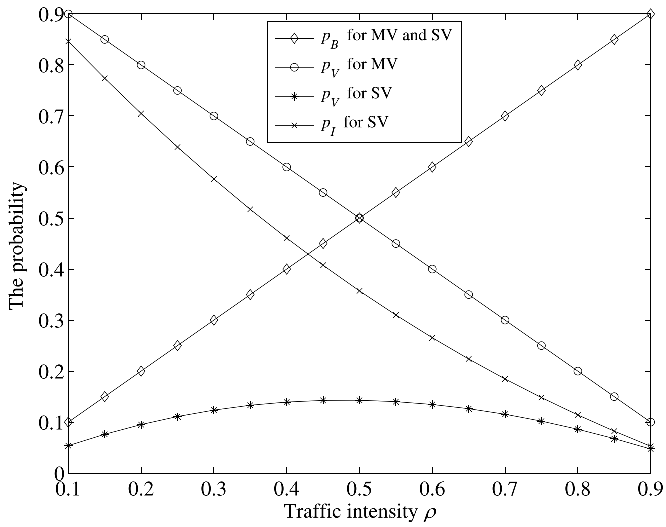

Figure 6: The probability versus traffic intensity ρ.

find that when ρ increases, the probability p_B that the server is in a busy state increases. This is because this probability p_B equals to ρ for two cases of vacation policy. When ρ increases, the probability p_V that the server is in a vacation state decreases for the case with multiple vacations (MV). This is because the higher the possibility that there will be customers arriving during the service cycle will be, therefore the mean vacation time will be shorter. When ρ increases, p_I that the server is an idle state decreases, while p_V for the case with singe vacation (SV) increases firstly, then decreases. This is because the higher the possibility that there will be customers arriving during the service cycle will be, almost there are no vacation time and idle time for the server.

To perform a sensitivity analysis on the optimal traffic intensity ρ^* and the cost function $F(E[V], \rho)$ with the changes of the traffic intensity ρ, we show the cost function $F(E[V], \rho)$ in Fig. 7. We assume $H \to \infty$ and $M \to \infty$, the queue turns into a Geom/G/1 queue with multiple vacations, and $C_1 = 10$, $C_2 = 7$, $C_3 = 2$.

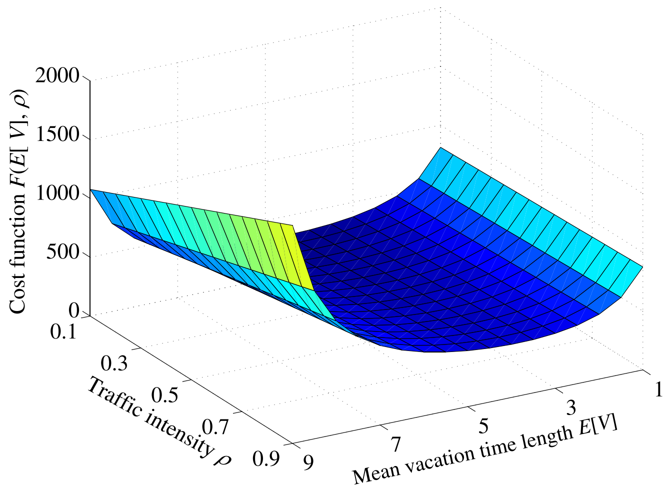

Figure 7: Mean cost function versus traffic intensity ρ.

Figure 7 shows how the cost function $F(E[V], \rho)$ changes with mean vacation time length $E[V]$ and the traffic intensity ρ. When the traffic intensity ρ is fixed, the cost function $F(E[V], \rho)$ firstly decreases as $E[V]$ increases, then increases as $E[V]$ increases. $E[V]$ significantly affects the cost function $F(E[V], \rho)$, and the minimum cost is $F(4.5, 0.1) = 337.37$. When the mean vacation time length $E[V]$ is fixed, the cost function $F(E[V], \rho)$ increases as the traffic intensity ρ increases. This is because the larger the traffic intensity ρ is, the more consecutive the traffic will be, then the less the total vacation time is, therefore $F(E[V], \rho)$ becomes larger. Figure 7 shows that the larger $F(E[V], \rho)$ is, the less the Quality of Service (QoS) of the system is.

These properties above provide the theoretical basis for the management and the optimization of the queueing system. We also can select the parameters C_1-C_3 as many other cases. For different parameters, we can show different results of the optimal service rate, the cost function $F(E[V], \rho)$ and different values.

6 Conclusions

In this paper, we presented in detail an analysis of a general limited service Geom/G/1 queue with multiple adaptive vacations. We derived the P.G.Fs. of the stationary queue length and the waiting time by using the embedded Markov chain method and the regeneration cycle approach. Furthermore, we obtained the probabilities of the server being in various states: busy, vacation and idle, respectively. We calculated their averages and higher moments of L_v and W_v by differentiating the P.G.Fs. to numerically compare different system cases, such as $D[L_v]$, $D[W_v]$, $CV(L_v)$ and $CV(W_v)$ for multiple vacations and single vacation. We also presented a cost function $F(E[V],\rho)$ that is to determine the optimal expected traffic intensity in the system. Therefore we can obtain the optimal value of service rate and the optimal management policy. In this model, in order to be simple, we assumed the service time and the vacation time to follow geometric distributions. In fact, they can also be assumed to follow Deterministic distribution, Logarithmic distribution and Poisson distribution and so on. Because the multiple vacation and the single vacation are the most typical vacation policies, we also studied two vacation policies to be the special cases of our model. Finally, we gave some special cases, in order to verify the conclusion that the choice of the vacation length and ρ is significant in improving performance of the system, we can also get the optimal service rate with the cost function $F(E[V],\rho)$.

References

[1] A. Banik, Stationary distributions and optimal control of queues with batch Markovian arrival process under multiple adaptive vacations, *Computers & Industrial Engineering* 65 (2013) 455–465.

[2] K. Bharath-Kumar, Discrete time queueing systems and their networks, *IEEE Transactions on Communications* COM-28 (1980) 260–263.

[3] O. Boxma, Workloads and waiting times in single-server with multiple customer class, *Queueing Systems* 5 (1989) 185–214.

[4] C. Chang, F. Chang and J. Ke, Economic application in a Bernoulli F-policy queueing system with server breakdown, *International Journal of Production Research* 52 (2014) 743–756.

[5] Y. Chang and W. Pearn, *Optimal management for infinite capacity N-policy M/G/1 queue with a removable service station, International Journal of Systems Science* 42 (2011) 1075–1083.

[6] S. Fuhrmann and R. Cooper, Stochastic decompositions in the M/G/1 queue with generalized vacations, *Operations Research* 33 (1985) 1117–1129.

[7] J. Hunter, *Mathematical Techniques of Applied Probability,* Academic Press, New York, 1983.

[8] H. Kobayashi and A. Konheim, Queueing models for computer communication system analysis, *IEEE Transactions on Communications* COM-25 (1977) 1–29.

[9] Z. Ma, W. Yue and N. Tian, Analysis of a Geom/G/1 queue with general limited service and MAVs, *Proc. of the 7th International Symposium on Operations Research and Its Applications*, 2008, pp. 359–368.

[10] T. Meisling, Discrete time queueing theory, *Operations Research* 6 (1985) 96–105.

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

[12] H. Takagi, Mean message waiting time in a symmetric polling system, *Performance Evaluation* 5 (1985) 293–302.

[13] H. Takagi, *Queueing Analysis, Vol. 3: Discrete-Time Systems,* Elsevier Science Publishers, Amsterdam, 1993.

[14] N. Tian and G. Zhang, *Vacation Queueing Models–Theory and Applications,* Springer, New York, 2006.

[15] E. Titchmarsh, *Theory of Functions,* Oxford University Press, London, 1952.

[16] K. Wang, Y. Liou and D. Yang, Cost optimization and sensitivity analysis of the machine repair problem with variable servers and balking, *Procedia-Social and Behavioral Sciences* 25 (2011) 178–188.

[17] E. Whittaker and G. Watson, *A Course of Modern Analysis,* Cambridge University Press, 1946.

[18] C. Wu, W. Lee, J. Ke and T. Liu, Optimization analysis of an unreliable multi-server queue with a controllable repair policy, *Computers & Operations Research* 49 (2014) 83–96.

[19] D. Yang and J. Ke, Cost optimization of a repairable M/G/1 queue with a randomized policy and single vacation, *Applied Mathematical Modelling* 38 (2014) 5113–5125.

[20] G. Zhang and N. Tian, Discrete time Geo/G/1 queue with multiple adaptive vacations, *Queueing Systems* 38 (2001) 419–429.

Manuscript received 12 May 2014
revised 31 August 2014, 8 November 2014
accepted for publication 10 November 2014

Zhanyou Ma
College of Science, Yanshan University
Qinhuangdao 066004 China
E-mail address: mzhy55@ysu.edu.cn

Wuyi Yue
Department of Intelligence and Informatics
Konan University, Kobe 658-8501 Japan
E-mail address: yue@konan-u.ac.jp

Li Chen
Liren College, Yanshan University
Qinhuangdao 066004 China
E-mail address: chenli@ysu.edu.cn