RAIRO-Oper. Res. 52 (2018) 35–54
https://doi.org/10.1051/ro/2017091

RAIRO Operations Research
www.rairo-ro.org

SENSITIVITY ANALYSIS OF AN M/G/1 RETRIAL QUEUEING SYSTEM WITH DISASTER UNDER WORKING VACATIONS AND WORKING BREAKDOWNS

P. Rajadurai*

Abstract. This paper deals with the new type of retrial queueing system with working vacations and working breakdowns. The system may become defective by disasters at any point of time when the regular busy server is in operation. The occurrence of disasters forces all customers to leave the system and causes the main server to fail. At a failure instant, the main server is sent to the repair and the repair period immediately begins. As soon as the orbit becomes empty at regular service completion instant or disaster occurs in the regular busy server, the server goes for a working vacation and working breakdown (called lower speed service period). During this period, the server works at a lower service rate to arriving customers. Using the supplementary variable technique, we analyze the steady state probability generating function of system size. Some important system performance measures are obtained. Finally, some numerical examples and cost optimization analysis are presented.

Mathematics Subject Classification. 60K25, 90B22, 68M20

Received December 24, 2016. Accepted December 19, 2017.

1. Introduction

In queueing theory, retrial queues have been intensive research topics for long time; we can find general models in retrial queues from Artalejo and Gomez-Corral [3] and Artalejo [2]. In retrial queueing system, retrial queues with repeated attempts are characterized by the fact that an arriving customer finds the server busy upon arrival is requested to leave the service area and join a retrial queue called orbit. After some time the customer in the orbit can repeat their request for service. An arbitrary customer in the orbit who repeats the request for service is independent of the rest of the customers in the orbit. Such queues play a special role in computer and telecommunication systems.

In a vacation queueing system, during the working vacation (WV) period, the server serves to customer at a lower service rate, but the server stops the service completely during the normal vacation period. This queueing system has major applications in providing network service, web service, file transfer service and mail service etc. In 2002, Servi and Finn [21] introduced an M/M/1 queueing system with working vacations. Wu and Takagi [23] extended the M/M/1/WV queue to an M/G/1/WV queue. Very recently, Arivudainambi *et al.* [1] have introduced an M/G/1 retrial queue with general retrial times and single working vacation. Chandrasekaran *et al.*

Keywords and phrases: Retrial queue, disaster, working vacations, working breakdowns.

Department of Mathematics, Srinivasa Ramanujan Centre, SASTRA University, Kumbakonam 612001, Thanjavur, Tamil Nadu, India.

* Corresponding author: psdurai17@gmail.com

[5] presented a short survey on working vacation queueing models. Furthermore, during the working vacation period, if there are customers at a service completion instant, the server can stop the vacation and come back to the regular busy state. This policy is called vacation interruption. Recently, authors like Gao *et al.* [9], Zhang and Hou [26], Gao and Liu [8], Zhang and Liu [27] and Rajadurai *et al.* [17, 19] analyzed a single server retrial queue with working vacations and vacation interruptions.

In queueing literature, mostly it is assumed that the server is available all the time on permanent basis. But the server failures which lead to service interruptions are quite common in many real life situations. It is well known that performance measures of unreliable queuing systems are heavily influenced by server failures. Boudali and Economou [4] have discussed the effect of catastrophes on the strategic customer behavior in queueing systems, where single server Markovian queue subject to Poisson generated catastrophes is considered. Whenever a catastrophe occurs, all customers are forced to abandon the system. Sudhesh *et al.* [22] have developed a model with N-policy queues with disastrous breakdown. Recently, Choudhury and Deka [6], Dimitriou [7], Lee *et al.* [14], Yang *et al.* [24] and Rajadurai *et al.* [16, 18] have considered the unreliable queueing systems with various features in which one of the assumptions is that as soon as failure occurs the server is sent for repair, during that time it stops providing service to the primary customers till service channel is repaired and customer who was just being served before server failure waits for the remaining service to complete.

This is the concept of the working breakdowns first introduced by Kalidass and Ramanath [12] in 2012. That is, the system may become defective by disasters at any point of time when the regular busy server is in operation, the system should be ready with a substitute (standby) server in preparation for possible main server failures. The substitute server renders services to customers while the main server is repaired. The service rate of the substitute server is different from (lower than) that of the main server. At the instant of the repair completion, the main server returns to the system and becomes available. Additionally, the working breakdown service can decrease complaints from the customers who should wait for the main server to be, repaired and reduces the cost of waiting customers. Therefore, the working breakdown service is a more reasonable repair policy for unreliable queueing systems. Kim and Lee [13] have discussed a model M/G/1 queueing system with disasters and working breakdown services. Recently, Jiang and Liu [11] have developed a model GI/M/1 queue in a multi-phase service environment with disasters and working breakdowns.

In view of the above, present investigation deals as the extension of Kalidass and Ramanath [12], Kim and Lee [13] by considering a single server retrial queueing system with multiple working vacations. To the author's best of knowledge, there are many works have been done in the literature for various queueing systems but not for working breakdown queues with retrials. Retrial queues, working vacations and working breakdowns are advanced level in queueing system. Motivated by above situations, this paper introduces and analyzes a new class of M/G/1 retrial queue with disasters under working breakdowns and working vacations. During the vacation and breakdown time also server works in different rate of services. Analytical results could provide the decision makers and the practitioners with very useful and helpful managerial information for designing a management policy. In addition, the discussed system has a potential application in the computer processing system, stochastic production and inventory systems with a multipurpose production facility and the machine replacement problems. The details of application example are introduced in Section 2.1.

The aim of this investigation is to find the queue size and orbit size distributions which are further employed to obtain the cost analysis and other performance measures of the system. The outline of the remaining sections is as follows: With the requisite assumptions, the brief mathematical model and the practical application are described in Section 2. In Section 3, the governing equations of our model, the steady state joint distribution of the server states and the number of customers in the orbit/system are obtained. Some important system performance measures are given in Section 4. In Section 5, the conditional stochastic decomposition law is demonstrated. Important special cases are given in Section 6. In Section 7, the cost optimization analysis is discussed. The effects of various parameters on the system performance are analyzed numerically in Section 8. Summary and conclusions of the work is presented in Section 9.

2. Description of the model

In this section, we consider a single server retrial queueing system with disasters under working vacations and working breakdowns. The detailed description of model is given as follows:

- ***The arrival process:*** Customers arrive at the system according to a Poisson process with rate λ.
- ***The retrial process:*** We assume that there is no waiting space and therefore if an arriving customer finds the server free, the customer begins his service immediately. Otherwise an arriving customer finds the server regular busy or lower speed service, the arrivals join the pool of blocked customers called an orbit in accordance with FCFS discipline. That is, only one customer at the head of the orbit queue is allowed access to the server. Inter-retrial times have an arbitrary distribution $R(t)$ with corresponding Laplace Stieltjes Transform (LST)$R^*(\vartheta)$.
- ***The regular service process:*** In normal busy period, the service time follows a general distribution. It is denoted by the random variable S_b with distribution function (d.f.) $S_b(t)$and LST $S_b^*(\vartheta)$.
- ***The lower speed service process:*** Here, both working vacation period and working breakdown period are considered as lower speed service period.
  (i) ***The multiple working vacations process:*** The server begins a working vacation each time when the orbit becomes empty and the vacation time follows an exponential distribution with parameter θ. If any customer arrives in a vacation period, the server continues to work at a lower speed service rate ($\mu_w < \mu_b$). The working vacation period is an operational period at a lower speed. If any customers are in the orbit at a lower speed service completion instant in the vacation period, the server will stop the vacation and come back to the normal busy period which means vacation interruption happens. Otherwise, it continues the vacation. When a vacation ends, if there are customers in the orbit, the server switches to the normal working level. Otherwise, the server begins another vacation.
  (ii) ***The working breakdown process:*** The system may become defective by disasters at any point of time when the regular busy server is in operation with exponentially distributed with a rate of α. The occurrence of disasters forces all present customers to leave the system and causes the main server to fail. At a failure instant, the main server is sent to the repair and the repair period immediately begins. The repair time follows an exponential distribution with rate of δ. The repaired server is assumed to be as good as a new server. However disaster occurs in the regular busy server, the server goes for a working breakdown. During working breakdown period, the substitute server works at a lower service rate to arriving customers ($\mu_w < \mu_b$). When a repair ends, if there are customers in the orbit, the server switches to the normal working level and will start a new busy period. Otherwise, it is idle and ready for serving new arrivals. During the working vacation and working breakdown periods (lower speed services), the service time follows a general random variable S_w with d.f. $S_w(t)$ and LST$S_w^*(\vartheta)$.
- Various stochastic processes involved in the system are assumed to be independent of each other.

2.1. Practical justifications of the suggested model

Our model has potential practical application in the area of computer processing system. In a computer processing system, the buffer size (orbit) used to store messages is finite and the messages (customers) arrive into the system one by one, and the processor (server) is in charge of processing messages. The working mail server may be affected by virus (breakdowns) and the system may slow down (working breakdown) the performance of the computer system. Then the main mail server sent for repair immediately. The computer system may still be able to perform various chores but at a considerably slower rate (lower service speed). If the processor is available indicating that it is not currently working on a task and then a message is processed. The messages are temporarily stored in a buffer to be served some time later (retrial time) according to FCFS if the processor is unavailable. To enhance the computer performance, whenever all messages are processed and no new messages arrive, the processor will perform a sequence of maintenance jobs, such as virus scan (working vacations). During the maintenance period, the processor can deal with the messages at the slower rate to economize the cost (working vacation period). Upon completion of the each maintenance, if no message is in the system then the processor may decide to go for another maintenance activity (multiple working vacations). This type of working vacation and working breakdown discipline is a good approximation of such computer processing system. This model finds other practical application in the operational model of stochastic production and inventory systems with a multipurpose production facility and the machine replacement problems.

3. Steady state analysis

In this section, we develop the steady state difference-differential equations for the retrial queueing system by treating the elapsed retrial times, the elapsed service times and the elapsed lower speed service times as supplementary variables. Then we derive the generating functions (GFs) of the orbit size for different server's states, the probability generating functions (PGFs) of number of customers in the system and orbit.

3.1. Notations and probabilities

In steady state, we assume that $R(0) = 0$, $R(\infty) = 1$, $S_b\ (0) = 0$, $S_b(\infty) = 1$, $S_w\ (0) = 0$, $S_w\ (\infty) = 1$ are continuous at $x = 0$. The following notations and probabilities are used in sequent sections:

- $a(x) \equiv$ the hazard rate (conditional completion rate) for retrial of $R(x)$; *i.e.*, $a(x)\mathrm{d}x = \frac{\mathrm{d}R(x)}{1-R(x)}$.
- $\mu_b(x) \equiv$ the hazard rate for service of $S_b\ (x)$; *i.e.*, $\mu_b(x)\mathrm{d}x = \frac{\mathrm{d}S_b(x)}{1-S_b(x)}$.
- $\mu_w(x) \equiv$ the hazard rate for lower rate service of $S_w\ (x)$; *i.e.*, $\mu_w(x)\mathrm{d}x = \frac{\mathrm{d}S_w(x)}{1-S_w(x)}$.
- $N(t) \equiv$ the number of customers in the orbit.
- $R^0(t) \equiv$ the elapsed retrial time.
- $S_b^0(t) \equiv$ the elapsed service time.
- $S_w^0(t) \equiv$ the elapsed lower rate service time.
- $P_0(t) \equiv$ the probability that the system is empty at time t.
- $Q_0(t) \equiv$ the probability that the system is empty at time t and the server is in working vacation and breakdown (lower speed service).
- $P_n(x,t) \equiv$ the probability that at time t there are exactly n customers in the orbit with the elapsed retrial time of the test customer undergoing retrial lying in between x and $x + \mathrm{d}x$.
- $\Pi_{b,n}(x,t) \equiv$ the probability that at time t there are exactly n customers in the orbit with the elapsed normal service time of the test customer undergoing service lying in between x and $x + \mathrm{d}x$.
- $\Pi_{w,n}(x,t) \equiv$ the probability that at time t there are exactly n customers in the orbit with the elapsed lower rate service time of the test customer undergoing service lying in between x and $x + \mathrm{d}x$.

3.2. The steady state equations

For further development of this retrial queueing model, let us define the random variable

$$C(t) = \begin{cases} 0, & \text{if the server is free and in working vacation and working breakdown period,} \\ 1, & \text{if the server is free and in regular service period,} \\ 2, & \text{if the server is busy and in regular service period on both phases at time } t, \\ 3, & \text{if the server is busy and in lower speed service period period at time } t. \end{cases}$$

Thus the supplementary variables $R^0(t)$, $S_b^0(t)$ and $S_w^0(t)$ are introduced in order to obtain a bivariate Markov process $\{C(t), N(t);\ t \geq 0\}$, where $C(t)$ denotes the server state (0, 1, 2, 3) depends on the server is free, regular busy and lower speed busy. If $C(t) = 1$ and $N(t) > 0$, then $R^0(t)$ represent the elapsed retrial time. If $C(t) = 2$ and $N(t) \geq 0$ then $S_b^0(t)$ corresponding to the elapsed time of the customer being served in normal busy period. If $C(t) = 3$ and $N(t) \geq 0$ then $S_w^0(t)$ corresponding to the elapsed time of the customer being served in lower rate service period.

Let $\{t_n; n = 1, 2, \ldots\}$ be the sequence of epochs at which either a normal service or lower service period completion occurs. The sequence of random vectors $Z_n = \{C\,(t_n+)\,,\ N\,(t_n+)\}$ forms a Markov chain which is embedded in the retrial queueing system. It follows from Appendix A that $\{Z_n;\ n \in N\}$ is ergodic if and only if $\rho < R^*(\lambda)$, for our system to be stable, where $\rho = \frac{\lambda}{\alpha}\left(1 - S_b^*(\alpha)\right)$.

For the process $\{N(t),\ t \geq 0\}$, we define the probabilities $P_0(t) = P\{C(t) = 0,\ N(t) = 0\}$ and $Q_0(t) = P\{C(t) = 0,\ N(t) = 0\}$ the probability densities

$$P_n(x,t)\mathrm{d}x = P\left\{C(t) = 1, N(t) = n, x \leq R^0(t) < x + \mathrm{d}x\right\}, \quad \text{for } t \geq 0,\ x \geq 0 \text{ and } n \geq 1.$$
$$\mathit{\Pi}_{b,n}(x,t)\mathrm{d}x = P\left\{C(t) = 2, N(t) = n, x \leq S_b^0(t) < x + \mathrm{d}x\right\}, \quad \text{for } t \geq 0,\ x \geq 0,\ n \geq 0.$$
$$\mathit{\Pi}_{w,n}(x,t)\mathrm{d}x = P\left\{C(t) = 4, N(t) = n, x \leq S_w^0(t) < x + \mathrm{d}x\right\}, \quad \text{for } t \geq 0,\ x \geq 0, \text{ and } n \geq 0.$$

We assume that the stability condition is fulfilled in the sequel and so that we can set $P_0 = \lim\limits_{t\to\infty} P_0(t)$ and $Q_0 = \lim\limits_{t\to\infty} Q_0(t)$ limiting densities for $(x,y) > 0$ and $n \geq 0$,

$$P_n(x) = \lim_{t\to\infty} P_n(x,t), \quad \mathit{\Pi}_{b,n}(x) = \lim_{t\to\infty} \mathit{\Pi}_{b,n}(x,t) \quad \textit{and} \quad \mathit{\Pi}_{w,n}(x) = \lim_{t\to\infty} \mathit{\Pi}_{w,n}(x,t).$$

Using the method of supplementary variable technique, we obtain the following system of equations that govern the dynamics of the system behavior.

$$\lambda P_0 = \delta Q_0. \tag{3.1}$$

$$(\lambda + \theta + \delta)\, Q_0 = \ \theta Q_0 + \int_0^\infty \mathit{\Pi}_{b,0}(x)\mu_b(x)\mathrm{d}x + \int_0^\infty \mathit{\Pi}_{w,0}(x)\mu_w(x)\mathrm{d}x + \alpha \int_0^\infty \mathit{\Pi}_{b,n}(x)\mathrm{d}x, \quad n \geq 0. \tag{3.2}$$

$$\frac{\mathrm{d}P_n(x)}{\mathrm{d}x} + (\lambda + a(x))\, P_n(x) = 0, \quad n \geq 1. \tag{3.3}$$

$$\frac{\mathrm{d}P\mathit{\Pi}_{b,0}(x)}{\mathrm{d}x} + (\lambda + \alpha + \mu_b(x))\, \mathit{\Pi}_{b,0}(x) = 0, \quad n = 0. \tag{3.4}$$

$$\frac{\mathrm{d}P\mathit{\Pi}_{b,n}(x)}{\mathrm{d}x} + (\lambda + \alpha + \mu_b(x))\, \mathit{\Pi}_{b,n}(x) = \lambda \mathit{\Pi}_{b,n-1}(x), \quad n \geq 1. \tag{3.5}$$

$$\frac{\mathrm{d}P\mathit{\Pi}_{w,0}(x)}{\mathrm{d}x} + (\lambda + \theta + \delta + \mu_w(x))\, \mathit{\Pi}_{w,0}(x) = 0, \quad n = 0. \tag{3.6}$$

$$\frac{\mathrm{d}P\mathit{\Pi}_{w,n}(x)}{\mathrm{d}x} + (\lambda + \theta + \delta + \mu_w(x))\, \mathit{\Pi}_{w,n}(x) = \lambda \mathit{\Pi}_{w,n-1}(x), \quad n \geq 1. \tag{3.7}$$

The steady state boundary conditions at $x = 0$ and $y = 0$ are

$$P_n(0) = \int_0^\infty \mathit{\Pi}_{b,n}(x)\mu_b(x)\mathrm{d}x + \int_0^\infty \mathit{\Pi}_{w,n}(x)\mu_w(x)\mathrm{d}x, \quad n \geq 1. \tag{3.8}$$

$$\mathit{\Pi}_{b,0}(0) = \int_0^\infty P_1(x)a(x)\mathrm{d}x \; + (\theta+\delta)\int_0^\infty \mathit{\Pi}_{w,0}(x)\mathrm{d}x + \lambda P_0, \quad n=0. \tag{3.9}$$

$$\mathit{\Pi}_{b,n}(0) = \int_0^\infty P_{n+1}(x)a(x)\mathrm{d}x + \lambda\int_0^\infty P_n(x)\mathrm{d}x + (\theta+\delta)\int_0^\infty \mathit{\Pi}_{w,n}(x)\mathrm{d}x, \quad n\geq 1. \tag{3.10}$$

$$\mathit{\Pi}_{w,n}(0) = \begin{cases} \lambda Q_0, & n=0 \\ 0, & n\geq 1. \end{cases} \tag{3.11}$$

The normalizing condition is

$$P_0 + Q_0 + \sum_{n=1}^{\infty}\int_0^\infty P_n(x)\mathrm{d}x + \sum_{n=0}^{\infty}\left(\int_0^\infty \mathit{\Pi}_{b,n}(x)\mathrm{d}x + \int_0^\infty \mathit{\Pi}_{w,n}(x)\mathrm{d}x\right) = 1. \tag{3.12}$$

3.3. The steady state solution

The steady state solution of the retrial queueing model is obtained by using the generating function technique. To solve the above equations, the GFs are defined for $|z|\leq 1$ as follows:

$$P(x,z) = \sum_{n=1}^{\infty} P_n(x)z^n; \quad P(0,z) = \sum_{n=1}^{\infty} P_n(0)z^n; \quad \mathit{\Pi}_b(x,z) = \sum_{n=0}^{\infty} \mathit{\Pi}_{b,n}(x)z^n;$$
$$\mathit{\Pi}_b(0,z) = \sum_{n=0}^{\infty} \mathit{\Pi}_{b,n}(0)z^n; \quad \mathit{\Pi}_w(x,z) = \sum_{n=0}^{\infty} \mathit{\Pi}_{w,n}(x)z^n \quad \text{and} \quad \mathit{\Pi}_w(0,z) = \sum_{n=0}^{\infty} \mathit{\Pi}_{w,n}(0)z^n.$$

Now multiplying the steady state equation and steady state boundary conditions from (3.2) to (3.11), by zn and summing over n, $(n=0,1,2,\ldots)$.

$$\frac{\partial P(x,z)}{\partial x} + (\lambda + a(x))\,P(x,z) = 0. \tag{3.13}$$

$$\frac{\partial \mathit{\Pi}_b(x,z)}{\partial x} + (\lambda(1-z) + \alpha + \mu_b(x))\,\mathit{\Pi}_b(x,z) = 0. \tag{3.14}$$

$$\frac{\partial \mathit{\Pi}_w(x,z)}{\partial x} + (\lambda(1-z) + \theta + \delta + \mu_w(x))\,\mathit{\Pi}_w(x,z) = 0. \tag{3.15}$$

$$P(0,z) = \int_0^\infty \mathit{\Pi}_b(x,z)\mu_b(x)\mathrm{d}x + \int_0^\infty \mathit{\Pi}_w(x,z)\mu_w(x)\mathrm{d}x - \alpha\int_0^\infty \mathit{\Pi}_b(x,z)\mathrm{d}x - (\lambda+\delta)Q_0. \tag{3.16}$$

$$\mathit{\Pi}_b(0,z) = \frac{1}{z}\int_0^\infty P(x,z)a(x)\mathrm{d}x + \lambda\int_0^\infty P(x,z)\mathrm{d}x + (\theta+\delta)\int_0^\infty \mathit{\Pi}_w(x,z)\mathrm{d}x + \lambda P_0. \tag{3.17}$$

$$\mathit{\Pi}_w(0,z) = \lambda Q_0. \tag{3.18}$$

Solving the partial differential equations (3.13)–(3.15), it follows that

$$P(x,z) = P(0,z)[1-R(x)]e^{-\lambda x}, \tag{3.19}$$

$$\mathit{\Pi}_b(x,z) = \mathit{\Pi}_b(0,z)[1-S_b(x)]e^{-A_b(z)x}, \tag{3.20}$$

$$\mathit{\Pi}_w(x,z) = \mathit{\Pi}_w(0,z)[1-S_w(x)]e^{-A_w(z)x}, \tag{3.21}$$

where $A_b(z) = (\alpha + \lambda(1-z))$ and $A_w(z) = (\theta + \delta + \lambda(1-z))$.

Inserting the equations (3.18)–(3.21) and (3.17) and making some calculation, finally we get,

$$\mathit{\Pi}_b(0,z) = \frac{P(0,z)}{z}\left[R^*(\lambda) + z\left(1-R^*(\lambda)\right)\right] + \lambda P_0 + \lambda Q_0 W(z), \tag{3.22}$$

where $W(z) = \frac{(\theta+\delta)[1-S_w^*(A_w(z))]}{(\theta+\delta)+\lambda(1-z)}$ and $B(z) = \frac{\alpha[1-S_b^*(A_b(z))]}{\alpha+\lambda(1-z)}$.

Using (3.19)–(3.21) and (3.22) in (3.16), we get

$$P(0,z) = \mathit{\Pi}_b(0,z)\left(S_b^*\left(A_b(z)\right) + B(z)\right) + \mathit{\Pi}_w(0,z)S_w^*\left(A_w(z)\right) - (\lambda+\delta)Q_0. \tag{3.23}$$

Using (3.18) and (3.22) in (3.23), we get

$$\begin{aligned}(z-(R^*(\lambda)+z(1-R^*(\lambda)))(S_b^*(A_b(z))+B(z)))P(0,z) &= zQ_0((S_b^*(A_b(z))+B(z))(\lambda W(z)+\delta)\\ &\quad +\lambda(S_w^*(A_w(z))-1)-\delta).\end{aligned} \tag{3.24}$$

From the above equation, we know that the key element for obtaining $P(0,z)$ is to find the zeros of $f(z) = z-\left(R^*(\lambda)+z(1-R^*(\lambda))\right)\left(S_b^*\left(A_b(z)\right)+B(z)\right) = 0$ in the range $0<z<1$ for the equation $f(z)=0$ (from [9]). To this end, we give the following lemma.

Lemma 3.1. *If $\rho < R^*(\lambda)$, the equation $z-\left(R^*(\lambda)+z(1-R^*(\lambda))\right)\left(S_b^*\left(A_b(z)\right)+B(z)\right)$ has no roots in the range $0<z<1$ and has the minimal nonnegative root $z=1$.*

Proof. We only need to prove that

$$u(z) \overset{\Delta}{=} \left(R^*(\lambda)+z(1-R^*(\lambda))\right)\left(S_b^*\left(A_b(z)\right)+B(z)\right)$$

is a probability generating function of the number of customers that arrive in the system. Denote by U the time period from the epoch a service completion occurs, leaving the orbit non-empty, to the next service completion epoch, by NU the number of primary customers that arrive during U and define

$$u_j(t)\mathrm{d}t = P\left(t < U \le t+\mathrm{d}t,\ N(U)=j\right).$$

Then,

$$u_j(t) = e^{-\lambda t}\alpha(t) * a_j(t) + (1-\delta_{j,0})\lambda e^{-\lambda t}(1-R(t)) * a_{j-1}(t), \quad j=0,1,2\ldots$$

where * means convolution, $\alpha(t)$ is the p.d.f. of inter-retrial times, $b(t)$ is the p.d.f. of normal service times and $a_j(t)\mathrm{d}t = e^{-\lambda t}\frac{(\lambda t)^j}{j!}b(t)$. Denote by $N_U(z)$ the probability generating function of N_U, we have that

$$\begin{aligned}
N_U(z) &= \sum_{j=0}^{\infty} z^j \int_0^{\infty} u_j(t)\mathrm{d}t \\
&= \sum_{j=0}^{\infty} z^j \int_0^{\infty} \left(e^{-\lambda t}\alpha(t) * a_j(t) + (1-\delta_{j,0})\lambda e^{-\lambda t}(1-R(t)) * a_{j-1}(t)\right)\mathrm{d}t \\
&= (R^*(\lambda) + z(1-R^*(\lambda)))\left(S_b^*\left(A_b(z)\right) + B(z)\right) \\
&= u(z),
\end{aligned}$$

which proves the expected result that $u(z) \overset{\Delta}{=} (R^*(\lambda) + z(1-R^*(\lambda)))\left(S_b^*\left(A_b(z)\right) + B(z)\right)$ is exactly a probability generating function. From assumption $\rho < R^*(\lambda)$, we have $E[N_u] = \frac{\mathrm{d}}{\mathrm{d}z}u(z)|_{z=1} = 1 - (R^*(\lambda) - \rho) < 1$ and the convex function $u(z)$ is a monotonically increasing function of z for $0 \le z \le 1$, and $u(0) = P\left(N_U = 0\right) < 1,\ u(1) = 1$. So we can easily prove the expected result of Lemma 3.1.

Then for $\rho < R^*(\lambda)$, $z - (R^*(\lambda) + z(1-R^*(\lambda)))\left(S_b^*\left(A_b(z)\right) + B(z)\right)$ never vanishes in the range $0 < z < 1$. From (3.24), we get

$$P(0,z) = \frac{Nr(z)}{Dr(z)} \tag{3.25}$$

$$\begin{aligned}
Nr(z) &= zQ_0\left(\left(S_b^*\left(A_b(z)\right) + B(z)\right)\left(\lambda W(z) + \delta\right) + \lambda\left(S_w^*\left(A_w(z)\right) - 1\right) - \delta\right) \\
Dr(z) &= \left\{z - (R^*(\lambda) + z(1-R^*(\lambda)))\left(S_b^*\left(A_b(z)\right) + B(z)\right)\right\}.
\end{aligned}$$

Using equation (3.25) in (3.22), we get

$$\Pi_b(0,z) = Q_0\left\{z\left(\lambda W(z) + \delta\right) + \left(\lambda\left(S_w^*\left(A_w(z)\right) - 1\right) - \delta\right)\left(R^*(\lambda) + z(1-R^*(\lambda))\right)\right\}/Dr(z). \tag{3.26}$$

Using (3.18) and (3.25)–(3.26) in (3.19)–(3.22), then the limiting probability generating functions $P(x,z)$, $\Pi_b(x,z)$ and $\Pi_w(x,z)$. Next we are interested in investigating the marginal orbit size distributions due to system state of the server. □

Theorem 3.2. *Under the stability condition $\rho < R^*(\lambda)$, the stationary distributions of the number of customers in the orbit when server being idle, regular busy, lower speed service and probability that server idle are given by*

$$P(z) = \frac{Nr(z)}{Dr(z)} \tag{3.27}$$

$$\begin{aligned}
Nr(z) &= zQ_0\frac{(1-R^*(\lambda))}{\lambda}\left(\left(S_b^*\left(A_b(z)\right) + B(z)\right)\left(\lambda W(z) + \delta\right) + \lambda\left(S_w^*\left(A_w(z)\right) - 1\right) - \delta\right) \\
Dr(z) &= \left\{z - (R^*(\lambda) + z(1-R^*(\lambda)))\left(S_b^*\left(A_b(z)\right) + B(z)\right)\right\}
\end{aligned}$$

$$\Pi_b(z) = Q_0\left(1 - S_b^*\left(A_b(z)\right)\right)\left\{z\left(\lambda W(z) + \delta\right) + \left(\lambda\left(S_w^*\left(A_w(z)\right) - 1\right) - \delta\right)\left(R^*(\lambda) + z(1-R^*(\lambda))\right)\right\}/A_b(z) \times Dr(z) \tag{3.28}$$

$$\Pi_w(z) = \left\{\lambda Q_0 W(z)/(\theta + \delta)\right\}, \tag{3.29}$$

where

$$Q_0 = \frac{R^*(\lambda) - \frac{\lambda}{\alpha}\left(1 - S_b^*(\alpha)\right).}{\left\{\left(1+\frac{\delta}{\lambda}\right) R^*(\lambda) + \frac{\lambda}{(\theta+\delta)}\left(1 - S_w^*(\theta+\delta)\right)\left(1 - S_w^*(\theta+\delta)\left(1 - R^*(\lambda)\right)\right) - \frac{\lambda}{\alpha} S_w^*(\theta+\delta)\left(1 - S_b^*(\alpha)\right)\right\}} \tag{3.30}$$

$$P_0 = \frac{R^*(\lambda) - \frac{\lambda}{\alpha}\left(1 - S_b^*(\alpha)\right).}{\frac{\lambda}{\delta}\left\{\left(1+\frac{\delta}{\lambda}\right) R^*(\lambda) + \frac{\lambda}{(\theta+\delta)}\left(1 - S_w^*(\theta+\delta)\right)\left(1 - S_w^*(\theta+\delta)\left(1 - R^*(\lambda)\right)\right) - \frac{\lambda}{\alpha} S_w^*(\theta+\delta)\left(1 - S_b^*(\alpha)\right)\right\}} \tag{3.31}$$

$$A_b(z) = (\alpha + \lambda(1-z)) \quad \text{and} \quad A_w(z) = (\theta + \delta + \lambda(1-z))\,.$$

Proof. Integrating the equations (3.19)–(3.22) with respect to x and define the partial probability generating functions as, $P(z) = \int_0^\infty P(x,z)\mathrm{d}x$, $\mathit{\Pi}_b(z) = \int_0^\infty \mathit{\Pi}_b(x,z)\mathrm{d}x$, $\mathit{\Pi}_w(z) = \int_0^\infty \mathit{\Pi}_w(x,z)\mathrm{d}x$. Using the normalizing condition, we can be determined the probability that the server is idle ($P0$) when no customer in the orbit by setting $z = 1$ in (3.27)–(3.29) and applying l'Hôspital's rule whenever necessary and we get $P_0 + Q_0 + P(1) + \mathit{\Pi}_b(1) + \mathit{\Pi}_v(1) = 1$. □

Theorem 3.3. *Under the stability condition $\rho < R^*(\lambda)$, the probability generating functions of number of customers in the system and orbit size distribution at stationary point of time are*

$$K_s(z) = \frac{Nr_s(z)}{Dr_s(z)} \tag{3.32}$$

$$Nr_s(z) = Q_0 \left\{ \begin{array}{l} A_b(z)\left\{z - \left(R^*(\lambda) + z(1 - R^*(\lambda))\right)\left(S_b^*\left(A_b(z)\right) + B(z)\right)\right\}\left(\left(\lambda z W(z)/(\theta+\delta)\right) + 1 + (\delta/\lambda)\right) \\ +zA_b(z)\left(1 - R^*(\lambda)\right)\left(\left(S_b^*\left(A_b(z)\right) + B(z)\right)\left(W(z) + (\delta/\lambda)\right) + \left(S_w^*\left(A_w(z)\right) - 1\right) - (\delta/\lambda)\right) \\ +z\left(1 - S_b^*\left(A_b(z)\right)\right)\left\{z\left(\lambda W(z) + \delta\right) + \left(\lambda\left(S_w^*\left(A_w(z)\right) - 1\right) - \delta\right)\left(R^*(\lambda) + z(1 - R^*(\lambda))\right)\right\} \end{array} \right\}$$

$$Dr_s(z) = A_b(z) \times \left\{z - \left(R^*(\lambda) + z(1 - R^*(\lambda))\right)\left(S_b^*\left(A_b(z)\right) + B(z)\right)\right\}.$$

$$K_o(z) = \frac{Nr_o(z)}{Dr_s(z)} \tag{3.33}$$

$$Nr_o(z) = Q_0 \left\{ \begin{array}{l} A_b(z)\left\{z - \left(R^*(\lambda) + z(1 - R^*(\lambda))\right)\left(S_b^*\left(A_b(z)\right) + B(z)\right)\right\}\left(\left(\lambda W(z)/(\theta+\delta)\right) + 1 + (\delta/\lambda)\right) \\ +zA_b(z)\left(1 - R^*(\lambda)\right)\left(\left(S_b^*\left(A_b(z)\right) + B(z)\right)\left(W(z) + (\delta/\lambda)\right) + \left(S_w^*\left(A_w(z)\right) - 1\right) - (\delta/\lambda)\right) \\ +\left(1 - S_b^*\left(A_b(z)\right)\right)\left\{z\left(\lambda W(z) + \delta\right) + \left(\lambda\left(S_w^*\left(A_w(z)\right) - 1\right) - \delta\right)\left(R^*(\lambda) + z(1 - R^*(\lambda))\right)\right\} \end{array} \right\} \tag{3.34}$$

where Q_0 is given in equation (3.30).

Proof. The probability generating function of the number of customer in the system ($K_s(z)$) and the probability generating function of the number of customer in the orbit ($K_o(z)$) is obtained by using $K_s(z) = P_0 + Q_0 + P(z) + z\left(\mathit{\Pi}_b(z) + \mathit{\Pi}_w(z)\right)$ and $K_o(z) = P_0 + Q_0 + P(z) + \mathit{\Pi}_b(z) + \mathit{\Pi}_w(z)$. Substituting the equations (3.27)–(3.31) in the above results, then the equations (3.32) and (3.33) can be obtained by direct calculation. □

4. System performance measures

In this section, we derive some interesting system probabilities when the system in different states, system performance measures, mean busy period and mean busy cycle of this model.

4.1. System state probabilities

From equations (3.27)–(3.29), by setting $z \to 1$ and applying l'Hôspital's rule whenever necessary, then we get the following results

(i) Let P be the steady state probability that the server is idle during the retrial,

$$P = P(1) = \frac{Q_0 \left(1 - R^*(\lambda)\right) \left\{ \left(1 - S_w^*(\theta+\delta)\right) \left[\frac{\lambda}{\alpha} \left(1 - S_b^*(\alpha)\right) + \frac{\lambda}{(\theta+\delta)} \left(1 - S_w^*(\theta+\delta)\right) \right] + \frac{\delta}{\alpha} \left(1 - S_b^*(\alpha)\right) \right\}}{\left(R^*(\lambda) - \rho\right)} \tag{4.1}$$

(ii) Let Π_b be the steady-state probability that the server is busy,

$$\Pi_b = \Pi_b(1) = \frac{Q_0 \left(\left(1 - S_b^*(\alpha)\right)/\alpha\right) \left\{ \lambda \left(1 - S_w^*(\theta+\delta)\right) \left(\frac{\lambda}{(\theta+\delta)} + R^*(\lambda) \right) + \delta R^*(\lambda) \right\}}{\left(R^*(\lambda) - \rho\right)} \tag{4.2}$$

(iii) Let Π_w be the steady state probability that the server is on lower speed service,

$$\Pi_w = \Pi_w(1) = \frac{\lambda Q_0 \left(1 - S_w^*(\theta+\delta)\right)}{(\theta+\delta)} \tag{4.3}$$

(iv) Let Π_w be the steady state probability that the server is on working vacations and working breakdown,

$$\Pi_{wv} = \Pi_w + Q_0 = \frac{Q_0 \left((\theta+\delta) + \lambda \left(1 - S_w^*(\theta+\delta)\right)\right)}{(\theta+\delta)} \tag{4.4}$$

(v) Let P_f be the steady state probability that server failure,

$$P_f = \alpha \times \Pi_b(1) = \frac{Q_0 \left(1 - S_b^*(\alpha)\right) \left\{ \lambda \left(1 - S_w^*(\theta+\delta)\right) \left(\frac{\lambda}{(\theta+\delta)} + R^*(\lambda) \right) + \delta R^*(\lambda) \right\}}{\left(R^*(\lambda) - \rho\right)} \tag{4.5}$$

4.2. Mean system size and orbit size

If the system is in steady state condition,

(i) The expected number of customers in the orbit (L_q) is obtained by differentiating (3.33) with respect to z and evaluating at $z = 1$

$$L_q = K_o'(1) = \lim_{z \to 1} \frac{\mathrm{d}}{\mathrm{d}z} K_o(z) = Q_0 \left[\frac{Nr_q''(1) Dr_q'(1) - Dr_q''(1) Nr_q'(1)}{2 \left(Dr_q'(1)\right)^2} \right] \tag{4.6}$$

$$\begin{aligned} Nr_q'(1) = {} & \alpha \left\{ \left(1 + \frac{\delta}{\lambda}\right) R^*(\lambda) + \frac{\lambda}{(\theta+\delta)} \left(1 - S_w^*(\theta+\delta)\right) \left(1 - S_w^*(\theta+\delta) \left(1 - R^*(\lambda)\right)\right) \right. \\ & \left. - \lambda S_w^*(\theta+\delta) \left(1 - S_b^*(\alpha)\right) \right\} \\ Dr_q'(1) = {} & \alpha \left(R^*(\lambda) - \frac{\lambda}{\alpha} \left(1 - S_b^*(\alpha)\right) \right) \end{aligned}$$

$$
\begin{aligned}
Nr_q''(1) = &\left(1 + \tfrac{\delta}{\lambda} + \tfrac{\lambda}{(\theta+\delta)}\left(1 - S_w^*(\theta+\delta)\right)\right) Dr_q''(1) + \tfrac{2\lambda\alpha B'(1)}{(\theta+\delta)}\left(R^*(\lambda) - \tfrac{\lambda}{\alpha}\left(1 - S_b^*(\alpha)\right)\right) \\
&- 2(\lambda-\alpha)\left(1 - R^*(\lambda)\right)\left\{\left(1 - S_w^*(\theta+\delta)\right)\left[\tfrac{\lambda}{\alpha}\left(1 - S_b^*(\alpha)\right) + \tfrac{\lambda}{(\theta+\delta)}\left(1 - S_w^*(\theta+\delta)\right)\right] + \tfrac{\delta}{\alpha}\left(1 - S_b^*(\alpha)\right)\right\} \\
&- 2\left(1 - R^*(\lambda)\right)\left[\left(B'(1) - \lambda S_b^{*'}(\alpha)\right)\left(\lambda\left(1 - S_w^*(\theta+\delta)\right) + \delta\right) - 2\lambda B'(1)\left(1 - S_b^*(\alpha) + (\alpha/(\theta+\delta))\right)\right] \\
&- 2\lambda\left(1 - S_b^*(\alpha)\right)\left[\lambda S_w^{*'}(\theta+\delta)\left(1 - R^*(\lambda)\right) - B'(1)\left(1 + 1/(\theta+\delta)\right)\right] \\
&- 2\lambda S_b^{*'}(\alpha)\left[\lambda\left(1 - S_w^*(\theta+\delta)\right)\left(\tfrac{\lambda}{(\theta+\delta)} + R^*(\lambda)\right) + \delta R^*(\lambda)\right] \\
Dr_q''(1) = &-2\lambda\left(R^*(\lambda) - B'(1) - \left(1 - R^*(\lambda) + \tfrac{\lambda}{\alpha}\right)\left(1 - S_b^*(\alpha)\right)\right)
\end{aligned}
$$

(ii) The expected number of customers in the system (L_s) is obtained by differentiating (3.32) with respect to z and evaluating at $z = 1$

$$
L_s = K_s'(1) = \lim_{z\to 1}\frac{\mathrm{d}}{\mathrm{d}z}K_s(z) = Q_0\left[\frac{Nr_s''(1)Dr_q'(1) - Dr_q''(1)Nr_q'(1)}{2\left(Dr_q'(1)\right)^2}\right] \tag{4.7}
$$

$$
\begin{aligned}
Nr_s'''(1) = Nr_q'(1) + 2\left(1 - S_b^*(\alpha)\right)&\left(\frac{\lambda}{(\theta+\delta)}\left(\alpha R^*(\lambda) - \lambda\left(1 - S_b^*(\alpha)\right) + \lambda\alpha\left(1 - S_w^*(\theta+\delta)\right)\right)\right. \\
&\left. + \left(\lambda\left(1 - S_w^*(\theta+\delta)\right) + \delta\right)R^*(\lambda)\right)
\end{aligned} \tag{4.8}
$$

(iii) The average time a customer spends in the system (W_s) and the average time a customer spends in the queue (W_q) are found by using the Little's formula $W_s = \frac{L_s}{\lambda}$ and $W_q = \frac{L_q}{\lambda}$, where $B'(1) = \frac{\lambda}{\alpha}\left(1 - S_b^*(\alpha) + \alpha S_b^{*'}(\alpha)\right)$; $W'(1) = \frac{\lambda}{(\theta+\delta)}\left(1 - S_w^*(\theta+\delta) + (\theta+\delta)S_w^{*'}(\theta+\delta)\right)$;

$$
S_b^{*'}(\alpha) = \int_0^\infty x e^{-\alpha x}\mathrm{d}S_b(x) \ \text{ and } \ S_w^{*'}(\theta+\delta) = \int_0^\infty x e^{-(\theta+\delta)x}\mathrm{d}S_w(x). \tag{4.9}
$$

4.3. Mean busy period and busy cycle

Let $E(Tb)$ and $E(Tc)$ be the expected length of busy period and busy cycle under the steady state conditions. The results follow directly by applying the argument of an alternating renewal process [9] which leads to

$$
P_0 = \frac{E(T_0)}{E(T_b) + E(T_0)};\ E(T_b) = \frac{1}{\lambda}\left(\frac{1}{P_0} - 1\right) \ \text{ and } \ E(T_c) = \frac{1}{\lambda P_0} = E(T_0) + E(T_b). \tag{4.10}
$$

where T_0 is the time length that the system in empty state. Since the inter-arrival time between two customers follows exponential distribution with parameter λ, we have that $E(T_0) = (1/\lambda)$. Inserting (3.31) into (4.10) and use the above results, then we can get

$$
E(T_b) = \frac{1}{\lambda}\frac{\left\{\left(\frac{\delta}{\lambda}\left(1 + \frac{\delta}{\lambda}\right) - 1\right)R^*(\lambda) + \frac{\delta}{(\theta+\delta)}\left(1 - S_w^*(\theta+\delta)\right)\left(1 - S_w^*(\theta+\delta)\left(1 - R^*(\lambda)\right)\right) - \frac{\delta}{\alpha}\left(S_w^*(\theta+\delta) - \lambda\right)\left(1 - S_b^*(\alpha)\right)\right\}}{R^*(\lambda) - \frac{\lambda}{\alpha}\left(1 - S_b^*(\alpha)\right)} \tag{4.11}
$$

$$
E(T_c) = \frac{\left\{\left(1 + \frac{\delta}{\lambda}\right)R^*(\lambda) + \frac{\lambda}{(\theta+\delta)}\left(1 - S_w^*(\theta+\delta)\right)\left(1 - S_w^*(\theta+\delta)\left(1 - R^*(\lambda)\right)\right) - \frac{\lambda}{\alpha}S_w^*(\theta+\delta)\left(1 - S_b^*(\alpha)\right)\right\}}{\delta\left(R^*(\lambda) - \frac{\lambda}{\alpha}\left(1 - S_b^*(\alpha)\right)\right)}. \tag{4.12}
$$

5. Conditional stochastic decomposition

In this section, we study the stochastic decomposition property of the system size distribution. The number of customers in the system is distributed as the sum of two independent random variables. In particular, in the context our system, we will discuss the conditional stochastic decomposition of the number of customers in the orbit give that the server is busy. Let N_b is the conditional orbit size of our retrial queuing system given that server is busy and N_0 is the conditional orbit size of the M/G/1 retrial queueing system is given that the server is busy which is discussed in Theorem 5.1.

Theorem 5.1. *The conditional orbit size N_b is given that the server is busy can be decomposed into the sum of two independent random variables $N_b = N_0 + N_c$. where N_0 has the generating function $N_0(z)$ as follows,*

$$N_0(z) = \frac{\left(S_b^*\left(A_b(z)\right) - 1\right)\left(R^*(\lambda) - \rho\right)}{\left(z - \left(R^*(\lambda) + z(1 - R^*(\lambda))\right) S_b^*\left(A_b(z)\right)\right)(\rho)}$$

and N_c is the additional queue length due to vacations with the probability generating function $N_c(z)$ as follows,

$$\begin{aligned} N_c(z) = &\left\{\frac{\left((\theta+\delta)\left(B(z)/\alpha\right)\left\{z\left(\lambda W(z)+\delta\right)+\left(\lambda\left(S_w^*\left(A_w(z)\right)-1\right)-\delta\right)\left(R^*(\lambda)+z(1-R^*(\lambda))\right)\right\}+\lambda W(z) Dr(z)\right)}{Dr(z)\times\left(S_b^*\left(A_b(z)\right)-1\right)}\right\} \\ &\times\frac{\rho\times\left(z-\left(R^*(\lambda)+z(1-R^*(\lambda))\right)S_b^*\left(A_b(z)\right)\right)}{\left((\theta+\delta)(\rho)\left\{\left(1-S_w^*(\theta+\delta)\right)\left(\frac{\lambda}{(\theta+\delta)}+R^*(\lambda)\right)+\frac{\delta}{\lambda}R^*(\lambda)\right\}+\lambda\left(1-S_w^*(\theta+\delta)\right)\left(R^*(\lambda)-\rho\right)\right)}. \end{aligned} \tag{5.1}$$

Proof. We observe that the generating function of system size distribution can be decomposed as follows:

The mathematical version of the stochastic decomposition law is $N_b(z) = N_0(z)\ N_c(z)$.

We know that for the M/G/1 retrial queueing system, the marginal function of the number of customers in the orbit when the server is busy is given by

$$\Phi(z) = \frac{\lambda P_0 R^*(\lambda)\left(S_b^*\left(A_b(z)\right) - 1\right)}{\left(z - \left(R^*(\lambda) + z(1 - R^*(\lambda))\right) S_b^*\left(A_b(z)\right)\right)}$$

and the probability that server is busy is given by

$$\Phi(1) = \frac{\lambda\rho P_0 R^*(\lambda)}{R^*(\lambda) - \rho},$$

then for the generating function $N_0(z)$, we have

$$N_0(z) = \frac{\Phi(z)}{\Phi(1)} = \frac{\left(S_b^*\left(A_b(z)\right) - 1\right)\left(R^*(\lambda) - \rho\right)}{\left(z - \left(R^*(\lambda) + z(1 - R^*(\lambda))\right) S_b^*\left(A_b(z)\right)\right)(\rho)}.$$

From the equations (3.28)–(3.29), we know that for our retrial system the generating function of Nb is given by

$$N_b(z) = \frac{\Pi_b(z) + \Pi_w(z)}{\Pi_b(1) + \Pi_w(1)}$$

$$N_b(z) = \frac{\left((\theta+\delta)\left(B(z)/\alpha\right)\left\{z\left(\lambda W(z)+\delta\right)+\left(\lambda\left(S_w^*\left(A_w(z)\right)-1\right)-\delta\right)\left(R^*(\lambda)+z(1-R^*(\lambda))\right)\right\}+\lambda W(z)Dr(z)\right)\left(R^*(\lambda)-\rho\right)}{Dr(z)\left((\theta+\delta)(\rho)\left\{\left(1-S_w^*(\theta+\delta)\right)\left(\frac{\lambda}{(\theta+\delta)}+R^*(\lambda)\right)+\frac{\delta}{\lambda}R^*(\lambda)\right\}+\lambda\left(1-S_w^*(\theta+\delta)\right)\left(R^*(\lambda)-\rho\right)\right)}.$$

$$N_b(z) = N_0(z) \times N_c(z)$$

From above stochastic decomposition law, we observe that $N_b(z) = N_0(z) \times N_c(z)$which conform that the decomposition results of Gao *et al.* [9], also valid for this special vacation system. □

6. Special cases

In this section, we analyze briefly some special cases of our model, which are consistent with the existing literature.

Case (i): *No disasters, No repair and No working breakdown*

Let $\alpha = \delta = 0$; our model can be reduced to a single server retrial queueing system with working vacations. In this case, $K_s(z)$ can be simplified to the following expression. The following result coincides with the result of Gao *et al.* [9].

$$K_s(z) = P_0 \frac{\left\{ \begin{array}{l} (1-z)\left[\begin{array}{l} (z-(R^*(\lambda)+z(1-R^*(\lambda)))\,S_b^*(\lambda-\lambda z))\,((\lambda W(z)/\theta)+1)+z\,(1-R^*(\lambda)) \\ \left[(S_w^*(\theta+\lambda-\lambda z)+W(z)S_b^*(\lambda-\lambda z))-1\right] \end{array} \right] \\ +(1-S_b^*(\lambda-\lambda z))\left[(S_b^*(\lambda-\lambda z)-1)\,(R^*(\lambda)+z(1-R^*(\lambda)))+zW(z)\right] \end{array} \right\}}{(1-z)\,(z-(R^*(\lambda)+z(1-R^*(\lambda)))\,S_b^*(\lambda-\lambda z))}$$

Case (ii): *No disasters, No repair, No working breakdown and No vacation interruption*

Let $(\alpha,\ \delta,\ \theta) \to (0,\ 0,\ 0)$, our model can be reduced to $M/G/1$ retrial queue with single working vacation. This model results coincide with the result of Arivudainambi *et al.* [1].

Case (iii): *No disasters, No working vacation and breakdown, No vacation interruption, No repair*

Let $(\alpha,\ \delta,\ \theta) \to (0,\ 0,\ 0)$ and $S_w^*(A_w(z)) \to 1$. Our model can be reduced to $M/G/1$ retrial queue with general retrial times. The following result coincides with the result of Gomes Corral [10].

$$K_s(z) = \frac{P_0 R^*(\lambda)(z-1)S_b^*(A_b(z))}{(z-(R^*(\lambda)+z(1-R^*(\lambda)))\,S_b^*(A_b(z)))} \ \text{ and } \ P_0 = \frac{R^*(\lambda)-\lambda E(S_b)}{R^*(\lambda)}.$$

7. Cost optimization analysis

In order to carry out cost analysis, the optimum design of a retrial queueing system is to determine the optimal system parameters, such as optimal mean service rate or optimal number of servers [25]. In this section, the optimal design of a single server retrial queueing system with disasters under working vacations and working breakdowns is addressed. Let us define,

Ch is holding cost per unit time for each customer present in the system; C_b is cost incurred per unit time when the server provides service during a normal busy period; C_v is cost incurred per unit time when the server provides service during a working vacation and working breakdown period; Cf is cost incurred per unit time when the server is in failure; C_1 is cost per customer served by the mean service rate μ_b; and C_2 is cost per customer served by the mean service rate μ_w.

Using the above cost parameters and corresponding system performance measures, the expected cost function per unit time under a linear cost structure is given by

$$TC = C_h L_s + C_b \varPi_b + C_v \varPi_{wv} + C_f P_f + C_1\mu_b + C_2\mu_w.$$

where L_s, $\varPi_b$, $\varPi_{wv}$, P_f are given in equations (4.7), (4.2), (4.4), (4.5) respectively. We assume exponential retrial times, service times, lower speed service times, vacation times and repair times. For the following values of the cost elements and other parameters like: $\lambda = 2,\ a = 3,\ \mu_b = 4,\ \mu_w = 2,\ \theta = 1,\ \alpha = 0.3,\ \delta = 8,\ C_h = \$5,\ C_b = \$350,\ C_v = \$100,\ C_f = \$200,\ C_1 = \$10$ and $C_2 = \$12$. We find the total expected cost per unit of time $\boldsymbol{TC} = \mathbf{\$241.5784.}$

Moreover, we examine the behavior of the expected cost function under different values of the cost parameters. System parameters are fixed as follows: $\lambda = 2,\ a = 3,\ \mu_b = 4,\ \mu_w = 2,\ \theta = 1,\ \alpha = 0.3,\ \delta = 8$; Tables 1–3 illustrate the effects of $(C_h,\ C_b)$, $(C_v,\ C_f)$ and $(C_1,\ C_2)$ on the expected cost function, respectively. It can be see that the expected cost function shows a linearly increasing trend with increasing cost parameters.

TABLE 1. Effects of (C_h, C_b) on the expected cost function TC with C_v = \$100, C_f= \$200, C_1 = \$10 and C_2= \$12.

(C_h, C_b)	(5, 350)	(10, 350)	(15, 350)	(5, 355)	(5, 360)
TC	241.5784	252.9014	264.2244	243.3840	245.1896

TABLE 2. Effects of (C_v, C_f) on the expected cost function TC with C_h= \$5, C_b= \$350, C_1= \$10 and C_2= \$12.

(C_v, C_f)	(100, 200)	(105, 200)	(110, 200)	(100, 205)	(100, 210)
TC	241.5784	241.9881	242.3979	242.1201	242.6618

TABLE 3. Effects of (C_1, C_2) on the expected cost function TC with C_h = \$5, C_b= \$350, Cv = \$100 and Cf = \$200.

(C_1, C_2)	(10, 12)	(15, 12)	(20, 12)	(10, 7)	(10, 2)
TC	241.5784	266.5784	291.5784	231.5784	221.5784

Next, a sensitivity analysis of some of the parameters on the system can be conducted. Fixing the base values given above, one parameter can be varied at a time and the corresponding objective function value computed. The following graphs from Figures 1–3 show the effect of some of the system parameters (a, μ_b, μ_v) on the total expected cost per unit of time.

Figure 1 shows that the expected cost function decreases for increasing the retrial rate (a). We observe from Figures 2 and 3 that the expected cost function first decreases as one of μ_b or μ_v increases; and then it increases thereafter. Thus, the expected cost function seems to be convex with respect to each one of the decision variables.

8. Sensitivity analysis and numerical examples

In this section, we present some numerical examples using MATLAB in order to illustrate the effect of various parameters in the system performance measures. We consider retrial times, service times, lower speed service times, vacation times and repair times are exponentially distributed. The arbitrary values to the parameters are so chosen such that they satisfy the stability condition. The following tables give the computed values of various characteristics of our model like, probability that the server is idle (P_0), the mean orbit size (L_q), the mean system size (L_s), probability that server is idle during retrial rime (P), regular busy (Π_b), working vacation and working breakdown (Π_w) and server failure (P_f) respectively. The exponential distribution is $f(x) = ve^{-vx}$, $x > 0$.

Algorithm to compute L_q:
Begin
Input: λ, a, μ_b, μ_w, θ, α and δ.
Compute: P from equation (4.1).
Compute: Π_b from equation (4.2).
Compute: Π_w from equation (4.3).
Compute: L_q from equation (4.6).
Output: L_q.

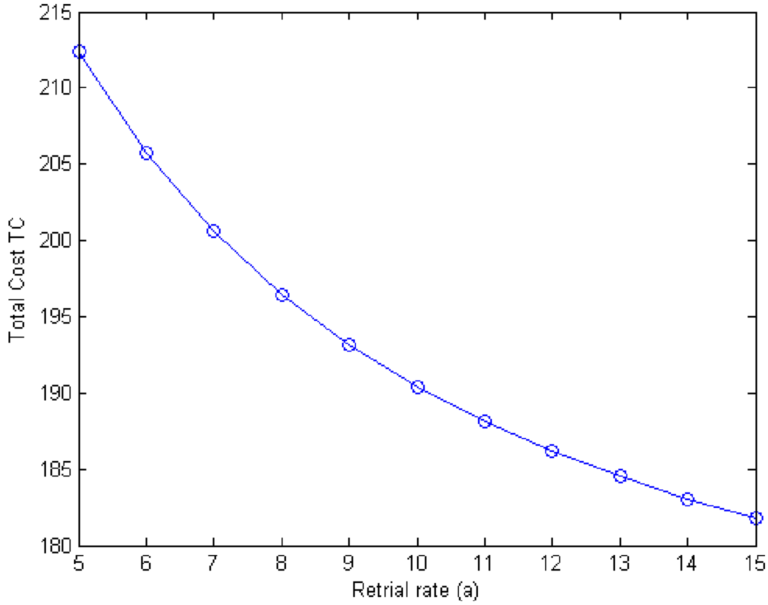

FIGURE 1. *TC versus a.* (Color online.)

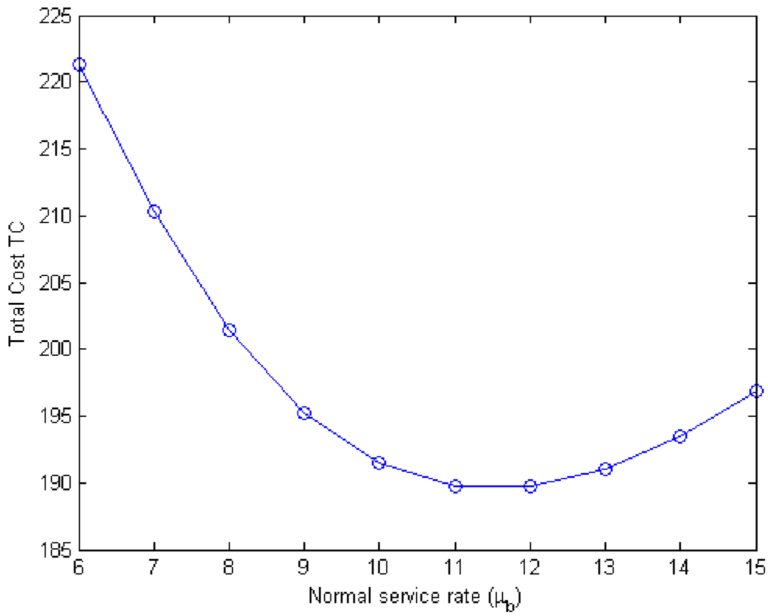

FIGURE 2. *TC versus* μ_b. (Color online.)

Tables 4–7 show the numerical results of the system performance measures with respect to the variation of the system parameters. From Table 4, we find that (i) L_s, L_q and P decreases as retrial rate (a) increases; (ii) P_0, Π_b, Π_{wv} and P_f increases as retrial rate (a) increases; it is found from Table 5 that (i) P_0, L_s, L_q Π_{wv} and Pf increases with increasing values of failure rate (α); (ii) P and Π_b decreases with increasing values of failure rate (α);

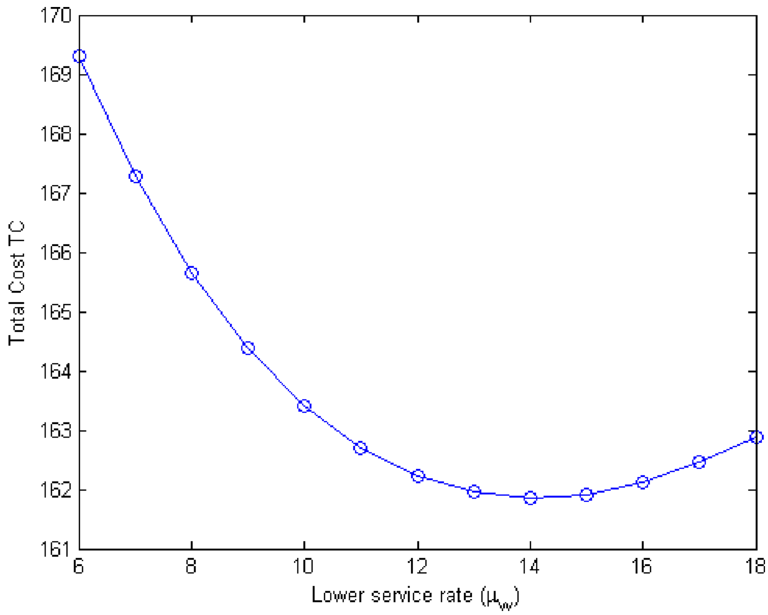

FIGURE 3. *TC versus* μ_v. (Color online.)

TABLE 4. The effect of retrial rate (a) on system performance measures for different values of $\lambda = 2,\ \mu_b = 4,\ \mu_w = 2,\ \theta = 1,\ \alpha = 0.4,\ \ \delta =\ 8.$

Retrial rate (a)	P_0	L_q	L_s	P	Π_b	Π_{wv}	P_f
3.00	0.1813	6.2096	5.7890	0.0649	0.4352	0.0536	0.1741
3.50	0.2144	3.2187	2.7969	0.0558	0.4353	0.0634	0.1741
4.00	0.2395	2.1178	1.6952	0.0490	0.4354	0.0708	0.1742
4.50	0.2590	1.7424	1.3191	0.0436	0.4355	0.0765	0.1742
5.00	0.2748	1.6814	1.2576	0.0393	0.4355	0.0812	0.1742

TABLE 5. The effect of failure rate (α) on system performance measures for different values of $\lambda = 2,\ a = 3,\ \mu_b = 4,\ \mu_w = 2,\ \theta = 1,\ \delta =\ 8.$

Failure rate (α)	P_0	L_q	L_s	P	Π_b	Π_{wv}	P_f
0.20	0.2686	0.6292	0.9754	0.0516	0.3617	0.0783	0.0723
0.30	0.2776	1.3613	1.7033	0.0509	0.3549	0.0810	0.1065
0.40	0.2863	1.6359	1.9738	0.0502	0.3483	0.0835	0.1393
0.50	0.2947	1.7383	2.0722	0.0496	0.3419	0.0860	0.1710
0.60	0.3028	1.7620	2.0922	0.0490	0.3358	0.0883	0.2015

As can be seen in Table 6, there is an increasing trend of P_0 as the lower speed service rate (μ_w) increases. However, the other measures such as L_s, L_q, P_0, Π_b, Π_{wv} and Pf decrease as the lower speed service rate (μ_w) increases. As expected from Table 7, increasing θ decreases the value of the L_s, L_q, P, Π_{wv} and other performance measures $P0$, Π_b and P_f increase. Based on the above, smaller for large values of μ_v and turns to zero when $\mu_w = \mu_b$. Another important case is $\mu_w = 0$, *i.e.*, the server cannot provide service during a vacation period; the effect of the vacation rate θ has a noticeable effect on the system performance and cannot be ignored.

Table 6. The effect of lower service rate (μ_w) on system performance measures for different values of $\lambda = 2,\ a = 3,\ \mu_b = 4,\ \theta = 1,\ \alpha = 0.3,\ \ \delta =\ 8$.

Lower service rate (μ_w)	P_0	L_q	L_s	P	Π_b	Π_{wv}	P_f
2.00	0.2944	2.0482	2.3873	0.0556	0.3480	0.0870	0.1740
3.00	0.2947	1.7383	2.0722	0.0496	0.3419	0.0860	0.1710
4.00	0.2949	1.4733	1.8029	0.0447	0.3367	0.0851	0.1684
5.00	0.2950	1.2444	1.5700	0.0406	0.3323	0.0843	0.1661
6.00	0.2951	1.0446	1.3669	0.0372	0.3283	0.0836	0.1642

Table 7. The effect of vacation rate (θ) on system performance measures for different values of $\lambda = 2,\ a = 3,\ \mu_b = 4,\ \mu_w = 2,\ \alpha = 0.3,\ \ \delta =\ 8$.

Vacation rate (θ)	P_0	L_q	L_s	P	Π_b	Π_{wv}	P_f
3.00	0.4115	1.1717	0.9828	0.0568	0.2005	0.1975	0.0601
4.00	0.4160	1.1285	0.9372	0.0561	0.2020	0.1966	0.0606
5.00	0.4197	1.0806	0.8873	0.0555	0.2032	0.1958	0.0610
6.00	0.4228	1.0340	0.8390	0.0549	0.2042	0.1951	0.0613
7.00	0.4254	0.9906	0.7941	0.0542	0.2051	0.1945	0.0615

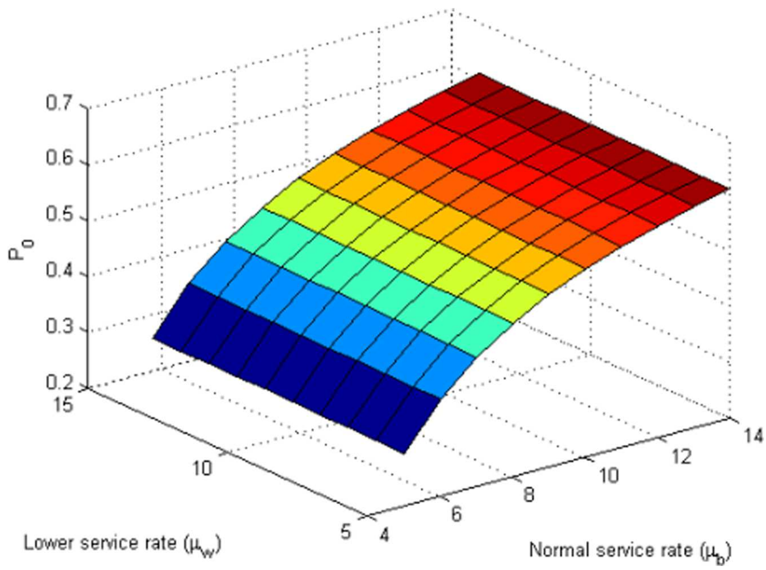

Figure 4. P_0 *versus* μ_w and μ_b. (Color online.)

For the effect of the parameters $a,\ \mu_b,\ \mu_w,\ \theta,\ \alpha$ and δ on the system performance measures, three dimensional graphs are illustrated in Figures 4–6. In Figure 4, the surface displays an upward trend as expected for increasing the value of regular service rate (μ_b) and lower service rate (μ_w) against idle probability (P_0). In Figure 5, we examine the behavior of the mean orbit size (L_q) decreases for increasing the value of retrial rate (a) and repair rate (δ). Figure 6 shows that the failure frequency (P_f) increases for increasing the value of breakdown rate (α) and vacation rate (θ).

From the above numerical examples, we can find the influence of parameters on the performance measures in the system and know that the results are coincident with the practical situations.

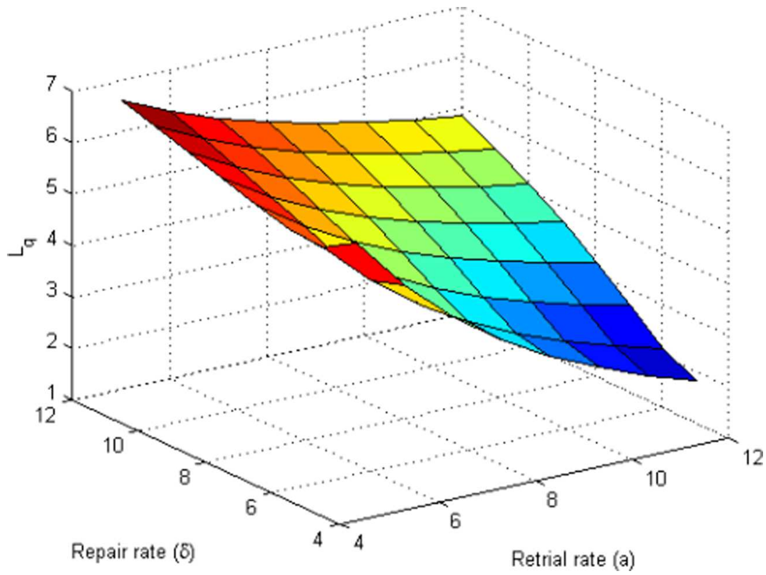

FIGURE 5. L_q *versus* a and δ. (Color online.)

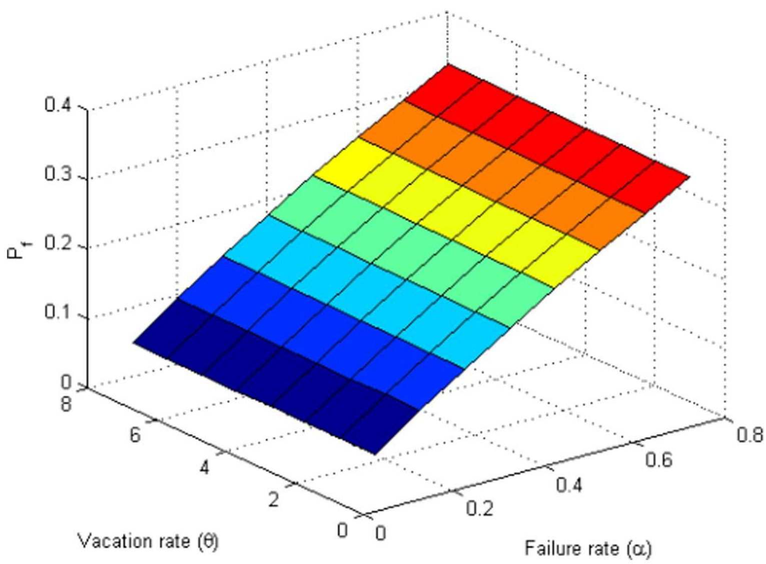

FIGURE 6. P_f *versus* θ and α. (Color online.)

9. Conclusion

In this work, we have investigated a single server retrial queueing system with disasters under working vacations and working breakdowns. The necessary and sufficient condition for the system to be stable is obtained. Using the probability generating function approach and the method of supplementary variable technique, the probability generating functions of the numbers of customers in the system and its orbit when it is free, regular busy, on lower speed service are derived. Various system performance measures and conditional stochastic decomposition law are discussed. The explicit expressions for the average queue length of orbit and system have been obtained. Finally, some numerical examples and cost optimization analysis are presented to study the impact of the system parameters and cost elements. The motivation for this model comes from wide range applications in many real time systems, for example in computer and communication network where messages are processed by a single server retrial queues in presence of working breakdowns and multiple working vacation policy. This proposed model has potential practical real life application in production to order system to enhance the performance of the production facility and to stop the production facility from becoming overloaded, in computer processing system and telephone consultation of medical service systems. Hopefully, this investigation will be great help to the system managers who can design a system with economic management and to make decisions regarding the size of the system and other factors in a well-to-do manner.

Appendix A

The embedded Markov chain $\{Z_n;\ n \in N\}$ is ergodic if and only if $\rho < R^*(\lambda)$, where $\rho = \frac{\lambda}{\alpha}\left(1 - S_b^*(\alpha)\right)$.

Proof. To prove the sufficient condition of ergodicity, it is very convenient to use Foster's criterion [15], which states that the chain$\{Z_n;\ n \in N\}$ is an irreducible and aperiodic Markov chain is ergodic if there exists a non-negative function $f(j)$, $j{\in}N$ and $\varepsilon{>}\ 0$, such that mean drift $\psi_j = E\left[f(z_{n+1}) - f(z_n)/z_n = j\right]$ is finite for all $j{\in}N$ and$\psi_j \le -\varepsilon$ for all $j{\in}N$, except perhaps for a finite number j's. In our case, we consider the function $f(j){=}\ j$. then we have

$$\psi_j = \begin{cases} \rho - 1, & \text{if } j = 0, \\ \rho - R^*(\lambda), & \text{if } j = 1, 2, \ldots \end{cases}$$

Clearly the inequality $\rho < R^*(\lambda)$ is sufficient condition for ergodicity.

To prove the necessary condition, as noted in Sennott *et al.* [20], if the Markov chain $\{Z_n;\ n \ge 1\}$ satisfies Kaplan's condition, namely, $\psi_j < \infty$ for all $j \ge 0$ and there exits $j_0 \in N$ such that $\psi_j \ge 0$ for $j \ge j_0$. Notice that, in our case, Kaplan's condition is satisfied because there is a k such that $m_{ij} = 0$ for $j < i - k$ and $i > 0$, where $\mathrm{M} = (m_{ij})$ is the one step transition matrix of $\{Z_n;\ n \in N\}$ .Then $\rho \ge R^*(\lambda)$ implies the non-ergodicity of the Markov chain. □

Acknowledgements. The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of this paper.

References

[1] D. Arivudainambi, P. Godhandaraman and P. Rajadurai, Performance analysis of a single server retrial queue with working vacation. *OPSEARCH* **51** (2014) 434–462.

[2] J.R. Artalejo, Accessible bibliography on retrial queues: progress in 2000–2009. *Math. Comput. Model.* **51** (2010) 1071–1081.

[3] J.R. Artalejo, A. Gomez-Corral, Retrial Queueing Systems. Springer, Berlin, Germany (2008).

[4] O. Boudali and A. Economou, The effect of catastrophes on the strategic customer behavior in queueing systems. *Naval Res. Logist.* **60** (2013) 571–587.

[5] V.M. Chandrasekaran, K. Indhira, M.C. Saravanarajan and P. Rajadurai, A survey on working vacation queueing models. *Int. J. Pure Appl. Math.* **106** (2016) 33–41.

[6] G. Choudhury and K. Deka, A single server queueing system with two phases of service subject to server breakdown and Bernoulli vacation. *Appl. Math. Model.* **36** (2012) 6050–6060.

[7] I. Dimitriou, A mixed priority retrial queue with negative arrivals, unreliable server and multiple vacations. *Appl. Math. Model.* **37** (2013) 1295–1309.

[8] S. Gao and Z. Liu, An M/G/1 queue with single working vacation and vacation interruption under Bernoulli schedule. *Appl. Math. Model.* **37** (2013) 1564–1579.

[9] S. Gao, J. Wang, W. Li, An M/G/1 retrial queue with general retrial times, working vacations and vacation interruption. *Asia-Pacific J. Oper. Res.* **31** (2014) 6–31.

[10] A. Gomez-Corral, Stochastic analysis of a single server retrial queue with general retrial times. *Naval Res. Logist.* **46** (1999) 561–581.

[11] T. Jiang and L. Liu, The GI/M/1 queue in a multi-phase service environment with disasters and working breakdowns. *Int. J. Comput. Math.* **94** (2017) 707–726.

[12] K. Kalidass and K. Ramanath, A queue with working breakdowns. *Comput. Ind. Eng.* **63** (2012) 779–783.
[13] B.K. Kim and D.H. Lee, The M/G/1 queue with disasters and working breakdowns. *Appl. Math. Model.* **38** (2014) 1788–1798.
[14] D.H. Lee, W.S. Yang and H.M. Park, Geo/G/1 queues with disasters and general repair times. *Appl. Math. Model.* **35** (2011) 1561–1570.
[15] A.G. Pakes, Some conditions for Ergodicity and recurrence of Markov chains. *Oper. Res.* **17** (1969) 1058–1061.
[16] P. Rajadurai, V.M. Chandrasekaran and M.C. Saravanarajan, Analysis of an M[X]/G/1 unreliable retrial *G*-queue with orbital search and feedback under Bernoulli vacation schedule. *OPSEARCH* **53** (2016) 197–223.
[17] P. Rajadurai, V.M. Chandrasekaran and M.C. Saravanarajan, Analysis of an unreliable retrial G-queue with working vacations and vacation interruption under Bernoulli schedule. To appear in *Ain Shams Eng. J.* (2016) DOI: 10.1016/j.asej.2016.03.008
[18] P. Rajadurai, M.C. Saravanarajan and V.M. Chandrasekaran, Analysis of an M[X]/(G1, G2)/1 retrial queueing system with balking, optional re-service under modified vacation policy and service interruption. *Ain Shams Eng. J.* **5** (2014) 935–950.
[19] P. Rajadurai, M.C. Saravanarajan and V.M. Chandrasekaran, A study on M/G/1 feedback retrial queue with subject to server breakdown and repair under multiple working vacation policy. To appear in *Alexandria Eng. J.* (2017) DOI: 10.1016/j.aej.2017.01.002
[20] L.I. Sennott, P.A. Humblet and R.L. Tweedi, Mean drifts and the non-Ergodicity of Markov chains. *Oper. Res.* **31** (1983) 783–789.
[21] L.D. Servi and S.G. Finn, M/M/1 queues with working vacations. *Perform. Eval.* **50** (2002) 41–52.
[22] R. Sudhesh, R. Sebasthi Priya and R.B. Lenin, Analysis of N-policy queues with disastrous breakdown. *TOP* **24** (2016) 612–634.
[23] D. Wu and H. Takagi, M/G/1 queue with multiple working vacations. *Perform. Eval.* **63** (2006) 654–681.
[24] S. Yang, J. Wu and Z. Liu, An M[X]/G/1 retrial G-queue with single vacation subject to the server breakdown and repair. *Acta Math. Appl. Sin. Engl. Ser.* **29** (2013) 579–596.
[25] D.Y. Yang and C.H. Wu, Cost-minimization analysis of a working vacation queue with N-policy and server breakdowns. *Comput. Ind. Eng.* **82** (2015) 151–158.
[26] M. Zhang and Z. Hou, M/G/1 queue with single working vacation. *J. Appl. Math. Comput.* **39** (2012) 221–234.
[27] M. Zhang and Q. Liu, An M/G/1 G-queue with server breakdown, working vacations and vacation interruption. *OPSEARCH* **52** (2015) 256–270.