Yeo, W. M., Yuan, X.-M. and Low, J. M. W. (2017) On M^{X}/G (M/H)/1 retrial system with vacation: service helpline performance measurement. Annals of Operations Research, 248(1-2), pp. 553-578.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

On $M^X/G(M/H)/1$ Retrial System with Vacation: Service Helpline Performance Measurement

Wee Meng Yeo[*], Xue-Ming Yuan[†], Joyce Mei Wan Low[‡]

Abstract

This paper analyzes an unreliable $M^X/G(M/H)/1$ retrial system with vacation. We present closed-form expressions for the important performance indicators of the system, and derive the optimal vacation policies for minimizing the average waiting time of orbiting customers. The performance metrics are relevant for helpline services rendered in non-government organizations. Numerical experiments are conducted to examine the effect of vacation policy on the queue length and busy period of the system.

Keywords: Retrial systems, vacations, batch arrivals, breakdown, reliability.

[*]Georgia Institute of Technology, Scheller College of Business, 800 West Peachtree Street GA30308. Email: yeowee.meng@scheller.gatech.edu. Tel: +1-(404)-924-9969

[†]Singapore Institute of Manufacturing Technology, 71 Nanyang Drive, Singapore 638075. Email: xmyuan@simtech.a-star.edu.sg. Tel: +65-67938407

[‡]Singapore Management University, 81 Victoria Street, Singapore 188065. Email: joycelow@smu.edu.sg. Tel: +65-68085299

1

On $M^X/G(M/H)/1$ Retrial System with Vacation: Service Helpline Performance Measurement

1 Introduction

We explore the application of retrial system to model helpline service provision in non-government organization (NGO). Operating under a resource-constrained environment, retrial systems with repeated attempts are prevalent and improving operational efficiency can be challenging. Three attributes are particularly salient: (i) repeated requests for urgent aid-support are often less prone to balking, (ii) allocation of resources are usually determined by NGO and (iii) support that is limited by manpower expertise. To motivate, we consider NGOs that manage helpline services focusing on domestic issues such as family violence to scarce resource distribution. The lack of manpower is a cruel reality that plagues NGOs and is in contrast to the commercial environment where large number of live agents are staffed. Furthermore, high level of neediness implies that blocking is less likely to happen. Constrained by manpower, organization can consider the following policy for the blocked victim: server will request for victim's number and searches for it when it is available, i.e., not serving other callers. Once these blocked distressed callers are cleared from the system, the helpline goes into a period of vacation. Such vacation can represent hiring of new helpline employees due to high attrition rate on a regular basis. In fact, training is necessary to ensure that the agent will be well-equipped to handle distressed calls. Motivated by the work of [7], our work lies in formulating performance metrics where callers do not balk and vacation is an important feature in such environment. The lack of financial resources and lack of manpower are the main drivers of blocking and backlogging the potential callers seeking assistance. The implication of these challenges is that service provision becomes provider-determined because needs fulfilment involves searching for and servicing blocked callers.

2 Literature Review

The extant literature on retrial systems is vast. For the keen reader, detailed overviews can be found in representative bibliographies such as Artalejo [2, 3] and the survey in Falin [11]. Generally, there are two types of retrial mechanisms: classical retrial policy and constant retrial policy. For systems with classical retrial policy, one can refer the works of Kulkarni and Choi [21], Yang and Li [32], Wang, Cao and Li [30], Li et al [24] and Falin [13, 14]. The rate of repeated attempt is $n\theta$ whenever there are n callers who independently seek service at a rate of θ. In our model, we adopt the constant retrial policy that is introduced by Fayolle [10] who investigate a telephone exchange model where caller at the head of the queue retries for service in an $M/M/1$ system. In this case, the retrial rate becomes $(1 - \delta_{0,n})\theta$, where $\delta_{0,n}$ is the Kronecker delta

symbol. Choi et al [8] study retrial queues motivated by communications protocols such as the ALOHA protocol and unslotted CSMA/CD. Atencia et al [7] note that such retrial policy is ubiquitous in communications and computer networks where repeated attempts are made by processor units independent to the number of messages stored in each node of the waiting server. Meanwhile, retrial systems with constant retrial policy are studied by Artalejo [5], Aissani [1], Li and Zhao [23], and Atencia et al [7]. There is a growing interest in retrial systems with vacation. Artalejo [5] considers an $M/G/1$ queue with multiple vacations and N-policy. Later, Aissani [1] extends the study to $M^X/G/1$ retrial system with multiple vacation. Krishnakumar and Vijayakumar [20] consider Bernoulli vacation schedule with general retrial times using the technique of supplementary variables. Using the same technique, Senthilkumar and Arumuganathan [22] consider the single server retrial system with batch arrivals and two phases of essential services under Bernoulli vacation policy.

We explore the application of $M^X/G(M/H)/1$ retrial system with vacation in resource-challenged environment where helpline programs are executed. First of all, the constant retrial rate that is independent of the orbit size is relevant to describe the autonomy in effort allocation induced by lack of manpower. For NGOs, such operational choice is appealing when system stability is critical in balancing the trade-off between manpower morale and maintaining high humanitarian impact. Next, the blocking of callers frequently arises in situations when server is busy, breaks down or unavailable due to vacation. Server unreliability models system malfunctioning from personnel falling ill and low morale to effectively carrying out humanitarian work. On the other hand, server unavailability under vacation models the key attribute of paper work clearance from authority at the receiving end. Thus, our vacation policy is very general that is based on multiple adaptive vacation (MAV) from Tian and Zhang [29].[1] We define j-adaptive vacation policy to be the MAV policy when the maximum number of vacations allowed is j.

Our contribution can be summarized can follows. Within the extant queueing literature, we extend the work of Atencia [7] who apply supplementary variables to obtain the k- busy period. Our collective marks approach allows us to "split" the k-busy into server idle (due to waiting for retrial) and working length. Subsequently, the busy period for the retrial system under vacation is derived. Compared to the work of Chang and Ke [15] that studies a batch arrival retrial system with generalized retrial times under same vacation policy, our work takes into account of server unreliability. The probable server breakdowns lead to service disruptions and impose inconveniences onto potential callers. Therefore, our study explicitly investigates the joint impact of vacation and the generalized repair times on the busy period.

---

[1] MAV policies are based on the number of vacations taken before the first caller arrives. The server takes vacation consecutively until at least one caller is found waiting in the system at the vacation completion instant given a maximum number of vacations allowed. If no caller is found in the system after the end of the last allowable vacation, server goes into idling and waits until one caller arrives. MAV policy is first proposed by Takagi [27] and later by Ke and Chu [16] and Ke [18] who refer it as "modified vacation policy".

The rest of the paper is organized as follows. Section 3 presents the model of the system. Section 4 studies the ergodicity of the embedded Markov chain which describes the number of callers in the system when a caller is departing. Using supplementary variables, Section 5 obtains the probability generating functions for the number of callers under various vacation policies. Section 7 derives the average waiting time of the callers and the average length of the regenerative cycle. Section 6 analyzes the busy period, which is a key concern for service providers in practice due to its direct implication on their utilization. Section 8 formulates the long run average cost, based on the derivation of the average length of the regenerative cycle in Section 7. We also determine the optimal vacation policy that is based on the cost parameters. The conditions for the optimal vacation policies are presented. Section 9 extensively conducts the numerical experiments. Section 10 suggests avenues for further research and concludes the study.

3   Model

We consider our model from the perspective of helpline implementation by means of a parsimonious single-server retrial system with no waiting space. The Laplace-Stieltjes transform for any cumulative distribution function $F(\cdot)$ with $F(0) = 0$ and $F(\infty) = 1$ is given by $\widetilde{F}(s) = \int_0^\infty e^{-st} dF(t)$.

<u>Arrivals.</u> The distressed callers arrive in batches forming a compound Poisson process with rate λ. With no waiting space, blocked callers are instructed to leave behind contact details and the server will search for these blocked callers after service completion. Under this searching mechanism, one can alternatively hold the view that the retrial rate induced from repeated attempts is $(1 - \delta_{0,n})\theta$. Let X be the random variable for the size of an arriving batch such that $P(X = k) = c_k$. Define $X(z)$ to be the p.g.f (probability generating function) of the batch size. For the nth moment of X, EX^n is denoted by γ_n.

<u>Server.</u> For simplicity, we assume that every caller will be served until service completes. The next service epoch begins during the earlier time between the searched caller and new distressed caller. Thus, competition for assistance comes from two sources: external arrival and orbit. Let B be the random variable for the service time of a caller with c.d.f $B(\cdot)$, $B(0) = 0$, and $B(\infty) = 1$. Denote EB to be the average caller's service time excluding repair. We assume that the server has an exponential lifetime with failure intensity of α. When the server fails, it is repaired immediately and the time required for repair is random variable R, with c.d.f $R(\cdot)$, $R(0) = 0$, and $R(\infty) = 1$. Let ER be the average repair time. On his arrival, the caller who finds that the server is broken down or busy will join the retrial group. No service will be rendered when the system breaks down. In the case when server fails during service, the caller will wait until the server is repaired. Let H be the random variable representing the generalized service time for each caller whose distribution function is given by

4

$H(t)$. From Tang [28], it can be shown that

$$H(t) = P(H \le t) = \sum_{l=0}^{\infty} \int_0^t R^{(l)}(t-u)e^{-\alpha u}\frac{(-\alpha u)^l}{l!}dB(u).$$

Using the argument found in Wang et al ([30], pp. 366) of taking Laplace transform, we have $EH = EB(1 + \alpha ER)$. Let M_n denote the number of callers arriving during the nth service time with $P(M_n = k) = m_{n,k}$. If $\widetilde{H}(s)$ denote the LST of distribution function H w.r.t s, then we can show that $\sum_{k=0}^{\infty} z^k P(M_n = k) = \widetilde{H}(\lambda - \lambda X(z))$.

<u>Vacation.</u> Depending on the vacation policy predetermined by the organization, there are two quantities of interests: the number of callers arriving during one vacation period and the total number of callers over the entire vacation. The length of one vacation period is described by random V with c.d.f $V(\cdot)$, $V(0) = 0$, and $V(\infty) = 1$. Under policy $\mathcal{F}_j$, the server will leave for another vacation when no callers are found in the orbit if $j > 1$ until a maximum j of such vacations is taken. In this case, $\{V_n : 1 \le n \le j\}$ is the associated i.i.d sequence of vacation period such that $V_n =_d V$, where $=_d$ means equal in distribution. Let $\overline{\nu}$ be the total number of callers at the end of the vacation. If we denote $\zeta_j(z) = \sum_{k=0}^{\infty} z^k P(\overline{\nu} = k)$, then $P\{\overline{\nu} = 0\} = \widetilde{V}(\lambda)^j$ whenever a maximum of j vacations is taken by the server. The p.g.f for the number of callers at the end of the vacation is

$$\zeta_j(z) = \widetilde{V}(\lambda)^j + \frac{1 - \widetilde{V}(\lambda)^j}{1 - \widetilde{V}(\lambda)}[\widetilde{V}(\lambda - \lambda X(z)) - \widetilde{V}(\lambda)]. \tag{1}$$

We note that (1) captures the retrial effect and is different from Takagi ([27], see Chapter 2). This departure is due to the absence of retrial phenomena for the conventional $M/G/1$ queue, implying that server idleness or non-idleness is fully determined by vacation terminal epoch. With server dedicated for searching blocked victims, this presents an additional window for idleness after a maximum number of allowable vacations. Therefore, we make a distinction between arrivals during a vacation period and arrivals during idle period. Thus, $E\overline{\nu} = \frac{1-\widetilde{V}(\lambda)^j}{1-\widetilde{V}(\lambda)}\lambda\gamma_1 EV$.

4 Ergodicity Results

We analyze the embedded Markov chain under $\mathcal{F}_j$ to determine the p.g.f for the queue length and its associated ergodicity result. The key is to obtain an embedded Markov Chain for the queue length using $\{S_n : n \in N\}$, the sequence of successive epoch of service completion. It is sufficient to study the process at service completion times. We define $Q_n = Q^{\mathcal{P}}(S_n+) = O^{\mathcal{P}}(S_n+)$ and $p_{ij} = P(Q_{n+1} = j|Q_n = i)$. Throughout this paper, let us define $\rho = \lambda\gamma_1 EH$, $r(\theta) = \frac{\lambda\gamma_1}{\lambda+\theta}$, and $\kappa(\theta) = \rho + r(\theta)$.

Lemma 1 *The probability generating function for the queue-length left by a departing caller can be written as*

$$Q_+^{\mathcal{F}_j}(z) = \frac{C_j(\theta)}{\lambda \gamma_1} \times \frac{\widetilde{H}(\lambda - \lambda X(z))}{(\lambda X(z) + \theta)\widetilde{H}(\lambda - \lambda X(z)) - (\lambda + \theta)z}$$
$$\times \{(1 - \widetilde{V}(\lambda - \lambda X(z)))(\lambda X(z) + \theta)g_j(\lambda) + \theta \widetilde{V}(\lambda)^j(1 - X(z))\}. \quad (2)$$

Furthermore, the embedded Markov chain at departure epoch for the process $\{Q(t) : t \geq 0\}$ is ergodic if and only if $\kappa(\theta) < 1$.

The necessary and sufficient condition for the ergodicity of the embedded Markov chain is derived using Kaplan's condition and Pake's Lemma. Our result says that $\lambda EX \left(EB(1 + \alpha ER) + \frac{1}{\lambda + \theta} \right) < 1$ is needed to guarantee the stability of the retrial system. Furthermore, this stability condition is independent in the choice of vacation policy. Due to the definition of the mean-drift function for the embedded Markov chain, the one-step drift removes the effect of callers arriving during vacation. Thus, the ergodicity result for the retrial system without vacation would have same condition (see Atencia et al [7]). Lemma 1 effectively generalizes the ergodicity result for the $M^X/G(M/H)/1$ queue by considering $\theta \to \infty$. This result coincides with Ke [17] where stability is dependent on the load factor $EB(1 + \alpha ER)$, the expected generalized service time. With the additional policy of searching for orbit callers, $\frac{1}{\lambda + \theta}$ represents the liability as a result of idleness that reduces in increasing rate of searching θ. In the efficienct management of the stability of the retrial system, the decision maker should note that the choice of vacation has a lesser impact than the rate of searching. Rather than worrying about the state of caller backlogging from vacation, higher rate of searching efforts and increasing service efficiency (which includes repairing breakdown) are more effective as levers in preventing system explosion. From now on, we shall suppose that the condition $\kappa(\theta) < 1$ holds.

5 Generating Functions for the System Sizes

We leverage on the ergodicity results to derive the steady state distribution of the retrial system under vacation policy $\mathcal{F}_j$. In particular, we are interested to establish the relationship between $Q^{\mathcal{F}_j}(z)$ and $Q_+^{\mathcal{F}_j}(z)$ so as to obtain the queue length result. These investigations lead to an important stochastic decomposition result for the stationary system size. During t, $C(t) = \nu, 0, 1, 2$ represents the server under vacation, idle, busy or under repair at time t. Let $\overline{\nu}(x) = \frac{dV(x)}{1 - V(x)}$, $\overline{b}(x) = \frac{dB(x)}{1 - B(x)}$, and $\overline{r}(x) = \frac{dR(x)}{1 - R(x)}$. Define the following for $x, y, k \geq 0; 1 \leq i \leq j$:

$\omega_{\nu,k,i}^{\mathcal{F}_j}(x)dx = \lim_{t \to \infty} P\{C(t) = \nu, O^{\mathcal{F}_j}(t) = k, V_-^i(t) \in (x, x + dx)\}.$
$\omega_{0,k}^{\mathcal{F}_j} = \lim_{t \to \infty} P\{C(t) = 0, O^{\mathcal{F}_j}(t) = k\}.$
$\omega_{1,k}^{\mathcal{F}_j}(x)dx = \lim_{t \to \infty} P\{C(t) = 1, O^{\mathcal{F}_j}(t) = k, X_-(t) \in (x, x + dx)\}$
$\omega_{2,k}^{\mathcal{F}_j}(x,y)dy = \lim_{t \to \infty} P\{C(t) = 2, O^{\mathcal{F}_j}(t) = k, X_-(t) = x, Y_- \in (y, y + dy)\}.$

For vacation policy $\mathcal{P}$, we denote $\omega_\nu^{\mathcal{P}}(z) = \sum_{i=1}^{j} \omega_{\nu,i}^{\mathcal{P}}(z)$. Our next result provides closed-form expressions for the p.g.f, allowing us to compute the long-run average fraction of the time when the server is under vacation, idle, busy and under repair. For our purpose, the following notations will be useful for stating this result, i.e., $g_j(\lambda) = \frac{1-\widetilde{V}(\lambda)^j}{1-\widetilde{V}(\lambda)}$ and $C_j(\theta) = \frac{1-\kappa(\theta)}{g_j(\lambda)EV + \frac{\theta\widetilde{V}(\lambda)^j}{\lambda(\lambda+\theta)}}$.

Theorem 1 *The stationary distribution of the process $\{O^{\mathcal{F}_j}(t) : t \geq 0\}$ has the following generating functions*

$$\omega_{\nu,i}^{\mathcal{F}_j}(z,x) = C_j(\theta)\widetilde{V}(\lambda)^{i-1}[1 - V(x)]e^{-(\lambda-\lambda X(z))x}, 1 \leq i \leq j \tag{3}$$

$$\omega_0^{\mathcal{F}_j}(z) = \mathcal{C}_j(\theta)\frac{\{\frac{\theta\widetilde{V}(\lambda)^j}{\lambda}[\widetilde{H}(\lambda - \lambda X(z)) - z] + z[1 - \widetilde{V}(\lambda - \lambda X(z))]g_j(\lambda)\}}{[(\lambda X(z) + \theta)\widetilde{H}(\lambda - \lambda X(z)) - (\lambda + \theta)z]} \tag{4}$$

$$\omega_1^{\mathcal{F}_j}(z,x) = C_j(\theta)\frac{\{(1 - \widetilde{V}(\lambda - \lambda X(z)))(\lambda X(z) + \theta)g_j(\lambda) + \theta\widetilde{V}(\lambda)^j[1 - X(z)]\}}{[(\lambda X(z) + \theta)\widetilde{H}(\lambda - \lambda X(z)) - (\lambda + \theta)z]}$$
$$\times [1 - B(x)]e^{-(\lambda-\lambda X(z)+\alpha-\alpha\widetilde{R}(\lambda-\lambda X(z)))x} \tag{5}$$
$$\omega_2^{\mathcal{F}_j}(z,y,x) = \alpha\omega_1^{\mathcal{F}_j}(z,x)[1 - R(y)]e^{-(\lambda-\lambda X(z))y}. \tag{6}$$

In particular, if we omit the elapsed vacation times, service times and repair times, the stationary distribution of the process $\{O^{\mathcal{F}_j}(t) : t \geq 0\}$ has the following generating functions

$$\omega_\nu^{\mathcal{F}_j}(z) = C_j(\theta)g_j(\lambda)\left(\frac{1 - \widetilde{V}(\lambda - \lambda X(z))}{\lambda - \lambda X(z)}\right)$$

$$\omega_1^{\mathcal{F}_j}(z) = C_j(\theta)\frac{(1 - \widetilde{V}(\lambda - \lambda X(z)))(\lambda X(z) + \theta)g_j(\lambda) + \theta\widetilde{V}(\lambda)^j[1 - X(z)]}{[(\lambda X(z) + \theta)\widetilde{H}(\lambda - \lambda X(z)) - (\lambda + \theta)z]} \times \frac{1 - \widetilde{H}(\lambda - \lambda X(z))}{\lambda - \lambda X(z) + \alpha - \alpha\widetilde{R}(\lambda - \lambda X(z))}$$

$$\omega_2^{\mathcal{F}_j}(z) = \alpha\omega_1^{\mathcal{F}_j}(z)\frac{1 - \widetilde{R}(\lambda - \lambda X(z))}{\lambda - \lambda X(z)}.$$

Using Theorem 1, we derive the steady-state probabilities by letting $z \to 1^-$.

Corollary 1 *The long-run average fraction of the time for the various server activities are given by the following:*

i. *the server takes vacation with probability $\omega_\nu^{\mathcal{F}_j}(1) = C_j(\theta)g_j(\lambda)EV$.*

ii. *the server is idle with probability $\omega_0^{\mathcal{F}_j}(1) = \frac{\frac{\theta}{\lambda(\lambda+\theta)}\widetilde{V}(\lambda)^j + \kappa(\theta)EVg_j(\lambda)}{EVg_j(\lambda) + \frac{\theta}{\lambda(\lambda+\theta)}\widetilde{V}(\lambda)^j} - \rho$.*

iii. *the server is busy with probability $\omega_1^{\mathcal{F}_j}(1) = \lambda\gamma_1 EB$.*

iv. *the server is under repair with probability $\omega_2^{\mathcal{F}_j}(1) = \alpha\omega_1^{\mathcal{F}_j}(1)ER$.*

7

For non-profits, the fraction of time when the server is busy is an important measure. Furthermore, it is interesting to note that the long run fraction of the busy period and server repairing time are independent of vacation policy type. Intuitively, this lead us to conclude the following. We are interested to obtain stochastic decomposition result for our retrial system and to motivate this, we first consider the related result under $\mathcal{F}_j$.

Corollary 2 *Let $\kappa(\theta) < 1$. Under vacation policy $\mathcal{F}_j$,*

i. *the p.g.f for the number of callers in orbit is given by*

$$O^{\mathcal{F}_j}(z) = C_j(\theta) \frac{(1-z)}{(\lambda - \lambda X(z))[(\lambda X(z) + \theta)\widetilde{H}(\lambda - \lambda X(z)) - (\lambda + \theta)z]}$$
$$\times \{(1 - \widetilde{V}(\lambda - \lambda X(z)))(\lambda X(z) + \theta)g_j(\lambda) + \theta\widetilde{V}(\lambda)^j[1 - X(z)]\} \qquad (7)$$

ii. *the p.g.f for the number of callers in the system is given by*

$$Q^{\mathcal{F}_j}(z) = O^{\mathcal{F}_j}(z)\widetilde{H}(\lambda - \lambda X(z)).$$

The p.g.f of the system size allows us to conclude that $Q^{\mathcal{F}_j}(z) = \frac{\gamma_1(1-z)}{1-X(z)}Q_+^{\mathcal{F}_j}(z)$. This holds in $M^X/G/1$ retrial system without vacation or $M^X/G/1$ with multiple adaptive vacation queue without retrial. Let $N^{\mathcal{P}}$ and $O^{\mathcal{P}}$ be the system size and orbit size of our retrial system at an arbitrary point under vacation policy $\mathcal{P}$. Collorary 2 implies that the system size of the retrial system with vacation is the sum of two independent random variables, i.e. $N^{\mathcal{P}} = O^{\mathcal{P}}+$Number of arrivals during generalized service time. Let $\Pi_0(z)$ denote the p.g.f of the number of callers in an unreliable single server retrial system under constant retrial policy. Atencia et al [7] have shown that

$$\Pi_0(z) = \frac{(\lambda + \theta)(1 - \kappa(\theta))(1 - z)H(\lambda - \lambda X(z))}{[(\lambda X(z) + \theta)H(\lambda - \lambda X(z)) - (\lambda + \theta)z]}.$$

Using $\Pi_0(z)$ and Corollary 2, we obtain the following crucial result under $\mathcal{F}_j$. For the general result, we consider the following. Denote N_R to be number of callers in the $M^X/G(M/H)/1$ retrial system without vacation and N_0 to be the stationary queue length of the $M^X/G(M/H)/1$ queue without retrial or vacation.

Theorem 2 *The stationary system size of the $M^X/G(M/H)/1$ retrial system under j-adaptive vacation can be written as $N_\nu^{\mathcal{F}_j} = N_R + N^{\mathcal{F}_j}$, where $N^{\mathcal{F}_j}$ is the additional queue length due to the vacation effect. The p.g.f of $N^{\mathcal{F}_j}$ is given by*

$$N^{\mathcal{F}_j}(z) = \frac{(1 - \widetilde{V}(\lambda - \lambda X(z)))(\lambda X(z) + \theta)g_j(\lambda) + \theta\widetilde{V}(\lambda)^j[1 - X(z)]}{\left((\lambda + \theta)EVg_j(\lambda) + \frac{\theta}{\lambda}\widetilde{V}(\lambda)^j\right)(\lambda - \lambda X(z))}. \qquad (8)$$

The stationary system size of $M^X/G(M/H)/1$ under j-adaptive vacation can be written as a sum of three random variables.

Theorem 2 allows us to conclude that the stationary system size of the $M^X/G(M/H)/1$ retrial system under j-adaptive vacation can be written as a sum of three random variables. The average waiting time is one of the most important performance measures of service quality. In the previous section, we obtain the important result from decomposing the number of callers of the retrial system under vacation into two independent random variables, namely, the system size for the unreliable batch arrival single server retrial system without vacation and system size due to the vacation effect. This decomposition is extremely useful in helping us to compute the average size of the system. By using the p.g.f of $N^{\mathcal{F}_j}(z)$ in Theorem 2, we can obtain the average number of callers in the system caused by the vacation effect (denoted by $EN^{\mathcal{F}_j}$). Applying L'Hospital rule and letting $z \to 1$, we get

$$EN^{\mathcal{F}_j} = \left[\frac{\lambda \gamma_1 g_j(\lambda)[EV + \frac{1}{2}(\lambda + \theta)(\mathrm{Var}(V) + (EV)^2)]}{(\lambda + \theta)EVg_j(\lambda) + \frac{\theta}{\lambda}\widetilde{V}(\lambda)^j} \right].$$

Denote $\Delta = \frac{EB^2}{(EB)^2}\rho^2 + \lambda \gamma_2 EH + \alpha(\lambda \gamma_1)^2 ERES$. Denote EN to be the mean number of callers in the retrial group for the batch retrial system without vacation. From the result of Atencia et al ([7], Corollary 2), we have the average number of the callers in the system is given by

$$EN = \frac{\lambda(\gamma_2 + 2\gamma_1 \rho)}{2(\lambda + \theta)(1 - \kappa(\theta))} + \frac{\Delta}{(1 - \kappa(\theta))} + \rho$$

Thus, the overall average system size will be the sum, i.e., $EN^{\mathcal{F}_j} + EN$. It is left to check that the total system size is increasing in j, the number of vacation taken. To this end, we shall present some special cases where the system takes multiple vacations, i.e. after coming from a vacation, the server takes another vacation as long as there are no callers in the orbit. Aissani [1] refers to this policy as the exhaustive vacation policy. Using Theorem 1, we obtain the results similar to the work of Aissani [1]. Finally, we consider the $M^X/G(M/H)/1$ retrial system where the server takes exactly one vacation immediately at the end of the busy period. If it finds no caller in the orbit upon returning from the vacation, it becomes idle until one caller arrives. To the best of our knowledge, there is no existing work that considers an unreliable single server retrial system taking exactly one vacation. Using applying $j = 1$ in Corollary 2, we shall state the result without proof.

Corollary 3 *Let $\mathcal{P}$ be the given vacation policy. For multiple-vacation policy,*

i. *the p.g.f for the number of callers in orbit is given by*

$$O^{\mathcal{M}}(z) = C_\infty(\theta) \frac{(1 - z)[1 - \widetilde{V}(\lambda - \lambda X(z))](\lambda X(z) + \theta)}{(\lambda - \lambda X(z))[(\lambda X(z) + \theta)\widetilde{H}(\lambda - \lambda X(z)) - (\lambda + \theta)z]},$$

ii. *the p.g.f for the number of callers in the system is*

$$Q^{\mathcal{M}}(z) = O^{\mathcal{M}}(z)\widetilde{H}(\lambda - \lambda X(z)).$$

9

Otherwise, we have $Q^{\mathcal{S}}(z) = O^{\mathcal{S}}(z)\widetilde{H}(\lambda - \lambda X(z))$ *and*

$$O^{\mathcal{S}}(z) = C_1(\theta)\frac{(1-z)\{(1-\widetilde{V}(\lambda - \lambda X(z)))(\lambda X(z) + \theta) + \theta\widetilde{V}(\lambda)[1 - X(z)]\}}{(\lambda - \lambda X(z))[(\lambda X(z) + \theta)\widetilde{H}(\lambda - \lambda X(z)) - (\lambda + \theta)z]},$$

$\diamond$

If we let $\theta \to \infty$, we will obtain the p.g.f of system size for the unreliable $M^X/G/1$ queue without retrial taking single vacation. Furthermore, if we consider a single batch arrival, i.e. $X(z) = z$ and $\alpha = 0$, then the p.g.f exactly coincides with that of Takagi [27].

6 Busy Period Analysis

In our retrial system, it is necessary to differentiate two types "idleness" for the working server driven by the presence of vacation. To see this, we appeal to the concept of $k-$busy period which is defined as the time period between starting with k callers and service completion. The notion of idleness in Atencia et al [7] is solely based on waiting for servicing the next arrival or caller in orbit under $k-$busy period. Beginning with k callers "idleness" can happen even after vacation. Thus, "server-waiting" period is different from "idle" period (driven by vacation) in our system. Using the method of collective marks, we analyze the busy period of our unreliable retrial system with vacation. Our approach deviates from Atencia et al [7] that allows us to "split" the busy period into value-adding working-length and efficiency-draining "server-waiting" plus "repair-servicing" periods. We define the busy period of our unreliable retrial system under $\mathcal{P}$ as the length between the instant when the server starts to serve the first caller after vacation to the instant when the system empties at the instant of last departure epoch, denoted by $L^{\mathcal{P}}$.

Once service begins, the server goes through periods of working, waiting (through caller searching), or under repair. After the vacation, the first caller served may arrive from the orbit or external sources. Our objective in this section is to compute the average length of the cumulative server-waiting period, working period, and repair period that make up $EL^{\mathcal{P}}$. Inspired by Falin and Templeton [12], we consider an auxiliary Poisson flow of "catastrophe" of rate s that is independent of the system functioning. For the given $s > 0$, the quantity random τ_s represents the time in which the catastrophe occurs. Indeed, we must have $P(\tau_s \le x) = 1 - e^{-sx}$. Let $\{S_n : n \ge 1\}$ be the sequence of epoch where callers depart. Denote $\pi_{ni}^{(k)}(s)$ to be the probability that at the time S_i (since the beginning of the k-busy period), no catastrophe has occurred and the k-busy period did not end before S_i. For $i = 1$, $\pi_{n1}^{(k)}(s)$ is the probability that at the first departure epoch, no catastrophe occurred, there are n callers and k-busy period did not end. This is equivalent to the probability that during the serving of the first caller, there are $n - k + 1$ arrivals and catastrophe did not end. We define $f^{(k)}(s, z, y) = \sum_{n=1}^{\infty} z^n \sum_{i=1}^{\infty} y^i \pi_{ni}^{(k)}(s)$.

10

Lemma 2 *The generating function $f^{(k)}(s,z,y)$ satisfies the functional equation*

$$\left\{1 - \left(\frac{y}{z}\right)\left(\frac{\lambda X(z) + \theta}{s + \lambda + \theta}\right)\widetilde{H}(s + \lambda - \lambda X(z))\right\}f^{(k)}(s,z,y) = yz^{k-1}\widetilde{H}(s + \lambda - \lambda X(z)).$$

(9)

The purpose of defining $\pi_{ni}^{(k)}(s)$ is to help us analyze the k-busy period (assuming that it starts at $t = 0$) through the following:

$$P_{0ni}^{(k)}(t) = P\{L^{(k)} > t, C(t) = 0, Q(t) = n, I(t) = i\}$$
$$P_{1ni}^{(k)}(t,x)dx = P\{L^{(k)} > t, C(t) = 1, \omega(t) \in (x, x + dx), Q(t) = n, I(t) = i\}$$
$$\varphi_{0ni}^{(k)}(s) = \int_0^\infty e^{-st}P_{0ni}^{(k)}(t)dt; \quad \varphi_{1ni}^{(k)}(s,x) = \int_0^\infty e^{-st}P_{1ni}^{(k)}(t,x)dt.$$

We consider two important transient distributions based on the system having n callers with exactly i services completed at time t. In particular, $P_{0ni}^{(k)}(t)$ represents the probability in which the $k-$busy period has not expired for the batch-arrival unreliable retrial system which is idle at t. On the other hand, $P_{1ni}^{(k)}(t, \cdot)$ represents the cumulative distribution for the elapsed service time when the $k-$busy period has not expired when server is busy at t. In order to understand the probabilistic interpretation of the transforms, we have

$$s\varphi_{0ni}^{(k)}(s) = P\{L^{(k)} > \tau_s, C(\tau_s) = 0, Q(\tau_s) = n, I(\tau_s) = i\};$$
$$s\varphi_{1ni}^{(k)}(s,x) = P\{L^{(k)} > \tau_s, C(\tau_s) = 1, \omega(\tau_s) \in (x, x + dx), Q(\tau_s) = n, I(\tau_s) = i\}.$$

For a given $s > 0$, $s\varphi_{0ni}^{(k)}(s)$ is the probability that at time τ_s when the first catastrophe occurred, the server was free, there were n callers in the system and exactly i had completed service but the $k-$busy period had not expired. The interpretation of $s\varphi_{1ni}^{(k)}(s,x)dx$ is similar and can be found in the work of Falin and Templeton [12]. To develop results of splitting the $k-$busy period, we define two additional generating functions $\varphi_0^{(k)}(s,z,y) = \sum_{i=1}^\infty y^i \sum_{n=1}^\infty z^n \varphi_{0ni}^{(k)}(s)$ and $\varphi_1^{(k)}(s,z,y,x) = \sum_{i=0}^\infty y^i \sum_{n=0}^\infty z^n \varphi_{1ni}^{(k)}(s,x)$. The exact expressions are found in the Appendix.

Theorem 3 *The expected length of a k-busy period, $EL^{(k)}$ is the sum of $L_0^{(k)}$ and $L_1^{(k)}$ given by*

$$L_0^{(k)} = \frac{k - 1 + \rho}{(\lambda + \theta)(1 - \rho) - \lambda\gamma_1}; L_1^{(k)} = \frac{EH(k(\lambda + \theta) - \lambda\gamma_1)}{(\lambda + \theta)(1 - \rho) - \lambda\gamma_1}.$$

Finally, $EL^{(k)} = L_0^{(k)} + L_1^{(k)} = \dfrac{[1 + (\lambda + \theta)EH]k - 1}{(\lambda + \theta)(1 - \rho) - \lambda\gamma_1}.$

Note that the length of the $k-$busy period increases in the number of starting callers and the reliability of the server. We can further refine $L_1^{(k)}$ into the sum of expected working length and repair length.

Corollary 4 *The quantity $L_1^{(k)}$ can be written as the sum of $L_w^{(k)}$ and $L_r^{(k)}$ given by*

$$L_w^{(k)} = \frac{EB(k(\lambda + \theta) - \lambda\gamma_1)}{(\lambda + \theta)(1 - \rho) - \lambda\gamma_1}; L_r^{(k)} = \frac{\alpha ERES(k(\lambda + \theta) - \lambda\gamma_1)}{(\lambda + \theta)(1 - \rho) - \lambda\gamma_1}.$$

6.1 Busy Period Under Vacation Policy $\mathcal{F}_j$

Applying the $k-$busy period developed earlier, we consider the busy period under $j-$adaptive vacation. With $\mathcal{F}_j$, we assume that the server takes at most j number of vacations repeatedly until at least one caller is found in the orbit when the server returns from the vacation. If no message is found by the end of the jth vacation, the server becomes idle. Again, we let $\widetilde{\pi}^{(k)}(s)$ be the L.S.T of the k-busy period. The event $\{\bar{\nu} = 0\}$ will generate $\widetilde{\pi}^{(m)}(s)$ with probability c_m. Suppose at the end of vacation, there are $k \geq 1$ callers in the orbit. With probability $\frac{\theta}{\theta + \lambda}$, $\widetilde{\pi}^{(k)}(s)$ is the L.S.T generated and with probability $\frac{\lambda c_m}{\lambda + \theta}$, $\widetilde{\pi}^{(k+m)}(s)$ is the L.S.T generated. Hence given that $\{\bar{\nu} = k\}$ for $k \geq 1$, the busy period generated has L.S.T $\frac{\theta}{\lambda + \theta}\widetilde{\pi}^{(k)}(s) + \frac{\lambda}{\lambda + \theta}\sum_{m=1}^{\infty} c_m\widetilde{\pi}^{(k+m)}(s)$. To facilitate our derivation, we define the following:

$$\varphi_0(s, z, y) = \sum_{k=1}^{\infty} \bar{\nu}_k \sum_{i=1}^{\infty} y^i \sum_{n=1}^{\infty} z^n \left\{ \frac{\theta}{\lambda + \theta}\varphi_{0ni}^{(k)}(s) + \frac{\lambda}{\lambda + \theta} \sum_{m=1}^{\infty} c_m\varphi_{0ni}^{(k+m)}(s) \right\};$$

$$\varphi_1(s, z, y, x) = \sum_{k=1}^{\infty} \bar{\nu}_k \sum_{i=1}^{\infty} y^i \sum_{n=1}^{\infty} z^n \left\{ \frac{\theta}{\lambda + \theta}\varphi_{1ni}^{(k)}(s, x) + \frac{\lambda}{\lambda + \theta} \sum_{m=1}^{\infty} c_m\varphi_{1ni}^{(k+m)}(s, x) \right\}.$$

The following generating functions will be critical for :

$$\psi_0(s, z, y) = \bar{\nu}_0 \sum_{m=1}^{\infty} c_m\varphi_0^{(m)}(s, z, y) + \varphi_0(s, z, y)$$

$$\psi_1(s, z, y, x) = \bar{\nu}_0 \sum_{m=1}^{\infty} c_m\varphi_1^{(m)}(s, z, y, x) + \varphi_1(s, z, y, x).$$

Theorem 4 *The expected length of an $\mathcal{F}_j$-busy period, $EL^{\mathcal{F}_j}$ is the sum of $L_0^{\mathcal{F}_j}$ and $L_1^{\mathcal{F}_j}$ given by*

$$L_0^{\mathcal{F}_j} = \frac{E\bar{\nu} - 1 + \kappa(\theta) + \frac{\widetilde{V}(\lambda)^j\theta\gamma_1}{\lambda + \theta}}{(\lambda + \theta)(1 - \kappa(\theta))}; L_1^{\mathcal{F}_j} = \rho \left[\frac{\frac{1 - \widetilde{V}(\lambda)^j}{1 - \widetilde{V}(\lambda)}EV + \frac{\theta}{\lambda(\lambda + \theta)}\widetilde{V}(\lambda)^j}{(1 - \kappa(\theta))} \right].$$

Furthermore, we have $EL^{\mathcal{F}_j} = -\frac{1}{\lambda + \theta} + \left(\frac{1 - \widetilde{V}(\lambda)^j}{1 - \widetilde{V}(\lambda)}EV + \frac{\widetilde{V}(\lambda)^j\theta}{\lambda(\lambda + \theta)} \right)\frac{\kappa(\theta)}{1 - \kappa(\theta)}.$ $\diamond$

We now state the main result as a corollary to Theorem 4.

Corollary 5 *$L_1^{\mathcal{F}_j}$ further be splitted up into $L_w^{\mathcal{F}_j}$ and $L_r^{\mathcal{F}_j}$ given by*

$$L_w^{\mathcal{F}_j} = \frac{\lambda\gamma_1 EB \left(\frac{1 - \widetilde{V}(\lambda)^j}{1 - \widetilde{V}(\lambda)}EV + \frac{\widetilde{V}(\lambda)^j\theta}{\lambda(\lambda + \theta)} \right)}{(1 - \kappa(\theta))}; L_r^{\mathcal{F}_j} = \alpha ERL_w^{\mathcal{F}_j}.$$

◇

Remark: In particular, when $j = 1$, one can show that $EL^{\mathcal{S}} = -\frac{1}{\lambda+\theta} + \frac{\kappa(\theta)}{C_1(\theta)}$.

7 Regenerative Cycle

Every epoch in which the busy period ends are the system's regeneration points. At those points, the number of callers in the system is zero. In order to formulate the long run average cost of the system per unit time under a given vacation and retrial policy, we are required to find the length of the regenerative cycle. This cycle is defined as the length of time between two successive regeneration points. We devote this section to computing the average length of the regenerative cycle. The generalized idle time of the retrial system will be defined as the total vacation period plus the first server-waiting period (if any). Let $C_\nu^{\mathcal{P}}$ be the random length of this cycle under the vacation policies $\mathcal{P} \in \{\mathcal{F}_j : j \geq 1\}$. Then $EC_\nu^{\mathcal{P}} = E\zeta_\nu^{\mathcal{P}} + EL^{\mathcal{P}}$, where $\zeta_\nu^{\mathcal{P}}$ is the generalized idle time due to vacation policy $\mathcal{P}$. Let $I' \overset{\mathrm{d}}{=} e^{\lambda+\theta}$, $I^o \overset{\mathrm{d}}{=} e^\lambda$, and $N(t)$ be the arrival process of the number of batches of callers during $(0, t]$. First of all, we derive the L.S.T of $\zeta_\nu^{\mathcal{F}_j}$, $\widetilde{I}_\nu^{\mathcal{F}_j}(s) = E[e^{-s\zeta_\nu^{\mathcal{F}_j}}]$.

Lemma 3 *The L.S.T of $\zeta_\nu^{\mathcal{F}_j}$ is given by*

$$\widetilde{I}_\nu^{\mathcal{F}_j}(s) = \left[\frac{1 - \widetilde{V}(s+\lambda)^j}{1 - \widetilde{V}(s+\lambda)}\right](\widetilde{V}(s) - \widetilde{V}(s+\lambda))\frac{\lambda+\theta}{s+\lambda+\theta} + \widetilde{V}(s+\lambda)^j \frac{\lambda}{s+\lambda}.$$
(10)

The average length of $\zeta_\nu^{\mathcal{F}_j}$ is given by $E\zeta_\nu^{\mathcal{F}_j} = \frac{1 - \widetilde{V}(\lambda)^j}{\lambda+\theta} + EV\left[\frac{1 - \widetilde{V}(\lambda)^j}{1 - \widetilde{V}(\lambda)}\right] +$

$\frac{1}{\lambda}\widetilde{V}(\lambda)^j$.

In order to see the proof of this result, we consider the following:

$$\zeta_\nu^{\mathcal{F}_j} = \begin{cases} V_1 + ... + V_n + I' & \text{w.p } P\{N(V_1) = 0, ..., N(V_n) > 0\}, 1 \leq n \leq j \\ V_1 + ... + V_j + I^o & \text{w.p } P\{N(V_1) = 0, ..., N(V_j) = 0\}. \end{cases}$$

The presence of caller in the vacation period determines the arrival rate for service. If any caller arrives during the nth vacation, then the first initiation of the busy period is exponentially distributed with mean $\frac{1}{\lambda+\theta}$ because of callers presence in the orbit. On the other hand, if no caller is present up to j vacations, then the first caller will arrive with mean $\frac{1}{\lambda}$. If any callers arrive during the nth vacation for $1 \leq n \leq j$, we have $E[e^{-s(V_1+...+V_n+I')}e^{-\lambda V_1}...e^{-\lambda V_{n-1}}(1 - e^{-\lambda V_n})] = [\widetilde{V}(s+\lambda)]^{n-1}(\widetilde{V}(s) - \widetilde{V}(s+\lambda))\frac{\lambda+\theta}{s+\lambda+\theta}$. If no caller arrive during vacation, we have $E[e^{-s(V_1+...+V_j+I^o)}e^{-\lambda V_1}...e^{-\lambda V_j}] = \widetilde{V}(s+\lambda)^j \frac{\lambda}{s+\lambda}$. Hence, we obtain

$\widetilde{I}_\nu^{\mathcal{F}_j}(s)$. We can easily derive $E\zeta_\nu^{\mathcal{F}_j}$ using $-\lim_{s\to 0+}\frac{d}{ds}\widetilde{I}_\nu^{FV}(s)$. Differentiating (10) w.r.t s and letting $s \to 0+$, we obtain $E\zeta_\nu^{\mathcal{F}_j}$. In particular, when $\theta \to \infty$, the L.S.T of the generalized idle period under vacation policy $\mathcal{F}_j$ is $\widetilde{I}_\nu(s) = \left[\frac{1-\widetilde{V}(s+\lambda)^j}{1-\widetilde{V}(s+\lambda)}\right](\widetilde{V}(s) - \widetilde{V}(s+\lambda)) + \widetilde{V}(s+\lambda)^j\frac{\lambda}{s+\lambda}$ for the ordinary queue without retrial. If we let $j \to \infty$, we obtain the L.S.T of the vacation of $M/G/1$ queue with multiple vacation $\widetilde{I}_\nu^{\mathcal{M}}(s) = \left[\frac{\widetilde{V}(s)-\widetilde{V}(s+\lambda)}{1-\widetilde{V}(s+\lambda)}\right]$. See Takagi ([27], pg. $123 - 124$). For the case when $j = 1$, the average length of the generalized idle period is $\frac{1}{\lambda+\theta} + \frac{1-\kappa(\theta)}{C_1(\theta)}$. Denote $L_\nu^{\mathcal{F}_j}$ be the length of the vacation taken by the server (over one regenerative cycle), its L.S.T can be derived by ignoring I^o and I' following the proof in Lemma 3 . In fact the L.S.T is easily shown to be

$$E[e^{-sL_\nu^{\mathcal{F}_j}}] = (\widetilde{V}(s) - \widetilde{V}(s+\lambda))\left[\frac{1-\widetilde{V}(s+\lambda)^j}{1-\widetilde{V}(s+\lambda)}\right] + \widetilde{V}(s+\lambda)^j.$$

Differentiating the above expression w.r.t s and letting $s \to 0$, we obtain $EL_\nu^{\mathcal{F}_j} = EVg_j(\lambda)$. Let $L_{idle}^{\mathcal{F}_j}$ to be the average idle period in one regenerative cycle. It is easy to verify that $L_{idle}^{\mathcal{F}_j} = E\zeta_\nu^{\mathcal{F}_j} - EL_\nu^{\mathcal{F}_j} + L_0^{\mathcal{F}_j} = \frac{\omega_0^{\mathcal{F}_j}(1)}{C_j(\theta)}$. For the retrial system under the multiple vacation policy, we have the following result.

Corollary 6 *The L.S.T of $\zeta_\nu^{\mathcal{M}}$ is given by*

$$\widetilde{I}_\nu^{\mathcal{M}}(s) = \left[\frac{\widetilde{V}(s) - \widetilde{V}(s+\lambda)}{1 - \widetilde{V}(s+\lambda)}\right]\frac{\lambda+\theta}{s+\lambda+\theta}. \tag{11}$$

The average length of $\zeta_\nu^{\mathcal{M}}$ is given by $E\zeta_\nu^{\mathcal{M}} = \frac{1}{\lambda+\theta} + \left[\frac{EV}{1-\widetilde{V}(\lambda)}\right]$.

The expression for $\widetilde{I}_\nu^{\mathcal{M}}(s)$ can be derived easily by letting $j \to \infty$. The derivation of $E\zeta_\nu^{\mathcal{M}}$ is similar. Using Lemma 3 and Corollary 6, we can easily provide the result for $EC_\nu^{\mathcal{F}_j}$ for $j \geq 1$. It turns out that the formula for the average length of the regenerative cycle is extremely simple.

Corollary 7 *The average length of the regenerative cycle under policy $\mathcal{F}_j$ is given by*

$$EC_\nu^{\mathcal{F}_j} = \frac{\left[EV\frac{1-\widetilde{V}(\lambda)^j}{1-\widetilde{V}(\lambda)} + \frac{V(\lambda)^j\theta}{\lambda(\lambda+\theta)}\right]}{1-\kappa(\theta)} = \frac{1}{C_j(\theta)}. \tag{12}$$

We provide some comparative studies on the average length of the busy period of the $M^{(x)}/G/1$ with and without vacations. Let EL^o be the average length of the busy period for the unreliable retrial system in Atencia et al [7]. They show that $EL^o = \frac{\kappa(\theta)/\lambda - 1}{(\lambda+\theta)(1-\kappa(\theta))}$. First we establish a monotonicity result of the busy period w.r.t the vacation policy of the busy period.

Lemma 4 *Suppose the condition $\kappa(\theta) < 1$ holds.*

1. *If $\frac{\theta}{\lambda(\lambda+\theta)} > \frac{EV}{1-\tilde{V}(\lambda)}$, the average length of the busy period $EL^{\mathcal{F}_j}$ decreases as j increases.*

2. *If $\frac{\theta}{\lambda(\lambda+\theta)} \leq \frac{EV}{1-\tilde{V}(\lambda)}$, the average length of the busy period $EL^{\mathcal{F}_j}$ increases as j increases.*

8 Long Run Average Cost and Vacation Policy

What is the optimal vacation policy we shall adopt given the cost parameters? The long run average cost (LRAC) will be used as our objective function. We show that either multiple vacation or single vacation policy is optimal. The cost considerations include setting a reward for taking vacations, cost for idling and repair. We also include $q > 0$ indicating the average fraction of idling or taking a break during vacation period. In real life, when the server is taking a vacation, it may include supplementary jobs that can provide value to the company such as preventive maintenance. Otherwise, server may be taking a break or on leave, incurring costs (of idling) to the firm. Our aim in this section is to investigate the relationship between the extent of idling during vacation and the optimal vacation policy of the call centre. Throughout this section, let us assume that for all retrial rates $\theta \in (0, \infty)$, we have $\kappa(\theta) = \rho + \frac{\lambda\gamma_1}{\lambda+\theta} < 1$.

Notations	Meaning
r_ν	reward per unit time taking vacation
c_0	cost per unit time of idling
c_h	cost per unit time of holding callers in orbit
c_r	cost of repair per unit time
q	average fraction of time idling during vacation, $q \in [0,1]$

From the notations above, it can be seen that $-r_\nu(1-q) + c_0 q$ is the reward per unit time gained by the system during vacation. It is easily seen that when $q = 0$, there is no idling and maximal value is added while $q = 1$ implies only idling cost is incurred. Using elementary renewal reward theory, the long run average cost of operating a server per unit time $G^{\mathcal{F}_j}(\theta, q)$ is given by

$$G^{\mathcal{F}_j}(\theta, q) = (-r_\nu(1-q) + c_0 q)\frac{L_\nu^{\mathcal{F}_j}}{EC_\nu^{\mathcal{F}_j}} + c_h \lambda EW^{\mathcal{F}_j} + c_0 \frac{L_{idle}^{\mathcal{F}_j}}{EC_\nu^{\mathcal{F}_j}} + c_r \frac{L_r^{\mathcal{F}_j}}{EC_\nu^{\mathcal{F}_j}}$$

$$= (-r_\nu(1-q) + c_0 q)\omega_\nu^{\mathcal{F}_j}(1) + c_h \lambda EW^{\mathcal{F}_j} + c_0 \omega_0^{\mathcal{F}_j}(1) + c_r \omega_2^{\mathcal{F}_j}(1).$$

Let us define the following notations:

$$\Delta_0 = [EV + \frac{1}{2}(\lambda + \theta)(\text{Var}(V) + (EV)^2)], Y_j = EV g_j(\lambda) + \frac{\theta}{\lambda(\lambda + \theta)}\widetilde{V}(\lambda)^j$$

$$\Delta_1(q) = (-r_\nu(1 - q) + c_0 q)(1 - \kappa(\theta))EV + c_0\kappa(\theta)EV + c_h r(\theta)\Delta_0,$$

$$\Delta_2 = -c_0\rho + \alpha c_r \lambda \gamma_1 ESER + c_h \left(\frac{\lambda(\gamma_2 + 2\gamma_1\rho)}{2(\lambda + \theta)(1 - \kappa(\theta))} + \frac{\Delta}{1 - \kappa(\theta)} + \rho \right),$$

$$\tag{13}$$

$$\phi_q(c_h) = \frac{c_h r(\theta)\Delta_0}{EV(1 - \kappa(\theta))(1 - q)}; \qquad \Gamma(c_0, q) = -c_0 + \phi_q(c_h).$$

Then, it can be shown that $G^{\mathcal{F}_j}(\theta, q)$ can be expressed as

$$G^{\mathcal{F}_j}(\theta, q) = \frac{g_j(\lambda)\Delta_1 + \frac{\theta c_0 \widetilde{V}(\lambda)^j}{\lambda(\lambda + \theta)}}{Y_j} + \Delta_2. \tag{14}$$

Using (14) and after some simplification, we obtain

$$G^{\mathcal{F}_{j+1}}(\theta, q) - G^{\mathcal{F}_j}(\theta, q) = \frac{\theta\widetilde{V}(\lambda)^j}{\lambda(\lambda + \theta)}\left(\frac{\Delta_1(q) - c_0 EV}{Y_j Y_{j+1}} \right). \tag{15}$$

Theorem 5 *Let (c_0, r_ν, c_h) be the given vector of cost and reward parameters, and q be the average fraction of idling time during vacation,*

i. *if $q \in [0, 1)$, policy $\mathcal{S}$ is optimal, when $r_\nu < \Gamma(c_0, q)$, otherwise $\mathcal{M}$ is optimal.*

ii. *if $q = 1$, policy $\mathcal{S}$ is always optimal.*

Proof: We give an outline of the proof.

i. Let $q \in [0, 1)$ be fixed and suppose $r_\nu < \Gamma(c_0, q)$. From (15), note that

$$c_0 EV - \Delta_1(q)$$
$$= c_0 EV + r_\nu(1 - q)(1 - \kappa(\theta))EV - c_0 q(1 - \kappa(\theta))EV - c_0\kappa(\theta)EV - c_h r(\theta)\Delta_0.$$
$$< c_0 EV + (-c_0 + \phi_q(c_h))(1 - q)(1 - \kappa(\theta))EV - c_0 q(1 - \kappa(\theta))EV$$
$$- c_0\kappa(\theta)EV - c_h r(\theta)\Delta_0 = 0.$$

Therefore for all $j \geq 1$, $G^{\mathcal{F}_{j+1}}(\theta, q) > G^{\mathcal{F}_j}(\theta, q)$ and so $\mathcal{S}$ is optimal. The proof that $\mathcal{M}$ is optimal when $r_\nu > \Gamma(c_0, q)$ is similar.

ii. If $q = 1$, we have $G^{\mathcal{F}_{j+1}}(\theta, q) > G^{\mathcal{F}_j}(\theta, q)$ for all $j \geq 1$ since $\Delta_1(1) = c_0 EV + c_h r(\theta)\Delta_0$. Thus, $c_0 EV - \Delta_1(1) = -c_h r(\theta)\Delta_0 < 0$, and so $\mathcal{S}$ is the optimal policy. $\diamond$

Figure **??** illustrates that the optimal vacation policy for the server is to either take single vacation or multiple vacations. It can shown that as $q \in [0, 1)$ (extent of idling) increases, the extent of adopting $\mathcal{S}$ becomes larger. But when $q = 1$ and fixed c_h, $\mathcal{S}$ is optimal for all (r_ν, c_0).

9 Numerical Experiment

In this section, we present some numerical examples to study the impact of parameters on the system performance such as queue length and busy period induced by the vacation policy. Parameters are set so that the ergodicity condition is met. In the first experiment study, we want to study the effect of variability of the vacation length on the system size. Specifically, if we allow the vacation length to be more variable, what is the impact on the average number of callers and busy period for the system? To proceed, we adopt the concept of relative variability using Definition 4.8 in Song [26].

Definition 1 *Consider two random variables X and Y having the same mean $EX = EY$, having distributions F and G with densities f and g. Suppose X and Y are either both continuous or both discrete. We say that X is more variable than Y, denoted by $X \geq_{var} Y$, if f crosses g exactly twice, first from above and then from below.*

To this end, we consider the model where $X \sim Geo(0.5), \lambda = 0.2, \theta = 3, \alpha = 0.05$. For convenience, we choose $EB = 0.4$ and $ER = 0.8$. Assuming that only the vacation length varies, the rest of the parameters are kept constant. We plot our results for the mean system size affected across the different values of vacation. Figure 1 shows the experiment with $V_1 \sim Erlang(2, 0.5), V_2 \sim Erlang(8, 2), V_3 \sim Erlang(1, 0.5), V_4 \sim Erlang(4, 2)$. It is easy to see that $V_1 \geq_{var} V_2$ and $V_3 \geq_{var} V_4$. Figure 1 shows that for vacation length that is more variable, the average number of callers in the system caused by any number of successive vacation taken is always larger.
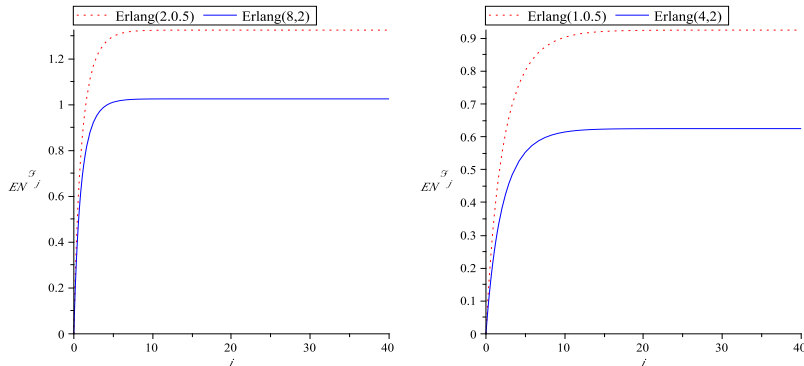

Figure 1: Plot of $EN^{\mathcal{F}_j}$ versus j.

Our second experiment investigates how the vacation policy affects the busy period. We want to numerically discuss the impact of a stochastically more variable vacation length on the busy period. Figure 2 shows that a more variable vacation length induces a greater mean busy period. Furthermore, the larger the number of vacations taken, the larger the mean busy period as well.
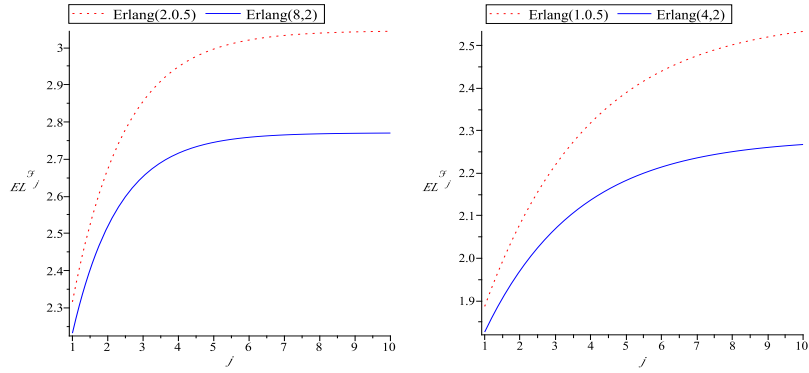
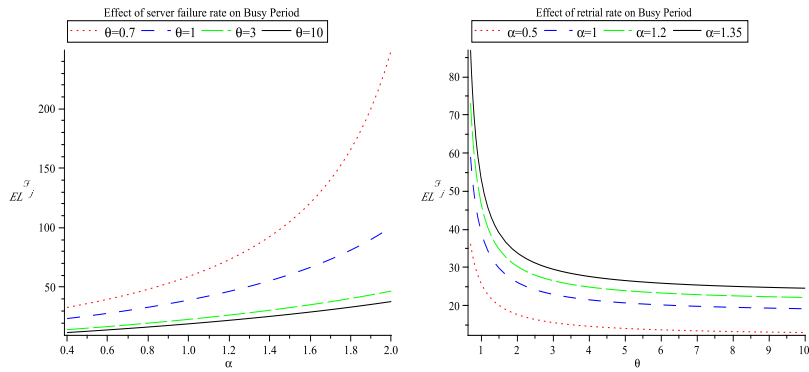17

Figure 2: Plot of $EL^{\mathcal{F}_j}$ versus j.
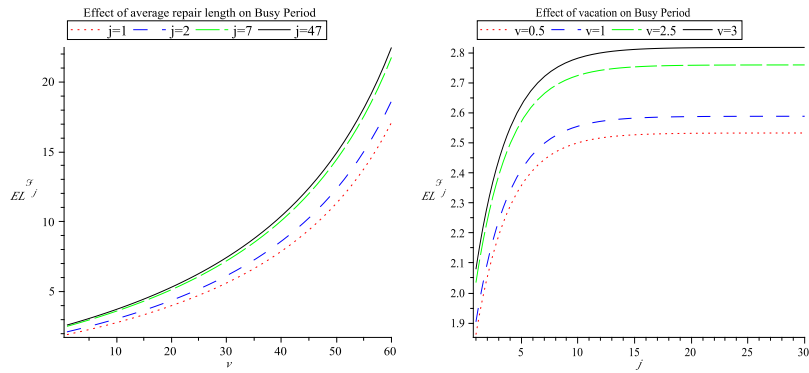

Figure 3: Plot of $EL^{\mathcal{F}_j}$ versus α and θ.

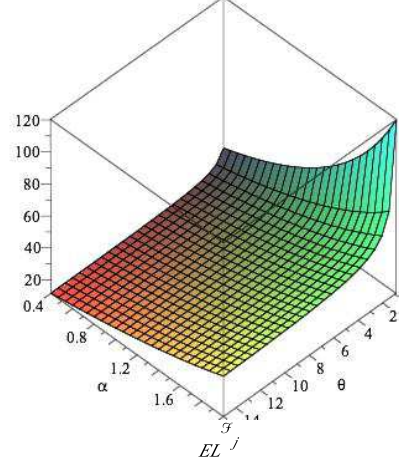Figure 4: Plot of $EL^{\mathcal{F}_j}$ versus j and v.

18

Figure 5: Joint effect of parameters on the busy period.

The third experiment discusses the impact of retrial rate and server failure rate on the busy period of the system with vacation. We assume that the server adopts single vacation policy with $V \sim Erlang(1, 0.5)$, $EB = 0.4$, $\lambda = 2$, $X \sim Geo(0.5)$ and $ER = 1$. Figure 3 shows that for each failure server rate α, the mean busy period is decreasing in θ, the retrial rate. As θ increases, the retrial system behaves more closely to those of $M^{(X)}/G/1$ queue with vacation and unreliable server. As such the busy period decreases because there is almost negligible retrial times and callers are queueing to be served. The right diagram in Figure 3 shows that for every retrial rate θ, the mean busy period increases in the server failure intensity, α which is again intuitive.

The last experiment involves fixing the vacation length $V \sim Erlang(1, 0.5)$. We want to study the joint impact of vacation policy and mean repair time on the busy period. Denote $v(= ER)$ and j to be the mean repair time and number of successive vacations taken by the server, respectively. The left diagram in Figure 4 shows that whenever the mean repair length increases, the busy period increases as well. Furthermore, we notice that as the successive number of vacation increases, the impact on the busy period gets reduced. The right diagram of Figure 4 shows that as the number of successive vacation increases, the busy period increases but stabilizes after a finite number of vacations. Numerically, we have establish the fact that for stochastically larger repair length of the server, the longer the mean busy period of the retrial system with vacation.

10   Discussions and Conclusions

Motivated by helpline services operating in resource-challenged environment, we described and analyzed an unreliable single server retrial system with batch ar-

rivals under multiple adaptive vacation policy. Our system generalizes a retrial system where the server takes either single vacation or multiple vacations. The constant retrial policy is adopted due to manpower shortage and searching for customers becomes important to handle the high level of stress and workload. We apply the technique of embedded Markov chain to derive the p.g.f of the system size at departure, denoted by $Q_+^{\mathcal{F}_j}(z)$. Furthermore, by using the technique of supplementary variables, we derive the p.g.f of system size at any time, denoted by $Q^{\mathcal{F}_j}(z)$. We establish the relationship between $Q^{\mathcal{F}_j}(z)$ and $Q_+^{\mathcal{F}_j}(z)$ which agrees in many $M/G/1$ queueing systems without vacations, including the retrial version. The stochastic decomposition of the system size is given and expressed as the sum of three random variables. Under this class of vacation policy, we show that single vacation policy minimizes caller's average waiting time. The technique of collective marks is applied to compute the system's average server-waiting length, average working length, and repair length. Finally, we derive the average length of a regenerative cycle to formulate an optimization model. Interestingly, our analysis reveals that either the single vacation or the multiple vacation policy is optimal to minimizing the long run average cost for the mobile business owner.

Congruent with intuitions, our experiment results show that variability of vacation length increases the average number of callers in the system. As the number of callers in orbit accumulates, the length of the busy period increases. For the same reason, increase in server failure rate and repair time also increase the length of the busy period. On the other hand, increase in retrial rate reduces the idle time of the computer. The net effect is the shortening of the busy period. We note that while an increase in the number of vacations taken by the server increases the busy period, the length of the busy period stabilizes rather quickly. We conjecture that this observation arises because service rate is relatively much higher than the caller arrival rate.

Our study assumes that the service rule is exhaustive. A natural extension of the model will be to consider vacation policy with non-exhaustive service. This means that the server can take vacation even when there are some callers in the orbit. So far, non-exhaustive service has only been studied in queues without retrial. One can also consider general inter-retrial times because it is more realistic to assume that callers have different behaviors. Recent efforts in the investigation of more interesting vacation policies have appeared in works of Ke [17, 19]. The work of Chang and Ke [15] that focuses on generalized retrial times can be included in our model. This extension might be significant because unreliable systems with generalized retrial times and repair times are not studied in literature. It is worthwhile to look at the joint impact of the generalized retrial times and repair times on the busy period.

11 Appendix

Notations	Meaning
$\mathcal{P}$	vacation policy
$M^X/G/1$	batch-arrival, general service-time queue
$M^X/G(M/H)/1$	unreliable batch-arrival, general service and repair-time after Markovian breakdown
$\mathcal{F}_j$	maximum allowable vacations to be taken is j
$\mathcal{S}$	single-vacation policy, i.e., $j = 1$
$\mathcal{M}$	multiple-vacation policy, i.e., $j \to \infty$
$Q^{\mathcal{P}}(t)$	transient number of callers both in service and in orbit under $\mathcal{P}$ and at t
$O^{\mathcal{P}}(t)$	transient number of callers in orbit under $\mathcal{P}$ and at t
$I^{\mathcal{P}}(t)$	transient number of callers in service under $\mathcal{P}$ and at t
$Q^{\mathcal{F}_j}(z)$	p.g.f of the system size at arbitrary epoch
$V_-^i(t)$	elapsed i^{th} vacation time at t
$X_-^i(t)$	elapsed i^{th} service time at t
$R_-^i(t)$	elapsed i^{th} repair time at t
$C(t)(= \nu, 0, 1, 2)$	state of the server at t, i.e., vacation, idle, busy or repair
$L^{(k)}$	$k-$busy period for the $M^X/G(M/H)/1$ retrial system without vacation
$L^{\mathcal{P}}$	busy period of $M^X/G(M/H)/1$ under $\mathcal{P}$ and $L^{\mathcal{P}} = L_0^{\mathcal{P}} + L_w^{\mathcal{P}} + L_r^{\mathcal{P}}$, where
$L_0^{\mathcal{P}}$	server-waiting time for the next busy epoch under $L^{\mathcal{P}}$
$L_1^{\mathcal{P}}$	generalized working length under $\mathcal{P}$ ($L_1^{\mathcal{P}} = L_w^{\mathcal{P}} + L_r^{\mathcal{P}}$)
$L_w^{\mathcal{P}}$	value-adding or working length under $L^{\mathcal{P}}$
$L_r^{\mathcal{P}}$	repair length during $L^{\mathcal{P}}$
$L_0^{(k)}$	server-waiting time for the next busy epoch under $L^{(k)}$
$L_1^{(k)}$	generalized working length under $k-$busy period ($L_1^{(k)} = L_w^{(k)} + L_r^{(k)}$)
$L_w^{(k)}$	value-adding or working length under $L^{(k)}$
$L_r^{(k)}$	repair length during $L^{(k)}$
$I(t)$	number of services completed at time t
$\omega(t)$	the elapsed service time of the work at time t.

Proof of Lemma 1: (i). The first part is to compute the p.g.f left by a departing caller. We divide our proofs into two cases.

Case 1: Given that $Q_n = 0$, the server goes into vacation immediately at S_n because it finds that the orbit is empty. Two events may happen, $\{Q_n = 0, \bar{\nu} = 0\}$ and $\{Q_n = 0, \bar{\nu} > 0\}$. On the event $\{Q_n = 0, \bar{\nu} = 0\}$, we must have $Q_{n+1} = X - 1 + M_{n+1}$. On the other hand, for the event $\{Q_n = 0, \bar{\nu} = i\}$ for $i \geq 1$, we have

$$Q_{n+1} = \begin{cases} i + M_{n+1} - 1 & \text{w.p } \frac{\theta}{\lambda + \theta} \\ i + M_{n+1} - 1 + X & \text{w.p } \frac{\lambda}{\lambda + \theta}. \end{cases}$$

$$P\{Q_{n+1} = k | Q_n = 0, \overline{\nu} = i\} = \begin{cases} \displaystyle\sum_{l=1}^{k+1} c_l m_{n+1,k-l+1} & \text{if } i = 0 \\ \displaystyle\frac{\lambda}{\lambda+\theta}\sum_{j=1}^{k-i+1} m_{n+1,k-j-i+1}c_j + \frac{\theta}{\lambda+\theta}m_{n+1,k-i+1} & \text{if } i \geq 1. \end{cases}$$

Thus, we have

$$p_{0k} = \sum_{i=0}^{\infty} P\{Q_{n+1} = k | Q_n = 0, \overline{\nu} = i\} P\{\overline{\nu} = i | Q_n = 0\}$$

$$= \widetilde{V}(\lambda)^j \sum_{l=1}^{k+1} c_l m_{n+1,k-l+1}$$

$$+ \sum_{i=1}^{k+1} \left\{ \frac{\lambda}{\lambda+\theta}\sum_{j=1}^{k-i+1} m_{n+1,k-j-i+1}c_j + \frac{\theta}{\lambda+\theta}m_{n+1,k-i+1} \right\} \overline{\nu}_i. \quad (16)$$

Case 2: For $Q_n \geq 1$, then we have similar argument. There are two types of customers that competes when $Q_n \geq 1$. One is the arrival of a new batch of customer, while the other is the random customer from the orbit. On $\{Q_n = i\}$, where $i \geq 1$, we have

$$Q_{n+1} = \begin{cases} i + M_{n+1} + X - 1 & \text{w.p } \frac{\lambda}{\lambda+\theta} \\ i + M_{n+1} - 1 & \text{w.p } \frac{\theta}{\lambda+\theta}. \end{cases}$$

Thus, we have for $k \geq i$,

$$p_{ik} = \frac{\lambda}{\lambda+\theta}\sum_{j=1}^{k-i+1} c_j m_{n+1,k-i+1-j} + \frac{\theta}{\lambda+\theta}m_{n+1,k-i+1}. \quad (17)$$

We want to compute the steady state distribution of the system size immediately after departure epochs. Since $\{Q_n : n \in \mathbf{N}\}$ constitutes a Markov chain, we can compute its steady state by using $\pi\mathbf{P} = \pi$ and $\pi\mathbf{1} = 1$. $\mathbf{P} = [p_{ij}]$ is the transition matrix of the embedded Markov Chain $\{Q_n\}$ and $\pi = (\pi_0, \pi_1\pi_2....)$, where π_k is the long run fraction when a departing customer leaves behind k customers in the system. Let $Q_+^{\mathcal{F}_j}(z)$ be the p.g.f of the system size that is left behind by a departing customer, then $Q_+^{\mathcal{F}_j}(z) = \sum_{k=0}^{\infty} \pi_k z^k$. Using (2), (3) and definition of $Q_+^{\mathcal{F}_j}(z)$, we have after some tedious algebraic manipulation,

$$Q_+^{\mathcal{F}_j}(z) = \pi_0 \frac{\{(1-\zeta_j(z))(\lambda X(z) + \theta) + \theta\widetilde{V}(\lambda)^j(1-X(z))\}\widetilde{H}(\lambda - \lambda X(z))}{(\lambda X(z) + \theta)\widetilde{H}(\lambda - \lambda X(z)) - (\lambda+\theta)z}. \quad (18)$$

From the normalization condition, $Q_+^{\mathcal{F}_j}(1) = 1$, we can find the constant π_0. Using (18) and L'Hospital rule, we obtain

22

$$\lim_{z \to 1^-} \Pi(z) = \pi_0 \left\{ \frac{\lambda\gamma_1 \left[\frac{1 - \widetilde{V}(\lambda)^j}{1 - \widetilde{V}(\lambda)} EV + \frac{\theta V(\lambda)^j}{\lambda(\lambda+\theta)} \right]}{1 - \kappa(\theta)} \right\} = 1$$

$$\Leftrightarrow \pi_0 = \frac{1 - \kappa(\theta)}{\lambda\gamma_1 \left[\frac{1 - \widetilde{V}(\lambda)^j}{1 - \widetilde{V}(\lambda)} EV + \frac{\theta V(\lambda)^j}{\lambda(\lambda+\theta)} \right]}.$$

Using (1) and (18), we obtain the p.g.f of the queue length left by a departing customer is given in the main paper.

(ii). **(Sufficiency)** The sufficiency of the ergodicity result can be shown using Pake's Lemma which is the statement as follows. If we have $Q(x) \geq 0$, and for all x, there exist $\epsilon > 0$ such that $E(Q_{t+1} - Q_t | Q_t = x) \leq -\epsilon < 0$ for all x except on a finite set C, then $\{Q_n\}$ is positive recurrent. To show sufficiency of our ergodicity result given any i and $\kappa(\theta) < 1$, we choose $\epsilon = \frac{1}{2}(1 - \kappa(\theta)) > 0$. Due to the recursive structure of the embedded sequence $\{Q_n : n \in \mathbf{N}\}$, we apply Pake's lemma as follows. The mean drift $y_i = E[Q_{n+1} - Q_n | Q_n = i]$ is calculated as follows.

$$\begin{aligned} y_i &= \frac{\theta}{\theta + \lambda} E[M_{n+1} - 1] + \frac{\lambda}{\theta + \lambda} E[X + M_{n+1} - 1] \\ &= EM_{n+1} - 1 + \frac{\lambda}{\theta + \lambda}\gamma_1 = \lambda\gamma_1 EH - 1 + \frac{\lambda}{\theta + \lambda}\gamma_1. \end{aligned}$$

Then, we have $y_i = -2\epsilon$, implying $y_i < -\epsilon$ for all states except for a finite number of states. Therefore, $\kappa(\theta) < 1$ is sufficient for the embedded chain to be ergodic. The necessary condition readily follows from Kaplan's condition (see Sennott et al ([25])), namely $y_i < \infty$ for all $i \geq 0$ and there exists $i_0 \in \mathbf{Z}^+$ such that $y_i \geq 0$ for all $i \geq i_0$. $\diamond$

Proof of Theorem 1: We define the generating functions

$$\omega_{\nu,i}(z, x) = \sum_{k=0}^{\infty} \omega_{\nu,k,i}(x)z^k; \omega_0(z) = \sum_{k=0}^{\infty} \omega_{0,k}(x)z^k;$$

$$\omega_1(z, x) = \sum_{k=0}^{\infty} \omega_{1,k}(x)z^k; \omega_0(z, y, x) = \sum_{k=0}^{\infty} \omega_{2,k}(x, y)z^k.$$

For notational parsimony, we omit the superscript $\mathcal{F}_j$ in the following discussion. Using the technique of supplementary variables, we obtain the following system

of equations for $x \geq 0, y \geq 0, k \geq 0$:

$$\left[\frac{d}{dx} + \lambda + \overline{\nu}(x)\right]\omega_{\nu,k,i}(x) = \lambda\sum_{l=1}^{k}c_j\omega_{\nu,k-l,i}(x), 1 \leq i \leq j \qquad (19)$$

$$\left[\frac{d}{dx} + \lambda + \overline{b}(x) + \alpha\right]\omega_{1,k}(x) = \lambda\sum_{l=1}^{k}c_j\omega_{1,k-l}(x) + \int_0^\infty \omega_{2,k}(x,y)\overline{r}(y)dy \quad (20)$$

$$\left[\frac{d}{dx} + \lambda + \overline{r}(x)\right]\omega_{2,k}(x,y) = \lambda\sum_{l=1}^{k}c_j\omega_{2,k-l}(x,y). \qquad (21)$$

For $k \geq 1$, we have

$$(\lambda + \theta)\omega_{0,k} = \int_0^\infty \omega_{1,k}(x)\overline{b}(x)dx + \sum_{i=1}^{j}\int_0^\infty \omega_{\nu,k,i}(x)\overline{\nu}(x)dx \qquad (22)$$

$$\lambda\omega_{0,0} = \int_0^\infty \overline{\nu}(x)\omega_{\nu,0,j}(x)dx. \qquad (23)$$

These equations are to be solved under the boundary conditions:

$$\omega_{\nu,0,1}(0) = \int_0^\infty \omega_{1,0}(x)\overline{b}(x)dx \qquad (24)$$

$$\omega_{\nu,0,i}(0) = \int_0^\infty \omega_{\nu,0,i-1}(x)\overline{\nu}(x)dx, 2 \leq i \leq j \qquad (25)$$

$$\omega_{\nu,k,i}(0) = 0, \quad k \geq 1, 1 \leq i \leq j \qquad (26)$$

$$\omega_{1,k}(0) = \lambda\sum_{j=1}^{k+1}c_j\omega_{0,k-j+1} + \theta\omega_{0,k+1}, k \geq 0 \qquad (27)$$

$$\omega_{2,k}(x,0) = \alpha\omega_{1,k}(x), k \geq 0. \qquad (28)$$

The normalizing condition is

$$\sum_{k=0}^{\infty}\left[\sum_{i=1}^{j}\int_0^\infty \omega_{\nu,k,i}(x)dx + \omega_{0,k} + \int_0^\infty \omega_{1,k}(x)dx + \int_0^\infty\int_0^\infty \omega_{2,k}(x,y)dydx\right] = 1.$$
$$(29)$$

Using generating functions, we can express system of equations to be

$$\left[\frac{d}{dx} + \lambda - \lambda X(z) + \overline{\nu}(x)\right]\omega_{\nu,i}(z,x) = 0, 1 \le i \le j \tag{30}$$

$$\left[\frac{d}{dx} + \lambda - \lambda X(z) + \overline{b}(x) + \alpha\right]\omega_1(z,x) = \int_0^\infty \omega_2(z,y,x)\overline{r}(y)dy \tag{31}$$

$$\left[\frac{d}{dy} + \lambda - \lambda X(z) + \overline{r}(y)\right]\omega_2(z,y,x) = 0. \tag{32}$$

$$\sum_{i=1}^{j}\omega_{\nu,0,i}(0) + (\lambda + \theta)\omega_0(z) - \theta\omega_{0,0}$$

$$= \int_0^\infty \omega_1(z,x)\overline{b}(x)dx + \sum_{i=1}^{j}\int_0^\infty \overline{\nu}(x)\omega_{\nu,i}(z,x)dx. \tag{33}$$

The boundary conditions can be expressed as

$$\omega_{\nu,i}(z,0) = \omega_{\nu,0,i}(0), \qquad 1 \le i \le j \tag{34}$$
$$z\omega_1(z,0) = (\lambda X(z) + \theta)\omega_0(z) - \theta\omega_{0,0} \tag{35}$$
$$\omega_2(z,0,x) = \alpha\omega_1(z,x). \tag{36}$$

To show $\omega_{\nu,i}^{\mathcal{F}_j}(z,x)$, (30) implies that

$$\omega_{\nu,i}(z,x) = \omega_{\nu,i}(z,0)[1 - V(x)]e^{-(\lambda-\lambda X(z))x}.$$

Next, (26), we have $\omega_{\nu,i}(z,0) = \omega_{\nu,0,i}(0)$. Using this, we have $\omega_{\nu,0,i}(x) = \omega_i(0,x) = \omega_{\nu,0,i}(0)[1 - V(x)]e^{-\lambda x}$. Combining it with (25), we have $\omega_{\nu,0,i}(0) = \omega_{\nu,0,i-1}(0)\widetilde{V}(\lambda)$. Recursively, we obtain for $1 \le i \le j$,

$$\omega_{\nu,i}(z,x) = \omega_{\nu,0,1}(0)\widetilde{V}(\lambda)^{i-1}[1 - V(x)]e^{-(\lambda-\lambda X(z))x}. \tag{37}$$

In particular, we get

$$\sum_{i=0}^{j}\omega_{\nu,0,i}(0) = g_j(\lambda)\omega_{\nu,0,1}(0); \lambda\omega_{0,0} = \omega_{\nu,0,1}(0)\widetilde{V}(\lambda)^j. \tag{38}$$

Next, (32) implies

$$\omega_2(z,y,x) = \omega_2(z,0,x)[1 - R(y)]e^{-(\lambda-\lambda X(z))y}. \tag{39}$$

Using both (39) and (36), we get the expression in (6). Substituting (39) into (31), we obtain the following:

$$\left[\frac{d}{dx} + \lambda - \lambda X(z) + \overline{b}(x) + \alpha - \alpha\widetilde{R}(\lambda - \lambda X(z))\right]\omega_1(z,x) = 0. \tag{40}$$

25

From (40), we get $\omega_1(z,x) = \omega_1(z,0)[1-B(x)]e^{(\lambda-\lambda X(z)+\alpha-\alpha\widetilde{R}(\lambda-\lambda X(z)))x}$. We substitute $\omega_1(z,x)$,(35),(37), and (38) into (33), we get

$$\omega_{\nu,0,1}(0)g_j(\lambda) + (\lambda+\theta)\omega_0(z) = \frac{\theta}{\lambda}\widetilde{V}(\lambda)^j\omega_{\nu,0,1}(0)$$
$$+ \frac{1}{z}[(\lambda X(z)+\theta)\omega_0(z) - \frac{\theta}{\lambda}\widetilde{V}(\lambda)^j\omega_{\nu,0,1}(0)] \quad (41)$$

After some re-arranging, we obtain

$$\omega_0^{\mathcal{F}_j}(z) = \omega_{\nu,0,1}(0)\frac{\{\frac{\theta\widetilde{V}(\lambda)^j}{\lambda}[\widetilde{H}(\lambda-\lambda X(z))-z]+z[1-\widetilde{V}(\lambda-\lambda X(z))]g_j(\lambda)\}}{[(\lambda X(z)+\theta)\widetilde{H}(\lambda-\lambda X(z))-(\lambda+\theta)z]}$$
$$(42)$$

Next using (35) and (42), we have

$$\omega_1(z,x) = \omega_{\nu,0,1}(0)\left(\frac{[1-\widetilde{V}(\lambda-\lambda X(z))](\lambda X(z)+\theta)g_j(y)+\theta\widetilde{V}(\lambda)^j[1-X(z)]}{(\lambda X(z)+\theta)\widetilde{H}(\lambda X(z)+\theta)-(\lambda+\theta)z}\right)$$
$$\times [1-B(x)]e^{(\lambda-\lambda X(z)+\alpha-\alpha\widetilde{R}(\lambda-\lambda X(z)))x}. \quad (43)$$

Integrating (43) w.r.t x, we obtain

$$\omega_1(z) = \omega_{\nu,0,1}(0)\left(\frac{[1-\widetilde{V}(\lambda-\lambda X(z))](\lambda X(z)+\theta)g_j(y)+\theta\widetilde{V}(\lambda)^j[1-X(z)]}{(\lambda X(z)+\theta)\widetilde{H}(\lambda X(z)+\theta)-(\lambda+\theta)z}\right)$$
$$\times \frac{1-\widetilde{H}(\lambda-\lambda X(z))}{\lambda-\lambda X(z)}. \quad (44)$$

Thus, we can obtain $\omega_2(z) = \alpha\omega_1(z)\frac{1-\widetilde{R}(\lambda-\lambda X(z))}{\lambda-\lambda X(z)}$. From the normalizing condition in (29), we have $\sum_{i=1}^j \omega_{\nu,i}(1)+\omega_0(1)+\omega_1(1)+\omega_2(1)=1$, and after some tedious algebra, we get $\omega_{\nu,0,1}(0) = C_j(\theta)$. $\diamond$

Proof of Corollary 1: We only show **i.** as the rest are similar. From Theorem 1, we consider $\omega_\nu^{\mathcal{F}_j}(1) = \lim_{z\to 1^-} C_j(\theta)g_j(\lambda)\left(\frac{1-\widetilde{V}(\lambda-\lambda X(z))}{\lambda-\lambda X(z)}\right)$ and applying L'Hospital rule, we obtain $\omega_\nu^{\mathcal{F}_j}(1) = C_j(\theta)g_j(\lambda)EV = \frac{[1-\kappa(\theta)]g_j(\lambda)EV}{g_j(\lambda)EV+\frac{\theta\widetilde{V}(\lambda)^j}{\lambda(\lambda+\theta)}} < 1$. It is ready to verify that $\omega_\nu^{\mathcal{F}_j}(1)+\omega_0^{\mathcal{F}_j}(1)+\omega_1^{\mathcal{F}_j}(1)+\omega_2^{\mathcal{F}_j}(1)=1$. $\diamond$

Proof of Corollary 2: Let $O^{\mathcal{F}_j}(z)$ and $Q^{\mathcal{F}_j}(z)$ be the p.g.f for the number of callers in orbit and system respectively. In order to prove the results for (i) and (ii), they follow easily from the fact that $O^{\mathcal{F}_j}(z) = \omega_\nu^{\mathcal{F}_j}(z)+\omega_0^{\mathcal{F}_j}(z)+\omega_1^{\mathcal{F}_j}(z)+\omega_2^{\mathcal{F}_j}(z)$ and $Q^{\mathcal{F}_j}(z) = \omega_\nu^{\mathcal{F}_j}(z)+\omega_0^{\mathcal{F}_j}(z)+z(\omega_1^{\mathcal{F}_j}(z)+\omega_2^{\mathcal{F}_j}(z))$. $\diamond$

Proof of Theorem 2: We want to show that the p.g.f for the number of customers in the orbit can be written as the sum of three random variables. Let M be the orbit size of the system based on constant retrial policy. First, we observe that

Atencia et al ([7], Theorem 3) has shown that $N_R = N_0 + M$. Finally the p.g.f for $N^{\mathcal{F}_j}(z)$ allows us to conclude that $N_\nu^{\mathcal{F}_j} = N^{\mathcal{F}_j} + N_0 + M$. $\diamond$

Proof of Corollary 7: We have $EC_\nu^{\mathcal{P}} = EL^{\mathcal{P}} + E\zeta_\nu^{\mathcal{P}}$ and using results in Theorem 4 and Corollary 6, we have the required result. $\diamond$

Proof of Lemma 2: To show Lemma 2, we consider the following. Denote $m_n(s) = P\{M = n, \tau_s > H\}$. For any $k \geq 1$, we have $\pi_{n,1}^{(k)}(s) = m_{n-k+1}(s)$ and for $i \geq 2$,

$$\pi_{n,i}^{(k)}(s) = \sum_{j=1}^n \pi_{n,i-1}^{(k)}(s) \sum_{l=1}^{n-j+1} \frac{\lambda c_l}{s+\lambda+\theta} m_{n-j-l+1}(s) + \sum_{j=1}^{n+1} \pi_{j,i-1}^{(k)}(s) \frac{\theta}{s+\lambda+\theta} m_{n-j+1}(s).$$

Finally, the proof is completed by showing that $\sum_{n=0}^\infty m_n(s)z^n = \widetilde{H}(s+\lambda - \lambda X(z))$. $\diamond$

For ease of exposition, we define $F(s,z,y,x) = (s+\lambda+\theta) - \frac{y}{z}(\lambda X(z) + \theta)\widetilde{H}(s+\lambda-\lambda X(z))$.

11.1 Proof of Theorem 3

The proof of Theorem 3 requires two further lemmas, i.e., Lemma 5 and Lemma 6. These results allow us to compute the working and server-waiting length of the k-busy period. Finally, we obtain $EL^{(k)}$ which agrees with Atencia et al [7] who use the technique of supplementary variables.

Lemma 5 *The generating function $\varphi_0^{(k)}(s,z,y)$ satisfies the functional equation*

$$F(s,z,y,x)\varphi_0^{(k)}(s,z,y) = yz^{k-1}\widetilde{H}(s+\lambda-\lambda X(z)). \tag{45}$$

Proof: Using the fact that $s\varphi_{0ni}^{(k)}(s) = P\{L^{(k)} > \tau_s, C(\tau_s) = 0, Q(\tau_s) = n, I(\tau_s) = i\} = \frac{s}{s+\lambda+\theta}\pi_{ni}^{(k)}(s)$. Thus, we have $\varphi_{0ni}^{(k)}(s) = \frac{\pi_{ni}^{(k)}(s)}{s+\lambda+\theta}$. The result follows from (9) since $\varphi_0^{(k)}(s,z,y) = \frac{f^{(k)}(s,z,y)}{s+\lambda+\theta}$. $\diamond$

Lemma 6 *The generating function $\varphi_1^{(k)}(s,z,y,x)$ satisfies the functional equation*

$$\varphi_1^{(k)}(s,z,y,x) = (1-H(x))e^{-(s+\lambda-\lambda X(z))x}$$
$$\times \left[z^{k-1} + \frac{1}{z}(\lambda X(z) + \theta)\varphi_0^{(k)}(s,z,y)\right]. \tag{46}$$

In addition, we have

$$\int_0^\infty \varphi_1^{(k)}(s,z,y,x)dx = \frac{1 - \widetilde{H}(s+\lambda-\lambda X(z))}{s+\lambda-\lambda X(z)}$$
$$\times \left[z^{k-1} + \frac{1}{z}(\lambda X(z) + \theta)\varphi_0^{(k)}(s,z,y)\right]. \tag{47}$$

Proof: Observe that $s\varphi_{1ni}^{(k)}(s,x) = P\{L^{(k)} > \tau_s, C(\tau_s) = 1, \omega(\tau_s) \in (x, x + dx), Q(\tau_s) = n, I(\tau_s) = i\}$. Following the arguments in Falin et al [12] or Artalejo et al [6], we obtain

$$\varphi_{1n0}^{(k)}(s,x) = [1 - H(x)]e^{-sx} \sum_{r=1}^{n-k+1} e^{-\lambda x} \frac{(\lambda x)^r}{r!} P\{X_1 + ... + X_r = n - k + 1\}$$

$$\varphi_{1ni}^{(k)}(s,x) = [1 - H(x)]e^{-sx}$$

$$\times \left\{ \sum_{j=1}^{n} \pi_{ji}^{(k)}(s) \sum_{l=1}^{n-j+1} \frac{\lambda c_l}{s + \lambda + \theta} \sum_{r=1}^{n-j+1-l} e^{-\lambda x} \frac{(\lambda x)^r}{r!} P\left(\sum_{t=1}^{r} X_t = n - j + 1 - l\right) \right.$$

$$\left. + \sum_{j=1}^{n+1} \pi_{ji}^{(k)}(s) \frac{\theta}{s + \lambda + \theta} \sum_{r=1}^{n-j+1} e^{-\lambda x} \frac{(\lambda x)^r}{r!} P\left(\sum_{t=1}^{r} X_t = n - j + 1\right) \right\}.$$

The result follows from the definition of $\varphi_1^{(k)}(s,z,y,x)$ after tedious algebraic manipulations. Finally, (47) follows from integrating (46) w.r.t x. $\diamond$

Proof of Theorem 3: In order to compute $L_0^{(k)}$, we need to compute $\lim_{z\to 1^-} \varphi(0,z,1)$. Let $s = 0, y = 1$ into (45), we obtain

$$\left[(\lambda + \theta) - \frac{1}{z}(\lambda X(z) + \theta)\widetilde{H}(\lambda - \lambda X(z))\right] \varphi_0^{(k)}(0,z,1) = z^{k-1}\widetilde{H}(\lambda - \lambda X(z)).$$

Differentiating the above equation w.r.t z and letting $z \to 1$, we obtain the desired result for $L_0^{(k)}$. The expected generalized working length of the k-busy period is given by $\lim_{z\to 1^-} \int_0^\infty \varphi_1^{(k)}(0,z,1,x)dx$. Note that $\lim_{z\to 1^-} \frac{1-\widetilde{H}(\lambda-\lambda X(z))}{\lambda-\lambda X(z)} = EH$. $\diamond$

Proof of Corollary 4: The proof is immediate from Theorem 3 and the fact that $EH = EB(1 + \alpha ER)$. $\diamond$

11.2 Proof of Theorem 4

To prove Theorem 4, we shall begin with a preliminary lemma.

Lemma 7 *The generating functions $\psi_0(s,z,y)$ and $\psi_1(s,z,y,x)$ satisfy the functional equations*

$$F(s,z,y,x)\psi_0(s,z,y) = \frac{y}{z}\left\{\zeta_j(z)\left(\frac{\lambda X(z) + \theta}{\lambda + \theta}\right) + \frac{\theta}{\lambda + \theta}\widetilde{V}(\lambda)^j[X(z) - 1]\right\}$$
$$\times \widetilde{H}(s + \lambda - \lambda X(z)).$$

$$\psi_1(s,z,y,x) = (1 - H(x))e^{-(s+\lambda-\lambda X(z))x}\frac{1}{z}$$
$$\times \left(\left\{\zeta_j(z)\left(\frac{\lambda X(z) + \theta}{\lambda + \theta}\right) + \frac{\theta}{\lambda + \theta}\widetilde{V}(\lambda)^j[X(z) - 1]\right\} \quad (48)$$
$$+ (\lambda X(z) + \theta)\psi_0(s,z,y)\right).$$

28

Proof: The results follow from applying Lemma 5 and 6. $\diamond$

Proof of Theorem 4: From Lemma 7, we let $s = 0, y = 1$, we have

$$\left[(\lambda + \theta) - \frac{1}{z}(\lambda X(z) + \theta)\widetilde{H}(\lambda - \lambda X(z))\right]\varphi_0(0, z, 1)$$
$$= \frac{\zeta(z)}{z}\left(\frac{\lambda X(z) + \theta}{\lambda + \theta}\right)\widetilde{H}(\lambda - \lambda X(z)).$$

Differentiate the above equation w.r.t z and let z approach 1, we obtain $L_0^{\mathcal{F}_j}$. Combining with (1), $L_1^{\mathcal{F}_j}$ is obtained by using $\lim_{z\to 1^-}\int_0^\infty \varphi(0, z, 1, x)dx$. Finally, $EL^{\mathcal{M}} = \varphi(0, 1, 1) + \lim_{z\to 1^-}\int_0^\infty \varphi(0, z, 1, x)dx$. $\diamond$

Proof of Lemma 4: To see this, we apply Theorem 4 and after some re-arranging, we have

$$EL^{\mathcal{F}_j} = -\frac{1}{\lambda + \theta} + \left[\widetilde{V}(\lambda)^j\left(\frac{\theta}{\lambda(\lambda + \theta)} - \frac{EV}{1 - \widetilde{V}(\lambda)}\right) + \frac{EV}{1 - \widetilde{V}(\lambda)}\right]\frac{\kappa(\theta)}{1 - \kappa(\theta)}.$$

Given that $\widetilde{V}(\lambda) \geq 0$, $\widetilde{V}(\lambda)^j$ is always increasing function in j. It is easy to see that the sign of $\frac{\theta}{\lambda(\lambda + \theta)} - \frac{EV}{1 - \widetilde{V}(\lambda)}$ determines the if $EL^{\mathcal{F}_j}$ is increasing or decreasing. $\diamond$

References

[1] A. Aissani, An $M^{(x)}/G/1$ retrial queue with exhaustive vacations, J. of Statistics and Management Systems. 3 (2000) 269-286.

[2] J.R. Artalejo, A classified bibliography of research on retrial queues: progress in 1990-1999, TOP. 7, 18(1) (1999a) 187-211.

[3] J.R. Artalejo, Accessible biblography on retrial queues, Mathematical and Computer Modelling. 30 (1999b) 1-6.

[4] J.R. Artalejo, A. Gomez-Corral, Retrial Queueing Systems: A computational approach, Springer, 2008.

[5] J.R. Artalejo, Analysis of an $M/G/1$ retrial queue with constant rate of repeated attempts and server vacations, Computers and Operations Research. 24 (1997) 493-504.

[6] J.R. Artalejo, G.I. Falin, M.J. Lopez-Herrero, A second order analysis of the waiting time in the $M/G/1$ retrial queue, Asia-Pacific Journal of Operational Research. 19(2) (2002) 131-148.

[7] I. Atencia, G. Bouza, P. Moreno, An $M^{(x)}/G/1$ retrial queue with server breakdowns and constant rate of repeated attempts, Annals of Operations Research. 157 (2008) 225-243.

[8] B.D. Choi, Y.W. Shin, W.C Ahn, Retrial queues with collision arising from unslooted CSMA/CD Protocols, Queueing Systems, 11 (1992), 335-356.

[9] K. Farahmand, Single line queue with repeated demands, Queueing Systems, 6 (1990) 223-228.

[10] G. Fayolle, A simple telephone exchange with delayed feedbacks, Chapter In Book "Teletraffic Analysis and Computer Performance Evaluation", pp. 245-253, Amsterdam: North Holland.

[11] G. Falin, A survey of retrial queues, Queueing Systems, 7 (1990) 127-168.

[12] G. Falin, J.G.C. Templeton, Retrial Queues, Monographs on Statistics and Applied Probability 75, Chapman and Hall, 1997.

[13] G. Falin, An $M/M/1$ retrial queue with retrial due to server failures, Queueing Systems. 58(3) (2008) 155-160.

[14] G. Falin, An $M/G/1$ retrial queue with an unreliable server and general repair times , Performance Evaluation. 67(7) (2010) 569-582.

[15] F.M. Chang, J.C. Ke, On a Batch Retrial Model with J Vacation, Journal of Computational and Applied Mathematics. 232 (2009) 402-414.

[16] J.C. Ke, Y.K. Chu, A modified vacation model $M^{(x)}/G/1$ system, Applied Stochastic Models in Business and Industry. 22 (2006) 1-16.

[17] J.C. Ke, An $M/G/1$ Queue Under Hysteretic Vacation Policy with an early Startup and Unreliable Server, Math. Meth. Oper. Res. 63 (2006) 357-369.

[18] J.C. Ke, Batch arrival queues under vacation policies with server breakdowns and startup/closedown times, Applied Mathematical Modelling, 31 (2007) 1282-1292.

[19] J.C. Ke, Optimal Strategy Policy in Batch Arrival Queue with Server Breakdowns and Multiple Vacations, Math. Meth. Oper. Res. 58 (2003) 41-56.

[20] B. Krishna Kumar, D. Arivudainambi, A. Vijayakumar, On the $M^{(x)}/G/1$ retrial queue with Bernoulli Schedules and general retrial times, Asia-Pacific Journal of Operational Research. 19(2) (2002) 177-194.

[21] V.G. Kulkarni, B.D. Choi, Retrial Queues with Server Subject to Breakdowns and Repairs, Queueing Systems. 7 (1990) 191-208.

[22] M. Senthilkumar, R. Arumuganathan, On the single server batch arrival retrial queue with general vacation time under Bernoulli schedule and two phases of heterogeneous service, Quality Technology and Quantitative Management. 5(2) (2008) 145-160.

[23] H. Li, Y.Q. Zhao, A retrial queue with constant retrial rate, server downs and impatient customers, Stochastic Models. 21 (2005) 531-550.

[24] Q. Li, Y. Ying, Y.Q. Zhao, A $BMAP/G/1$ Retrial Queue with a Server Subject to Breakdowns and Repairs, Annals of Operations Research. 141 (2006) 233-270.

[25] L.I. Senott, R.L. Humblet, R.L. Tweedie Mean drifts and non-ergodicity of Markov chain, Operations Research. 31 (1983) 783-789.

[26] J.S. Song, "The Effect of Leadtime Uncertainty in a Simple Stochastic Inventory Model", Management Science. 40(5) (1994) 603-613.

[27] H. Takagi, Queueing Analysis, Vol. 1: Vacation and Priority Systems, North-Holland Elsevier, Amsterdam, 1991.

[28] Y.H. Tang, A single server $M/G/1$ queueing system subject to breakdowns-some reliability and queueing problem, Microelectronics and Reliability. 37 (1997) 315-321.

[29] N.S. Tian, Z.G. Zhang Vacation Queueing Models, Springer, 2006.

[30] J. Wang, J. Cao, Q. Li, Reliability Analysis of the Retrial Queue with Server Breakdowns and Repairs, Queueing Systems. 38(4) (2001) 363-380.

[31] T. Yang, J.G.C. Templeton, A survey of retrial queues, Queueing Systems. 2 (1987) 201-233.

[32] T. Yang, H. Li, The M/G/1 Retrial Queue with the Server Subject to Starting Failures, Queueing Systems. 16 (1994) 83-96.